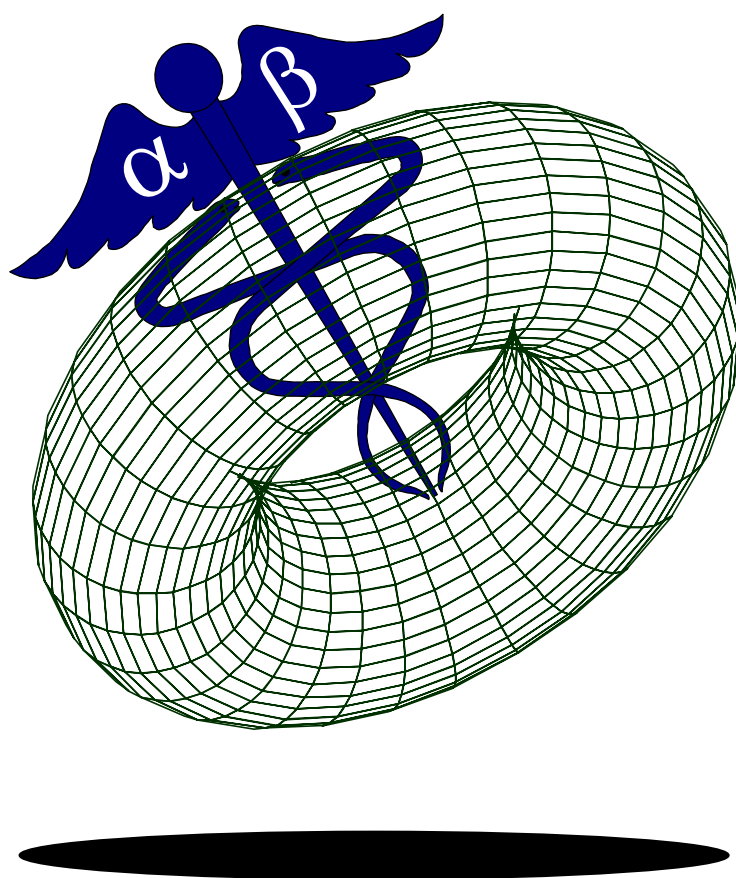


คู่มือปฏิบัติการชีวสถิติ (Workbook for Biostatistics)

สำหรับเรียนรู้ชีวสถิติด้วยตนเอง



โดย บัณฑิต กีนดำรพ
ภาควิชาชีวสถิติและประชากรศาสตร์
คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น

คู่มือปฏิบัติการชีวสถิติ

(Workbook for Biostatistics)

โดย

รองศาสตราจารย์ ดร. บัณฑิต อินคำรพ

วท.บ. (สาธารณสุขศาสตร์) เกียรตินิยม

M.P.H. (Epidemiology)

Grad. Dip. Med. Stat.

Ph.D. (Statistics)

ภาควิชาชีวสถิติและประชากรศาสตร์

คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น

II

พิมพ์ครั้งที่ 1 พ.ศ. 2541	500	เล่ม
พิมพ์ครั้งที่ 2 พ.ศ. 2543	1,000	เล่ม

หจก. โรงพิมพ์คลังนานาวิทยา

232/199 หมู่ที่ 6 ถนนศรีจันทร์ ตำบลในเมือง อำเภอเมือง จังหวัดขอนแก่น 40000

โทรศัพท์ (043) 328589-91 แฟกซ์ (043) 328592

การคัดลอก ถ่ายเอกสาร หรือพิมพ์เผยแพร่หนังสือเล่มนี้

สามารถทำได้กรณีเพื่อการศึกษาและไม่หวังผลกำไร

การทำเผยแพร่คราวละมากๆ ต้องได้รับการอนุญาตจากผู้เขียนเป็นลายลักษณ์อักษรก่อน

คำนำ

วิชาสถิติ มีการเรียนการสอนในหลักสูตรการศึกษาตั้งแต่ระดับมัธยมศึกษาจนถึงระดับปริญญาเอก สิ่งนี้บ่งชี้ถึงความสำคัญของศาสตร์นี้ แม้มีการเรียนซ้ำหลายครั้งแต่มีข้อบ่งชี้หลายอย่างที่แสดงให้เห็นว่าบุคลากรทางด้านการแพทย์และสาธารณสุขยังขาดความเข้าใจในหลักสถิติ อันเป็นรากฐานที่สำคัญ ข้อบ่งชี้หนึ่งที่ชัดเจนคือพบว่ามีการใช้สถิติที่ไม่เหมาะสมมากกว่าครึ่งในบทความวิจัยที่ตีพิมพ์ในวารสารต่าง ๆ มีการนำเสนอที่ไม่เหมาะสม และที่พบว่าไม่เหมาะสมหรือผิดมากที่สุดคือการใช้และแปลความหมายการทดสอบความมีนัยสำคัญทางสถิติ (Significant) โดยรวมแล้วมักพบว่าปัญหาด้านคำนวณมีน้อยกว่าการเลือกใช้และการแปลความหมาย ไม่ใช่เฉพาะในประเทศไทยแต่เป็นปัญหาในระดับโลก เช่นเรื่องการแปลความหมายและใช้การทดสอบสมมติฐานทางสถิติที่ผิดพลาดเป็นมาตลอดระยะเวลากว่าครึ่งศตวรรษและยังพบมากกระทั่งปัจจุบัน ความเข้าใจถึงหลักการและแนวคิดของสถิติ จะทำให้สามารถแก้ปัญหาข้างต้นและเป็นพื้นฐานสำคัญสำหรับการเรียนรู้ด้วยตนเองในระดับที่สูงขึ้นได้

สถิติเป็นทั้งความรู้พื้นฐานที่พึงมี เป็นสิ่งที่ใช้เป็นเครื่องมือแสวงหาคำตอบความรู้ ทั้งที่ทำได้สามารถอ่านผลงานการศึกษาวิจัยได้อย่างมีวิจารณญาณ และใช้เป็นเครื่องมือที่ใช้เปลี่ยนข้อมูลข่าวสารเป็นองค์ความรู้เมื่อทำการศึกษาวิจัย จึงนับว่ามีความสำคัญมากในยุคนี้ซึ่งในยุคของข้อมูลข่าวสารซึ่งในยุคที่ความรู้คืออำนาจที่แท้จริง ความเข้าใจที่ว่า สถิติคือตัวเลข เป็นความเข้าใจที่ผิด แต่สถิติเป็นการผสมอย่างกลมกลืนระหว่างคณิตศาสตร์ ตรรกศาสตร์ และวิจารณญาณ ที่ว่ายากนั้นไม่ใช่คณิตศาสตร์ แต่เป็นตรรกศาสตร์และวิจารณญาณ เป็นเรื่องการรู้จักเหตุรู้จักผล การใช้เหตุผลในการสรุป และการพิจารณาไตร่ตรอง แต่สิ่งเหล่านี้ไม่สามารถเกิดได้จากการอ่านเพียงอย่างเดียว ดังคำกล่าวที่ว่า "การอ่านตำราว่ายนน้ำเพียงอย่างเดียวไม่สามารถทำให้ผู้อ่านว่ายน้ำได้" ในทำนองเดียวกัน "จะเข้าใจสถิติต้องทำวิจัย" แต่คนส่วนมากหาโอกาสเช่นนั้นได้ยาก เฉพาะอย่างยิ่งในตอนเริ่มต้น นอกจากนั้นตำราและการจัดการเรียนการสอนโดยทั่วไปมักแยกสิ่งเหล่านี้เป็นส่วน ๆ จึงไม่เอื้อให้ผู้เรียนสัมฤทธิ์ผล แม้จะเรียนเรื่องเดิมหลายครั้ง จึงเป็นที่มาของคู่มือปฏิบัติการชีวสถิติ

คู่มือนี้ได้รับการออกแบบที่เอื้อให้ผู้เรียนเข้าใจหลักสถิติ เริ่มจากเรื่องง่าย ๆ ตัวเลขไม่มากจนเกินไป หลีกเลี่ยงการใช้สูตร พิสูจน์ให้เห็นว่าแนวคิดทางสถิติใช้ได้อย่างไร ทำให้แนวคิดสถิติเป็นสิ่งที่ผู้เรียนสัมผัสได้ คือใช้สามัญสำนึกตามได้ จัดประสบการณ์การเรียนรู้ที่เป็นบูรณาการ ใช้ตัวอย่างจากข้อมูลชุดเดียวที่ครอบคลุมวิธีการทางสถิติอย่างกว้างขวาง

ผู้เขียนขอขอบคุณอาจารย์ในภาควิชาชีวสถิติและประชากรศาสตร์ทุกท่านที่สนับสนุนและให้คำแนะนำที่มีคุณค่าในหลายด้าน ขอขอบคุณนักศึกษาคณะเภสัชศาสตร์ ปีที่ 4 ปีการศึกษา 2541 ซึ่งเป็นปีแรกที่ใช้คู่มือปฏิบัติการนี้ ที่อนุญาตให้ใช้ข้อมูลสำหรับคู่มือนี้ และศิษย์อีกหลาย ๆ รุ่นต่อมาที่สะท้อนข้อคิดเห็นจากการใช้ ซึ่งนำมาสู่การเปลี่ยนแปลงจนเป็นฉบับนี้

ผู้เขียนใคร่ขอกราบขอบพระคุณบรรพคณาจารย์ที่มีอายุกล่าวนามได้หมด ทั้งที่ได้พร่ำสอนในหลักสูตรทุกระดับที่ผู้เขียนได้ร่ำเรียนมา และท่านผู้แต่งตำราต่าง ๆ ที่ผู้เขียนใช้ประกอบการจัดทำคู่มือนี้ ความผิดพลาดที่อาจเกิดขึ้นได้ ถือเป็นความผิดพลาดของผู้เขียน และใคร่ขอน้อมรับด้วยความยินดี

บัณฑิต ถิ่นคำรพ

มิถุนายน 2543

สารบัญ

	หน้า
คำนำ	I
คำแนะนำการใช้คู่มือปฏิบัติการชีวสถิติ.....	III
บทที่ 1 ภาพรวมของวิชาชีวสถิติ.....	1
บทที่ 2 การเก็บรวบรวมข้อมูล	17
บทที่ 3 การตรวจสอบการแจกแจงของข้อมูล.....	23
ส่วนที่ 1 กรณีข้อมูลต่อเนื่อง.....	23
ส่วนที่ 2 กรณีข้อมูลแจกแจงนับ	28
บทที่ 4 สถิติเชิงพรรณนาเพื่อสรุปลักษณะตัวอย่าง	29
บทที่ 5 การอนุมานทางสถิติเพื่อตอบคำถามวิจัย.....	32
กิจกรรมที่ 5.1 : การแจกแจงค่าสถิติ และ Central Limit Theorem.....	32
กิจกรรมที่ 5.2 : การใช้ประโยชน์จากการแจกแจงแบบปกติ	41
กิจกรรมที่ 5.4 : การสรุปผลการวิเคราะห์ข้อมูลในทางปฏิบัติ.....	47
บทที่ 6 การประยุกต์แนวคิดสถิติในการวิจัย	48
กิจกรรมที่ 6.1 : สรุปการอนุมานทางสถิติ.....	48
กิจกรรมที่ 6.2 : ประยุกต์แนวคิดทางสถิติกับคำถามวิจัยต่าง ๆ	58
กิจกรรมที่ 6.3 : แนวทางการเลือกใช้สถิติกับคำถามวิจัยต่าง ๆ	63
บรรณานุกรม.....	64
ภาคผนวก 1 เฉลยคู่มือปฏิบัติการ	67
คำเฉลยบทที่ 1 ภาพรวมของวิชาชีวสถิติ.....	67
คำเฉลยบทที่ 2 การเก็บรวบรวมข้อมูล	72
คำเฉลยบทที่ 3 การตรวจสอบการแจกแจงของข้อมูล	76
คำเฉลยบทที่ 4 สถิติเชิงพรรณนาเพื่อสรุปลักษณะตัวอย่าง.....	80
คำเฉลยบทที่ 5 การอนุมานทางสถิติเพื่อตอบคำถามวิจัย	82
คำเฉลยบทที่ 6 การประยุกต์แนวคิดสถิติในการวิจัย	99
ภาคผนวก 2 การวิเคราะห์ข้อมูลตามที่ระบุในเฉลยคู่มือโดยใช้ STATA.....	114
ภาคผนวก 3 ข้อมูลการสำรวจนักศึกษา	144
ภาคผนวก 4 แบบสอบถามนักศึกษา	149

คำแนะนำการใช้คู่มือปฏิบัติการชีวสถิติ

ปรัชญาของการทำคู่มือนี้คือ “เมื่อผู้เรียนเข้าใจแนวคิดพื้นฐานของสถิติ และเกิดความประทับใจในประโยชน์ที่ได้จากการประยุกต์ใช้ จะทำให้ผู้เรียนสามารถทำความเข้าใจกับความรู้สถิติในระดับที่สูงขึ้นได้ และเกิดการเรียนรู้ที่ต่อเนื่องและยั่งยืน” ตัวอย่างที่แสดงให้เห็นถึงความสำเร็จในการใช้คู่มือนี้เช่น คำกล่าวที่สะท้อนออกมาหนึ่งปีหลังจบหลักสูตรว่า “การสอนของอาจารย์จบสิ้นไปนานแล้ว แต่การตบยังกำลังเรียนอยู่ และเห็นเพื่อนร่วมรุ่นหลายคนซื้อหาตำราสถิติมาอ่านมากมาย”

คู่มือนี้มีได้สมบูรณ์ในตัวเอง เนื่องจากได้ให้ความสำคัญการการผูกโยงเนื้อหาสถิติเข้ากับงานวิจัยให้เห็นเป็นภาพรวม จึงมีการกล่าวถึงวิธีการทางสถิติมากมายแต่ไม่ลงรายละเอียด ขอให้ผู้อ่านศึกษารายละเอียดวิธีการทางสถิติที่ตนสนใจเพิ่มเติมจากเอกสารอ้างอิงที่ให้ไว้เพื่อให้เกิดประโยชน์สูงสุด

ตลอดการฝึกปฏิบัติการตามคู่มือนี้ โปรแกรมสถิติที่ใช้คือ STATA ซึ่งเล็ก ง่าย เร็ว ราคาไม่แพงและครอบคลุมวิธีการทางสถิติที่กว้างขวางมาก อีกทั้งมีการเขียนโปรแกรมย่อยโดยกลุ่มนักสถิติจากทั่วทุกมุมโลก เพื่อคำนวณค่าสถิติที่ไม่สามารถหาได้ในโปรแกรมอื่น แล้วแจกจ่ายกันใช้ฟรีผ่านทางอินเทอร์เน็ต เป็นระยะ (ท่านสามารถสั่งซื้อโปรแกรมได้โดยตรงที่ <http://www.stata.com>)

เมื่อเสร็จสิ้นการใช้คู่มือนี้ ผู้เรียนควรมีความสามารถดังนี้

1. อธิบายหลักสถิติได้อย่างถูกต้อง
2. อธิบายบทบาทของสถิติในงานวิจัยทางวิทยาศาสตร์สุขภาพได้
3. อธิบายความหมาย บอกความสำคัญ และจำแนกความแตกต่างระหว่างสิ่งต่อไปนี้ได้
 - 4.1 ข้อมูล และตัวแปร
 - 4.2 ค่าสถิติ และค่าพารามิเตอร์
 - 4.3 สถิติเชิงพรรณนา และสถิติเชิงอนุมาน
 - 4.4 การประมาณค่าช่วงเชื่อมั่น และการทดสอบนัยสำคัญทางสถิติ
4. บอกแนวคิดของการเลือกใช้สถิติในการวิจัยได้
5. ใช้คอมพิวเตอร์วิเคราะห์ข้อมูลเบื้องต้นได้
6. บอกหลักการนำเสนอค่าสถิติจากผลที่ได้จากคอมพิวเตอร์ในการสรุปผลการวิจัยได้

ผู้เรียนควรพยายามทำแบบฝึกหัดก่อน แล้วเปรียบเทียบคำตอบที่ได้กับเฉลยที่ให้ไว้ในภาคผนวก การใช้คู่มือนี้ในการอบรมที่มีผู้เข้ารับการอบรมมากพอ เช่น 100 คนขึ้นไป ขอแนะนำว่าให้ใช้ข้อมูลจริงของกลุ่ม เพราะจะสนุกตื่นเต้นพร้อมกับได้เรียนรู้ด้วย แฟ้มข้อมูลที่ใช่แสดงเป็นตัวอย่างในเฉลยคู่มือสามารถดาวน์โหลดได้ที่อินเทอร์เน็ตที่โฮมเพจของผู้เขียน คือที่ <http://web.kku.ac.th/~bandit/data> นอกจากนี้ในนิตยสารยังมีแหล่งทรัพยากรอื่นๆ อีกมากมายที่เป็นประโยชน์ต่อการเรียนรู้สถิติ



ภาพรวมของวิชาชีวสถิติ (Overview of biostatistics)

1

1.1 องค์ความรู้ที่ได้จากการอ่านผลการวิจัย : อ่านแล้วรู้อย่างไร? (Pre test)

จากบทสรุปงานวิจัยหนึ่ง ความว่า

“การศึกษาเชิงทดลองทางการแพทย์โดยมีกลุ่มควบคุมเพื่อทราบว่ายาใหม่ขนานหนึ่งมีประสิทธิภาพในการรักษาโรคหรือไม่ พบว่ากลุ่มทดลองซึ่งได้รับยาขนานดังกล่าวมีอัตราการหายจากโรคสูงกว่ากลุ่มควบคุมที่ใช้ยาหลอก (placebo) อย่างมีนัยสำคัญทางสถิติ ($p\text{-value} < 0.05$)”

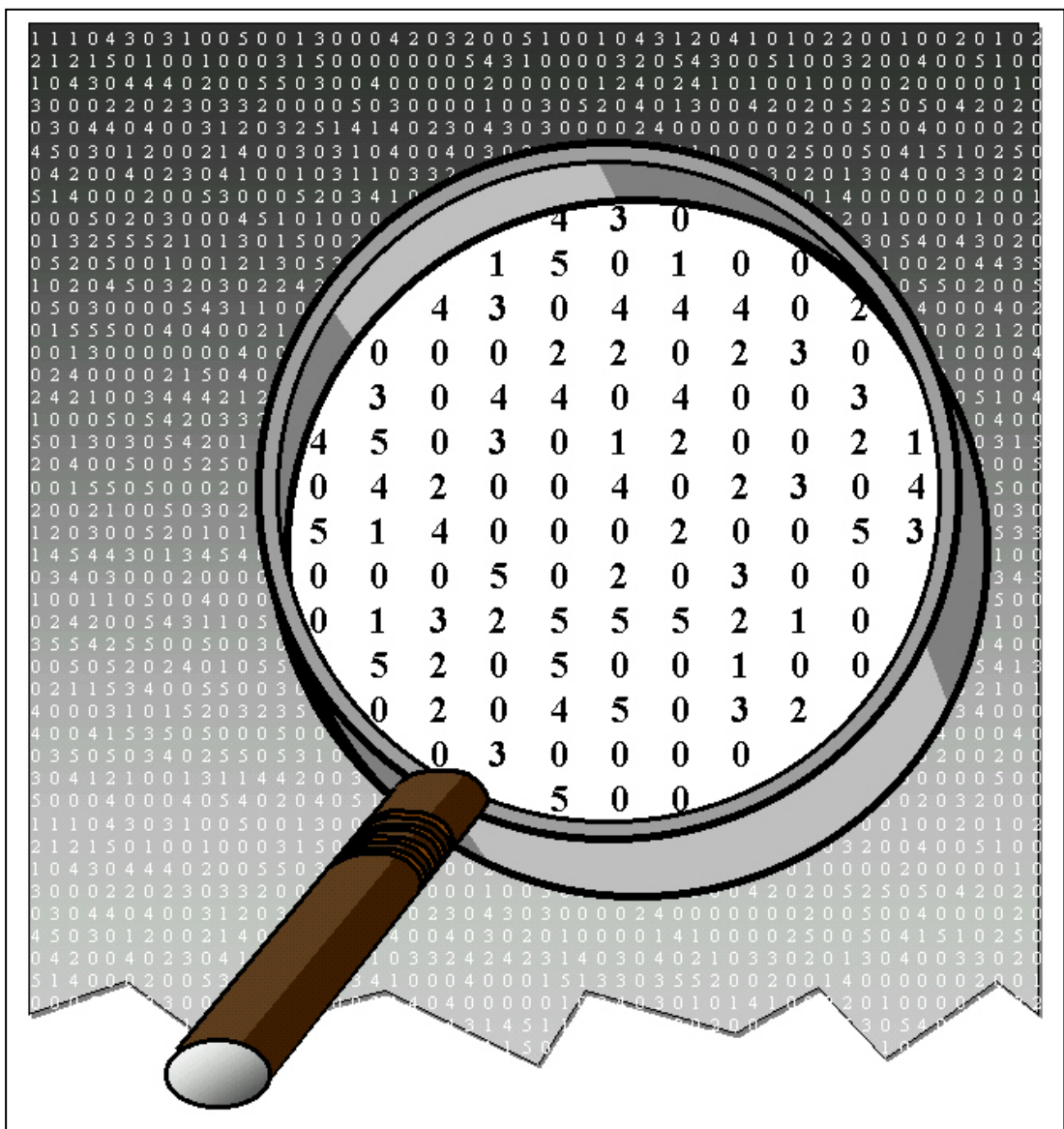
ท่านเห็นด้วยกับคำกล่าวใดต่อไปนี้

- [] 1. การศึกษานี้ได้พิสูจน์ให้เห็นแล้วว่ายาใหม่นี้ดีกว่ายาหลอก
- [] 2. ถ้าความเป็นจริงคือ “ยาใหม่ไม่มีประสิทธิภาพในการรักษาโรค” แล้ว โอกาสที่จะได้ผลการศึกษาตามที่กล่าวข้างต้นมีน้อยกว่า 5%
- [] 3. ประสิทธิภาพในการรักษาโรคของยาใหม่ที่ได้จากการศึกษาครั้งนี้สูงมาก กล่าวคือโอกาสที่จะพบว่ายานี้ไม่ดีไปกว่ายาหลอกมีเพียง 5 % เท่านั้น
- [] 4. ไม่รู้จริงๆ ว่า $p\text{-value}$ คืออะไร จึงไม่อยากจะเดาคำตอบ

ดัดแปลงจาก Wulff, H.R., Andersen, B., Brandenhoff, P., and Guttler, F. (1987) What do doctors know about statistics?. *Statistics in Medicine*, **6**, 3–10.

อภิปราย

- 1.2 ต่อไปนี้เป็นภาพจำลองของข้อมูลชุดหนึ่ง สมมติให้เป็นข้อมูลจำนวนบุหรืที่สูบ (มวนต่อวัน) ของคนงานทุกคนในทุกโรงงานอุตสาหกรรมในจังหวัดหนึ่ง ในกรอบพื้นที่ดำตัวหนังสือสีขาวสมมติเป็นชุดข้อมูลที่ยาวมาก ให้ถือเป็นประชากร (Population) ถ้ามีนักวิจัยคนหนึ่งทำการวิจัยเกี่ยวกับการสูบบุหรี่ในคนงานโรงงานอุตสาหกรรมในจังหวัดนี้ จะเรียกว่านี่คือประชากรเป้าหมาย (Target population) สมมติให้ข้อมูลเหล่านี้จัดเรียงโดยสุ่ม และให้ข้อมูลในภาพแว่นขยายคือที่เราที่อยู่จริงในมือ เช่นรายชื่อคนงานที่ทำงานในโรงงานอุตสาหกรรมที่จดทะเบียนกับอุตสาหกรรมจังหวัด หรือที่มีชื่อในบัญชีกองทุนเงินทดแทนประกันสังคม เป็นต้น เราเรียกบัญชีรายชื่อที่พร้อมจะให้สุ่มเลือกตัวอย่างนี้ว่ากรอบตัวอย่าง (Sampling frame) เราจะสุ่มตัวอย่างอย่างง่าย (Simple random sample) จากกรอบตัวอย่างนี้ จำนวน 10 ตัวอย่าง



1.3 เขียนข้อมูลที่สุ่มได้ลงในกรอบต่อไปนี้ (n = 10)

--

1.4 แสดงข้อมูลให้อยู่ในรูปตารางต่อไปนี้

เลขที่	จำนวนบุหรี่ที่สูบ (มวนต่อวัน)
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

- 1.5 ในการเก็บข้อมูลครั้งนี้ ถ้าเราได้สอบถามอายุและเพศด้วย ตารางข้อมูลควรมีลักษณะที่แสดงข้างล่างนี้ คราวนี้เราเขียนหัวตารางด้วยคำสั้นๆ เป็นภาษาอังกฤษ เช่น ID ย่อมาจาก Identification number คือเลขที่แบบสอบถาม CIGAR มาจาก Cigarette เพื่อแทนข้อความจำนวนบุหรี่ที่สูบ (มวนต่อวัน) นั่นเอง นอกจากนั้นที่เพิ่มมาคือ AGE และ SEX แทนอายุและเพศตามลำดับ ทั้งนี้เพื่อง่ายต่อการจัดการกับข้อมูลในขั้นต่อไป ซึ่งต้องใช้คอมพิวเตอร์ (มีความจำเป็นที่ต้องทำตามที่เราพูดเพราะโปรแกรมวิเคราะห์ข้อมูลทางสถิติไม่เข้าใจคำสั่งภาษาไทย ชื่อตัวแปรต้องเป็นภาษาอังกฤษ)

คำถาม :

หัวตารางเหล่านี้เรียกว่า ([] ข้อมูล / [] ตัวแปร).

ตัวเลขที่กรอกลงในตารางเรียกว่า ([] ข้อมูล / [] ตัวแปร).

ID	CIGAR	AGE	SEX
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

นี่คือลักษณะข้อมูลที่แสดงให้เห็นในจอคอมพิวเตอร์เมื่อเราสั่งให้แสดงรายการข้อมูล (List) แต่ละแถวของตารางคือข้อมูลที่ได้จากแบบสอบถาม 1 ชุด เรียกว่า 1 Record ในที่นี้มี 10 Records หรืออีกนัยหนึ่งคือ “การวิจัยนี้มีขนาดตัวอย่างเท่ากับ 10” เราเขียนแทนด้วย “ $n = 10$ ”

1.6 ข้อมูลแต่ละชุดจำเป็นต้องมีคำอธิบายที่บันทึกไว้อย่างเป็นระบบ เพื่อให้การประมวลผลข้อมูลมีประสิทธิภาพ ไม่สับสน เราเรียกคำอธิบายเหล่านี้ว่า Data dictionary ถือว่าเป็นสิ่งจำเป็นมาก

ชื่อตัวแปร	คำอธิบาย	รหัส
ID	เลขที่แบบสอบถาม	1 ถึง n
CIGAR	จำนวนบุหรี่ที่สูบต่อวัน	จำนวนมวนต่อวัน
AGE	อายุ	อายุเป็นปี
SEX	เพศ	1 = ชาย; 2 = หญิง

1.7 นอกจาก Data dictionary ตามที่กล่าวข้างต้นแล้ว การอธิบายลักษณะข้อมูลควรมีการแสดง ลักษณะข้อมูลตัวอย่างจำนวนหนึ่ง สิ่งนี้แสดงรูปแบบของแฟ้มข้อมูล เรียกว่า Data format หรือ Data layout ดังตัวอย่างต่อไปนี้ ซึ่งแสดงรูปแบบข้อมูลจาก Data dictionary ตามที่กล่าวข้างต้น

ID	CIGAR	AGE	SEX
1	0	40	1
2	5	23	1
3	0	35	2
--- ข้าม n - 4 ชุดข้อมูล ---			
n	0	42	2

1.8 ให้ออกแบบสอบถามที่ยังผลให้ได้ข้อมูลตาม Data layout และ Data dictionary ข้างต้น

1.9 การแสดงทั้ง Data dictionary และ Data layout ควบคู่กันมีความสำคัญทั้งต่อการวางแผน วิเคราะห์ข้อมูล และต่อการวิเคราะห์ข้อมูลโดยใช้คอมพิวเตอร์ ช่วยให้สามารถเลือกใช้คำสั่งได้ ถูกต้อง บางคำสั่งจำเป็นต้องมี Data structure ที่จำเพาะ นอกจากนั้นยังช่วยให้เข้าใจคำว่า "กลุ่มอิสระต่อกัน" และ "กลุ่มไม่อิสระต่อกัน" ที่กล่าวในแผนภูมิที่ 2 และ 3 ถัดจากนี้ไป เช่นจำนวนบุหรี่ที่สูบจำแนกตามเพศตามที่กล่าวข้างต้นเป็นสองกลุ่มอิสระต่อกัน จำนวนบุหรี่ที่สูบก่อนและหลัง โครงการรณรงค์เลิกสูบบุหรี่เป็นสองกลุ่มที่ไม่เป็นอิสระต่อกัน ประเด็นนี้มีความสำคัญและคนส่วนมากไม่เข้าใจ จึงแสดงให้เห็นดูเป็นตัวอย่าง ดังต่อไปนี้ (โปรดสังเกตความแตกต่างของ Data dictionary และ Data layout ระหว่างสองรูปแบบ จากแบบสอบถามเดียวกัน)

1.9.1 แบบสอบถาม

ID:.....	
แบบสำรวจการสูบบุหรี่ตามโครงการณรงค์เลิกบุหรี่ในโรงงานอุตสาหกรรม	
คำถามและคำตอบ	รหัส
1. ชื่อ	
2. อายุ ปี	AGE[][]
3. เพศ [] 1. ชาย [] 2. หญิง	SEX[]
4. ก่อนโครงการรณรงค์ฯ จำนวนบุหรี่ที่สูบต่อวันโดยเฉลี่ย มวน	CIGAR1[][]
5. หลังโครงการรณรงค์ฯ จำนวนบุหรี่ที่สูบต่อวันโดยเฉลี่ย มวน	CIGAR2[][]

1.9.2 Data dictionary และ Data layout ทั้งสองแบบต่อไปนี้

1.9.2.1 แบบกว้าง (Wide format)

Questionnaire		
ชื่อตัวแปร	คำอธิบาย	รหัส
ID	เลขที่แบบสอบถาม	1 ถึง n
AGE	อายุ	อายุเป็นปี
SEX	เพศ	1 = ชาย; 2 = หญิง
CIGAR1	จำนวนบุหรี่ที่สูบต่อวัน (ก่อนโครงการ)	จำนวนมวนต่อวัน
CIGAR2	จำนวนบุหรี่ที่สูบต่อวัน (หลังโครงการ)	จำนวนมวนต่อวัน

Data Layout				
ID	AGE	SEX	CIGAR1	CIGAR2
1	40	1	0	0
2	23	1	5	0
3	35	2	0	0
--- ข้าม n - 4 ชุดข้อมูล ---				
n	42	2	0	5

1.9.2.2 แบบยาว (Long format)

Questionnaire		
ชื่อตัวแปร	คำอธิบาย	รหัส
ID	เลขที่แบบสอบถาม	1 ถึง n
CIGAR	จำนวนบุหรี่ที่สูบต่อวัน	จำนวนมวนต่อวัน
VISIT	การเยี่ยมชมสภาพ	1 = ก่อนโครงการ 2 = หลังโครงการ
AGE	อายุ	อายุเป็นปี
SEX	เพศ	1 = ชาย; 2 = หญิง

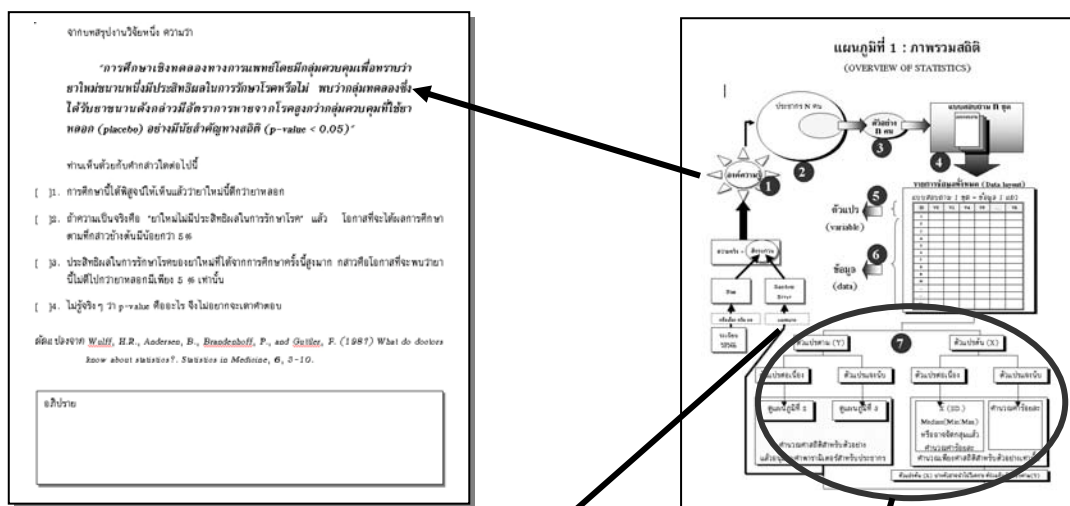
ให้ใช้ข้อมูลจาก 1.9.2.1 เติมลงในตารางข้อมูลต่อไปนี้

Data Layout				
ID	CIGAR	VISIT	AGE	SEX
1	-----	1	-----	-----
1	-----	2	-----	-----
2	-----	1	-----	-----
2	-----	2	-----	-----
3	-----	1	-----	-----
3	-----	2	-----	-----
--- ข้าม n - 4 ชุดข้อมูล ---				
n	0	1	42	2
n	5	2	42	2

1.10 แผนภูมิที่ 1 ต่อไปนี้ สรุปกิจกรรมที่ทำมาแล้วทั้งหมด เชื่อมโยงกับแนวคิดสถิติในภาพรวม

1.11 คำอธิบายส่วนที่สำคัญของแผนภูมิที่ 1

- 1.11.1 **1** คือองค์ความรู้ (Body of knowledge) เป็นเป้าหมายหลักของการวิจัย รูปธรรมขององค์ความรู้คือรายงานผลการวิจัย ตัวอย่างเช่น โจทย์ Pretest ในข้อ 1.1 ท่านพยายามแปลความหมายผลการวิจัย เพื่อจะตัดสินใจว่าจะเปลี่ยนจากยาเก่าไปใช้ยาใหม่หรือไม่ นั่นคือพอได้องค์ความรู้ จะนำไปสู่การปฏิบัติ ภาพข้างล่างแสดงบทบาทของสถิติต่อการได้มาซึ่งองค์ความรู้ และแสดง 2 ส่วนประกอบหลักของสถิติคือสถิติเชิงพรรณนา และการอนุมานทางสถิติ



การอนุมานทางสถิติ (Statistical inference):

อธิบายลักษณะของประชากร มี 2 องค์ประกอบ

(i) การประมาณค่า (Estimation)

รายงานผลเป็นค่าช่วงเชื่อมั่น (95% Confidence intervals)

(ii) การทดสอบสมมติฐาน (Hypothesis testing)

รายงานผลเป็นค่า $p\text{-value}$ (หรือมีนัยสำคัญ/ไม่มีนัยสำคัญ Significant / Non-significant)

สถิติเชิงพรรณนา

(Descriptive statistics):

อธิบายลักษณะของตัวอย่าง

คำถาม :

1.11.1.1 จากรายงานผลการวิจัยในข้อ 1.1 นั้น

สถิติที่รายงานมีอะไรบ้าง?

รายงานดังกล่าวขาดอะไร?

1.11.1.2 ถ้าหากท่านต้องตัดสินใจว่าจะใช้ยาใหม่หรือไม่ ข้อมูลเท่าที่รายงานในผลการวิจัยนั้น ท่านจะตัดสินใจอย่างไร?

1.11.2 **2** และ **3** เกี่ยวข้องกับประชากร (Population) และตัวอย่าง (Sample) ตามลำดับ

คำถาม : ในการสำรวจการสูบบุหรี่ตามที่กล่าวข้างต้น สมมติให้วัตถุประสงค์การวิจัยคือ “เพื่อประมาณค่าความชุกของการสูบบุหรี่ (ร้อยละของคนที่สูบบุหรี่) ของคนงานในโรงงานอุตสาหกรรมในจังหวัดหนึ่ง” กล่าวคือคนที่สูบอย่างน้อยหนึ่งมวนต่อวัน ให้นับเป็นผู้สูบบุหรี่ (ตัวแปร CIGAR ในที่นี่มีค่า 1 = สูบ และ 0 = ไม่สูบ)

1.11.2.1 ประชากรเป้าหมายคือ

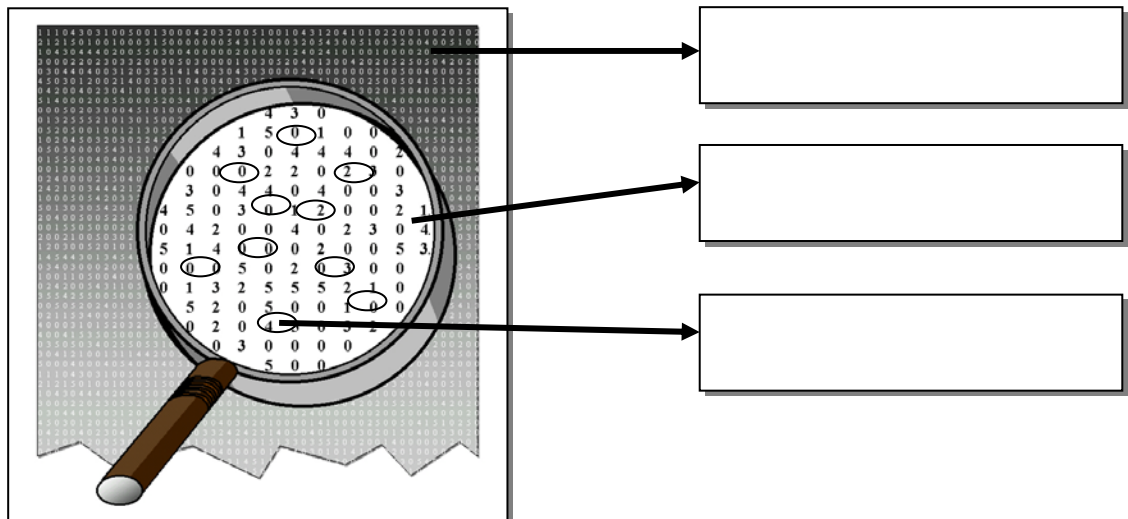
1.11.2.2 กรอบการสุ่มตัวอย่างคือ

1.11.2.3 ตัวอย่างคือ

1.11.2.4 ขนาดตัวอย่างเท่ากับ

1.11.2.5 วิธีการสุ่มตัวอย่างคือ

1.11.2.6 ให้กรอกคำอธิบายลงกรอบ ให้สอดคล้องกับตำแหน่งภาพจำลองประชากรและตัวอย่าง



1.11.3 **4** ถึง **6** เกี่ยวข้องกับการเก็บรวบรวมข้อมูล และลักษณะข้อมูลที่พร้อมนำไปวิเคราะห์ในขั้นตอนต่อไป

1.11.4 **7** เป็นภาพหยาบๆ ของแนวทางการเลือกใช้สถิติในการวิเคราะห์ข้อมูล

ขอใช้เรื่องการสำรวจการสูบบุหรี่เป็นเรื่อง ถ้าถามว่า "จะวิเคราะห์ข้อมูลชุดนี้อย่างไร?" เราต้องตอบคำถามให้ได้ก่อนว่า "วัตถุประสงค์ของการวิจัยคืออะไร?"

สมมติว่าเรามีข้อมูลสำหรับตัวแปร CIGAR AGE และ SEX ต่อไปนี้เป็นวัตถุประสงค์วิจัยแต่ละงานวิจัยที่เป็นไปได้ ตามด้วยชุดคำถามที่สามารถนำไปสู่การเลือกสถิติสำหรับวิเคราะห์ข้อมูลได้ โดย 5 คำถามแรก (ก. ถึง จ.) ใช้แผนภูมิที่ 1 ส่วนคำถามสุดท้าย ฉ. ให้พลิกไปดูแผนภูมิที่ 2 และ 3 (แสดงไว้ท้ายข้อ 1.11.4.5) ที่สอดคล้องกับหัวคำถามแรก

1.11.4.1 เพื่อประมาณค่าเฉลี่ยจำนวนบุหรี่ที่สูบต่อวัน (หมายเหตุ: การสูบบุหรี่ในที่นี้วัดเป็นจำนวนมวนที่สูบต่อวัน)

- ก. ตัวแปรตาม: (☐ CIGAR ☐ AGE ☐ SEX)
- ข. ประเภทตัวแปรตาม:..... (☐ Continuous ☐ Categorical)
- ค. ตัวแปรต้น: (☐ CIGAR ☐ AGE ☐ SEX ☐ ไม่มี)
- ง. ประเภทตัวแปรต้น:..... (☐ Continuous ☐ Categorical ☐ ไม่เกี่ยวข้อง)
- จ. แผนภูมิที่สอดคล้องกับงานวิจัยนี้คือแผนภูมิที่
- ฉ. โปรดระบุสถิติที่เหมาะสม

1.11.4.2 เพื่อหาขนาดความสัมพันธ์ระหว่างการสูบบุหรี่กับอายุ (หมายเหตุ: การสูบบุหรี่ในที่นี้วัดเป็นจำนวนมวนที่สูบต่อวัน)

- ก. ตัวแปรตาม: (☐ CIGAR ☐ AGE ☐ SEX)
- ข. ประเภทตัวแปรตาม:..... (☐ Continuous ☐ Categorical)
- ค. ตัวแปรต้น: (☐ CIGAR ☐ AGE ☐ SEX ☐ ไม่มี)
- ง. ประเภทตัวแปรต้น:..... (☐ Continuous ☐ Categorical ☐ ไม่เกี่ยวข้อง)
- จ. แผนภูมิที่สอดคล้องกับงานวิจัยนี้คือแผนภูมิที่
- ฉ. โปรดระบุสถิติที่เหมาะสม

1.11.4.3 เพื่อเปรียบเทียบความชุก (ร้อยละ) ของการสูบบุหรี่ระหว่างเพศชายและเพศหญิง (หมายเหตุ: การสูบบุหรี่ในที่นี้วัดเป็นสูบ กับไม่สูบ)

- ก. ตัวแปรตาม: (☐ CIGAR ☐ AGE ☐ SEX)
- ข. ประเภทตัวแปรตาม:..... (☐ Continuous ☐ Categorical)
- ค. ตัวแปรต้น: (☐ CIGAR ☐ AGE ☐ SEX ☐ ไม่มี)
- ง. ประเภทตัวแปรต้น:..... (☐ Continuous ☐ Categorical ☐ ไม่เกี่ยวข้อง)
- จ. แผนภูมิที่สอดคล้องกับงานวิจัยนี้คือแผนภูมิที่
- ฉ. โปรดระบุสถิติที่เหมาะสม

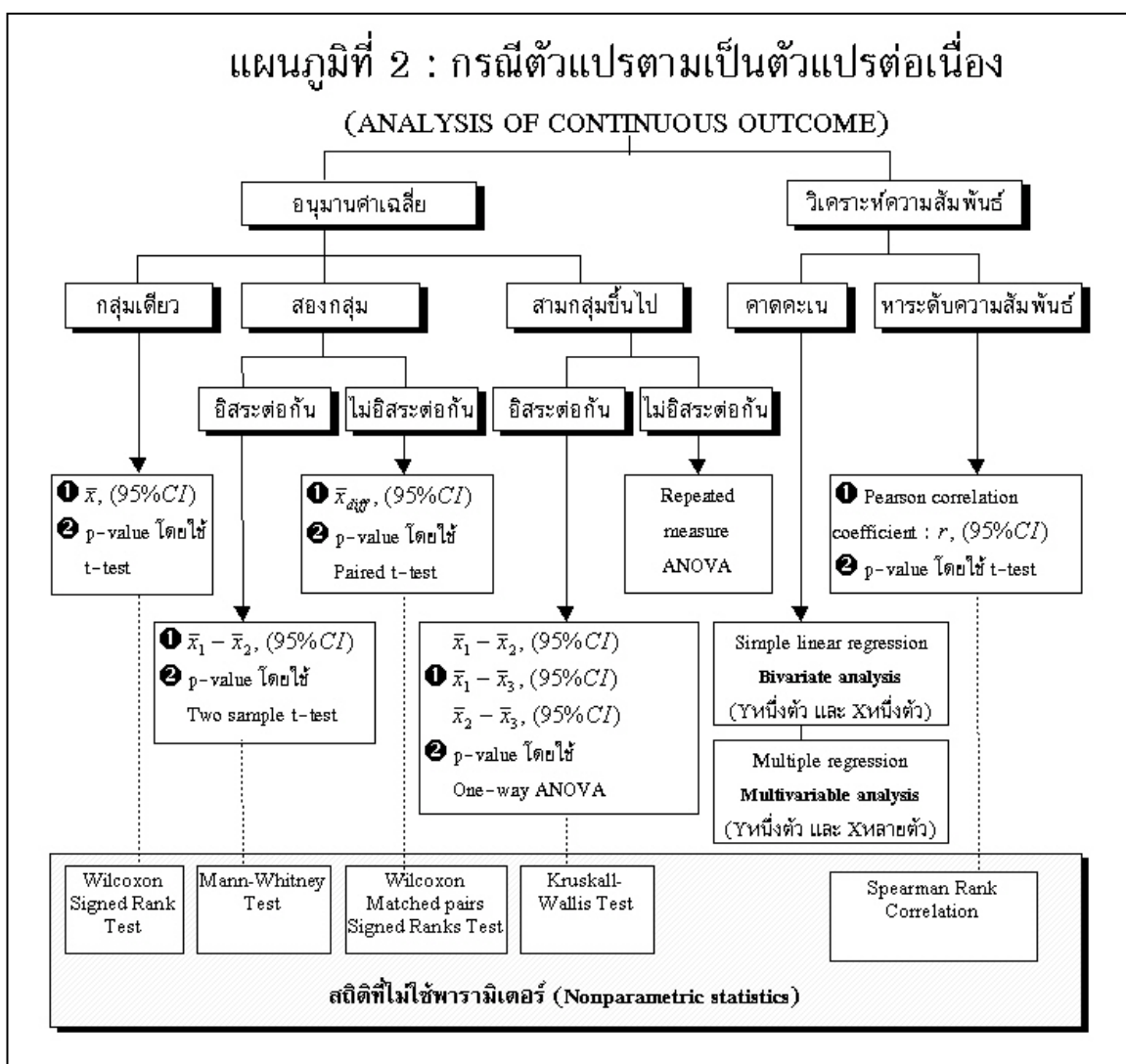
1.11.4.4 เพื่อหาปัจจัยที่มีผลต่อการสูบบุหรี่ (หมายเหตุ: การสูบบุหรี่ในที่นี้วัดเป็นสูบ กับไม่สูบ)

- ก. ตัวแปรตาม: (☐ CIGAR ☐ AGE ☐ SEX)
- ข. ประเภทตัวแปรตาม:..... (☐ Continuous ☐ Categorical)
- ค. ตัวแปรต้น: (☐ CIGAR ☐ AGE ☐ SEX ☐ ไม่มี)
- ง. ประเภทตัวแปรต้น:..... (☐ Continuous ☐ Categorical ☐ ไม่เกี่ยวข้อง)
- จ. แผนภูมิที่สอดคล้องกับงานวิจัยนี้คือแผนภูมิที่
- ฉ. โปรดระบุสถิติที่เหมาะสม

1.11.4.5 เพื่อหาปัจจัยที่มีผลต่อการสูบบุหรี่ (หมายเหตุ: การสูบบุหรี่ในที่นี้วัดเป็นจำนวนวันที่สูบบุหรี่ต่อวัน)

- ก. ตัวแปรตาม: (☐ CIGAR ☐ AGE ☐ SEX)
- ข. ประเภทตัวแปรตาม:..... (☐ Continuous ☐ Categorical)
- ค. ตัวแปรต้น: (☐ CIGAR ☐ AGE ☐ SEX ☐ ไม่มี)
- ง. ประเภทตัวแปรต้น:..... (☐ Continuous ☐ Categorical ☐ ไม่เกี่ยวข้อง)
- จ. แผนภูมิที่สอดคล้องกับงานวิจัยนี้คือแผนภูมิที่
- ฉ. โปรดระบุสถิติที่เหมาะสม

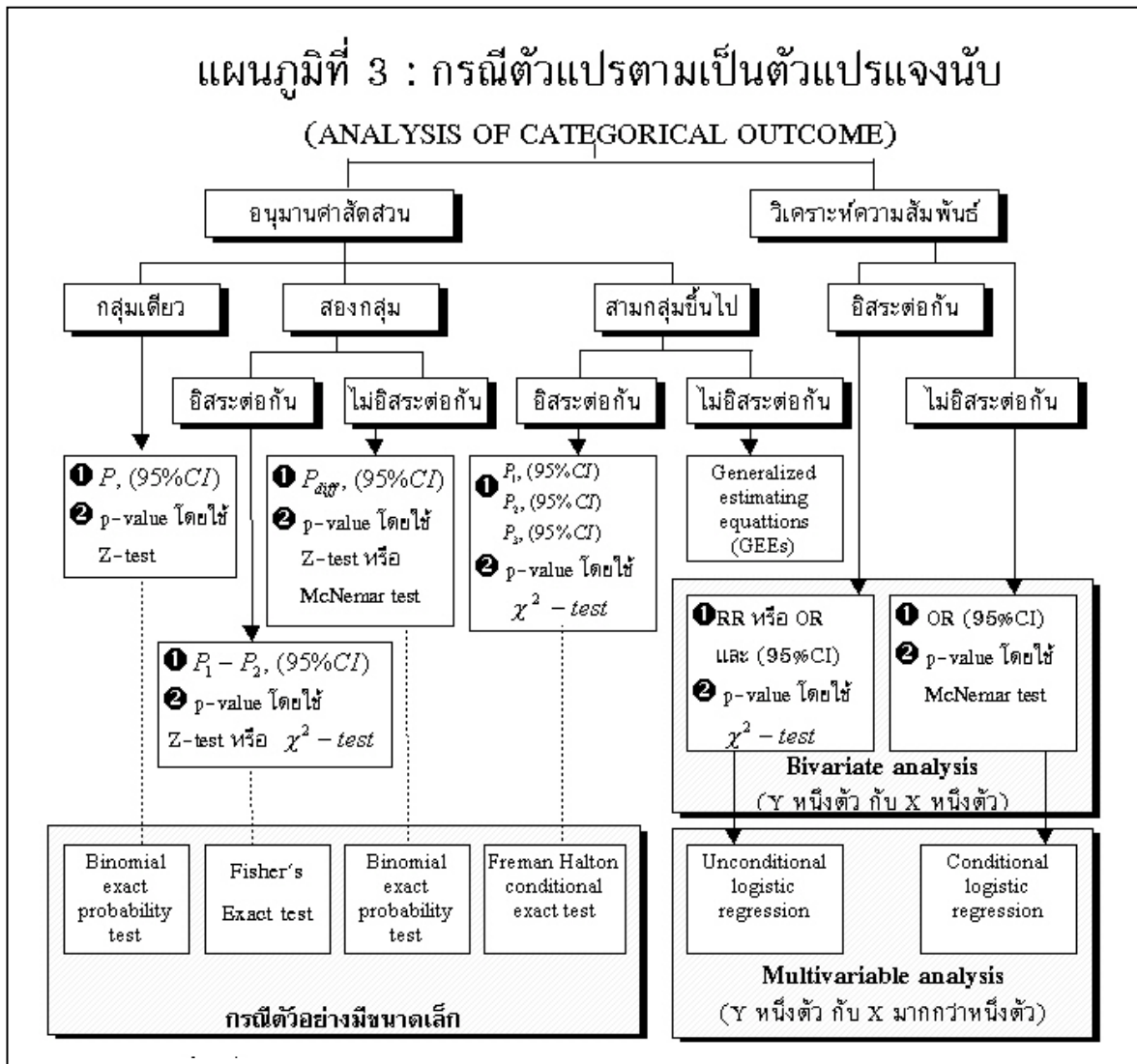
แผนภูมิที่ 2 : กรณีตัวแปรตามเป็นตัวแปรต่อเนื่อง



เอกสารอ่านประกอบ

1. อรุณ จิรวัฒน์กุล, มาลินี เหล่าไพบูลย์, จิราพร เขียวอยู่, ยุพา ถาวรพิทักษ์, จารุวรรณ โชคคณาพิทักษ์, บัณฑิต ถิ่นคำพร, นิคม ถนอมเสียง. (2542). *ชีวสถิติ*. ขอนแก่น. โรงพิมพ์คลังน่านวิทยา.
2. Everitte BS. (1995) The analysis of repeated measures: a practical reviews with examples. *The Statistician*, 44(1):113-35.
3. Wonnacott TH., Wonnacott RJ. (1990), *Introductory statistics*. 5th Edition. New York: John Willey & Sons .

แผนภูมิที่ 3 : กรณีตัวแปรตามเป็นตัวแปรแจกแจง



เอกสารอ่านประกอบ

1. อรุณ จิรวัฒนกุล, มาลินี เหล่าไพบูลย์, จิราพร เขียวอยู่, ชุพา ถาวรพิทักษ์, จารุวรรณ โชคคณาพิทักษ์, บัณฑิต ถิ่นคำพร, นิคม ถนอมเสียง. (2542). *ชีวสถิติ*. ขอนแก่น. โรงพิมพ์คลังน่านาวิทยา.
2. Altman DG. (1991) *Practical statistics for medical research*. London, Chapman and Hall.
3. Fleiss JL. (1981) *Statistical methods for rate and proportion*. Toronto: John Wiley & Sons.
4. Kleinbaum D.G. (1994) *Logistic Regression: A self-learning text*. New York, Springer-Verlag.
5. Neuhaus J. (1992) Statistical methods for longitudinal and clustered designs with binary responses. *Statistical Methods in Medical Research*. 1:249-37.
6. Rosner B. (1990) *Fundamental of biostatistics*. 3rd ed. Boston, Massachusetts: PWS-KENT Publishing Company.

1.12 สรุปภาพรวมสถิติ (ท่านที่เข้าใจสิ่งที่กล่าวข้างต้นดีแล้วสามารถข้ามไปบทที่ 2 ได้
โดยไม่ขาดความต่อเนื่อง)

กิจกรรมข้างต้น สามารถสรุปได้ตามตารางต่อไปนี้ โดยคอลัมน์ซ้ายมือสรุปเนื้อหา ส่วนคอลัมน์
ขวามือให้รายละเอียดเพิ่มเติม ผู้อ่านที่พื้นฐานดีแล้วจึงอาจไม่จำเป็นต้องอ่านคอลัมน์ขวามือ

เนื้อหาหลัก	คำอธิบาย หรือ ขยายความ
(1) สถิติ เป็นเครื่องมือที่สำคัญในกระบวนการแสวงหาความรู้	1) องค์ความรู้ คือเป้าหมายสูงสุด มิใช่ค่าสถิติ 2) การได้มาซึ่งองค์ความรู้มีกระบวนการแสวงหา หรือสร้างองค์ความรู้ที่เรียกว่า การวิจัย ซึ่งเกือบทุกขั้นตอนของการวิจัยนั้น ต้องใช้สถิติ แต่ส่วนมากเกี่ยวข้องกับการเก็บรวบรวม นำเสนอ วิเคราะห์ และแปลผลข้อมูล 3) ผลที่ได้จากกระบวนการดังกล่าวมีเพียงตัวเลขที่หาความหมายไม่ได้ หากแต่ต้องแปลความหมายได้ ซึ่งหมายถึงองค์ความรู้ที่จะนำไปใช้ประโยชน์ต่อไป
(2) องค์ความรู้มุ่งอธิบายลักษณะปรากฏการณ์ หรือเรื่องที่ศึกษา ในประชากร	องค์ความรู้ที่เป็นสากล มีประโยชน์ต่อมวลมนุษยชาติ อธิบายปรากฏการณ์ที่ศึกษาในกลุ่มที่มีขอบเขตกว้าง ใหญ่กว่ากลุ่มที่ศึกษา ในทางสถิติเรียกกลุ่มที่เราต้องการนำผลจากการศึกษาไปใช้อธิบายนี้ว่าประชากร
(3) เราสามารถได้องค์ความรู้เกี่ยวกับประชากรได้โดยอาศัยข้อมูลที่ได้จากกลุ่มที่เราสุ่มมาเพียงบางส่วนจากประชากร เรียกว่า ตัวอย่าง	การศึกษาทุกหน่วยของประชากร เป็นเรื่องที่เป็นไปได้ยาก เนื่องจากไม่ทราบขอบเขต หรือแม้ทราบขอบเขตก็ทำไม่ได้เพราะสิ้นเปลืองทรัพยากรมาก สถิติเข้ามาช่วยตรงที่ว่า แม้ทราบขอบเขต หรือมีทรัพยากรไม่จำกัด ก็ไม่จำเป็นต้องทำทุกหน่วยประชากร แต่ทำเฉพาะตัวอย่างที่สุ่มมาเพียงบางส่วนจากประชากรก็พอเพียง สถิติจึงกล่าววิธีการอธิบายลักษณะของประชากร โดยใช้ข้อมูลที่ได้จากตัวอย่าง
(4) การอธิบายเพียงลักษณะของกลุ่มตัวอย่าง เรียกว่า สถิติเชิงพรรณนา (Descriptive statistics) ค่าที่ใช้สรุปลักษณะกลุ่มตัวอย่าง เรียกว่า ค่าสถิติ (Statistic) ส่วนนี้มีความสำคัญยิ่งที่จะอธิบายสิ่งที่ได้มาจริง ๆ แต่	ค่าสถิติ เป็นค่าที่เราได้มาโดยตรงจากการศึกษา แต่ยังไม่ใช่เป้าหมายหลัก ในขณะที่ค่าพารามิเตอร์ซึ่งเป็นเป้าหมายหลักนั้น ไม่มีใครทราบ แต่สามารถนำค่าสถิติไปใช้ประมาณค่าพารามิเตอร์ได้ เมื่อค่าสถิติได้มาจากตัวอย่าง จึงเกี่ยวข้องกับการสุ่มตัวอย่าง เมื่อ

เนื้อหาหลัก	คำอธิบาย หรือ ขยายความ
<p>ไม่สมบูรณ์ เพราะเป้าหมายคือต้องอธิบายลักษณะประชากรโดยใช้ข้อมูลที่ได้ การอธิบายลักษณะของประชากร โดยใช้ข้อมูลที่ได้จากตัวอย่างนี้เรียกว่า การอนุมานทางสถิติ (Inferential statistics) ค่าที่ใช้สรุปลักษณะประชากร เรียกว่า ค่าพารามิเตอร์ (Parameter)</p>	<p>มีการสุ่ม จึงเกี่ยวข้องกับโอกาสที่ถูกสุ่ม ยังผลให้ค่าสถิติที่ได้นั้น แปรเปลี่ยนไปตามตัวอย่างที่สุ่มมาได้ ค่าที่ได้จากการศึกษาหลายครั้ง หรือหลายคนศึกษาเรื่องเดียวกัน จึงอาจไม่เท่ากันได้ เพียงเพราะ “โอกาส” หรือ “ความน่าจะเป็น” ที่สำคัญคือเราจะสรุปผลจากการศึกษาเพียงครั้งเดียว ดังนั้น การใช้ค่าสถิติไปอนุมานค่าประชากร จึงต้องประยุกต์ใช้ทฤษฎีความน่าจะเป็นเข้าช่วย</p>
<p>(5) การอนุมานทางสถิติ มี 2 องค์ประกอบคือ การประมาณค่า (Parameter estimation) และการทดสอบสมมติฐาน (Hypothesis testing)</p>	<p>ผลการวิจัยใดๆ ควรมีการประมาณค่าเสมอ ส่วนการทดสอบสมมติฐานนั้น อาจมี หรือไม่มีก็ได้ อย่างไรก็ตาม ให้ขึ้นอยู่กับวัตถุประสงค์ของการวิจัย</p>
<p>(6) ตัวอย่างผลการศึกษาวิจัย</p> <p>“จากการศึกษาน้ำหนักแรกเกิดของทารก 420 คนในพื้นที่ที่มีโครงการรณรงค์ทางโภชนาการ (พื้นที่ทดลอง) และ 450 คนในพื้นที่ปกติ (พื้นที่ควบคุม) พบว่าค่าเฉลี่ยน้ำหนักแรกเกิดในพื้นที่ทดลองเท่ากับ 3,800 กรัม (ส่วนเบี่ยงเบนมาตรฐาน = 150) และในพื้นที่ควบคุมเท่ากับ 3,100 กรัม (ส่วนเบี่ยงเบนมาตรฐาน = 140) ค่าเฉลี่ยน้ำหนักแรกเกิดในพื้นที่ทดลองสูงกว่าพื้นที่ควบคุม 700 กรัม (95% ช่วงเชื่อมั่น = 200 ถึง 1,200) ซึ่งมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($p\text{-value} = 0.007$)”</p>	<p>ภายในเครื่องหมายคำพูดทั้งหมดคือ องค์ความรู้ตัวอักษรเรียงทั้งหมดเป็นการพรรณนาทางสถิติ ส่วนตัวอักษรเรียงและทึบนั่น เป็นการอนุมานทางสถิติ ซึ่งประกอบด้วย การประมาณค่า (แสดงด้วยค่า 95% ช่วงเชื่อมั่น) และการทดสอบสมมติฐาน (แสดงด้วยค่า $p\text{-value}$) มักแปลความหมายว่า มีหรือไม่มีนัยสำคัญ (Significant) จึงมักเรียกกันว่า การทดสอบนัยสำคัญทางสถิติ</p>



การเก็บรวบรวมข้อมูล (Data collection)

2

จากนี้เป็นต้นไป โปรดสมมติตนเองเป็นนักวิจัย ทำการวิจัยในประชากรกลุ่มหนึ่งซึ่งเป็นนักศึกษาเภสัชศาสตร์ชั้นปีที่ 4 มหาวิทยาลัยขอนแก่น ปี พ.ศ. 2541 มีจำนวนทั้งสิ้น 143 คน ข้อมูลที่ใช้ต่อไปนี้เป็นข้อมูลจริงทั้งหมด ยกเว้นข้อมูลเกี่ยวกับน้ำหนักตัวซึ่งมีการแก้ไขเล็กน้อยเพื่อให้ได้สถานการณ์การเรียนรู้สมบูรณ์ขึ้น

2.1 สร้างกรอบตัวอย่าง (Sampling frame)

ในทางปฏิบัติ ใช้บัญชีรายชื่อนักศึกษา แล้วเขียนหมายเลขลงหน้าชื่อ โดยให้หมายเลข 1 กับนักศึกษาชื่อแรก หมายเลข 2 สำหรับคนมีสอง เรื่อยไปจนครบทุกคน แต่ในปฏิบัติการนี้ ให้ใช้หมายเลขที่อยู่ภายใต้คอลัมน์ "เลขที่" เท่านั้น (ขอสงวนรายชื่อไว้เป็นความลับ)

2.2 สุ่มตัวอย่าง (Sampling)

ขั้นที่ 1 : เลือกนักศึกษาที่จะทำการศึกษา จำนวน 30 คน โดยการสุ่มตัวอย่างอย่างง่าย (Simple random sampling) และใช้ตารางเลขสุ่ม ต่อไปนี้คือตารางเลขสุ่มที่ให้สำหรับใช้เลือกนักศึกษาตัวอย่าง

คำแนะนำ : ให้สุ่มเลือกตัวเลขเริ่มต้นก่อน จากนั้นเขียนวงกลมล้อมรอบเลขที่สุ่มได้เพียงตัวเดียวคือ จากนั้นนับรวมเอาตัวเลขที่อยู่ด้านขวาอีก ตัว เพื่อให้ได้เป็นตัวเลข หลัก แล้วทำเครื่องหมายค้นไว้เป็นช่วงๆ ติดต่อกันไปในทิศทางจากขวาไปซ้าย และจากบนลงล่าง (ถ้าสุดหน้ากระดาษ ให้เริ่มต้นบรรทัดบนสุด) ได้ตัวเลขชุดใดที่มีค่าอยู่ระหว่าง 1 ถึง 143 ให้บันทึกไว้ในตารางที่ 2 ถ้าซ้ำให้ข้าม จนได้ครบ 30 หมายเลข จึงถือว่าเสร็จสิ้นการสุ่มตัวอย่าง

ตารางที่ 1 ตารางเลขสุ่ม

11164	36318	75061	37674	26320	75100	10431	20418	19228	91792
21215	91791	76831	58678	87054	31687	93205	43685	19732	08468
10438	44482	66558	37649	08882	90870	12462	41810	01806	02977
36792	26236	33266	66583	60881	97395	20461	36742	02852	50564
73944	04773	12032	51414	82384	38370	00249	80709	72605	67497
04563	12872	14063	93104	78483	72717	68714	18078	25005	04151
64208	48237	41701	73117	33242	42314	83049	21933	92813	04763
51486	72875	38605	29341	80749	80151	33835	52602	79147	08868
99756	26360	64516	17971	48478	09610	04638	17141	09227	10606
71325	55217	13015	79207	00431	45117	33827	92873	02953	85474
65285	97198	12138	53010	94601	15838	16805	61004	43516	17020
17264	57327	38224	29301	31381	38109	34976	65692	98566	29550
95639	99754	31199	92558	68368	04985	51092	37780	40261	14479
61555	76404	86210	11808	12841	45147	97438	60022	12645	62000
78137	98768	04689	87130	79225	08153	84967	64539	79493	74917
62490	99215	84987	28759	19177	14733	24550	28067	68894	38490
24216	63444	21283	07044	92729	37284	13211	37485	10415	36457
16975	95428	33226	55903	31605	43817	22250	03918	46999	98501
59138	39542	71168	57609	91510	77904	74244	50940	31553	62562
29478	59652	50414	31966	87912	87154	12944	49862	96566	48825
96155	95009	27429	72918	08457	78134	48407	26061	58754	05326
29621	66583	62966	12468	20245	14015	04014	35713	03980	03024
12639	75291	71020	17265	41598	64074	64629	63293	53307	48766
14544	37134	54714	02401	63228	26831	19386	15457	17999	18306
83403	88827	09834	11333	68431	31706	26652	04711	34593	22561
10011	75004	86054	41190	10061	19660	03500	68412	57812	57929
92420	65431	16530	05547	10683	88102	30176	84750	10115	69220
35542	55865	07304	47010	43233	57022	52161	82976	47981	46588
86595	26247	18552	29491	33712	32285	64844	69395	41387	87195
72115	34985	58036	99137	47482	06204	24138	24272	16196	04393
40603	16152	83235	37361	98783	24838	39793	80954	76865	32713
40941	53585	69958	60916	71018	90561	84505	53980	64735	85140
73505	83472	55953	17957	11446	22618	34771	25777	27064	13526
39412	16013	11442	89320	11307	49396	39805	12249	57656	88686
57994	76748	54627	48511	78646	33287	35524	54522	08795	56273

ขั้นที่ 2 : นำหมายเลขที่ได้จากขั้นที่ 1 เขียนลงคอลัมน์แรกสุด (เลขที่) ในตารางข้างล่างนี้
ตารางข้อมูลนักศึกษาที่ถูกสุ่มเป็นตัวอย่าง 30 คน

เลขที่	ID	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
	1													
	2													
	3													
	4													
	5													
	6													
	7													
	8													
	9													
	10													
	11													
	12													
	13													
	14													
	15													
	16													
	17													
	18													
	19													
	20													
	21													
	22													
	23													
	24													
	25													
	26													
	27													
	28													
	29													
	30													

ขั้นที่ 3 เก็บรวบรวมข้อมูล (Data collection)

ในเดือนมิถุนายน 2541 ได้ใช้แบบสอบถามเป็นเครื่องมือในการเก็บรวบรวมข้อมูล แบบสอบถามที่ใช้ให้ชื่อว่า "แบบสอบถามนักศึกษา" ในภาคผนวก 4 นักศึกษาทุกคนได้รับการดูแลให้กรอกข้อมูลให้ตรงตามความจริงที่สุด ใช้เครื่องชั่งน้ำหนักและวัดความสูงสำหรับกรอกข้อมูลน้ำหนักและความสูง และกรณีการเก็บข้อมูลอัตราการเต้นชีพจรก่อนและหลังวิ่งนั้น ให้วิ่งกับที่จริงๆ และวิ่งให้

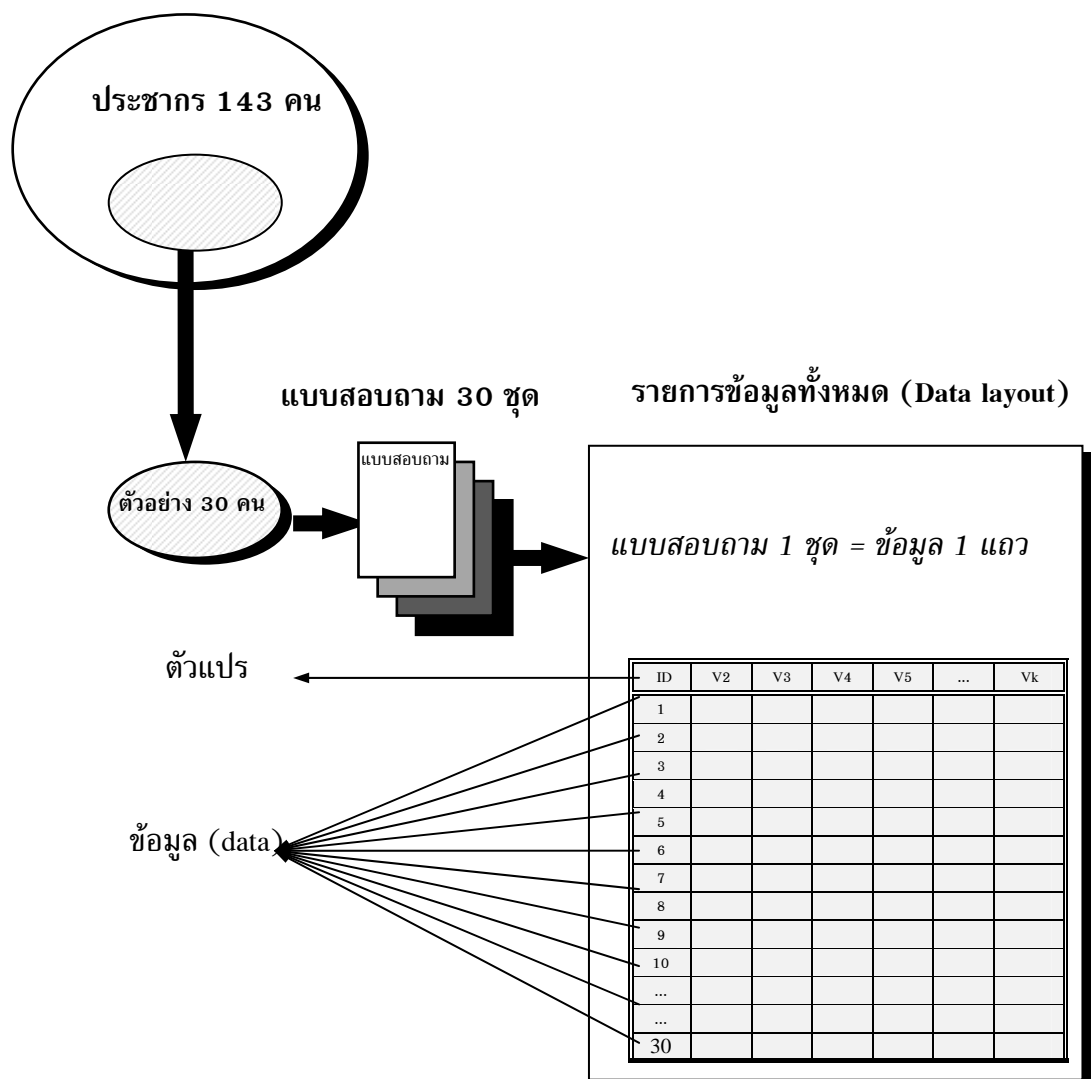
เร็วที่สุดเท่าที่ตนเองทำได้ ข้อมูลทั้งหมดได้รับการป้อนเข้าคอมพิวเตอร์และพิมพ์ออกมาตาม
รายละเอียดที่แนบในภาคผนวกที่ 3

การปฏิบัติ :

- (1) เพื่อให้คุ้นเคยกับแบบสอบถามที่ใช้ตลอดปฏิบัติการนี้ จึงแนะนำให้กรอกข้อมูลของตัวเองลงในแบบสอบถามท้ายเล่ม
- (2) นำข้อมูลจากแบบสอบถามไปกรอกลงในแถวแรกของตารางข้อมูลในภาคผนวก 3 (ที่ ID = 0)
- (3) นำข้อมูลจากภาคผนวก 3 เฉพาะรายชื่อที่ถูกสุ่มเป็นตัวอย่าง บันทึกลงในตารางที่ 2 ที่แสดงไว้หน้าที่ผ่านมา

หมายเหตุ : การเก็บข้อมูลครั้งนี้ ให้ชื่อว่า "การสำรวจนักศึกษา" และจะใช้ชื่อนี้อ้างอิงตลอดการปฏิบัติการ

ต่อไปนี้เป็นแผนภูมิสรุปกิจกรรมที่ท่านได้ทำแล้ว



2.3 คำถาม

2.3.1 การสุ่มตัวอย่างตามที่นักศึกษาได้ทำไปแล้วในข้อ 2.1.2 นั้น จัดเป็นประเภทใด

- [] 1. Simple random sampling
- [] 2. Systematic random sampling
- [] 3. Cluster random sampling
- [] 4. Stratified random sampling

2.3.2 ตัวเลือกใดข้างต้นที่ไม่ถูกเลือก มีวิธีการปฏิบัติอย่างไร จงอธิบายโดยยกตัวอย่างสถานการณ์ข้างต้น

2.3.3 จงระบุสเกลการวัดของตัวแปรจากแบบสอบถามที่ใช้

คำถาม	สเกลการวัด
2. เพศ (1= ชาย, 2= หญิง)	
3. อายุ.....ปี	
4. ความสูง.....เซนติเมตร	
5. น้ำหนัก.....กิโลกรัม	
6. เกรดเฉลี่ยสะสมเมื่อจบการศึกษาระดับมัธยมปลาย.....	
7. เกรดเฉลี่ยสะสมจนถึงเทอมที่ผ่านมา.....	
8. ท่านสอบเข้ามหาวิทยาลัยโดยระบบใด (1= โควต้า, 2 = เอนทรานซ์)	
9. ภูมิลำเนาอยู่ภาคใด (1= เหนือ, 2= ตะวันออกเฉียงเหนือ, 3= กลาง, 4= ใต้)	
10. อัตราการเต้นของชีพจรก่อนวิ่ง ครั้ง ต่อ นาที	
11. อัตราการเต้นของชีพจรหลังวิ่ง ครั้ง ต่อ นาที	
12. สุขภาพจิตของท่านก่อนเข้าเรียนมหาวิทยาลัย (1 = มีความเครียด, 2 = ไม่มีความเครียด)	
13. สุขภาพจิตของท่านหลังเข้าเรียนมหาวิทยาลัยแล้ว (1 = มีความเครียด, 2 = ไม่มีความเครียด)	
14. สถานสุขภาพ (1= เยี่ยม, 2= ดี, 3= ปานกลาง, 4= แย่)	

2.3.4 จงระบุประเภทตัวแปรจากแบบสอบถามที่ใช้ โดยเขียนหมายเลขลงช่อง "ประเภทตัวแปร"

เขียนหมายเลข 1 ถ้าเป็นตัวแปรแจกแจง (Categorical variable)

เขียนหมายเลข 2 ถ้าเป็นตัวแปรต่อเนื่อง (Continuous variable)

คำถาม	ประเภท
2. เพศ (1= ชาย, 2= หญิง)	
3. อายุ.....ปี	
4. ความสูง.....เซนติเมตร	
5. น้ำหนัก.....กิโลกรัม	
6. เกรดเฉลี่ยสะสมเมื่อจบการศึกษาระดับมัธยมปลาย.....	
7. เกรดเฉลี่ยสะสมจนถึงเทอมที่ผ่านมา.....	
8. ท่านสอบเข้ามหาวิทยาลัยโดยระบบใด (1= โควต้า, 2 = เอนทรานซ์)	
9. ภูมิลำเนาอยู่ภาคใด (1= เหนือ, 2= ตะวันออกเฉียงเหนือ, 3= กลาง, 4= ใต้)	
10. อัตราการเต้นของชีพจรก่อนวิ่ง ครั้ง ต่อ นาที	
11. อัตราการเต้นของชีพจรหลังวิ่ง ครั้ง ต่อ นาที	
12. สุขภาพจิตของท่านก่อนเข้าเรียนมหาวิทยาลัย (1 = มีความเครียด, 2 = ไม่มีความเครียด)	
13. สุขภาพจิตของท่านหลังเข้าเรียนมหาวิทยาลัยแล้ว (1 = มีความเครียด, 2 = ไม่มีความเครียด)	
14. สถานสุขภาพ (1= เยี่ยม, 2=ดี, 3=ปานกลาง, 4=แย่)	

2.3.5 ตัวแปรตามข้อ 3 ในแบบสอบถามคือ อายุ นั้น สามารถสร้างคำถามรูปแบบใหม่ตามที่ปรากฏข้างล่างนี้ จงระบุประเภทตัวแปรดังกล่าว

คำถาม	ประเภท
3. อายุ []1. น้อยกว่า 5 ปี []2. 5 ถึง 25 ปี []3. 26 - 54 ปี []4. 55 ปีขึ้นไป	



การตรวจสอบการแจกแจงของข้อมูล (Examining distribution of the data)

3

ส่วนที่ 1 กรณีข้อมูลต่อเนื่อง

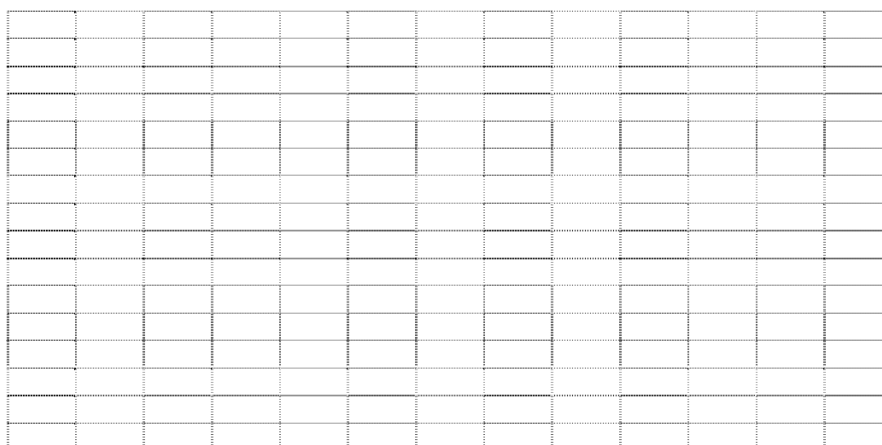
3.1 จงเขียนข้อมูลดิบค่าน้ำหนักทั้ง 30 ค่า ที่ได้จากการสำรวจโดยตัวท่านเอง

--

3.2 จงสร้างตารางแจกแจงความถี่ (Frequency table) ของน้ำหนักนักศึกษา โดยใช้ข้อมูลจากการสำรวจ

น้ำหนัก	ความถี่	ความถี่สัมพัทธ์	ความถี่สะสมสัมพัทธ์
รวม			

3.3 จงใช้ข้อมูลจากข้อ 3.2 สร้างฮิสโตแกรม (Histogram) พร้อมกับสร้างโค้งความถี่ (Frequency curve) ซ้อนทับ Histogram นั้นด้วย



3.4 จากข้อ 3.3 จงอธิบายลักษณะการกระจายของข้อมูลน้ำหนักนักศึกษา

[] สมมาตร (symmetry)

[] เบ้ซ้าย (left skew)

[] เบ้ขวา (right skew)

[] อื่นๆ ระบุ

3.5 ถ้าคู่มือนี้ใช้ในห้องเรียนที่มีกลุ่มผู้เรียนหลายคน วิทยากรจะถามผู้รับการอบรมทุกคนว่า “ข้อมูลน้ำหนัก มีการแจกแจงแบบใด” ผู้รับการอบรมยกมือเมื่อตรงกับตัวเลือกในตารางที่วิทยากรอ่านให้ฟัง วิทยากรนับเสร็จสิ้นแล้วจะแจ้งทุกคนในห้องให้บันทึกค่าลงตารางข้างล่างนี้ (กรณีท่านอ่านคู่มือนี้และฝึกทำเองคนเดียว ท่านสามารถจำลองเหตุการณ์ได้โดยการสุ่มด้วยคอมพิวเตอร์ตามที่แสดงในภาคผนวก 2 โดยสั่งคำสั่ง gr v5 หลังจากสั่งสุ่มตัวอย่างแต่ละครั้ง อย่างไรก็ตามท่านสามารถข้ามการทดลองนี้ได้โดยไม่สูญเสียความต่อเนื่องของเนื้อหา)

การแจกแจง	จำนวน	ร้อยละ
สมมาตร		
เบ้ซ้าย		
เบ้ขวา		

สรุป:

3.6 คำกล่าวต่อไปนี้ สมมติขึ้นเพื่อการศึกษาเท่านั้น ไม่เกี่ยวข้องกับข้อมูลการสำรวจนักศึกษา และอาจไม่เป็นจริงตามนั้น

“จากการศึกษารายได้ต่อหัวต่อปีพบว่า ประชาชนไทย 80% เป็นผู้ที่มีรายได้ต่ำ (มีรายได้ต่อหัวต่อปีไม่เกิน 20000 บาท) พบผู้มีรายได้ต่ำสุดเท่ากับ 500 บาทต่อปี และสูงสุดหลายพันล้านบาท”

3.6.1 โค้งความถี่รายได้คนไทยตามคำกล่าวข้างต้น มีลักษณะแบบใด

จำนวนประชากร



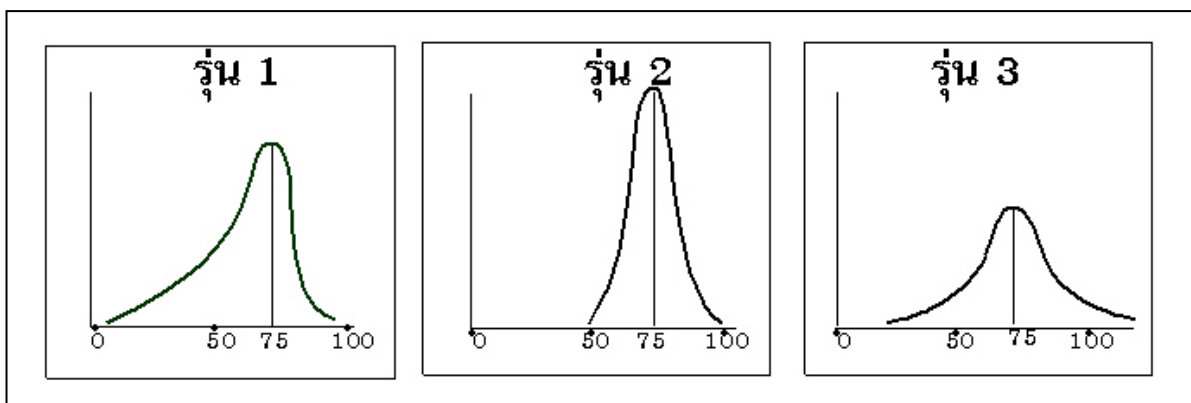
รายได้ต่อหัวต่อปี

3.6.2 ข้อสังเกต

ก. พื้นที่ใต้โค้งความถี่ 80% ของทั้งหมด คลุมค่าข้อมูลรายได้ตั้งแต่ 500 จนถึง 20,000 บาท
(นี่คือรูปลักษณะของ "การประมาณค่า" ที่จะกล่าวในรายละเอียดต่อไป)

ข. โอกาสที่จะพบกับผู้มีรายได้ตั้งแต่ 20,000 บาทขึ้นไป เท่ากับ
(นี่คือรูปลักษณะของ "การทดสอบสมมติฐาน" ที่จะกล่าวในรายละเอียดต่อไป)

3.7 จากโค้งความถี่แสดงค่าคะแนนผลการสอบวิชาชีวสถิติ ของนักศึกษา 3 รุ่น มีลักษณะต่อไปนี้



3.7.1 แกนตั้งคือ แกนนอนคือ

3.7.2 ค่าเฉลี่ย (Mean) คะแนนของรุ่นที่ 1 = รุ่นที่ 2 = รุ่นที่ 3 =

3.7.3 จงอธิบายลักษณะการแจกแจงข้อมูลตามความหมายของคะแนนจากการสอบ

รุ่นที่ 1.

รุ่นที่ 2.

รุ่นที่ 3.

3.7.4 จงเปรียบเทียบการแจกแจงข้อมูลคะแนนรุ่น 2 กับ รุ่น 3

.....

3.7.5 จงเขียนค่าสถิติที่จะใช้สรุปค่าคะแนนดังกล่าว ([] ค่าเฉลี่ย หรือ [] ค่ามัธยฐาน)

รุ่นที่ 1.

รุ่นที่ 2.

รุ่นที่ 3.

3.8 จงสรุปความสำคัญของลักษณะการแจกแจงของข้อมูลต่อการวิเคราะห์ข้อมูล

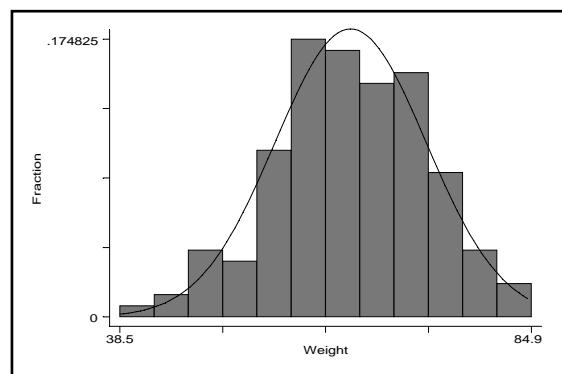
3.9 การตรวจสอบการแจกแจงแบบปกติโดยใช้ STATA เพื่อตรวจสอบข้อมูลน้ำหนัก (v5) (นี่คือโลกความเป็นจริง คือเราไม่นั่งทำด้วยมือตามที่กล่าวข้างต้น แต่ใช้คอมพิวเตอร์ทำให้)

3.9.1 ดูลักษณะการกระจายจากฮิสโตแกรม

คำสั่งคือ

```
. use ps143.dta
```

```
. gr v5, bin(12) norm
```

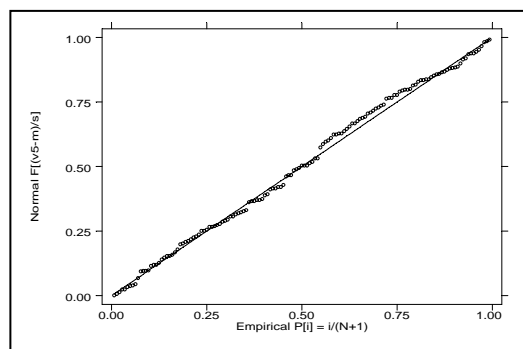


หมายเหตุ : ข้อมูลที่ใช้วิเคราะห์นี้ คือ $N = 143$ ราย ทั้งนี้เพื่อจะได้ใช้ประโยชน์ตามข้อ 5.4.1 ต่อไป

3.9.2 ดูว่าข้อมูลมีการแจกแจงแบบปกติหรือไม่จาก Standardized normal probability (P-P) plot (ดูรายละเอียดใน StataCorp, 1999)

คำสั่งคือ

```
pnorm v5
```



ข้อมูลกระจาย
ใกล้เคียงกับเส้น
ทะแยงแสดงว่า
ข้อมูลแจกแจงแบบ
ปกติ

3.9.3 ทดสอบทางสถิติว่าข้อมูลมีการแจกแจงแบบปกติหรือไม่จาก Shapiro-Wilk W test

คำสั่งคือ

swilk v5

Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Pr > z
v5	143	0.99317	0.763	-0.611	0.72926

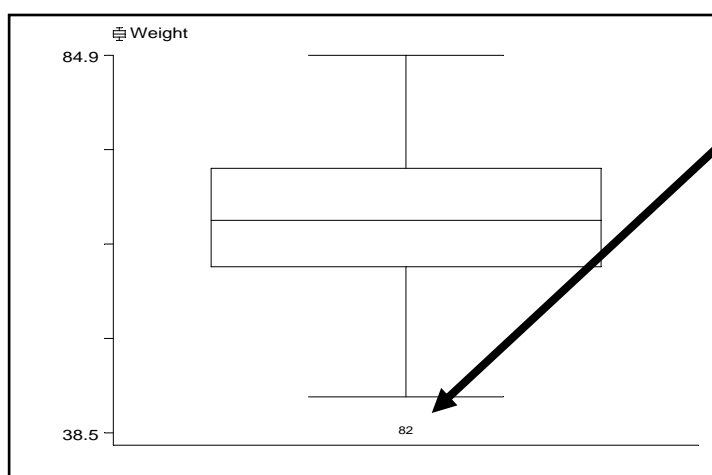
ค่า p-value > 0.05 สรุปว่าข้อมูลมีการแจกแจงแบบปกติ

3.10 การตรวจหา Outliers --> ข้อมูลน้ำหนัก (v5)

นำเสนอโดย Box-and-Whisker plot

คำสั่งคือ

graph v5, box symbol([id])



ข้อมูลนักศึกษาที่มี ID = 82
มีค่าต่ำกว่าปกติ หรือเป็น
Outlier หากต้องการทราบ
ข้อมูล สั่งโดย

list v5 if id == 82

พบว่าเท่ากับ 38.5 กก.

ส่วนที่ 2 กรณีข้อมูลแจกนับ

ตรวจสอบค่านอกช่วง เช่นเพศต้องมีได้เพียง 2 ค่า ใช้คำสั่งต่อไปนี้


คำสั่งคือ

tab v2

Sex	Freq.	Percent	Cum.
1	30	20.98	20.98
2	113	79.02	100.00
Total	143	100.00	

3.11 ตรวจสอบข้อมูล ระบบการสอบเข้ามหาวิทยาลัย (v8)

คำสั่งคือ



ผล

v8	Freq.	Percent	Cum.
1	48	33.57	33.57
2	47	32.87	66.43
3	48	33.57	100.00
Total	143	100.00	

ข้อสังเกตและเสนอแนะ :



สถิติเชิงพรรณนาเพื่อสรุปลักษณะตัวอย่าง (Descriptive statistics for describing the study samples)

4

กิจกรรมนี้ มักประกอบด้วยหลักการ (Principle) และลีลา (Style) จึงจำแนกกล่าวเป็น 2 ตอนดังนี้

ตอนที่ 1 หลักการ (ค่าสถิติใดบ้างที่สามารถคำนวณได้จากตัวแปรที่ระบุ)

4.1 จงกา ✓ ลงช่องที่สอดคล้องกับค่าสถิติที่เหมาะสมสำหรับใช้สรุปลักษณะของข้อมูลแต่ละตัวแปร

ค่าสถิติ	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
1. มาตรวัดแนวโน้มส่วนกลาง													
ค่าเฉลี่ย													
ค่ามัธยฐาน													
ค่าฐานนิยม													
2. มาตรวัดการกระจาย													
ค่าความแปรปรวน													
ค่าส่วนเบี่ยงเบนมาตรฐาน													
ค่ามัธยฐาน													
ค่าต่ำสุด													
ค่าสูงสุด													
3. ค่าร้อยละ													

ตอนที่ 2 ลีลา (ค่าใดบ้างที่จะนำเสนอ และมีรูปแบบที่นิยมนำเสนอเป็นอย่างไร)

4.2 ต่อไปนี้คือรูปแบบที่นิยมในการนำเสนอข้อมูลสำหรับพรรณนาลักษณะของตัวอย่างที่ศึกษา จงนำผลที่ได้ข้างต้นเติมลงในช่องตารางตามที่เหมาะสม (วิเคราะห์จากข้อมูล $n = 30$)

ตารางที่ 1. ลักษณะทางประชากรของกลุ่มที่ศึกษา

ตัวแปร	จำนวน	ร้อยละ
1. เพศ		
ชาย		
หญิง		
รวม		
2. อายุ		
20 ปี หรือน้อยกว่า		
21 - 22 ปี		
23 ปีขึ้นไป		
รวม		
ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน)		
ค่ามัธยฐาน(ค่าต่ำสุด:ค่าสูงสุด)		
3. ความสูง		
น้อยกว่า 150 ซม.		
150 - 159 ซม.		
160 ซม. ขึ้นไป		
รวม		
ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน)		
ค่ามัธยฐาน(ค่าต่ำสุด:ค่าสูงสุด)		
4. น้ำหนัก		
น้อยกว่า 50 กก.		
50 - 59 กก.		
60 กก. ขึ้นไป		
รวม		
ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน)		
ค่ามัธยฐาน(ค่าต่ำสุด:ค่าสูงสุด)		

จสรุปผลจากตารางข้างต้น ให้ได้ใจความที่สามารถนำไปบรรจุในรายงานผลการวิจัยได้

ถ้าหากท่านได้รับมอบหมายให้ทำการสำรวจเช่นเดียวกันนี้ที่มหาวิทยาลัยมหิดล ด้วยขนาดตัวอย่าง 30 ราย เท่ากัน เพื่อเปรียบเทียบกัน จงออกแบบตารางเพื่อนำเสนอข้อมูลชุดที่นำเสนอข้างต้น (แสดงเฉพาะหัวตารางและหัวคอลัมน์เฉพาะข้อมูลเพศเพียงตัวแปรเดียว โดยเว้นข้อมูลไว้)





การอนุมานทางสถิติเพื่อตอบคำถามวิจัย (Statistical inference for answering the research question)

5

กิจกรรมที่ 5.1 : การแจกแจงค่าสถิติ และ Central Limit Theorem

5.1 ข้อมูล ค่าสถิติ หรือค่าพารามิเตอร์

5.1.1 จากปฏิบัติการที่ 2 น้ำหนักของท่าน (ค่า V5 ของ ID = 0) = กิโลกรัม
(ค่านี้คือ ☐ ข้อมูล ☐ ค่าสถิติ ☐ ค่าพารามิเตอร์)

5.1.2 จากปฏิบัติการที่ 4 ค่าเฉลี่ยน้ำหนัก = กิโลกรัม
(ค่านี้คือ ☐ ข้อมูล ☐ ค่าสถิติ ☐ ค่าพารามิเตอร์)

5.1.3 สิ่งที่เราต้องการทราบคือ ค่าเฉลี่ยน้ำหนักของนักศึกษาทั้งหมดที่เรียนในห้องนี้
(ค่านี้คือ ☐ ข้อมูล ☐ ค่าสถิติ ☐ ค่าพารามิเตอร์)

เราสามารถหาค่าตามข้อ 5.1.3 ได้หากเก็บข้อมูลน้ำหนักของนักศึกษาทุกคน
แต่ในชีวิตจริงเราไม่ทำ เพราะ

- i)
- ii)
- iii)

5.2 คำที่ตอบคำถามข้อ 5.1.3 (ค่าเฉลี่ยน้ำหนักของนักศึกษาทั้งหมดที่เรียนในห้องนี้) ได้ใกล้เคียงที่สุดคือค่าในข้อต่อไปนี้ใช่หรือไม่

- ก. ข้อ 5.1.1 คือขนาดตัวอย่างเท่ากับ 1 ☐ ใช่ ☐ ไม่ใช่
- ข. ข้อ 5.1.2 คือขนาดตัวอย่างเท่ากับ 30 ☐ ใช่ ☐ ไม่ใช่
- ค. เพิ่มขนาดตัวอย่างเป็น 50 ☐ ใช่ ☐ ไม่ใช่
- ง. เพิ่มขนาดตัวอย่างเป็น 100 ☐ ใช่ ☐ ไม่ใช่

5.3 ค่าจากข้อ 5.1.2 ซึ่งเป็นค่าเฉลี่ยน้ำหนัก สามารถเปลี่ยนแปลงได้ตามตัวอย่างที่สุ่มมาได้ ดังนี้

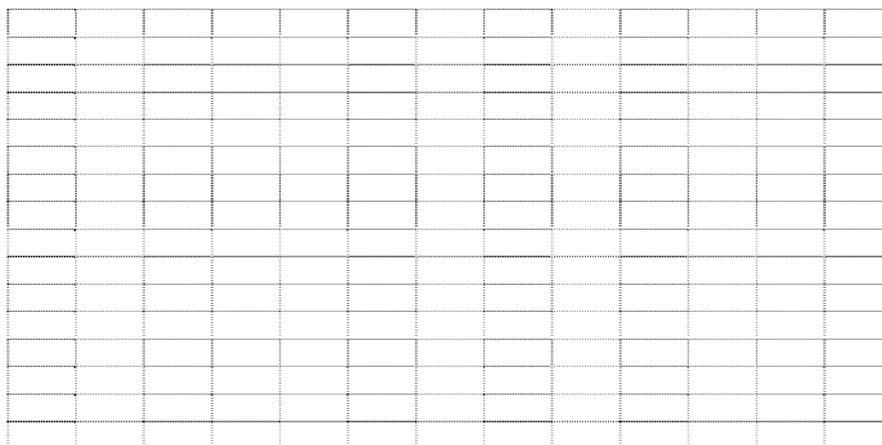
5.3.1 ให้เขียนค่าเฉลี่ยน้ำหนักที่ตนเองคำนวณได้จากการสุ่มตัวอย่าง $n=30$ ลงในช่องแรกของตารางข้างล่างนี้ จากนั้นให้เขียนค่าเฉลี่ยดังกล่าวที่ผู้รับการอบรมคนอื่น ๆ คำนวณได้ ลงช่องถัดไปจนครบทุกคน (ดูแนวทางการจำลองเหตุการณ์นี้โดยใช้คอมพิวเตอร์ในภาคผนวก 2)

ข้อมูลนี้คือค่าเฉลี่ยน้ำหนักจากหลายการวิจัยที่ทำในเรื่องเดียวกัน ในประชากรเดียวกัน และขนาดตัวอย่างเท่ากัน (Repeated samples)

5.3.2 จงคำนวณค่าเฉลี่ยจากค่าเฉลี่ยในตารางข้างต้น =

ค่านี้เรียกว่า ค่าเฉลี่ยของค่าเฉลี่ยของตัวอย่าง ($\bar{\bar{X}}$) หรือ Mean of the sampling means

5.3.3 สร้าง Histogram แสดงการแจกแจงของค่าเฉลี่ย พร้อมเขียนโค้งความถี่ทับลงด้วย



5.3.4 ค่าที่แกนนอนของกราฟ ณ จุดที่แบ่งพื้นที่ใต้กราฟออกเป็นสองส่วนเท่า ๆ กัน เท่ากับ

5.3.5 เปรียบเทียบค่าจากข้อ 5.3.4 ที่ได้จากกราฟนี้กับค่าเฉลี่ยที่คำนวณได้จากข้อมูลค่าเฉลี่ยของตัวอย่างตามข้อ 5.3.2 ว่าเท่ากันหรือไม่ ตอบ:

5.3.6 ข้อมูลน้ำหนักของนักศึกษาทั้งหมด ($N = 143$) นำมาคำนวณค่าเฉลี่ยได้เท่ากับ กิโลกรัม (ดูภาคผนวก 2) ค่านี้คือค่าพารามิเตอร์ แทนด้วย μ

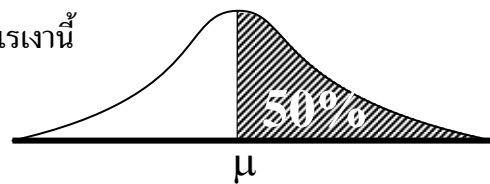
ถ้ามีการศึกษาวิจัยได้สรุปผลการศึกษาโดยใช้เพียงค่าสถิติที่ได้จากตัวอย่างเดียวเท่านั้น เช่น ได้ค่าเฉลี่ยน้ำหนักเท่ากับ 61 กิโลกรัม แล้วสรุปว่าสูงหรือต่ำกว่าค่าบางค่า เช่น นำไปเปรียบเทียบกับที่มีการวิจัยเช่นเดียวกันนี้ในมหาวิทยาลัยอื่นซึ่งเท่ากับ 65 กิโลกรัม จึงสรุปว่านักศึกษาในมหาวิทยาลัยที่ตนศึกษาวิจัยนั้นมีน้ำหนักเฉลี่ยต่ำกว่า ท่านจะให้ข้อคิดเห็นอย่างไร กับนักวิจัยนั้น

5.3.7 เปรียบเทียบค่าเฉลี่ยของค่าเฉลี่ยของตัวอย่าง (\bar{X}_x) กับค่า μ และให้ความเห็น

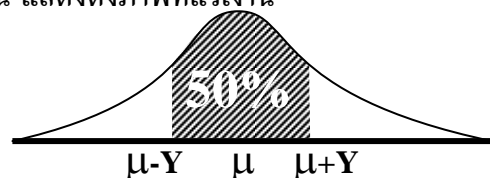
5.3.8 มีการวิจัยจำนวนเท่าใดที่ได้ค่าเฉลี่ย (\bar{X}) มากกว่าค่า \bar{X}_x ตอบ
คิดเป็น% ของจำนวนการวิจัยทั้งหมด

5.3.9 จากข้อ 5.3.8 เรากล่าวอีกนัยหนึ่งได้ว่า "การวิจัยใดๆ จะได้ค่าเฉลี่ยมากกว่าค่า μ มีโอกาส 0.5"

นั่นคือ 50% ของพื้นที่ใต้กราฟ คือค่าความน่าจะเป็นเท่ากับ
พื้นที่ใต้กราฟที่กล่าวถึงนี้ แสดงดังภาพที่แรเงา



5.3.10 ถ้านำค่า Y บวกเข้าและลบออกจากค่า μ แล้วทำให้ได้ช่วงของค่าเฉลี่ยที่คลุมพื้นที่ใต้กราฟ 50% แล้ว พื้นที่ใต้กราฟที่กล่าวถึงนี้ แสดงดังภาพที่แรเงา



5.3.11 จากข้อ 5.3.10 ค่า $\mu-Y$ ถึง $\mu+Y$ คือช่วงที่มีโอกาสเท่ากับ 0.5 ที่จะคลุมค่า μ

5.3.12 จากข้อ 5.3.10 ค่า Y เท่ากับเท่าใดนั้น สามารถหาได้ถ้าทราบทุกค่าบนแกนนอน พร้อมกับค่าความน่าจะเป็นของทุกช่วงของค่าบนแกนนอน (ค่าความน่าจะเป็น คือค่าที่บอกขนาดของ“โอกาส” คือพื้นที่ใต้กราฟ) ดังนั้นกราฟใดที่รู้คุณสมบัติการแจกแจงแล้ว (Known distribution) ถ้าบอกค่าบนแกนนอนจะทำให้รู้ค่าความน่าจะเป็น และในทางกลับกันถ้าบอกความน่าจะเป็นก็จะทำให้รู้ค่าบนแกนนอนได้

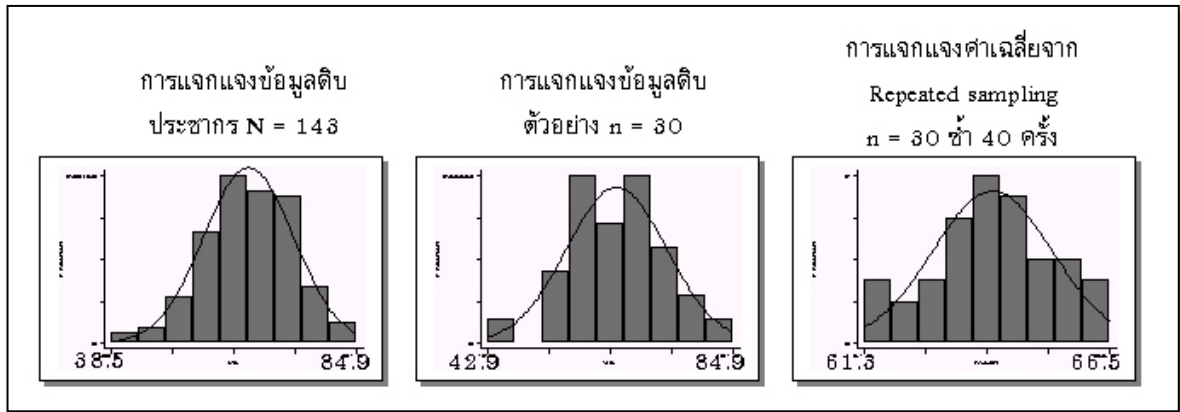
5.3.13 ดังนั้น เพื่อให้สามารถประมาณค่า μ ได้ เราต้องรู้คุณสมบัติการแจกแจงค่า \bar{X} หรือการแจกแจงอื่นใดที่แทนกันได้

บทบรรยายเรียงลำดับประเด็นสำคัญ :

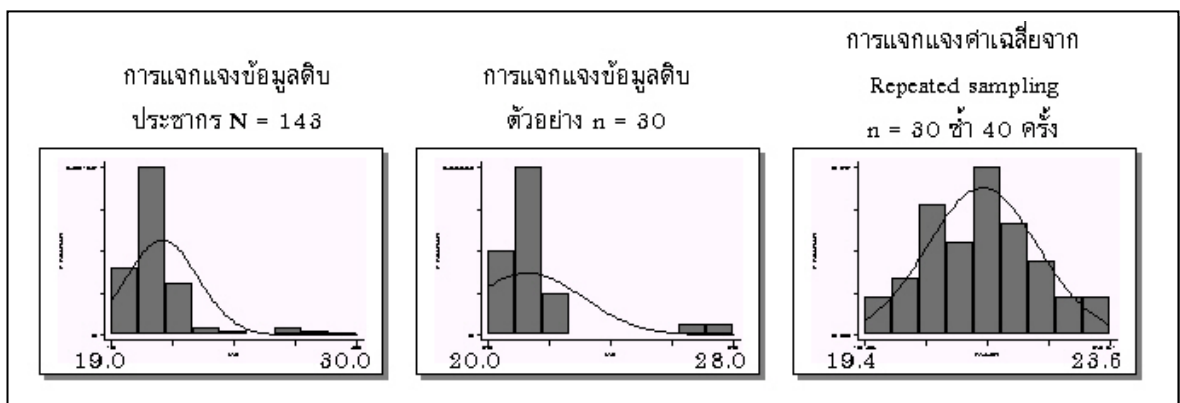
1. จากข้อมูลตัวอย่าง เราสามารถคำนวณได้ค่าเฉลี่ยของตัวอย่าง เรียกว่า ค่าสถิติ (บางกรณีค่าสถิติอาจเป็นค่าผลต่างของค่าเฉลี่ย ค่าร้อยละ ค่าสัมประสิทธิ์สหสัมพันธ์ ค่าRelative risk ค่าOdds ratio ฯลฯ)
2. จากข้อมูลประชากร เราสามารถคำนวณได้ค่าเฉลี่ยของประชากร เรียกว่า ค่าพารามิเตอร์
3. จากข้อมูลค่าเฉลี่ยจากการสุ่มตัวอย่างซ้ำๆ กัน (Repeated samples) เสมือนทำวิจัยเรื่องเดียวกันในประชากรเดียวกันหลายๆ ครั้ง เราสามารถคำนวณได้ค่าเฉลี่ยของค่าของค่าเฉลี่ยจากแต่ละการวิจัย ค่าดังกล่าวใกล้เคียงกับค่าพารามิเตอร์มาก
4. การวิจัยมีเป้าหมายที่ต้องการทราบคือค่าพารามิเตอร์ เช่นในตัวอย่างคือน้ำหนักเฉลี่ยของประชากรทั้ง 143 คน แต่ในความเป็นจริง เราไม่สามารถทำทุกหน่วยประชากรได้ จึงไม่มีใครทราบค่าพารามิเตอร์ แต่พบว่า Repeated samples สามารถได้ค่าที่ใกล้เคียงกับค่าพารามิเตอร์มากที่สุด นั่นคือ**มีหนทางที่ให้ค่าพารามิเตอร์โดยไม่ต้องเก็บข้อมูลทุกหน่วยประชากร**
5. ในความเป็นจริงก็ไม่มีใครทำ Repeated samples อีกเช่นกัน เราทำวิจัยเพียงครั้งเดียว สิ่งที่มีอยู่ในมือจึงเป็นค่าสถิติค่าเดียว แต่เราก็**ไม่แน่ใจ**ค่าดังกล่าวว่าจะใกล้เคียงค่าพารามิเตอร์หรือไม่ เพียงใด ดังนั้น เราจึงต้องบอกระดับความ**แน่ใจ**ว่าค่าพารามิเตอร์เป็นเท่าใด โดยบอกเป็นช่วงที่เป็นไปได้สูงที่ครอบคลุมค่าพารามิเตอร์ เช่น **“มีความเชื่อมั่น 95% ว่าค่าพารามิเตอร์จะอยู่ระหว่าง..... ถึง”** เป็นต้น นี่คือการประมาณค่า แต่ถ้าต้องการบอกเป็นโอกาสที่จะได้ผลการศึกษาเท่ากับค่าสถิติหรือมากกว่านั้นถ้าสมมติฐานเป็นจริง จะรายงานเป็นค่า **p -value** ส่วนนี้คือการทดสอบสมมติฐาน
6. ทั้งสองส่วนนี้คือการอนุมานทางสถิติ ซึ่งจะเห็นว่า**มีความน่าจะเป็น**เข้ามาเกี่ยวข้อง แต่ความน่าจะเป็นนี้ยังต้องได้จากพื้นที่ใต้กราฟของการแจกแจงค่าสถิติ ซึ่งในความเป็นจริงไม่มีใครทำ แต่โชคดียังดีที่มี Central Limit Theorem มาช่วย

5.4 ทำความเข้าใจ Central Limit Theorem จากข้อมูลจริง

5.4.1 ต่อไปนี้คือกราฟแสดงการแจกแจงข้อมูลน้ำหนักนักศึกษาที่ทำมาแล้วทั้งหมด



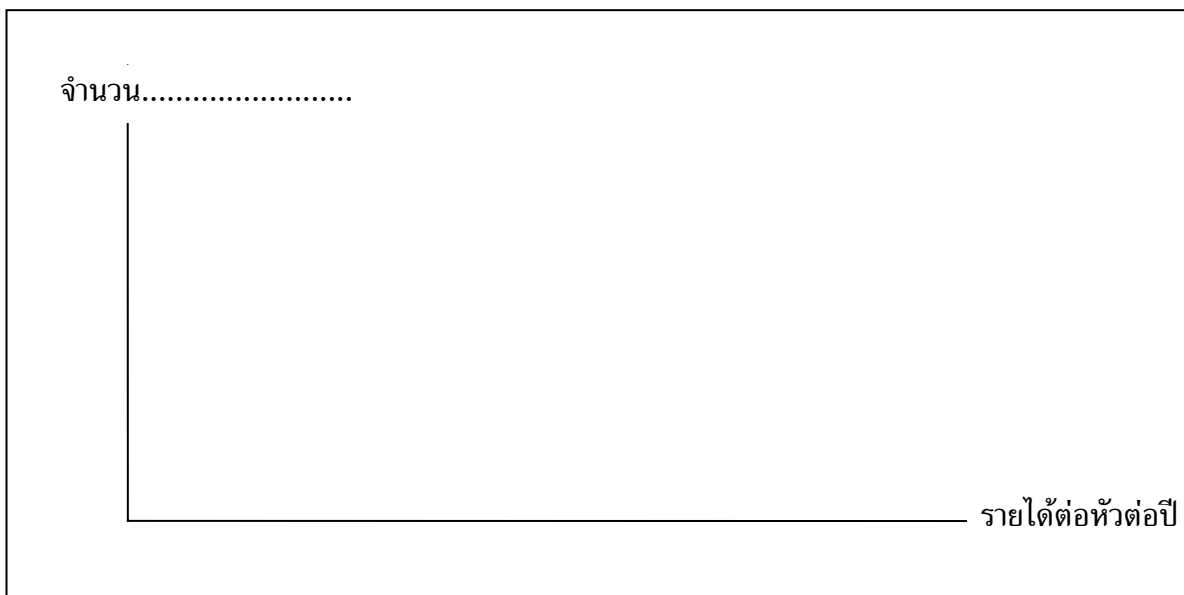
จากนั้นทำเช่นเดียวกันนี้กับข้อมูลอายุนักศึกษา (V2) ท่านสามารถทดลองเองได้โดยทำตามขั้นตอนที่กล่าวข้างต้น ได้ผลดังนี้



หมายเหตุ: ท่านสามารถพิสูจน์สิ่งนี้กับข้อมูลชุดใดก็ได้ การสุ่มตัวอย่างจากข้อมูลมีแสดงไว้ในภาคผนวก 2 โดยใช้ STATA คำสั่งคือ `sampling` แล้วตามด้วยค่า % ที่ท่านระบุว่าจะสุ่มมาจากข้อมูลชุดนั้นเป็นร้อยละเท่าไร

จงเขียนสรุปสิ่งที่สังเกตได้

5.4.2 ต่อคำพูดที่ว่า “จากการศึกษารายได้ต่อหัวต่อปีพบว่า 80% ของคนไทยยากจน มีเศรษฐกิจพื้นฐานไม่ก็ตระกูล” โคว์ความถี่รายได้คนไทยจะมีลักษณะตามที่ปรากฏในปฏิบัติการที่ 3 ข้อ 3.6.1 นั้น ถ้ามีการศึกษาเหมือนกัน 100 การศึกษา ด้วยขนาดตัวอย่างห้าหมื่นคนต่อครั้งการสำรวจ แล้วนำค่าเฉลี่ยจากทุกการศึกษามาดูการกระจายข้อมูลค่าเฉลี่ยนั้นตามวิธีการข้างต้น โคว์ความถี่จะมีลักษณะใด



5.4.3 ถ้าทำเช่นเดียวกันกับที่กล่าวในข้อ 5.4.2 แต่เพิ่มขนาดตัวอย่างเป็นห้าแสนคนต่อครั้งการสำรวจ ลักษณะของกราฟแจกแจงค่าเฉลี่ยที่ได้ เมื่อเปรียบเทียบกับที่ได้จากขนาดตัวอย่างเป็นห้าหมื่นคนต่อครั้งการสำรวจ

5.4.4 กราฟแสดงการแจกแจงข้อมูลดิบ อย่างเช่นในข้อ 3.3 ในบทที่ 3 เรียกว่า Empirical distribution ส่วนกราฟแสดงการแจกแจงค่าสถิติ อย่างเช่นในข้อ 5.3.3 ในบทที่ 5 นี้เรียกว่า Sampling distribution ต่อไปนี้คือประเด็นสำคัญเกี่ยวกับการแจกแจงเหล่านี้

- ✓ Empirical distribution ของประชากร (เช่นจากข้อมูลทั้งหมด $N=143$) เป็นอย่างไร Empirical distribution ของตัวอย่าง (เช่นจากข้อมูลตัวอย่าง $n=30$) มักคล้ายกัน ถ้าประชากรเบ้ ตัวอย่างก็มักเบ้ตาม เป็นต้น
- ✓ อย่างไรก็ตาม Sampling distribution จะสมมาตรเสมอ และถ้าขนาดตัวอย่างใหญ่พอ Sampling distribution จะมีคุณสมบัติที่เป็นมาตรฐาน เรียกว่า Normal distribution
- ✓ คุณสมบัติที่เป็นมาตรฐาน ของ Normal distribution จึงถือเป็นหัวใจของการอนุมานทางสถิติ
- ✓ เราเรียกกราฟที่แสดงการกระจายที่เป็น Normal distribution ว่า Normal curve

- ✓ กราฟที่แสดงการกระจายนี้ อธิบายได้ด้วยค่า 2 ค่าที่บอกตำแหน่งกึ่งกลาง คือค่าเฉลี่ย (หมายถึงค่าที่แกนนอนของกราฟ ณ จุดที่แบ่งกราฟออกเป็นสองส่วนเท่าๆ กัน) และค่าที่บอกการกระจาย หรือบอกความกว้างแคบของฐานกราฟ ค่าเฉลี่ยคำนวณได้ง่ายตรงไปตรงมา ส่วนค่าการกระจายต้องการการทำความเข้าใจอย่างถ่องแท้

ข้อ 5.5 ถึง 5.7 มีไว้เพื่อทำความเข้าใจกับที่มาของการคำนวณค่าการกระจายซึ่งเป็นหัวใจของสถิติ ขอให้ทำความเข้าใจให้สามารถเชื่อมโยงกับสามัญสำนึกให้ได้

5.5 ฐานของกราฟแสดงการแจกแจงที่กว้างหรือแคบ บอกถึงการกระจายข้อมูลชุดนั้นว่ามากหรือน้อยตามลำดับ

สำหรับ Empirical distribution นั้น มาตรวัดการกระจายคือ ส่วนเบี่ยงเบนมาตรฐาน (Standard deviation บางครั้งย่อเป็น SD หรือ S) ขอให้พิจารณาวิธีการคำนวณเพื่อเกิดความเข้าใจว่า วัดการกระจายอย่างไร ขั้นแรกคำนวณค่าเฉลี่ย (คือเอาข้อมูลทุกค่ารวมกันแล้วหารด้วยขนาดตัวอย่าง หรือ n) ขั้นที่สองนำข้อมูลแต่ละค่าไปลบออกจากค่าเฉลี่ย ผลต่างที่ได้ทุกค่ากลายเป็นข้อมูลอีกชุดหนึ่งที่บอกขนาดการกระจายข้อมูลแต่ละค่าจากค่ากลางของมัน ถ้านำผลต่างที่ได้ทุกค่าดังกล่าวมาหาค่าเฉลี่ยก็น่าจะบอกการกระจายได้เลย แต่มีปัญหาทางคณิตศาสตร์เพราะพอรวมกันจะได้ค่าเป็นศูนย์ ศูนย์นำไปหารด้วย n เพื่อหาค่าเฉลี่ยยังผลให้ได้ค่าเฉลี่ยเป็นศูนย์ทุกกรณีซึ่งไม่ถูกต้อง ดังนั้นจึงแก้ปัญหานี้ เป็นขั้นที่สามของการคำนวณคือยกกำลังสองผลต่างแต่ละค่าขึ้นก่อนแล้วจึงนำมารวมกัน (ผลต่างที่ได้ค่าติดลบ พอยยกกำลังสองจะเป็นบวก) จากนั้นหารด้วย n-1 [เพราะเราไม่มีอิสระในการเลือกค่าเพื่อการคำนวณนี้ 1 ค่า (ที่ค่าสุดท้ายไม่มีตัวอื่นเหลือให้เลือก คือไม่มี Variation สำหรับนำมาคำนวณหาค่าอะไรต่อไปได้แล้ว) เราเรียก n-1 นี้ว่า *ขั้นของความอิสระ (Degree of freedom)*] ค่าที่ได้ก็คือค่าเฉลี่ยผลต่างกำลังสองของข้อมูลแต่ละค่ากับค่าเฉลี่ยของข้อมูลชุดนั้น นั่นเอง ให้ชื่อค่านี้ว่า แวเหเรียนซ์ (Variance) แต่แปลความหมายยากเพราะติดค่ากำลังสองอยู่ ดังนั้นขั้นสุดท้าย จึงถอดราก (Square root ก็คือยกกำลัง 1/2 จึงเป็นวิธีการที่ทำ

ให้กำลังสองหายไป) ได้ค่าเฉลี่ยการกระจายข้อมูลแต่ละค่าที่กระจายจากค่ากลางของกลุ่มข้อมูลนั้นตามต้องการ นี่คือนิยามส่วนเบี่ยงเบนมาตรฐาน แต่เขียนเป็นประโยคจะยืดยาวยาว จึงแทนด้วยสูตร ดังนี้

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

ดังนั้น ค่า SD ของการสำรวจน้ำหนักนักศึกษาของท่านเท่ากับ
(นี่คือ SD สำหรับ $n = 30$ หรือค่าที่บอกขนาดความกว้างแคบของฐานกราฟ Empirical distribution นั่นเอง)

5.6 ใช้ข้อมูลค่าสถิติจาก Repeated sampling ในข้อ 5.3.1 คำนวณค่า SD ได้เท่ากับ

ค่านี้เรียกว่า Standard error ย่อเป็น SE กล่าวอีกนัยหนึ่ง *SE คือ SD ของค่าสถิติ* เป็นค่าที่บอกขนาดความกว้างแคบของฐานกราฟ Sampling distribution

5.7 ข้อ 5.6 เป็น SE สำหรับ Repeated sampling เมื่อ $n = 30$ ให้จินตนาการข้อมูลน้ำหนักนักศึกษาที่เราสุ่มจาก $N = 143$ ถ้าหากทำ Repeated sampling ใหม่ด้วย $n = 100$ ค่า SE จะมีขนาดเป็นอย่างไรเทียบกับเมื่อ $n = 30$ ตอบ(มากกว่า / น้อยกว่า)

นั่นคือกราฟ Sampling distribution มีฐานที่ (กว้างกว่า / แคบกว่า) เมื่อ $n = 30$

กล่าวอีกนัยหนึ่ง ค่า SE ต้องมี $1/n$ เป็นผลคูณ

คำถามคือ $1/n$ ไปคูณกับอะไร คำตอบคือ คูณกับ SD เพราะค่า SE ขึ้นอยู่กับทั้ง n และ SD

จึงคำนวณ SE จาก Variance ซึ่งเป็นค่าดั้งเดิมของการวัดการกระจาย ทหารด้วย n

จากที่เรารู้ว่า Variance ติดค่ายกกำลังสอง พอถอดรู้ที่จะได้ SD ในทางกลับกันพอยกกำลังสอง SD เราจะได้ Variance หรือ $SD^2 = \text{Variance}$ ดังนั้น $SE = SD^2/n$ เพื่อง่าย จึงติด Square root ให้ SD^2/n กลายเป็นสูตรต่อไปนี้

$$SE = \frac{SD}{\sqrt{n}}$$

5.8 จากข้อมูลการสำรวจน้ำหนักนักศึกษา $n = 30$ ของท่าน ค่า SD =

ดังนั้น ค่า SE =

เปรียบเทียบกับค่า SD ที่ได้จาก Repeated samples ในข้อ 5.6 ซึ่งเท่ากับ

จะเห็นว่าได้ค่าใกล้เคียงกัน (โอกาสที่จะไม่ได้ผลตามนี้มีได้น้อย ท่านสามารถพิสูจน์โดยทำซ้ำหลายๆ ครั้งหรือสอบถามจากเพื่อนที่ใช้คู่มือนี้)

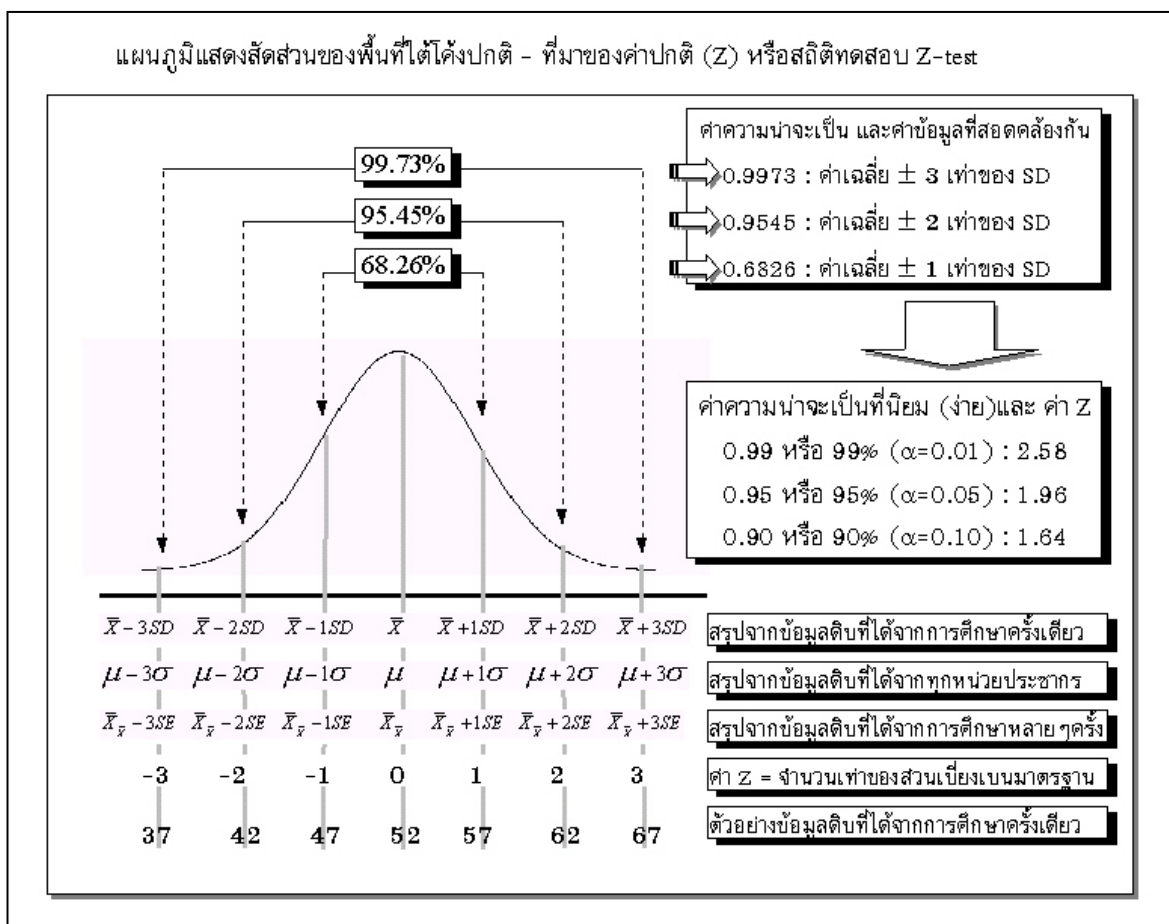
ดังนั้น เรามีหนทางได้มาซึ่ง SE จากการสุ่มตัวอย่างวิจัยเพียงครั้งเดียว โดยไม่จำเป็นต้องทำซ้ำเช่นเดียวกันนี้หลายๆ ครั้ง

นั่นหมายถึงเราสามารถรู้คุณสมบัติการแจกแจงของค่าสถิติได้โดยใช้ผลจากการวิจัยเพียงครั้งเดียวของเรา ดังนั้น การประมาณค่าหรือทดสอบสมมติฐานเกี่ยวกับค่าพารามิเตอร์จึงสามารถทำได้จากการวิจัยเพียงครั้งเดียว

บทบรรยายเรียงลำดับประเด็นสำคัญ (ต่อ):

7. จาก Central limit theorem เราบอกได้ว่า ไม่ว่าการแจกแจงข้อมูลของประชากรจะเป็นแบบใด การแจกแจงค่าสถิติจาก Repeated sampling ที่มีขนาดตัวอย่างใหญ่พอ จะเป็น Normal distribution เสมอ
8. เมื่อรู้ว่า Sampling distribution มีการแจกแจงแบบปกติ (Normal distribution) เราก็สามารถใช้คุณสมบัติของการแจกแจงนั้นในการอนุมานค่าสถิติเป็นค่าพารามิเตอร์ได้
9. การอนุมานดังกล่าวจะถูกต้องถ้าการแจกแจงเป็นไปตามที่เชื่อนั้น การระบุไปว่าเราเชื่อว่าการแจกแจงแบบใดนั้น เราเรียกว่า ข้อกำหนดเบื้องต้น (Assumption) บางครั้งอาจต้องทดสอบก่อนทำการอนุมานทางสถิติว่า ที่เชื่อเช่นนั้นเป็นจริงตามนั้นหรือไม่ หรือที่เรียกว่า ทดสอบข้อกำหนดเบื้องต้น ถ้าไม่ ผู้วิเคราะห์มักจะพยายามแปลงข้อมูลให้เป็นสเกลอื่น เช่น แปลงเป็นค่า Logarithm ก็เรียกว่า Log transformation จนข้อมูลได้ตาม Assumption ก่อนการอนุมาน อนุมานเสร็จจึง Anti-logarithm กลับเป็นค่าดั้งเดิม แต่ถ้าไม่สามารถทำให้ได้ตาม Assumption นั้นก็ไปเลือกใช้ Non-parametric statistics
10. การอนุมานทางสถิติที่อาศัยการแจกแจงที่รู้คุณสมบัติแล้ว เรียกว่า Parametric statistics ส่วนการอนุมานทางสถิติที่ไม่อาศัยการแจกแจงใดๆ เลย เรียกว่า Non-parametric statistics
11. Parametric statistics จะมีประสิทธิภาพสูงกว่า Non-parametric statistics ถ้าเป็นไปตาม Assumption
12. การรู้คุณสมบัติการแจกแจงต่าง ๆ จึงมีความสำคัญ และที่พบบ่อยที่สุดคือ Normal distribution

กิจกรรมที่ 5.2 : การใช้ประโยชน์จากการแจกแจงแบบปกติ



ความเข้าใจเกี่ยวกับโค้งปกติ เป็นพื้นฐานที่สำคัญสำหรับการทำความเข้าใจสถิติ เนื่องจากวิธีการทางสถิติส่วนมาก มีข้อกำหนดเบื้องต้นเกี่ยวข้องกับการแจกแจงแบบปกติ นอกจากนั้น การแจกแจงแบบอื่นก็อาศัยหลักการเดียวกันนี้ เกิดเป็นสถิติทดสอบที่มีชื่อต่างกันไป เช่นกรณีข้อมูลต่อเนื่องถ้าข้อมูลมีการแจกแจงแบบปกติ สถิติทดสอบคือ Z-test แต่ถ้าขนาดตัวอย่างไม่ใหญ่พอพบว่าการแจกแจง t สถิติทดสอบกรณีนี้คือ t-test กรณีข้อมูลแจกแจงนับก็มีลักษณะการแจกแจงหลายรูปแบบเช่นการแจกแจงทวินาม (Binomial) การแจกแจงปัวซอง (Poisson) นอกจากนั้นยังมีการแจกแจงค่าสถิติอื่นๆ เช่น การแจกแจงไคสแควร์ การแจกแจง F สถิติทดสอบก็ให้ชื่อตามการแจกแจงนั้นๆ เป็นต้น

กรณีข้อมูลต่อเนื่อง ค่าสถิติเหล่านี้คือ “จำนวนเท่าของ SE” ที่ห่างจากค่าเฉลี่ย เช่นจากรูปข้างต้น เมื่อ $Z = 2$ หมายถึงค่าใดๆ ที่ห่างจากค่าเฉลี่ย 2 เท่าของ SE ถ้าค่าสถิติ (ค่าเฉลี่ย) เท่ากับ 52 ค่า SE เท่ากับ 5 เรารู้ค่าสถิติ ที่ $Z = 2$ ซึ่งเท่ากับ 62 อีกด้านหนึ่งของโค้งความถี่ที่สมมาตรกันคือ $Z = -2$ ได้ค่าสถิติเท่ากับ 42 พื้นที่ที่ครอบคลุมโดยค่า 42 ถึง 62 คิดเป็น 95.45% ของพื้นที่ทั้งหมด คำตอบคือช่วงของข้อมูลที่ครอบคลุมพื้นที่ตามที่ระบุ นี่ก็คือการประมาณค่า ในทางตรงข้าม ถ้าสนใจว่า “ถ้าค่าเฉลี่ยจริง (ค่าพารามิเตอร์) เท่ากับ 52 แล้ว โอกาสที่ทำการวิจัยแล้วได้ค่าสถิติเท่ากับ

หรือน้อยกว่า 37 เป็นเท่าใด?” คำตอบจะเป็นค่าความน่าจะเป็น หรือ p-value นี่คือการทดสอบสมมติฐาน

ข้อ 5.9 และข้อ 5.10 ให้ใช้ข้อมูลจากการสำรวจน้ำหนักนักศึกษาโดยตัวท่านตามที่กล่าวข้างต้น

5.9 ใช้ประโยชน์จากการแจกแจงข้อมูลดิบ (Empirical distribution)

ถ้าข้อมูลมีการแจกแจงแบบปกติ (นี่คือการกล่าวข้อตกลงเบื้องต้น หรือ Assumption)

ก. ค่าที่ครอบคลุม 95% ของจำนวนข้อมูลทั้งหมดคือ

ข. ถ้าทราบว่านักศึกษาที่มหาวิทยาลัยมหิดลก็มีลักษณะข้อมูลน้ำหนักเช่นเดียวกับที่ท่านสำรวจพบในมหาวิทยาลัยขอนแก่น จากนักศึกษา 500 คน จะมีกี่คนที่มีน้ำหนัก 66 กิโลกรัมขึ้นไป

5.10 ใช้ประโยชน์จากการแจกแจงค่าสถิติ (Sampling distribution)

ถ้าข้อมูลมีการแจกแจงแบบปกติ (นี่คือการกล่าวข้อตกลงเบื้องต้น หรือ Assumption)

ก. ค่าที่ครอบคลุม 95% ของจำนวนข้อมูลทั้งหมด (ในที่นี้คือจำนวนครั้งของการสำรวจ) คือ

ข. ถ้าค่าเฉลี่ยน้ำหนักของประชากรเท่ากับ 60 กิโลกรัม การที่จะสุ่มสำรวจหนึ่งครั้งแล้วได้ค่าเฉลี่ยน้ำหนักของนักศึกษาเท่ากับ “.....” (ใส่ค่าเฉลี่ยที่ได้จากการสำรวจโดยท่าน) กิโลกรัมหรือสุดขั้วไปทางที่เป็นไปได้น้อย (More extreme) มีโอกาสเท่ากับ

5.11 จากข้อ 5.9 ค่าที่ได้จากกราฟของการแจกแจงข้อมูลดิบ X จะนำไปสู่ค่า

จากข้อ 5.10 ค่าที่ได้จากกราฟของการแจกแจงค่าสถิติ \bar{X} จะนำไปสู่ค่า

กิจกรรมที่ 5.3 : การอนุมานทางสถิติ (การประมาณค่า และการทดสอบสมมติฐาน)

5.12 การประมาณค่า (Estimation)

5.12.1 ปฏิบัติการเพื่อทำความเข้าใจช่วงเชื่อมั่น

5.12.1.1 การประมาณค่าพารามิเตอร์ โดยอาศัยข้อมูลจากตัวอย่าง ในที่นี้ ใช้ข้อมูลน้ำหนักนักศึกษาเป็นกรณีตัวอย่าง จงคำนวณช่วงเชื่อมั่น 95% ของค่าเฉลี่ยน้ำหนักนักศึกษา (ให้ทศนิยม 1 ตำแหน่ง)

สูตรที่ใช้ : $\bar{X} \pm t_{\alpha=0.025; df=n-1} (S / \sqrt{n})$ เมื่อ $df = n - 1 = 30 - 1 = 29$

ด้วยตัวอย่างขนาดเล็ก การแจกแจงไม่เป็นแบบ Normal แต่เป็นแบบ t ซึ่งฐานกราฟกว้างกว่าเล็กน้อย จึงใช้ t แทน Z ถ้าขนาดตัวอย่างใหญ่ใช้ Z ได้ ในที่นี้หาค่า t จากตาราง t -distribution ในบทที่ 6

STATA: `su v5` --> เพื่อได้ค่าส่วนเบี่ยงเบนมาตรฐาน (S) และค่าเฉลี่ย (\bar{X})
ของน้ำหนัก

แทนค่า :

95% CI. :

STATA: `ci v5` --> เพื่อได้ช่วงเชื่อมั่น 95% ของค่าเฉลี่ยน้ำหนัก

แปลความหมาย :

5.12.1.2 ให้ผู้รับการอบรมเขียนช่วงเชื่อมั่นค่าเฉลี่ยน้ำหนัก ที่ตนเองคำนวณได้จากการสุ่มตัวอย่าง
นักศึกษา 30 คน ลงในช่องแรก ของตารางข้างล่างนี้ จากนั้นให้เขียนช่วงเชื่อมั่นค่าเฉลี่ย
ดังกล่าว ที่เพื่อนผู้รับการอบรมคนอื่นๆ คำนวณได้จากการสุ่มตัวอย่างนักศึกษา 30 คน
(กรณีท่านทำคนเดียว โปรดศึกษาการจำลองเหตุการณ์นี้ในค่าเฉลี่ยที่ภาคผนวก 2)

5.12.1.3 สร้างแผนภูมิแสดงช่วงเชื่อมั่นจากตารางข้างต้น

ค่าพารามิเตอร์(μ) =

จากช่วงเชื่อมั่นทั้งหมด ช่วง คร่อมค่าพารามิเตอร์ ช่วง คิดเป็น %

5.12.1.4 จากข้อ 5.12.1.3 จงพิจารณาหาแนวทางในการแปลความหมายช่วงเชื่อมั่น แล้วเขียนแปลความหมายค่าช่วงเชื่อมั่นที่ท่านคำนวณได้

5.12.2 ข้อ 5.12.1.1 นั้น เป็นการประมาณค่าเฉลี่ยในประชากรกลุ่มเดียว ถ้าเป็นการประมาณค่าเฉลี่ยในประชากรสองกลุ่มที่อิสระต่อกัน ตัวอย่างคำถามวิจัยคือ จงคำนวณช่วงเชื่อมั่น 95% ของค่าความแตกต่างค่าเฉลี่ยน้ำหนักระหว่างนักศึกษาชายและหญิง
ค่าที่ต้องใช้ในการคำนวณ :

	ชาย	หญิง
\bar{X}
S.D.
n

STATA: `su v5 if v2 == 1` --> เพื่อได้ค่าส่วนเบี่ยงเบนมาตรฐาน (S) และค่าเฉลี่ย (\bar{X})
ของน้ำหนักนักศึกษาชาย
`su v5 if v2 == 2` --> เพื่อได้ค่าส่วนเบี่ยงเบนมาตรฐาน (S) และค่าเฉลี่ย (\bar{X})
ของน้ำหนักนักศึกษาหญิง

สูตรที่ใช้ : $(\bar{x}_1 - \bar{x}_2) \pm t_{0.025} \times se(\bar{x}_1 - \bar{x}_2)$ เมื่อ $se(\bar{x}_1 - \bar{x}_2) = s \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ และ

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$
 และ
$$df = n_1 + n_2 - 2$$

ท่านสามารถแทนค่าลงในสูตรคำนวณด้วยมือแล้วเปรียบเทียบผลที่ได้กับผลจาก STATA ซึ่งจะพบว่าเท่ากัน (บางครั้งการแทนค่าลงในสูตรและคำนวณด้วยมือทำให้เข้าใจสถิติยิ่งขึ้น) แต่นับจากนี้เราจะใช้ผลจาก STATA ซึ่งในความเป็นจริง เราจะทำแบบนี้

STATA: `ttest v5, by(v2)`

95% CI. :

แปลความหมาย :

5.12.3 การประมาณค่าเฉลี่ยในประชากรสองกลุ่มที่ไม่อิสระต่อกัน ตัวอย่างคำถามวิจัยคือ "ผลของการวิ่งกับที่หนึ่งนาที่ต่ออัตราชีพจรของนักศึกษา" หรืออาจเขียนเป็นวัตถุประสงค์ว่า "เพื่อประมาณค่าเฉลี่ยผลต่างอัตราการเต้นของชีพจรก่อนและหลังวิ่งหนึ่งนาที่" จึงแสดงแนวทางที่ท่านวิเคราะห์ข้อมูลเพื่อตอบคำถามวิจัยนี้

5.13 การทดสอบสมมติฐาน (Hypothesis testing)

5.13.1 การทดสอบสมมติฐานค่าเฉลี่ยในประชากรกลุ่มเดียว

โจทย์ : น้ำหนักเฉลี่ยนักศึกษามหาวิทยาลัยขอนแก่นเท่ากับ 55 กิโลกรัม ผลการสำรวจนักศึกษาของท่าน ให้ผลสอดคล้องกับข้อมูลดังกล่าวหรือไม่

ก. ตั้งสมมติฐาน

ข. กำหนดระดับนัยสำคัญ

ค. คำนวณค่าสถิติทดสอบ จากสูตร:
$$t = \frac{\bar{x} - k}{\frac{s}{\sqrt{n}}}$$

ง. หาค่า p-value

STATA : `ttest v5 = 55`

จ. ตัดสินใจและสรุปผล

5.13.2 การทดสอบสมมติฐานค่าเฉลี่ยในประชากรสองกลุ่มที่อิสระต่อกัน

โจทย์ : น้ำหนักเฉลี่ยนักศึกษาชายแตกต่างจากน้ำหนักเฉลี่ยนักศึกษาหญิงหรือไม่

ก. ตั้งสมมติฐาน

ข. กำหนดระดับนัยสำคัญ

ค. คำนวณค่าสถิติทดสอบ จากสูตร: $t = \frac{\bar{x}_1 - \bar{x}_2}{se(\bar{x}_1 - \bar{x}_2)}$

ง. หาค่า p-value

STATA : `ttest v5, by(v2)`

จ. ตัดสินใจและสรุปผล

5.13.3 การทดสอบสมมติฐานค่าเฉลี่ยในประชากรสองกลุ่มที่ไม่อิสระต่อกัน ตัวอย่างคำถามคือ ค่าเฉลี่ยอัตราการเต้นชีพจร (ครั้งต่อนาที) ก่อนและหลังการวิ่งแตกต่างกันหรือไม่ จงแสดง แนวทางที่ท่านวิเคราะห์ข้อมูลเพื่อตอบคำถามวิจัยนี้

STATA : `ttest v11 = v10`

กิจกรรมที่ 5.4 : การสรุปผลการวิเคราะห์ข้อมูลในทางปฏิบัติ

5.14 จงสรุปผลจากการวิเคราะห์ข้อมูลเพื่อเปรียบเทียบน้ำหนักระหว่างนักศึกษาชายและหญิง โดยใช้แนวทางตามที่เขียนเป็นตัวอย่างในข้อ (6) ของปฏิบัติการที่ 1 (ภายใต้หัวข้อที่ 1.12)

--

5.15 ข้อความสรุปผลในข้อ 5.14 อยู่ในส่วนใดของรายงานวิจัย

--

5.16 จงแสดงรูปแบบการนำเสนอผลการวิเคราะห์ในรายงานวิจัยโดยใช้กรณีเปรียบเทียบค่าน้ำหนักเฉลี่ยระหว่างชายกับหญิงเป็นคำถามวิจัย

--



การประยุกต์แนวคิดสถิติในการวิจัย (Application of statistical concepts in research practice)

6

นี่คือบทสุดท้ายของคู่มือนี้ สองกิจกรรมในบทนี้เป็นการสรุปการอนุมานทางสถิติซึ่งถือเป็นหัวใจของคู่มือนี้ และการประยุกต์แนวคิดสถิติสำหรับการวิจัย กิจกรรมแรกเป็นเพียงการสรุป เพื่อเน้นประเด็นที่สำคัญโดยนำมาเรียบเรียงให้ดูง่ายขึ้น พร้อมกับโยงกับสิ่งที่พบบ่อยในตำราทางสถิติคือตารางค่าสถิติ และคำถามเพื่อต่อยอดความเข้าใจ อีกทั้งเป็นเป็นเสมือน Post test ของคู่มือนี้ด้วย

ส่วนกิจกรรมที่สองเป็นการขยายความเข้าใจทั้งหมดจากที่ทำมาโดยตลอดในคู่มือนี้ แล้วนำไปประยุกต์ใช้กับคำถามวิจัยที่หลากหลาย ครอบคลุมลักษณะงานวิจัยที่พบบ่อย มีสรุปแนวทางวิเคราะห์ในภาคผนวก 1 ส่วนแนวทางการวิเคราะห์จากข้อมูลจริงที่ใช้เดินเรื่องในคู่มือนี้โดยใช้ STATA พร้อมตัวอย่างการนำเสนอค่าสถิติจากผลที่ได้จากคอมพิวเตอร์ นั้น มีสรุปในภาคผนวก 2

กิจกรรมที่ 6.1 : สรุปการอนุมานทางสถิติ

6.1 สรุปสูตรทั่วไปของการอนุมานทางสถิติ

6.1.1 การประมาณค่า (Parameter estimation) คือคำนวณช่วงเชื่อมั่น (Confidence interval)

สูตรทั่วไปเขียนเป็น

$$\text{ช่วงเชื่อมั่น} = \text{ค่าสถิติจากตัวอย่างที่ศึกษา} \pm [(\text{ค่าสัมประสิทธิ์}) \times (\text{SE ของค่าสถิติจากตัวอย่างที่ศึกษา})]$$

หรือใช้สัญลักษณ์ทางสถิติเป็น

$$[(1-\alpha) \times 100]\% \text{ CI} = \text{ค่าสถิติ} \pm [(\text{ค่าสัมประสิทธิ์}_{\alpha/2}) \times (\text{SE ของค่าสถิติจากตัวอย่างที่ศึกษา})]$$

ผู้วิจัยกำหนดค่า α ซึ่งหมายถึงโอกาสผิดพลาดในการประมาณค่า ทำให้รู้ค่าสัมประสิทธิ์ของการแจกแจงข้อมูลชุดนั้น (โดยเปิดตารางค่าสถิติ) สุดท้ายได้คำตอบเป็นค่าล่างและบน เป็นช่วงเชื่อมั่น

6.1.2 การทดสอบสมมติฐาน (Hypothesis testing) คือหาค่า p-value

สูตรทั่วไปเขียนเป็น

$$\text{ค่าสถิติทดสอบ} = \frac{\text{ค่าสถิติจากตัวอย่างที่ศึกษา} - \text{ค่าที่กำหนดไว้ใน } H_0}{\text{SE ของค่าสถิติจากตัวอย่างที่ศึกษา}}$$

จากสูตร จะเห็นว่ามีความที่กำหนัดไว้ใน H_0 เช่นในศาลยุติธรรม ต้องเริ่มจากการกล่าวหาโดยโจทก์ว่าจำเลยผิด ศาลจะเชื่อว่าจำเลยไม่ผิดไว้ก่อน (H_0) แล้วพยานหลักฐานจากโจทก์ จะใช้เป็นข้อพิสูจน์ว่าจำเลยผิดตามข้อกล่าวหา (H_A) ในงานวิจัยก็ทำนองเดียวกันนี้

ข้อสังเกต:

- (1) การ “พิสูจน์” ในศาล ถ้าได้ข้อสรุปว่า จำเลยไม่ผิดตามข้อกล่าวหา หมายถึงจำเลยอาจบริสุทธิ์ หรือผิดจริงแต่พยานหลักฐานไม่เพียงพอก็ได้

การ “ตรวจ” สิ่งส่งตรวจทางห้องปฏิบัติการ (ห้อง Lab) ให้ผลลบ หมายถึงตรวจไม่พบเชื้อที่ขอให้ตรวจ แต่สิ่งส่งตรวจนั้นอาจมีหรือไม่มีเชื้อนั้นก็ได

การ “ทดสอบ” สมมติฐานทางสถิติที่ให้ผล Non-significant หมายถึง เราไม่มีหลักฐานเพียงพอที่จะปฏิเสธสมมติฐาน ถ้าสมมติฐานเป็นจริง

แม้ทั้งสามกรณีให้ผลตรงข้ามกับที่กล่าว คือมีพยานหลักฐานเอาผิดจำเลยได้ ผล Lab เป็นผลบวก หรือทดสอบสมมติฐานให้ผล Significant ตามลำดับ แต่ข้อสรุปดังกล่าวอาจผิดก็ได้ คือจำคุกคนไม่ผิด ผล Lab เป็นผลบวกสูง (False positive) และปฏิเสธสมมติฐานที่เป็นจริง ตามลำดับ

ในการวิจัย สถิติจะช่วยบอกขนาดของ “โอกาสที่จะสรุปผิด” การสรุปผิดกรณีปฏิเสธสมมติฐานที่เป็นจริงเรียกว่า “ α Error” การสรุปผิดกรณียอมรับสมมติฐานที่ไม่เป็นจริงเรียกว่า “ β Error” แต่เนื่องจากเราไม่ได้ทดสอบว่าสมมติฐานถูก จึงไม่มีการตัดสินใจยอมรับสมมติฐาน ดังนั้นจึงบอกแทนด้วยโอกาสที่ตัดสินใจถูกกรณีสมมติฐานไม่เป็นจริงแล้วตัดสินว่าไม่เป็นจริงตาม คำนวณได้จาก “ $1 - \beta$ ” เรียกว่า “อำนาจการทดสอบ (Power)”

- (2) ถ้าไม่มีการเปรียบเทียบ ก็ไม่มีการทดสอบสมมติฐาน และไม่มีการกล่าวถึงอำนาจการทดสอบ กล่าวอีกนัยหนึ่ง การศึกษาที่มีวัตถุประสงค์หลักเพื่อประมาณค่า เช่น การสำรวจหาระดับน้ำตาลในเลือดโดยเฉลี่ยของคนไทย นั้น ไม่มีการกำหนดอำนาจของการศึกษาในการคำนวณขนาดตัวอย่าง และไม่มีค่า p-value ในรายงานผลการศึกษา

6.2 เปรียบเทียบแนวคิดของการประมาณค่าและการทดสอบสมมติฐาน

ในการประมาณค่าผู้วิจัยกำหนด α ซึ่งเป็นโอกาสผิดพลาดในการประมาณค่า แล้วหาค่าพารามิเตอร์ แต่ในการทดสอบสมมติฐานนั้นผู้วิจัยกำหนด α ซึ่งเป็นโอกาสผิดพลาดในการตัดสินใจปฏิเสธสมมติฐานที่เป็นจริง เพื่อเป็นเกณฑ์ในการตัดสินใจ โดยต้องมีการกำหนดหรือสมมติค่าพารามิเตอร์ขึ้นเพื่อเป็นตัวตั้งสำหรับการทดสอบ แล้วใช้ข้อมูลจากการวิจัยไปทดสอบ เพื่อได้คำตอบว่า ถ้าสมมติฐานเป็นจริงแล้ว โอกาสที่จะได้ข้อมูลตามที่ผู้วิจัยได้มาหรือสุดขั้วกว่านั้น เป็นเท่าใด (โอกาส คือ ความน่าจะเป็น คือ p-value) นำค่า p-value นี้ไปเปรียบเทียบกับเกณฑ์ α แล้วตัดสินใจปฏิเสธ หรือไม่ปฏิเสธสมมติฐาน	
ดังนั้น	การประมาณค่า มีทิศทางจากค่าความน่าจะเป็นไปหาค่าพารามิเตอร์
ส่วน	การทดสอบสมมติฐาน มีทิศทางจากค่าพารามิเตอร์ไปหาค่าความน่าจะเป็น

6.3 ค่า α และค่าสัมประสิทธิ์ของการแจกแจงแบบต่าง ๆ ที่ใช้บ่อย

ตารางแสดงค่า α และ Z ที่ใช้บ่อย และค่า t ที่ใช้ในตัวอย่างที่ผ่านมา			
ระดับนัยสำคัญ (α)	ระดับความเชื่อมั่น ($1 - \alpha$)	ค่าสัมประสิทธิ์ (Z)	ค่าสัมประสิทธิ์ (t ที่ $df=29$)
0.10	0.90	1.64	1.70
0.05	0.95	1.96	2.04
0.01	0.99	2.58	2.76

โปรดดูตัวเลขในวงรีเส้นคู่ในตารางข้างล่างนี้ เพื่อประกอบการทำความเข้าใจถึงที่มา

ตาราง Normal Distribution

(A) Z	(B) AREA BETWEEN MEAN AND Z	(C) AREA BEYOND Z	(A) Z	(B) AREA BETWEEN MEAN AND Z	(C) AREA BEYOND Z	(A) Z	(B) AREA BETWEEN MEAN AND Z	(C) AREA BEYOND Z
0.00	0.0000	0.5000	0.35	0.1368	0.3632	0.70	0.2580	0.2420
0.01	0.0040	0.4960	0.36	0.1406	0.3594	0.71	0.2611	0.2389
0.02	0.0080	0.4920	0.37	0.1443	0.3557	0.72	0.2642	0.2358
0.03	0.0120	0.4880	0.38	0.1480	0.3520	0.73	0.2673	0.2327
0.04	0.0160	0.4840	0.39	0.1517	0.3483	0.74	0.2704	0.2296
0.05	0.0199	0.4801	0.40	0.1554	0.3446	0.75	0.2734	0.2266
0.06	0.0239	0.4761	0.41	0.1591	0.3409	0.76	0.2764	0.2236
0.07	0.0279	0.4721	0.42	0.1628	0.3372	0.77	0.2794	0.2206
0.08	0.0319	0.4681	0.43	0.1664	0.3336	0.78	0.2823	0.2177
0.09	0.0359	0.4641	0.44	0.1700	0.3300	0.79	0.2852	0.2148
0.10	0.0398	0.4602	0.45	0.1736	0.3264	0.80	0.2881	0.2119
0.11	0.0438	0.4562	0.46	0.1772	0.3228	0.81	0.2910	0.2090
0.12	0.0478	0.4522	0.47	0.1808	0.3192	0.82	0.2939	0.2061
0.13	0.0517	0.4483	0.48	0.1844	0.3156	0.83	0.2967	0.2033
0.14	0.0557	0.4443	0.49	0.1879	0.3121	0.84	0.2995	0.2005
0.15	0.0596	0.4404	0.50	0.1915	0.3085	0.85	0.3023	0.1977
0.16	0.0636	0.4364	0.51	0.1950	0.3050	0.86	0.3051	0.1949
0.17	0.0675	0.4325	0.52	0.1985	0.3015	0.87	0.3078	0.1922
0.18	0.0714	0.4286	0.53	0.2019	0.2981	0.88	0.3106	0.1894
0.19	0.0753	0.4247	0.54	0.2054	0.2946	0.89	0.3133	0.1867
0.20	0.0793	0.4207	0.55	0.2088	0.2912	0.90	0.3159	0.1841
0.21	0.0832	0.4168	0.56	0.2123	0.2877	0.91	0.3186	0.1814
0.22	0.0871	0.4129	0.57	0.2157	0.2843	0.92	0.3212	0.1788
0.23	0.0910	0.4090	0.58	0.2190	0.2810	0.93	0.3238	0.1762
0.24	0.0948	0.4052	0.59	0.2224	0.2776	0.94	0.3264	0.1736

(A) Z	(B) AREA BETWEEN MEAN AND Z	(C) AREA BEYOND Z	(A) Z	(B) AREA BETWEEN MEAN AND Z	(C) AREA BEYOND Z	(A) Z	(B) AREA BETWEEN MEAN AND Z	(C) AREA BEYOND Z
0.25	0.0987	0.4013	0.60	0.2257	0.2743	0.95	0.3289	0.1711
0.26	0.1026	0.3974	0.61	0.2291	0.2709	0.96	0.3315	0.1685
0.27	0.1064	0.3936	0.62	0.2324	0.2676	0.97	0.3340	0.1660
0.28	0.1103	0.3897	0.63	0.2357	0.2643	0.98	0.3365	0.1635
0.29	0.1141	0.3859	0.64	0.2389	0.2611	0.99	0.3389	0.1611
0.30	0.1179	0.3821	0.65	0.2422	0.2578	1.00	0.3413	0.1587
0.31	0.1217	0.3783	0.66	0.2454	0.2546	1.01	0.3438	0.1562
0.32	0.1255	0.3745	0.67	0.2486	0.2514	1.02	0.3461	0.1539
0.33	0.1293	0.3707	0.68	0.2517	0.2483	1.03	0.3485	0.1515
0.34	0.1331	0.3669	0.69	0.2549	0.2451	1.04	0.3508	0.1492
0.35	0.1368	0.3632	0.70	0.2580	0.2420	1.05	0.3531	0.1469
1.05	0.3531	0.1469	1.40	0.4192	0.0808	1.75	0.4599	0.0401
1.06	0.3554	0.1446	1.41	0.4207	0.0793	1.76	0.4608	0.0392
1.07	0.3577	0.1423	1.42	0.4222	0.0778	1.77	0.4616	0.0384
1.08	0.3599	0.1401	1.43	0.4236	0.0764	1.78	0.4625	0.0375
1.09	0.3621	0.1379	1.44	0.4251	0.0749	1.79	0.4633	0.0367
1.10	0.3643	0.1357	1.45	0.4265	0.0735	1.80	0.4641	0.0359
1.11	0.3665	0.1335	1.46	0.4279	0.0721	1.81	0.4649	0.0351
1.12	0.3686	0.1314	1.47	0.4292	0.0708	1.82	0.4656	0.0344
1.13	0.3708	0.1292	1.48	0.4306	0.0694	1.83	0.4664	0.0336
1.14	0.3729	0.1271	1.49	0.4319	0.0681	1.84	0.4671	0.0329
1.15	0.3749	0.1251	1.50	0.4332	0.0668	1.85	0.4678	0.0322
1.16	0.3770	0.1230	1.51	0.4345	0.0655	1.86	0.4686	0.0314
1.17	0.3790	0.1210	1.52	0.4357	0.0643	1.87	0.4693	0.0307
1.18	0.3810	0.1190	1.53	0.4370	0.0630	1.88	0.4699	0.0301
1.19	0.3830	0.1170	1.54	0.4382	0.0618	1.89	0.4706	0.0294
1.20	0.3849	0.1151	1.55	0.4394	0.0606	1.90	0.4713	0.0287
1.21	0.3869	0.1131	1.56	0.4406	0.0594	1.91	0.4719	0.0281
1.22	0.3888	0.1112	1.57	0.4418	0.0582	1.92	0.4726	0.0274
1.23	0.3907	0.1093	1.58	0.4429	0.0571	1.93	0.4732	0.0268
1.24	0.3925	0.1075	1.59	0.4441	0.0559	1.94	0.4738	0.0262
1.25	0.3944	0.1056	1.60	0.4452	0.0548	1.95	0.4744	0.0256
1.26	0.3962	0.1038	1.61	0.4463	0.0537	1.96	0.4750	0.0250
1.27	0.3980	0.1020	1.62	0.4474	0.0526	1.97	0.4756	0.0244
1.28	0.3997	0.1003	1.63	0.4484	0.0516	1.98	0.4761	0.0239
1.29	0.4015	0.0985	1.64	0.4495	0.0505	1.99	0.4767	0.0233
1.30	0.4032	0.0968	1.65	0.4505	0.0495	2.00	0.4772	0.0228
1.31	0.4049	0.0951	1.66	0.4515	0.0485	2.01	0.4778	0.0222
1.32	0.4066	0.0934	1.67	0.4525	0.0475	2.02	0.4783	0.0217
1.33	0.4082	0.0918	1.68	0.4535	0.0465	2.03	0.4788	0.0212
1.34	0.4099	0.0901	1.69	0.4545	0.0455	2.04	0.4793	0.0207
1.35	0.4115	0.0885	1.70	0.4554	0.0446	2.05	0.4798	0.0202
1.36	0.4131	0.0869	1.71	0.4564	0.0436	2.06	0.4803	0.0197
1.37	0.4147	0.0853	1.72	0.4573	0.0427	2.07	0.4808	0.0192
1.38	0.4162	0.0838	1.73	0.4582	0.0418	2.08	0.4812	0.0188
1.39	0.4177	0.0823	1.74	0.4591	0.0409	2.09	0.4817	0.0183
1.40	0.4192	0.0808	1.75	0.4599	0.0401	2.10	0.4821	0.0179

(A) Z	(B) AREA BETWEEN MEAN AND Z	(C) AREA BEYOND Z	(A) Z	(B) AREA BETWEEN MEAN AND Z	(C) AREA BEYOND Z	(A) Z	(B) AREA BETWEEN MEAN AND Z	(C) AREA BEYOND Z
2.10	0.4821	0.0179	2.40	0.4918	0.0082	2.90	0.4981	0.0019
2.11	0.4826	0.0174	2.41	0.4920	0.0080	2.92	0.4982	0.0018
2.12	0.4830	0.0170	2.42	0.4922	0.0078	2.94	0.4984	0.0016
2.13	0.4834	0.0166	2.43	0.4925	0.0075	2.96	0.4985	0.0015
2.14	0.4838	0.0162	2.44	0.4927	0.0073	2.98	0.4986	0.0014
2.15	0.4842	0.0158	2.45	0.4929	0.0071	3.00	0.4987	0.0013
2.16	0.4846	0.0154	2.46	0.4931	0.0069	3.02	0.4987	0.0013
2.17	0.4850	0.0150	2.47	0.4932	0.0068	3.04	0.4988	0.0012
2.18	0.4854	0.0146	2.48	0.4934	0.0066	3.06	0.4989	0.0011
2.19	0.4857	0.0143	2.49	0.4936	0.0064	3.08	0.4990	0.0010
2.20	0.4861	0.0139	2.50	0.4938	0.0062	3.10	0.4990	0.0010
2.21	0.4864	0.0136	2.52	0.4941	0.0059	3.12	0.4991	0.0009
2.22	0.4868	0.0132	2.54	0.4945	0.0055	3.14	0.4992	0.0008
2.23	0.4871	0.0129	2.56	0.4948	0.0052	3.16	0.4992	0.0008
2.24	0.4875	0.0125	2.58	0.4951	0.0049	3.20	0.4993	0.0007
2.25	0.4878	0.0122	2.60	0.4953	0.0047	3.25	0.4994	0.0006
2.26	0.4881	0.0119	2.62	0.4956	0.0044	3.30	0.4995	0.0005
2.27	0.4884	0.0116	2.64	0.4959	0.0041	3.35	0.4996	0.0004
2.28	0.4887	0.0113	2.66	0.4961	0.0039	3.40	0.4997	0.0003
2.29	0.4890	0.0110	2.68	0.4963	0.0037	3.45	0.4997	0.0003
2.30	0.4893	0.0107	2.70	0.4965	0.0035	3.50	0.4998	0.0002
2.31	0.4896	0.0104	2.72	0.4967	0.0033	3.55	0.4998	0.0002
2.32	0.4898	0.0102	2.74	0.4969	0.0031	3.60	0.4998	0.0002
2.33	0.4901	0.0099	2.76	0.4971	0.0029	3.65	0.4999	0.0001
2.34	0.4904	0.0096	2.78	0.4973	0.0027	3.70	0.4999	0.0001
2.35	0.4906	0.0094	2.80	0.4974	0.0026	3.75	0.4999	0.0001
2.36	0.4909	0.0091	2.82	0.4976	0.0024	3.80	0.4999	0.0001
2.37	0.4911	0.0089	2.84	0.4977	0.0023	3.85	0.4999	0.0001
2.38	0.4913	0.0087	2.86	0.4979	0.0021	3.90	0.4999	0.00005
2.39	0.4916	0.0084	2.88	0.4980	0.0020	3.95	0.4999	0.00004
2.40	0.4918	0.0082	2.90	0.4981	0.0019	4.00	0.4999	0.00003

ตาราง t-Distribution

DF	Level of Significance					
	0.10	0.05	0.025	0.01	0.005	0.0005
1	3.078	6.314	12.706	31.821	63.657	63.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587

DF	Level of Significance					
	0.10	0.05	0.025	0.01	0.005	0.0005
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.992
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

ตาราง χ^2 Distribution

DF	Probability (α)										
	0.995	0.990	0.975	0.950	0.900	0.500	0.100	0.050	0.025	0.010	0.005
1	0.00	0.00	0.00	0.00	0.02	0.45	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	1.39	4.61	5.99	7.38	19.21	10.60
3	0.07	0.11	0.22	0.35	0.58	2.37	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.38	28.41	31.41	34.17	37.57	40.00

DF	Probability (α)										
	0.995	0.990	0.975	0.950	0.900	0.500	0.100	0.050	0.025	0.010	0.005
21	8.03	8.90	10.28	11.50	13.24	20.38	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.80	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.42	104.22
100	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

6.4 ความสัมพันธ์ระหว่าง Z-distribution และ t-distribution และ χ^2 -distribution

ตัวอย่าง : ความน่าจะเป็นทั้งสองทาง (two-tailed probability) จาก 3 ตารางค่าสถิติข้างต้น

$$\chi^2_{\alpha=0.05; df=1} = Z^2_{\alpha=0.05/2} = t^2_{\alpha=0.05/2; df=\infty} = 1.96^2 = 3.84$$

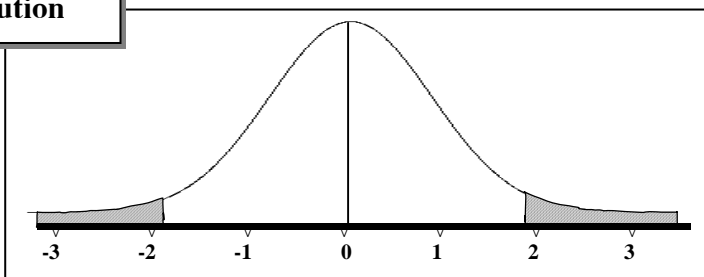
คำถาม : กราฟทั้งสามข้างล่างนี้ เป็นการแจกแจงความน่าจะเป็นของสามค่าสถิติ พื้นที่แรเงาของแต่ละกราฟมีสัดส่วนเท่ากับ 5% ของพื้นที่ทั้งหมดของกราฟนั้น

แกนตั้งแทน แกนนอนแทน

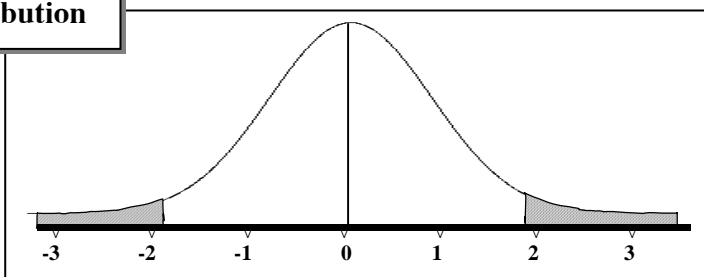
จงเขียนกราฟทับ t-distribution โดยให้ $df = 29$

จงเขียนกราฟทับ χ^2 -distribution โดยให้ $df = 100$

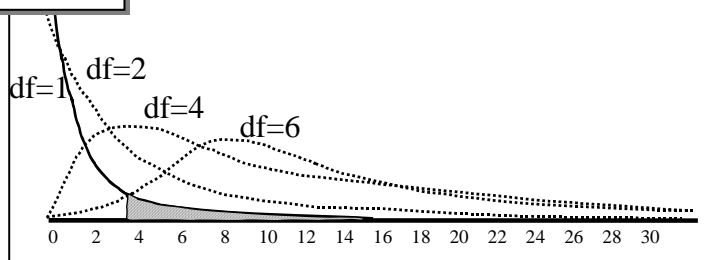
Z-distribution



$t_{df=\infty}$ -distribution



χ^2 -distribution

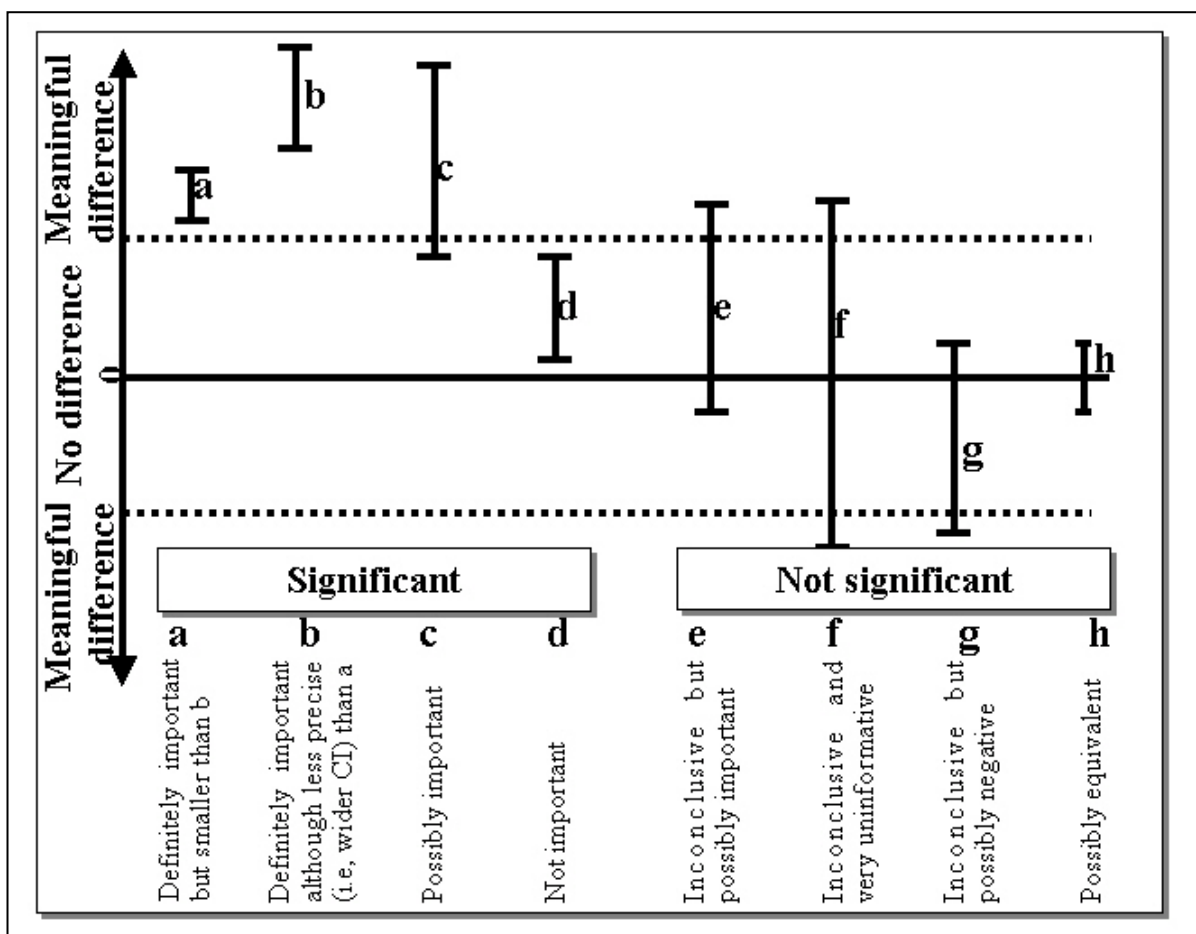


6.5 การประมาณค่า การทดสอบสมมติฐาน และความสำคัญทางคลินิกและการแพทย์และสาธารณสุข

แผนภูมิแสดงช่วงเชื่อมั่น (95% CI) ของผลต่างของสองสัดส่วน จาก 8 การศึกษา

สมมติว่าต่อไปนี้เป็น 8 ผลการศึกษาที่เป็นไปได้ (a ถึง h) จากการศึกษาประสิทธิภาพในการรักษาโรคของยาสองขนาน เพื่อเปรียบเทียบอัตราการหายจากโรค ค่าสถิติที่นำมาอนุมานคือผลต่างอัตราการหายจากการใช้ยาใหม่ลบด้วยอัตราการหายจากการใช้ยาเก่า

(Adapted from: Armitage, P. and Berry, G. *Statistical methods in medical research*. 3rd edition. Blackwell Scientific Publications, Oxford. 1994. page 99)



อภิปราย

6.6 ทดสอบความเข้าใจการประมาณค่า การทดสอบสมมติฐาน (Post test)

จากบทสรุปงานวิจัยหนึ่ง ความว่า

“การศึกษาเชิงทดลองทางการแพทย์โดยมีกลุ่มควบคุมเพื่อทราบว่ายาใหม่ชนิดหนึ่งมีประสิทธิผลในการรักษาโรคหรือไม่ พบว่ากลุ่มทดลองซึ่งได้รับยาชนิดดังกล่าวมีอัตราการหายจากโรคสูงกว่ากลุ่มควบคุมที่ใช้ยาหลอก (placebo) อย่างมีนัยสำคัญทางสถิติ ($p\text{-value} < 0.05$)”

ท่านเห็นด้วยกับคำกล่าวใดต่อไปนี้

- [] 1. การศึกษานี้ได้พิสูจน์ให้เห็นแล้วว่ายาใหม่นี้ดีกว่ายาหลอก
- [] 2. ถ้าความเป็นจริงคือ “ยาใหม่ไม่มีประสิทธิผลในการรักษาโรค” แล้ว โอกาสที่จะได้ผลการศึกษาตามที่กล่าวข้างต้นมีน้อยกว่า 5%
- [] 3. ประสิทธิผลในการรักษาโรคของยาใหม่ที่ได้จากการศึกษาครั้งนี้สูงมาก กล่าวคือโอกาสที่จะพบว่ายานี้ไม่ดีไปกว่ายาหลอกมีเพียง 5 % เท่านั้น
- [] 4. ไม่รู้จริงๆ ว่า $p\text{-value}$ คืออะไร จึงไม่อยากจะเดาคำตอบ

ดัดแปลงจาก Wulff, H.R., Andersen, B., Brandenhoff, P., and Guttler, F. (1987) What do doctors know about statistics?. *Statistics in Medicine*, **6**, 3–10.

อภิปราย

ท่านได้รับการร้องขอให้ช่วยเขียนสรุปผลดังกล่าว โปรดเขียนให้อยู่ในรูปที่ผู้วิจัยสามารถตัดแปะในสรุปผลการศึกษาของเขาได้ โดยสมมติตัวเลขใส่ลงไปในงานด้วย

กิจกรรมที่ 6.2 : ประยุกต์แนวคิดทางสถิติกับคำถามวิจัยต่าง ๆ

ต่อไปนี้เป็นตัวอย่างวัตถุประสงค์การวิจัย สมมติให้แต่ละวัตถุประสงค์มาจากแต่ละงานวิจัย และใช้แบบสอบถามนักศึกษาตามที่แนบท้ายนี้ ในการตอบคำถามให้ใช้ชื่อตัวแปรย่อด้วยเพื่อโยงได้ง่ายกับคำสั่ง STATA ท่านสามารถพลิกดูวิธีการวิเคราะห์ โดยใช้ STATA ในภาคผนวก 2 ในการนี้ใช้ข้อมูลนักศึกษาทั้ง 143 คนเป็นตัวอย่าง

1. เพื่อประมาณค่าเกรดเฉลี่ยสะสมของนักศึกษาในมหาวิทยาลัย

- ตัวแปรตามคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน
- ตัวแปรต้นคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ถ้าเป็นตัวแปรแฉกนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)
- แนวทางการวิเคราะห์โดยสังเขป:
.....

2. เพื่อศึกษาว่าเกรดเฉลี่ยสะสมในมหาวิทยาลัยของนักศึกษามากกว่า 2.0 หรือไม่

- ตัวแปรตามคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน
- ตัวแปรต้นคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ถ้าเป็นตัวแปรแฉกนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)
- แนวทางการวิเคราะห์โดยสังเขป:
.....

3. เพื่อศึกษาว่าเกรดเฉลี่ยสะสมในมหาวิทยาลัยของนักศึกษาชายและหญิงแตกต่างกันหรือไม่

➤ ตัวแปรตามคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉงนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน

➤ ตัวแปรต้นคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉงนับ
ถ้าเป็นตัวแปรแฉงนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)

➤ แนวทางการวิเคราะห์โดยสังเขป:
.....

4. เพื่อศึกษาว่าอัตราการเต้นของชีพจรของนักศึกษาก่อนและหลังวิ่งกับที่หนึ่งนาทีแตกต่างกันหรือไม่

➤ ตัวแปรตามคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉงนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน

➤ ตัวแปรต้นคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉงนับ
ถ้าเป็นตัวแปรแฉงนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)

➤ แนวทางการวิเคราะห์โดยสังเขป:
.....

6. เพื่อศึกษาความสัมพันธ์ระหว่างน้ำหนักกับอัตราชีพจรขณะพัก

➤ ตัวแปรตามคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉงนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน

➤ ตัวแปรต้นคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉงนับ
ถ้าเป็นตัวแปรแฉงนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)

➤ แนวทางการวิเคราะห์โดยสังเขป:
.....

7. เพื่อศึกษาว่าน้ำหนักสามารถใช้ทำนายอัตราชีพจรขณะพักได้หรือไม่ อย่างไร

➤ ตัวแปรตามคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉงนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน

➤ ตัวแปรต้นคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉงนับ
ถ้าเป็นตัวแปรแฉงนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)

➤ แนวทางการวิเคราะห์โดยสังเขป:
.....

8. ผู้วิจัยสร้างตัวแปรขึ้นมาใหม่ชื่อ "การเจ็บป่วย" ให้มีค่าเป็น *ไม่ป่วย* กับ *ป่วย* โดยถือว่านักศึกษา *ไม่ป่วย* เมื่อตัวแปรในข้อ 14 ของแบบสอบถาม คือสถานสุขภาพมีค่า 1 (เยี่ยม) และ 2 (ดี) และ *ป่วย* คือสถานสุขภาพมีค่า 3 (ปานกลาง) และ 4 (แย่) วัตถุประสงค์การวิจัยคือ เพื่อประมาณค่าความชุก (Prevalence) ของการป่วย (ค่าร้อยละ)

- ตัวแปรตามคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน
- ตัวแปรต้นคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ถ้าเป็นตัวแปรแฉกนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)
- แนวทางการวิเคราะห์โดยสังเขป:
.....

9. เพื่อศึกษาว่าความชุกการป่วยของนักศึกษาแตกต่างจากของประชากรทั่วไป ซึ่งเท่ากับ 5% หรือไม่

- ตัวแปรตามคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน
- ตัวแปรต้นคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ถ้าเป็นตัวแปรแฉกนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)
- แนวทางการวิเคราะห์โดยสังเขป:
.....

10. เพื่อศึกษาว่าความชุกการเจ็บป่วยในกลุ่มนักศึกษาชายและกลุ่มนักศึกษาหญิงแตกต่างกันหรือไม่

- ตัวแปรตามคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน
- ตัวแปรต้นคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ถ้าเป็นตัวแปรแฉกนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)
- แนวทางการวิเคราะห์โดยสังเขป:
.....

11. เพื่อศึกษาความสัมพันธ์ระหว่างการเจ็บป่วยกับเพศ

- ตัวแปรตามคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน
- ตัวแปรต้นคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ถ้าเป็นตัวแปรแฉกนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)
- แนวทางการวิเคราะห์โดยสังเขป:
.....

12. เพื่อศึกษาว่าสุขภาพจิต (เกิดอาการเครียด) ของนักศึกษาก่อนและหลังเข้าเรียนมหาวิทยาลัยแตกต่างกันหรือไม่

- ตัวแปรตามคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน
- ตัวแปรต้นคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ถ้าเป็นตัวแปรแฉกนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)
- แนวทางการวิเคราะห์โดยสังเขป:
.....

13. เพื่อศึกษาว่าภูมิลำเนาที่มีความสัมพันธ์กับความชุกการเจ็บป่วยหรือไม่

- ตัวแปรตามคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน
- ตัวแปรต้นคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ถ้าเป็นตัวแปรแฉกนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)
- แนวทางการวิเคราะห์โดยสังเขป:
.....

14. เพื่อศึกษาปัจจัยที่มีผลต่อเกรดเฉลี่ยสะสมในมหาวิทยาลัย

- ตัวแปรตามคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน
- ตัวแปรต้นคือ:
.....
- แนวทางการวิเคราะห์โดยสังเขป:
.....

15. เพื่อศึกษาปัจจัยที่มีผลต่อ การเจ็บป่วย

- ตัวแปรตามคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน
- ตัวแปรต้นคือ:
- แนวทางการวิเคราะห์โดยสังเขป:
.....

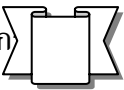
16. ประยุกต์แนวคิดทางสถิติกับการคำนวณขนาดตัวอย่าง

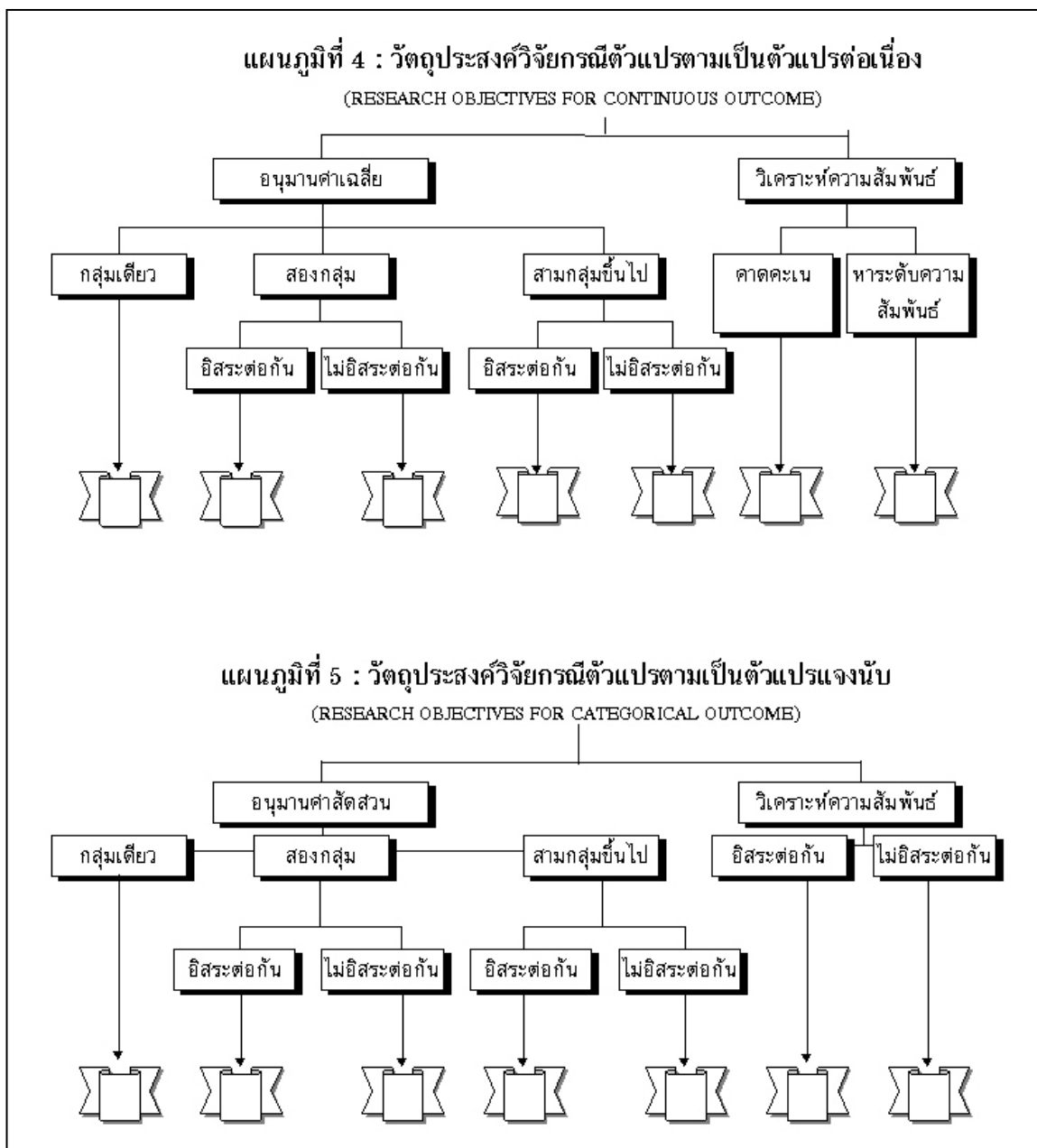
โจทย์: ผู้เข้ารับการอบรมได้รับมอบหมายให้ทำการศึกษาวิจัยเพื่อเปรียบเทียบน้ำหนักของนักศึกษาชายกับนักศึกษาหญิงที่มหาวิทยาลัยเชียงใหม่ จงคำนวณขนาดตัวอย่างสำหรับการศึกษาวิจัยดังกล่าว

- ตัวแปรตามคือ:..... เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน
- ตัวแปรต้นคือ: เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ
ถ้าเป็นตัวแปรแฉกนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)
- แนวทางการคำนวณขนาดตัวอย่างโดยสังเขป:
.....

หมายเหตุ : ข้อสุดท้ายนี้ คำถามเกี่ยวกับ แนวทางการคำนวณขนาดตัวอย่างโดยสังเขปนั้น โปรดบอกเพียงแนวทางการค้นหาสูตรการคำนวณขนาดตัวอย่างที่เหมาะสม

กิจกรรมที่ 6.3 : แนวทางการเลือกใช้สถิติกับคำถามวิจัยต่าง ๆ

จากตัวอย่างวัตถุประสงค์การวิจัยข้อ 1 ถึง 15 ในกิจกรรมที่ 6.2 จงนำหมายเลขหน้า
วัตถุประสงค์การวิจัยดังกล่าว ใส่ใน  สำหรับแต่ละกรณีในแผนภูมิที่ 4 และ 5 ต่อไปนี้



บรรณานุกรม

- บัณฑิต ถิ่นคำรพ. (2543). การวิเคราะห์ข้อมูลการวิจัยทางวิทยาศาสตร์สุขภาพ โดยใช้การถดถอยลอจิสติก. ขอนแก่น. คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น.
- บัณฑิต ถิ่นคำรพ. (2540). เครือข่ายอินเทอร์เน็ต : ชุมทรัพย์ทางปัญญาในยุคโลกาภิวัตน์. วารสารระบดวิทยาภาคตะวันออกเฉียงเหนือ. 2(2): 1 - 5
- บัณฑิต ถิ่นคำรพ. (2540). ระบดวิทยา : เครื่องมืออันทรงพลังทางปัญญา. วารสารระบดวิทยาภาคตะวันออกเฉียงเหนือ. 2(1): 1 - 3
- บัณฑิต ถิ่นคำรพ. (2540). สถิติสำหรับการวิจัยทางสาธารณสุข : ตอนที่ 1 แนวทางการเขียนเรื่องที่เกี่ยวข้องกับสถิติในโครงร่างวิจัย. วารสารระบดวิทยาภาคตะวันออกเฉียงเหนือ. 2(3): 17 - 22
- บัณฑิต ถิ่นคำรพ. (2541.). สถิติสำหรับการวิจัยทางสาธารณสุข : ตอนที่ 2 วิธีการนำเสนอข้อมูลทางสถิติในรายงานวิจัย (ภาคที่ 1). วารสารระบดวิทยาภาคตะวันออกเฉียงเหนือ. 3(1): 17 - 20
- บัณฑิต ถิ่นคำรพ. (2541). สถิติสำหรับการวิจัยทางสาธารณสุข : ตอนที่ 3 วิธีการนำเสนอข้อมูลทางสถิติในรายงานวิจัย (ภาคที่ 2). วารสารระบดวิทยาภาคตะวันออกเฉียงเหนือ. 3(2):26-29
- บัณฑิต ถิ่นคำรพ. (2542). การวางแผนประมวผล และวิเคราะห์ข้อมูล. วารสารจักษุวิทยาสาธารณสุข. 13(1): 85 - 96.
- บัณฑิต ถิ่นคำรพ. (2542). การอ่านบทความวิจัยอย่างมีวิจารณ์ญาณ. วารสารระบดวิทยาภาคตะวันออกเฉียงเหนือ. 4(2) : 1-5.
- บัณฑิต ถิ่นคำรพ. (2541). ความสำคัญและความจำเป็นของการวิเคราะห์ข้อมูลตัวแปรเชิงพหุในการวิจัยทางสาธารณสุข. วารสารระบดวิทยาภาคตะวันออกเฉียงเหนือ. 3(3): 1-13
- บัณฑิต ถิ่นคำรพ. ภาพรวมสถิติ. วารสารระบดวิทยาภาคตะวันออกเฉียงเหนือ, 2542. 4(1) : 1-11.
- อรุณ จิรวัดน์กุล, มานี เหล่าไพบูลย์, จิราพร เขียวอยู่, ยุพา ถาวรพิทักษ์, จารุวรรณ โชคคณาพิทักษ์, บัณฑิต ถิ่นคำรพ, นิคม ถนอมเสียง. (2542). ชีวสถิติ. ขอนแก่น. โรงพิมพ์คลังนานาวิทยา.

- Armitage, P., and Berry, G. (1994). *Statistical methods in medical research*. Oxford: Blackwell Scientific Publications.
- Bland, M. (1998). *An introduction of medical statistics*. Oxford: Oxford University Press.
- Daniel, W. W. (1995). *Biostatistics: A foundation for analysis in the health sciences*. 6th ed. New York: John Wiley & Sons.
- Daniel, W.W. (1991). *Biostatistics: A foundation for analysis in the health sciences*. 5th ed. New York: John Wiley & Sons.
- Dawson-Souder, B., and Trapp, R.G. (1990). *Basic and clinical biostatistics*. Norwalk, Connecticut: Appleton & Lange.
- Diggle, P.J., Liang, K-Y, and Zeger, S.L. (1994). *Analysis of longitudinal data*. New York: Oxford University Press.
- Everitt, B. S. (1995). The analysis of repeated measures: A practical review with examples. *The Statistician*, **44**(1), 113-135.
- Everitt, B.S. (1977). *The Analysis of Contingency Tables*. London: Chapman and Hall.
- Everitt, B.S. (1994). *Statistical methods for medical investigations*. 3rd edition. London: Edward Arnold.
- Fienberg, S.E. (1980). *The analysis of cross-classified categorical data*. 2nd edition. Cambridge: The MIT Press.
- Fleiss, J.L. (1981). *Statistical methods for rates and proportions*. 2nd edition. New York: John Willey & Sons.
- Guyatt, G., Jaeschke, R., Heddle, N., Cook, D., Shannon, H., and Walter S. (1995). Interpreting study results: confidence intervals. *Canadian Medical Association Journal*. **152**:169-173.
- Hosme, D.W. Lemeshow, S. (1989). *Applied Logistic Regression*. New York: John Wiley & Sons.
- Jaeschke, R., Guyatt, G., Shannon, H., Walter, S. Cook, D. Heddle, N. (1995). Assessing the effects of treatment: measures of association . *Canadian Medical Association Journal*. **152**: 351-357
- Kleinbaum, D.G. (1994). *Logistic Regression: A self-learning text*. New York Springer-Verlag.
- Kleinbaum, D.G. (1996). *Survival analysis: A self-learning text*. New York: Springer-Verlag.
- Kleinbaum, D.G., Kupper, L.L., and Morgenstern, H. (1982). *Epidemiologic research: principles and qualitative methods*. London: Lifetime Learning Publications.
- Kleinbaum, D.G., Kupper, L.L., Muller, K.E, and Nizam, A. (1998). *Applied regression analysis and other multivariable methods*. Pacific Grove: Duxbury Press.
- Landis, J.R. and Cock, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*. **33**:159-174.

- Lang, TA., Secic, M. (1997). *How to report statistics in medicine: annotated guidelines for authors, editors, and reviewers*. Philadelphia: American College of Physician.
- Liang, K-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**(1):13-22.
- Meinert, C.L. and Tonascia, S. (1986). *Clinical trials, design, conduct, and analysis*. New York: Oxford University Press.
- Neuhaus J. (1992) Statistical methods for longitudinal and clustered designs with binary responses. *Statistical Methods in Medical Research*. 1:249-37.
- Rabe-Hesketh, S. and Everitt, B. (1998). *A handbook of statistical analysis using Stata*. London: Chapman & Hall/CRC.
- Rosner, Bernard. (1995). *Fundamentals of Biostatistics*. 4th ed. Belmont, California: Duxbury Press.
- StataCorp. (1999). *Stata statistical software: Release 6.0*. College Station. TX: Stata Corporation.
- Wonnacott TH., Wonnacott RJ. (1990). *Introductory statistics*. 5th Edition. New York: John Willey & Sons .
- Wulff, H.R., Andersen, B., Brandenhoff, P., and Guttler, F. (1987). What do doctors know about statistics?. *Statistics in Medicine*, **6**, 3-10.

ภาคผนวก 1

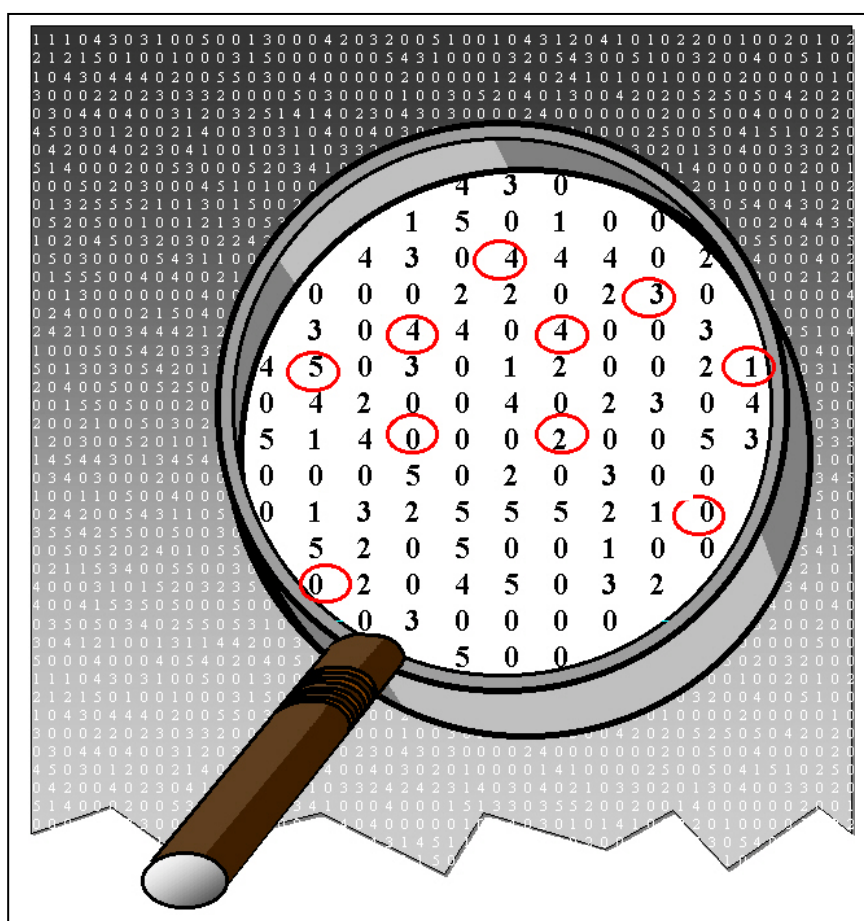
เฉลยคู่มือปฏิบัติการ

ในภาคผนวก 1 นี้ บางข้อจำเป็นต้องมีข้อความเดิม ดังนั้นจึงให้อักษรลักษณะที่ต่างไปจากข้อความเดิมแสดงค่าเฉลย การวิเคราะห์ข้อมูล เฉพาะอย่างยิ่งสำหรับคำตอบของทั้ง 16 คำถามของบทที่ 6 ปฏิบัติการที่ 6.2 นั้น คูรายละเอียดได้จากภาคผนวก 2

คำเฉลยบทที่ 1 ภาพรวมของวิชาชีวสถิติ

1.1 ดูคำตอบและคำอธิบายบทที่ 6 ข้อ 6.6

1.2 ข้อมูลที่ถูกสุ่มเป็นตัวอย่าง คือที่มีวงกลมล้อมรอบ จำนวน 10 ข้อมูล



1.3 เขียนข้อมูลที่สุ่มได้ลงในกรอบต่อไปนี้ ($n = 10$)

4 3 4 4 5 1 0 2 5 5

1.4 แสดงข้อมูลให้อยู่ในรูปตารางต่อไปนี้

เลขที่	จำนวนบุหรี่ที่สูบ (มวนต่อวัน)
1	4
2	3
3	4
4	4
5	5
6	1
7	0
8	2
9	0
10	0

1.5

หัวตารางเหล่านี้เรียกว่า (☐ ข้อมูล / ☒ ตัวแปร)

ตัวเลขที่กรอกลงในตารางเรียกว่า (☒ ข้อมูล / ☐ ตัวแปร)

ID	CIGAR	AGE	SEX
1	4	25	1
2	3	32	1
3	4	28	1
4	4	29	1
5	5	22	1
6	1	20	2
7	0	54	2
8	2	23	2
9	0	44	2
10	0	52	2

สมมติ
ตัวเลข
ข้าง
เอง

1.6 -

1.7 -

1.8

ID:.....	
แบบสำรวจการสูบบุหรี่ตามโครงการรณรงค์การเลิกบุหรี่ในโรงงานอุตสาหกรรม	
คำถามและคำตอบ	รหัส
ชื่อ	
1. จำนวนบุหรี่ที่สูบต่อวันโดยเฉลี่ย มวน	CIGAR[][]
2. อายุ ปี	AGE[][]
3. เพศ []1. ชาย []2. หญิง	SEX[]

1.9.1 -

1.9.2 -

1.9.2.1 -

1.9.2.2 ให้ใช้ข้อมูลจาก 1.9.2.1 เติมลงในตารางข้อมูลต่อไปนี้

ID	AGE	SEX	CIGAR1	CIGAR2
1	40	1	0	0
2	23	1	5	0
3	35	2	0	0
--- ข้าม n - 4 ชุดข้อมูล ---				
n	42	2	0	5



ID	CIGAR	VISIT	AGE	SEX
1	0	1	40	1
1	0	2	40	1
2	5	1	23	1
2	0	2	23	1
3	0	1	35	2
3	0	2	35	2
--- ข้าม n - 4 ชุดข้อมูล ---				
n	0	1	42	2
n	5	2	42	2

1.10 -

1.11 -

1.11.1 -

1.11.1.1

สถิติที่รายงานมีอะไรบ้าง? ค่า p-value อย่างเดียว

รายงานดังกล่าวขาดอะไร? สิ่งที่สำคัญที่ขาดไปคือขนาดของผลการรักษา

(Magnitude of effect)

คืออัตราการหายจากการใช้ยาใหม่สูงกว่ายาหลอกมากน้อยเท่าใด ดังนั้นอย่างน้อยใน รายงานสรุปผลการศึกษาคควรมี 95% Confidence interval ของผลต่างอัตราการหาย ตามด้วย ค่า p-value

1.11.1.2 ไม่สามารถตัดสินใจได้ เพราะในการศึกษาที่มีขนาดตัวอย่างที่ใหญ่มากๆ มักจะให้ผล Significant แม้อัตราการหายจากยาใหม่ต่างจากยาหลอกเพียงเล็กน้อย จนไม่มีความสำคัญทางการแพทย์

1.11.2 -

1.11.2.1 ประชากรเป้าหมายคือ คนงานทุกคนในทุกโรงงานอุตสาหกรรมในจังหวัด

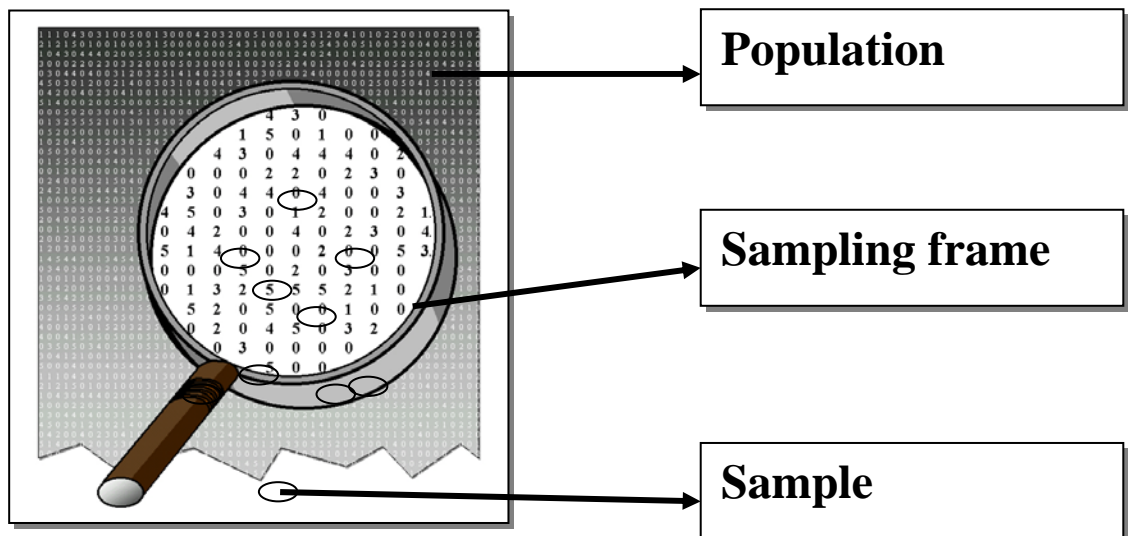
1.11.2.2 กรอบการสุ่มตัวอย่างคือ รายชื่อคนงานของทุกโรงงานที่จดทะเบียนกับอุตสาหกรรมจังหวัด

1.11.2.3 ตัวอย่างคือ คนงานที่ถูกสุ่มเลือก

1.11.2.4 ขนาดตัวอย่างเท่ากับ 10

1.11.2.5 วิธีการสุ่มตัวอย่างคือ การสุ่มอย่างง่าย (Simple random sampling)

1.11.2.6 ให้กรอกคำอธิบายลงกรอบ ให้สอดคล้องกับตำแหน่งภาพจำลองประชากรและตัวอย่าง



1.11.3 -

1.11.4 -

1.11.4.1 ก. (☒)CIGAR []AGE []SEX)

ข. (☒)Continuous []Categorical)

ค. ([]CIGAR []AGE []SEX ☒)ไม่มี)

ง. ([]Continuous []Categorical ☒)ไม่เกี่ยวข้อง)

จ. แผนภูมิที่ 2

ฉ. ① \bar{x} , (95% CI) ② p -value โดยใช้ t -test

1.11.4.2

ก. (☒ CIGAR [] AGE [] SEX)

ข. (☒ Continuous [] Categorical)

ค. ([] CIGAR ☒ AGE [] SEX [] ไม่มี)

ง. (☒ Continuous [] Categorical [] ไม่เกี่ยวข้อง)

จ. แผนภูมิที่ 2 ① Pearson correlation coefficient : r , (95% CI)

② p -value โดยใช้ t -test

1.11.4.3

ก. (☒ CIGAR [] AGE [] SEX)

ข. ([] Continuous ☒ Categorical)

ค. ([] CIGAR [] AGE ☒ SEX [] ไม่มี)

ง. ([] Continuous ☒ Categorical [] ไม่เกี่ยวข้อง)

จ. แผนภูมิที่ 3

ฉ. ① $P_1 - P_2$, (95% CI) ② p -value โดยใช้ Z-test หรือ χ^2 -test

1.11.4.4

ก. (☒ CIGAR [] AGE [] SEX)

ข. ([] Continuous ☒ Categorical)

ค. ([] CIGAR ☒ AGE ☒ SEX [] ไม่มี)

ง. (☒ Continuous ☒ Categorical [] ไม่เกี่ยวข้อง)

จ. แผนภูมิที่ 3

ฉ. Unconditional logistic regression

1.11.4.5

ก. (☒ CIGAR [] AGE [] SEX)

ข. (☒ Continuous [] Categorical)

ค. ([] CIGAR ☒ AGE ☒ SEX [] ไม่มี)

ง. (☒ Continuous ☒ Categorical [] ไม่เกี่ยวข้อง)

จ. แผนภูมิที่ 2

ฉ. Multiple regression

คำเฉลยบทที่ 2 การเก็บรวบรวมข้อมูล

2.1 -

2.2

ขั้นที่ 1 :

คำแนะนำ : ให้สุ่มเลือกตัวเลขเริ่มต้นก่อน จากนั้นเขียนวงกลมล้อมรอบเลขที่สุ่มได้เพียงตัวเดียว คือ **8**. จากนั้นนับรวมเอาตัวเลขที่อยู่ด้านขวาอีก **2**.... ตัว เพื่อให้ได้เป็นตัวเลข **3**.... หลัก แล้วทำเครื่องหมายคั่นไว้เป็นช่วงๆ ติดต่อกันไปในทิศทางจากขวาไปซ้าย และจากบนลงล่าง (ถ้าสุดหน้ากระดาษ ให้เริ่มต้นบรรทัดบนสุด) ได้ตัวเลขชุดใดที่มีค่าอยู่ระหว่าง 1 ถึง 143 ให้บันทึกไว้ในตารางที่ 2 ถ้าซ้ำให้ข้าม จนได้ครบ 30 หมายเลข จึงถือว่าเสร็จสิ้นการสุ่มตัวอย่าง

ตารางที่ 1 ตัวเลขเรียงคือชุดของการสุ่ม แบ่งชุดละ 3 หลักด้วยเครื่องหมาย , ตัวที่บคือชุดที่เป็นตัวเลขอยู่ในช่วง 1 ถึง 143

11164	36318	75061	37674	26320	75100	10431	20418	19228	91792
21215	91791	76831	58678	87054	31687	93205	43685	19732	08468
10438	เริ่มจาก 8 ในทิศทาง ซ้าย ไปขวา บนลงล่าง			08882	90870	12462	41810	01806	02977
36792				60881	97395	20461	36742	02852	50564
73944				82384	38370	00249	80709	72605	67497
4563	12872	93104	78483	72717	68714	18078	25005	04151	
64208	48255	73117	33242	42314	83049	21933	92813	04763	
51486	72875	3860,5	29,341,	807,49	8,015,1	33,835,	526,02	7,914,7	08,868,
997,56	2,636,0	64,516,	179,71	4,847,8	09,610,	046,38	1,714,1	09,227,	106,06
7,132,5	55,217,	130,15	7,920,7	00,431,	451,17	3,382,7	92,873,	029,53	8,547,4
65,285,	971,98	1,213,8	53,010,	946,01	1,583,8	16,805,	610,04	4,351,6	17,020,
172,64	5,732,7	38,224,	293,01	3,138,1	38,109,	349,76	6,569,2	98,566,	295,50
9,563,9	99,754,	311,99	9,255,8	68,368,	049,85	5,109,2	377,80	4,026,1	14,479,
615,55	7,640,4	86,210,	118,08	1,284,1	45,147,	974,38	6,002,2	12,645,	620,00
7,813,7	98,768,	046,89	8,713,0	79,225,	081,53	8,496,7	64,539,	794,93	7,491,7
62,490,	992,15	8,498,7	28,759,	191,77	1,473,3	24,550,	280,67	6,889,4	38,490,
242,16	6,344,4	21,283,	070,44	9,272,9	37,284,	132,11	3,748,5	10,415,	364,57
1,697,5	95,428,	332,26	5,590,3	31,605,	438,17	2,225,0	03,918,	469,99	9,850,1
59,138,	395,42	7,116,8	57,609,	915,10	7,7904	74244	50940	31553	62562
29478	59652	50414	31966	87912	87154	12944	49862	96566	48825
96155	95009	27429	72918	08457	78134	48407	26061	58754	05326
29621	66583	62966	12468	20245	14015	04014	35713	03980	03024

ชั้นที่ 2 :

เลขที่	ID	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
15	1	2	21	170.0	67.7	3.50	2.85	2	2	60	84	1	2	3
133	2	2	22	155.0	54.5	3.16	2.18	3	2	72	90	1	1	2
27	3	2	20	164.0	62.1	1.51	3.11	2	3	84	108	2	1	2
64	4	1	21	169.5	67.2	2.99	1.97	1	2	64	132	2	1	2
46	5	1	20	172.0	73.7	3.00	3.30	2	2	72	88	2	2	2
109	6	2	21	160.0	58.8	2.78	1.80	3	2	72	120	2	2	2
106	7	2	21	155.0	58.8	3.40	2.08	3	2	64	112	2	2	2
67	8	2	21	155.0	57.7	3.15	2.50	1	2	60	82	2	2	3
132	9	2	20	153.0	60.2	2.85	2.30	3	2	84	124	2	1	2
130	10	2	22	164.0	64.3	2.85	2.41	3	2	74	122	2	1	2
29	11	1	21	166.0	74.6	2.09	2.45	2	1	80	146	2	2	2
10	12	2	21	153.0	59.3	3.30	2.44	2	2	76	96	2	1	2
11	13	2	21	155.0	42.9	3.45	2.87	2	2	62	120	2	1	2
44	14	2	21	169.0	69.7	3.25	2.68	2	2	80	120	1	1	2
20	15	1	21	168.0	84.9	2.33	2.38	2	3	72	128	1	1	2
13	16	2	21	149.0	53.5	2.53	2.83	2	2	64	108	2	1	4
138	17	2	20	150.0	56.0	2.69	2.17	3	3	80	112	2	1	3
49	18	2	21	155.0	68.2	3.48	2.66	1	2	80	112	1	1	2
26	19	2	22	160.0	62.8	3.05	2.77	2	2	72	100	2	2	2
114	20	2	21	147.5	64.2	3.30	2.67	3	2	84	116	1	2	3
118	21	1	21	173.0	73.9	2.30	1.80	3	2	72	110	2	1	2
81	22	2	21	160.0	68.8	3.45	2.85	1	2	68	92	2	2	2
2	23	2	20	168.0	65.2	2.98	2.35	2	2	72	140	1	2	2
7	24	2	21	168.4	67.5	2.89	2.21	2	2	80	112	2	2	2
79	25	2	22	153.0	76.4	2.50	2.77	1	2	76	100	1	1	2
70	26	2	21	154.0	60.6	3.04	2.93	1	2	96	143	2	1	1
113	27	2	21	169.0	67.2	2.79	2.17	3	2	80	136	1	1	2
3	28	2	21	169.0	78.7	3.64	2.56	2	3	72	100	2	2	2
116	29	2	21	157.5	58.9	3.20	1.90	3	3	72	91	2	1	3
107	30	2	23	162.0	71.5	3.25	1.79	3	1	80	112	1	1	3



ผลจากการสุ่มเลือกตัวอย่างโดยใช้
ตารางเลขสุ่ม ตามขั้นตอนที่ 2

ผลจากการนำข้อมูลจากภาคผนวก 3 เฉพาะรายชื่อ
ถูกสุ่มเป็นตัวอย่างมาลงตาราง ตามขั้นตอนที่ 3

ชั้นที่ 3 -

2.3

-

2.3.1 [✓] 1. Simple random sampling

[] 2. Systematic random sampling

[] 3. Cluster random sampling

[] 4. Stratified random sampling

2. Systematic random sampling เช่น จากบัญชีรายชื่อนักศึกษาทั้ง 143 คน หาดด้วย 30 ได้ 4.8 ปิดขึ้นเป็น 5 นั่นคือเราจะสุ่มเลือก 1 คนจากทุกๆ 5 คน จากนั้นสุ่มหมายเลขเริ่มต้นของ 5 คนแรก สมมติได้หมายเลข 3 นี่ก็คือคนที่ถูกเลือกเป็นตัวอย่าง คนต่อไปคือคนที่ $3+5$ คือคนหมายเลข 8 คนต่อไปคือหมายเลข 13 เรื่อยไปจะได้ครบ 30 คนพอดีเมื่อถึงคนสุดท้ายในบัญชีรายชื่อ

3. Cluster random sampling เช่น คณะจัดแบ่งนักศึกษาทั้ง 143 คนออกเป็น 14 กลุ่ม กลุ่มละประมาณ 10 คนเพื่อมอบหมายอาจารย์ที่ปรึกษา ผู้วิจัยไม่มีบัญชีรายชื่อนักศึกษาแต่มีรายชื่ออาจารย์ที่ปรึกษาทั้ง 10 คน จึงสุ่มอย่างง่ายมา 3 คน ได้นักศึกษา 30 คน

4. Stratified random sampling เช่น ผู้วิจัยต้องการให้ได้ตัวอย่างนักศึกษาจากทุกภาคของประเทศ ขั้นแรกจึงนำรายชื่อนักศึกษาทุกคนมาจำแนกตามภาคเหนือ อีสาน กลาง และใต้ สมมติได้ 20 80 20 และ 23 คน ตามลำดับ ซึ่งรวมได้ 143 คน ต้องการตัวอย่าง 30 คน นั่นคือ ภาคละ 21% (คือ $30/143$) นำ 0.21 คูณเข้ากับจำนวนนักศึกษาในแต่ละภาคได้ตัวอย่างประมาณ 4 17 4 และ 5 คนตามลำดับ

2.3.3

คำถาม	สเกลการวัด
2. เพศ (1= ชาย, 2= หญิง)	Nominal scale (Dichotomous variable)
3. อายุ.....ปี	Ratio scale
4. ความสูง.....เซนติเมตร	Ratio scale
5. น้ำหนัก.....กิโลกรัม	Ratio scale
6. เกรดเฉลี่ยสะสมเมื่อจบการศึกษาระดับมัธยมปลาย.....	Interval scale
7. เกรดเฉลี่ยสะสมจนถึงเทอมที่ผ่านมา.....	Interval scale
8. ท่านสอบเข้ามหาวิทยาลัยโดยระบบใด (1= โควต้า, 2 = เอนทรานซ์)	Nominal scale

9. ภูมิลำเนาอยู่ภาคใด (1= เหนือ, 2= ตะวันออกเฉียงเหนือ, 3= กลาง, 4= ใต้)	Nominal scale (Polytomous variable)
10. อัตราการเต้นของชีพจรก่อนวิ่ง ครั้ง ต่อ นาที	Ratio scale
11. อัตราการเต้นของชีพจรหลังวิ่ง ครั้ง ต่อ นาที	Ratio scale
12. สุขภาพจิตของท่านก่อนเข้าเรียนมหาวิทยาลัย (1 = มีความเครียด, 2 = ไม่มีความเครียด)	Nominal scale (Dichotomous variable)
13. สุขภาพจิตของท่านหลังเข้าเรียนมหาวิทยาลัยแล้ว (1 = มีความเครียด, 2 = ไม่มีความเครียด)	Nominal scale (Dichotomous variable)
14. สถานสุขภาพ (1= แย่, 2= ดี, 3= ปานกลาง, 4= แ่)	Ordinal scale

2.3.4

คำถาม	ประเภท
2. เพศ (1= ชาย, 2= หญิง)	Categorical variable
3. อายุ.....ปี	Continuous variable
4. ความสูง.....เซนติเมตร	Continuous variable
5. น้ำหนัก.....กิโลกรัม	Continuous variable
6. เกรดเฉลี่ยสะสมเมื่อจบการศึกษาระดับมัธยมปลาย.....	Continuous variable
7. เกรดเฉลี่ยสะสมจนถึงเทอมที่ผ่านมา.....	Continuous variable
8. ท่านสอบเข้ามหาวิทยาลัยโดยระบบใด (1= โควต้า, 2 = เอนทรานซ์)	Categorical variable
9. ภูมิลำเนาอยู่ภาคใด (1= เหนือ, 2= ตะวันออกเฉียงเหนือ, 3= กลาง, 4= ใต้)	Categorical variable
10. อัตราการเต้นของชีพจรก่อนวิ่ง ครั้ง ต่อ นาที	Continuous variable
11. อัตราการเต้นของชีพจรหลังวิ่ง ครั้ง ต่อ นาที	Continuous variable
12. สุขภาพจิตของท่านก่อนเข้าเรียนมหาวิทยาลัย (1 = มีความเครียด, 2 = ไม่มีความเครียด)	Categorical variable
13. สุขภาพจิตของท่านหลังเข้าเรียนมหาวิทยาลัยแล้ว (1 = มีความเครียด, 2 = ไม่มีความเครียด)	Categorical variable
14. สถานสุขภาพ (1= แย่, 2=ดี, 3=ปานกลาง, 4=แ่)	Categorical variable

2.3.5

คำถาม	ประเภท
3. อายุ []1. น้อยกว่า 5 ปี []2. 5 ถึง 25 ปี []3. 26 - 54 ปี []4. 55 ปีขึ้นไป	Categorical variable

คำเฉลยบทที่ 3 การตรวจสอบการแจกแจงของข้อมูล

ส่วนที่ 1

3.1

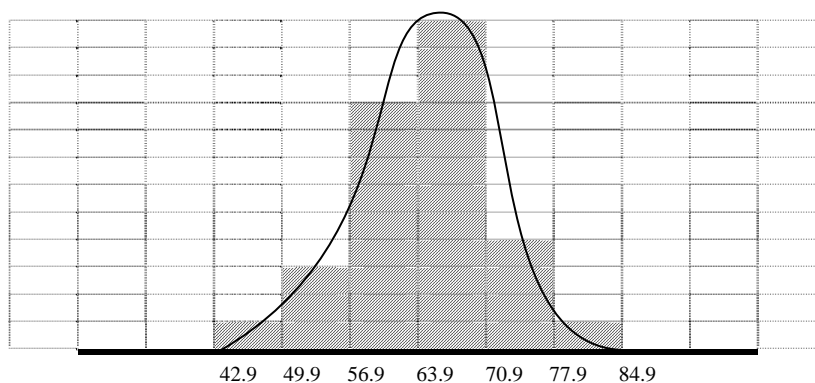
67.7	54.5	62.1	67.2	73.7	58.8	58.8	57.7	60.2	64.3
74.6	59.3	42.9	69.7	84.9	53.5	56.0	68.2	62.8	64.2
73.9	68.8	65.2	67.5	76.4	60.6	67.2	78.7	58.9	71.5

3.2

ค่าต่ำสุด 42.9
ค่าสูงสุด 84.9
จัดแบ่งเป็น 6 กลุ่ม
ได้ความกว้างชั้น
 $[(84.9-42.9)/6]=7$
จากนั้นจัดชั้นคะแนน
42.9 ถึง 42.9+7 ฯลฯ

น้ำหนัก	ความถี่	ความถี่ สัมพัทธ์ (%)	ความถี่ สัมพัทธ์ สะสม (%)
42.90 - 49.90	1	3.3	3.3
49.91 - 56.90	3	10.0	13.3
56.91 - 63.90	9	30.0	43.3
63.91 - 70.90	12	40.0	83.3
70.91 - 77.90	4	13.3	96.6
77.91 - 84.90	1	3.3	99.9
รวม	30	99.9	

3.3

3.4 ☒ สมมาตร (symmetry)☐ เบ้ซ้าย (left skew)☐ เบ้ขวา (right skew)☐ อื่นๆ ระบุ

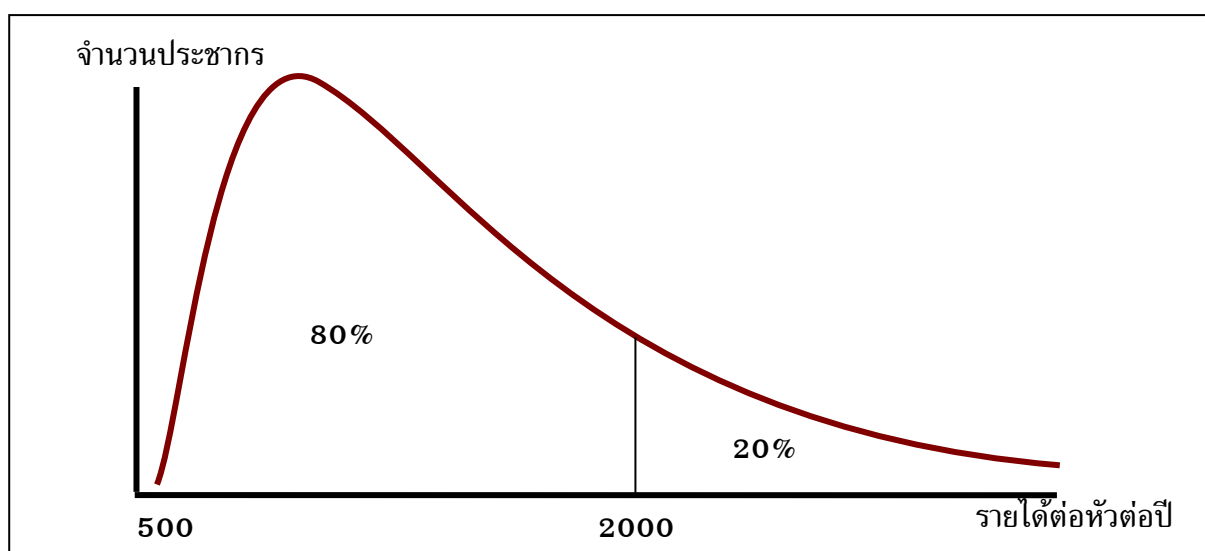
3.5

สรุป: ส่วนมากจะได้โค้งสมมาตร เพราะการกระจายของข้อมูลของประชากรทั้งหมด 143 คนเป็นแบบสมมาตร (ดูกราฟในข้อ 3.9.1 ในแบบฝึกหัด)

3.6

-

3.6.1



- 3.6.2 ก. -
 ข. 0.2
 3.7 -

3.7.1 แกนตั้งคือ จำนวนนักศึกษา แกนนอนคือ คะแนน

3.7.2 ค่าเฉลี่ย (Mean) คะแนนของรุ่นที่ 1 < 75 รุ่นที่ 2 = 75 รุ่นที่ 3 = 75

3.7.3 จงอธิบายลักษณะการแจกแจงข้อมูลตามความหมายของคะแนนจากการสอบ

รุ่นที่ 1. นักศึกษาได้คะแนนตั้งแต่ 0 ถึง 80 ส่วนมากได้คะแนนสูง แต่ยังมีบางส่วนที่ซึ่งเป็นส่วนน้อยได้คะแนนต่ำมากๆ

รุ่นที่ 2. นักศึกษาได้คะแนนตั้งแต่ 50 ถึง 100 ส่วนมากได้ 75 ที่ได้สูงหรือต่ำมากๆ เมื่อเทียบกับกลุ่มแล้วจะมีน้อยมาก

รุ่นที่ 3. นักศึกษาได้คะแนนตั้งแต่ 25 ถึง 100 ส่วนมากได้ 75 ที่ได้สูงหรือต่ำมากๆ เมื่อเทียบกับกลุ่มแล้วจะมีน้อยมาก

3.7.4 จงเปรียบเทียบการแจกแจงข้อมูลรุ่น 2 กับ รุ่น 3 ค่าเฉลี่ยเท่ากัน แต่รุ่น 3 มีการกระจายข้อมูลมากกว่า

3.7.5 จงเขียนคำสถิติที่เหมาะสมที่จะใช้สรุปค่าคะแนนดังกล่าว โดยให้พิจารณาลักษณะการกระจายข้อมูลประกอบการตัดสินใจ [จากรูปกราฟ จะสังเกตได้ว่า ต้องอธิบายทั้งค่ากลางและค่าการกระจาย]

รุ่นที่ 1. เพราะกราฟเบ้ จึงใช้ค่ามัธยฐาน ร่วมกับพิสัย (ต่ำสุด:สูงสุด)

รุ่นที่ 2. เพราะกราฟสมมาตร จึงใช้ค่าเฉลี่ย ร่วมกับค่าส่วนเบี่ยงเบนมาตรฐาน

รุ่นที่ 3. เพราะกราฟสมมาตร จึงใช้ค่าเฉลี่ย ร่วมกับค่าส่วนเบี่ยงเบนมาตรฐาน

3.8 จงสรุปความสำคัญของลักษณะการแจกแจงของข้อมูลต่อการวิเคราะห์ข้อมูล

ข้อมูลต่อเนื่อง ควรมีการตรวจสอบลักษณะการแจกแจงข้อมูลก่อนการวิเคราะห์เสมอ เพราะการจะใช้สถิติอะไรในการสรุปข้อมูล ล้วนต้องอาศัยความรู้เกี่ยวกับลักษณะการแจกแจงข้อมูล การสรุปต้องบอกทั้งค่ากลางและค่าการกระจายจึงจะครบถ้วน เพราะค่ากลางเดียวกัน ค่าการกระจายอาจแตกต่างกันได้ ถ้าเพียงการสรุปเชิงพรรณนา จะมีหลักหยาบๆ คือหากข้อมูลแจกแจงแบบเบ้ ให้ใช้ค่ามัธยฐาน ร่วมกับพิสัย (ต่ำสุด:สูงสุด) หากสมมาตร ให้ใช้ค่าเฉลี่ย ร่วมกับค่าส่วนเบี่ยงเบนมาตรฐาน

3.9 -

3.9.1 คุณลักษณะการกระจายจากฮิสโตแกรม [ดูภาคผนวก 2]

3.9.2 -

3.9.3 -

3.10 -

ส่วนที่ 2

3.11. คำสั่งคือ tab v8

ข้อสังเกตและเสนอแนะ : ตัวแปรนี้ระบุในแบบสอบถามว่ามี 2 ค่าคือ 1. โดยสอบเอ็นทรานซ์ และ 2. โดยสอบโควต้า แต่พบจากการวิเคราะห์เบื้องต้นว่ามีกลุ่มที่ 3 ผู้วิเคราะห์ต้องตรวจสอบว่าลงรหัสผิด หรือป้อนข้อมูลผิด แล้วไปค้นแบบเก็บรวบรวมข้อมูลเพื่อตรวจสอบและแก้ไขให้ถูกต้องต่อไป แต่ถ้าพบว่ามีผิด กลุ่มที่ 3 คือนักศึกษาภาคพิเศษ นั่นหมายถึงผู้วิเคราะห์ต้องแก้ไขแบบสอบถามเพื่อให้สอดคล้องกันในรายงานวิจัยต่อไป ตัวอย่างนี้ ข้อเท็จจริงคือมีกลุ่มที่ 3 คือนักศึกษาภาคพิเศษ

คำเฉลยบทที่ 4 สถิติเชิงพรรณนาเพื่อสรุปลักษณะตัวอย่าง

ตอนที่ 1

4.1

ค่าสถิติ	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
1. มาตรการแนวโน้มนิสัย													
ค่าเฉลี่ย		✓	✓	✓	✓	✓			✓	✓			
ค่ามัธยฐาน		✓	✓	✓	✓	✓			✓	✓			
ค่าฐานนิยม		✓	✓	✓	✓	✓			✓	✓			
2. มาตรการกระจาย													
ค่าความแปรปรวน		✓	✓	✓	✓	✓			✓	✓			
ค่าส่วนเบี่ยงเบนมาตรฐาน		✓	✓	✓	✓	✓			✓	✓			
ค่ามัธยฐาน		✓	✓	✓	✓	✓			✓	✓			
ค่าต่ำสุด		✓	✓	✓	✓	✓			✓	✓			
ค่าสูงสุด		✓	✓	✓	✓	✓			✓	✓			
3. ค่าร้อยละ	✓						✓	✓			✓	✓	✓

ตอนที่ 2)

4.2

ตารางที่ 1. ลักษณะทางประชากรของกลุ่มที่ศึกษา

ตัวแปร	จำนวน	ร้อยละ
1. เพศ		
ชาย	5	16.7
หญิง	25	83.3
รวม	30	100.0
2. อายุ		
20 ปี หรือน้อยกว่า	5	16.7
21 - 22 ปี	24	80.0
23 ปีขึ้นไป	1	3.3
รวม	30	100.0
ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน)	21.0 (0.7)	
ค่ามัธยฐาน(ค่าต่ำสุด:ค่าสูงสุด)	21 (20:23)	

ตัวแปร	จำนวน	ร้อยละ
3. ความสูง		
น้อยกว่า 150 ซม.	2	6.7
150 - 159 ซม.	11	36.7
160 ซม. ขึ้นไป	17	56.7
รวม	30	100.1
ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน)	160.8 (7.6)	
ค่ามัธยฐาน (ค่าต่ำสุด:ค่าสูงสุด)	160.0 (147.5:173.0)	
4. น้ำหนัก		
น้อยกว่า 50 กก.	1	3.3
50 - 59 กก.	8	26.7
60 กก. ขึ้นไป	21	70.0
รวม	30	100.0
ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน)	65.0 (8.6)	
ค่ามัธยฐาน (ค่าต่ำสุด:ค่าสูงสุด)	64.8 (42.9:84.9)	

จึงสรุปผลจากตารางข้างต้น ให้ได้ใจความที่สามารถนำไปบรรจุในรายงานผลการวิจัยได้

จากการสำรวจนักศึกษาคณะเภสัชศาสตร์จำนวน 30 คน ร้อยละ 83.3 เป็นเพศหญิง (ตารางที่ 1) มีอายุเฉลี่ย 21.0 ปี (ส่วนเบี่ยงเบนมาตรฐาน = 0.7) มากกว่าครึ่งเล็กน้อยที่มีความสูงมากกว่า 160 เซนติเมตร มีน้ำหนักตั้งแต่ 42.9 ถึง 84.9 กิโลกรัม (มัธยฐาน = 64.8)

ถ้าหากท่านได้รับมอบหมายให้ทำการสำรวจเช่นเดียวกันนี้ที่มหาวิทยาลัยมหิดล ด้วยขนาดตัวอย่าง 30 ราย เท่ากัน เพื่อเปรียบเทียบกัน จึงออกแบบตารางเพื่อนำเสนอข้อมูลชุดที่น่าเสนอข้างต้น (แสดงเฉพาะหัวตารางและหัวคอลัมน์เฉพาะข้อมูลเพศเพียงตัวแปรเดียว โดยเว้นข้อมูลไว้)

ตารางที่ 2. ลักษณะทางประชากรของกลุ่มที่ศึกษา จำแนกตามมหาวิทยาลัย

ตัวแปร	มหาวิทยาลัยขอนแก่น		มหาวิทยาลัยมหิดล	
	จำนวน	ร้อยละ	จำนวน	ร้อยละ
1. เพศ				
ชาย				
หญิง				
รวม				
2. อายุ				
20 ปี หรือน้อยกว่า				
21 - 22 ปี				
23 ปีขึ้นไป				
รวม				
ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน)				
ค่ามัธยฐาน(ค่าต่ำสุด:ค่าสูงสุด)				

ค่าเฉลี่ยบทที่ 5 การอนุมานทางสถิติเพื่อตอบคำถามวิจัย

กิจกรรมที่ 5.1 :

5.1 ข้อมูล ค่าสถิติ หรือค่าพารามิเตอร์

5.1.1 จากปฏิบัติการที่ 2 น้ำหนักของท่าน (ค่า V5 ของ ID = 0) = **59** กิโลกรัม

(ค่านี้คือ ☒ ข้อมูล ☐ ค่าสถิติ ☐ ค่าพารามิเตอร์)

5.1.2 จากปฏิบัติการที่ 4 ค่าเฉลี่ยน้ำหนัก = **65** กิโลกรัม

(ค่านี้คือ ☐ ข้อมูล ☒ ค่าสถิติ ☐ ค่าพารามิเตอร์)

- 5.1.3 สิ่งที่เราต้องการทราบคือ ค่าเฉลี่ยน้ำหนักของนักศึกษาทั้งหมดที่เรียนในห้องนี้
(ค่านี้คือ ☐ ข้อมูล ☐ ค่าสถิติ ☒ ค่าพารามิเตอร์)

เราสามารถหาค่าตามข้อ 5.1.3 ได้หากเก็บข้อมูลน้ำหนักของนักศึกษาทุกคน
แต่ในชีวิตจริงเราไม่ทำ เพราะ

- i) ประชากรมีขนาดใหญ่ทำให้สิ้นเปลืองมาก โอกาสผิดพลาดมีสูงตามปริมาณงาน
- ii) ไม่ทราบขอบเขตประชากรทั้งหมด เช่นศึกษาพฤติกรรมผู้ติดเชื้อเอดส์
- iii) มีวิธีการได้มาซึ่งค่าพารามิเตอร์โดยไม่ต้องศึกษาทุกหน่วยในประชากร

- 5.2 ค่าที่ตอบคำถามข้อ 5.1.3 (ค่าเฉลี่ยน้ำหนักของนักศึกษาทั้งหมดที่เรียนในห้องนี้)
ได้ใกล้เคียงที่สุดคือค่าในข้อต่อไปนี้ใช่หรือไม่

- ก. ข้อ 5.1.1 คือขนาดตัวอย่างเท่ากับ 1 ☐ ใช่ ☒ ไม่ใช่
- ข. ข้อ 5.1.2 คือขนาดตัวอย่างเท่ากับ 30 ☐ ใช่ ☒ ไม่ใช่ แต่ยังดีกว่า 1
- ค. เพิ่มขนาดตัวอย่างเป็น 50 ☐ ใช่ ☒ ไม่ใช่ แต่ยังดีกว่า 30
- ง. เพิ่มขนาดตัวอย่างเป็น 100 ☐ ใช่ ☒ ไม่ใช่ แต่ยังดีกว่า 50

- 5.3 ค่าจากข้อ 5.1.2 ซึ่งเป็นค่าเฉลี่ยน้ำหนัก สามารถเปลี่ยนแปลงได้ตามตัวอย่างที่สุ่มมาได้ ดังนี้

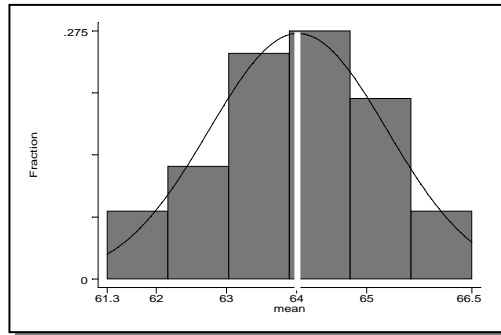
- 5.3.1 ให้เขียนค่าเฉลี่ยน้ำหนักที่ตนเองคำนวณได้จากการสุ่มตัวอย่าง $n=30$ ลงในช่องแรกของตารางข้างล่างนี้ จากนั้นให้เขียนค่าเฉลี่ยดังกล่าวที่ผู้รับการอบรมคนอื่น ๆ คำนวณได้ ลงช่องถัดไปจนครบทุกคน [เพื่อให้สมจริง ค่าเฉลี่ยต่อไปนี้ได้มาจากการสุ่มโดยคอมพิวเตอร์ เสมือนกับมีนักวิจัยหลายคนมาทำวิจัยในประชากรเดียวกันนี้ ด้วยขนาดตัวอย่างเท่าๆ กันคือ 30 ดูตัวอย่างการสุ่มและการวิเคราะห์ในภาคผนวก 2]

65.0	63.7	63.6	64.9	65.4	63.2	63.2	64.6	66.4	65.4
64.1	62.9	66.5	61.7	63.5	65.4	62.2	63.9	63.4	64.3
62.4	64.3	63.7	62.5	65.6	64.9	63.0	64.7	64.1	61.3
63.4	66.4	64.1	64.6	64.2	64.3	61.6	63.7	63.7	65.1

5.3.2 จงคำนวณค่าเฉลี่ยจากค่าเฉลี่ยในตารางข้างต้น = **64.0**

ค่านี้เรียกว่า ค่าเฉลี่ยของค่าเฉลี่ยของตัวอย่าง ($\bar{\bar{X}}$) หรือ Mean of the sampling means

5.3.3 สร้าง Histogram แสดงการแจกแจงของค่าเฉลี่ย พร้อมเขียนโค้งความถี่ทับลงด้วย
ทำเช่นเดียวกันกับบทที่ 3 ข้อ 3.2 หรือใช้ STATA ทำดังนี้



5.3.4 เท่ากับ 64

5.3.5 ตอบ: เท่ากัน

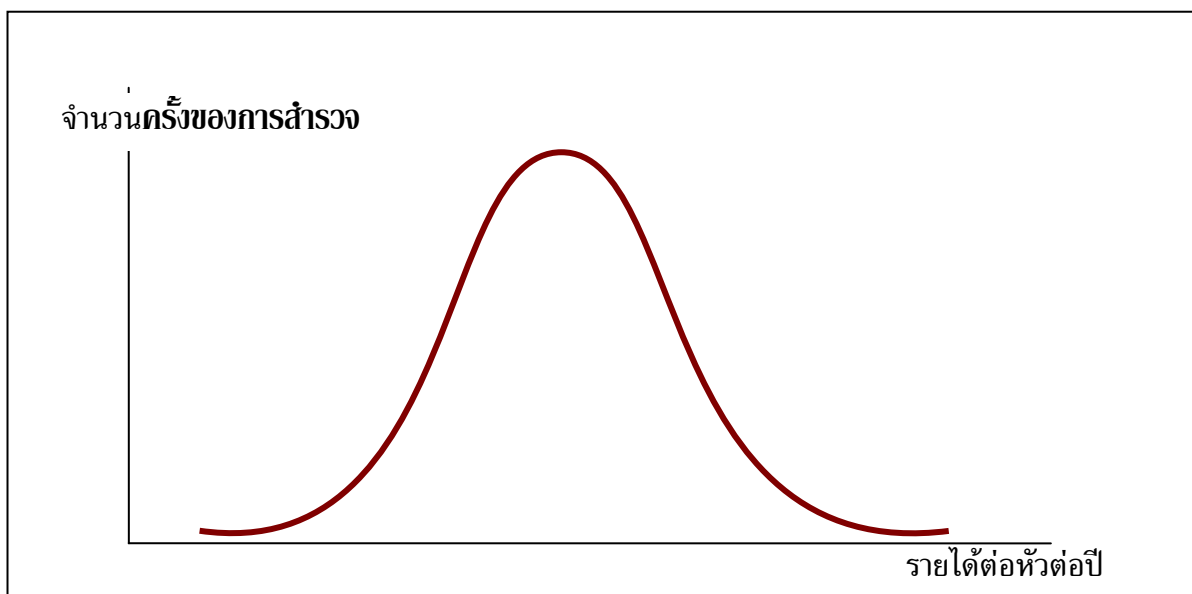
5.3.6 ข้อมูลน้ำหนักของนักศึกษาทั้งหมด ($N = 143$) นำมาคำนวณค่าเฉลี่ยได้เท่ากับ **64.5** กิโลกรัม (วิทยากรให้คำตอบ) ค่านี้คือค่าพารามิเตอร์ แทนด้วย μ

ถ้ามีการศึกษาวิจัยใดสรุปผลการศึกษาโดยใช้เพียงค่าสถิติที่ได้จากตัวอย่างเดียวเท่านั้น เช่น ได้ค่าเฉลี่ยน้ำหนักเท่ากับ 61 กิโลกรัม แล้วสรุปว่าสูงหรือต่ำกว่าค่าบางค่า เช่น นำไปเปรียบเทียบกับที่มีการวิจัยเช่นเดียวกันนี้ในมหาวิทยาลัยอื่นซึ่งเท่ากับ 65 กิโลกรัม **จึงสรุปว่า นักศึกษาในมหาวิทยาลัยที่ตนศึกษาวิจัยนั้นมีน้ำหนักเฉลี่ยต่ำกว่า** ท่านจะให้ข้อคิดเห็นอย่างไร กับนักวิจัยนั้น เห็นว่า 61 ต่ำกว่า 65 จริง แต่นี้ได้จากการสุ่มตัวอย่างเพียงครั้งเดียว ข้อมูลจาก **Repeated sampling** ข้างต้นบ่งชี้ชัดเจนว่า แม้ค่าจริงของประชากร (ในโลกความเป็นจริงไม่มีใครรู้ แต่ตัวอย่างนี้เรารู้) เท่ากับ 64.5 แต่การสุ่มตัวอย่างอาจทำให้ได้ค่าเฉลี่ยตั้งแต่ 61.3 ถึง 66.5 ดังนั้นการสุ่มตัวอย่างทำให้มีโอกาสได้ค่าเฉลี่ย 65 โดยบังเอิญได้ ทั้งที่ความจริงแล้ว น้ำหนักเฉลี่ยจริงๆ ของประชากรทั้งสองกลุ่มอาจไม่แตกต่างกัน จึงไม่สามารถสรุปได้ว่าต่ำกว่า ($\bar{\bar{X}}$)

- 5.3.7 เปรียบเทียบค่าเฉลี่ยของค่าเฉลี่ยของตัวอย่าง กับค่า μ และให้ความเห็น
เท่ากัน ($64.0 \approx 64.5$)
- 5.3.8 มีการวิจัยจำนวนเท่าใดที่ได้ค่าเฉลี่ย () มากกว่าค่า \bar{X}_x ตอบ
คิดเป็น 50% ของจำนวนการวิจัยทั้งหมด
- กล่าวคือ จาก Repeated sampling 40 ตัวอย่าง มี 20 ตัวอย่างที่มีค่ามากกว่า 64
- 5.3.9 ค่าความน่าจะเป็นเท่ากับ 0.50
- 5.3.10 -
- 5.3.11 -
- 5.3.12 -
- 5.3.13 -
- 5.4 -
- 5.4.1 สรุปสิ่งที่สังเกตได้

การแจกแจงข้อมูลประชากรเป็นแบบใด การแจกแจงข้อมูลตัวอย่างมักจะเป็นแบบนั้น
แต่การแจกแจงค่าสถิติมักสมมาตรเสมอ เช่น ข้อมูลอายุ แม้การแจกแจงข้อมูลประชากร หรือ
ตัวอย่างจะเบ้ขวา แต่การแจกแจงค่าเฉลี่ยของอายุจะสมมาตร นอกจากนั้นยังสังเกตได้ว่า การ
กระจายของการแจกแจงค่าสถิติจะน้อยกว่าของประชากรหรือตัวอย่างเสมอ (ดูจากขนาดความ
แตกต่างระหว่างค่าต่ำสุดและสูงสุดของแต่ละกราฟ)

5.4.2



5.4.3 ถ้าทำเช่นเดียวกันกับที่กล่าวในข้อ 5.4.2 แต่เพิ่มขนาดตัวอย่างเป็นห้าแสนคนต่อครั้งการสำรวจ ลักษณะของกราฟแจกแจงค่าเฉลี่ยที่ได้ เมื่อเปรียบเทียบกับที่ได้จากขนาดตัวอย่างเป็นห้าหมื่นคนต่อครั้งการสำรวจ จะได้กราฟสมมาตรเช่นกัน แต่ฐานแคบกว่า กล่าวคือ ค่าเฉลี่ยจากหลายการสำรวจกรณีแต่ละการสำรวจมีขนาดตัวอย่างใหญ่ จะมีการกระจายน้อยกว่ากรณีตัวอย่างขนาดเล็ก

5.4.4 -

5.5 ค่า SD ของการสำรวจน้ำหนักนักศึกษาของท่านเท่ากับ 8.6 (ดูได้จากผลการวิเคราะห์ด้วยคำสั่ง summarize v5 โดย STATA หรือแทนค่าสูตรข้างต้น โดยใช้ข้อมูลจากตารางข้อมูลนักศึกษาที่ถูกสุ่มเป็นตัวอย่าง ในบทที่ 2 ดังตัวอย่าง

5.6 SD ได้เท่ากับ 1.27

5.7 SE น้อยกว่า

ฐาน แคบกว่า

5.8 $SD = 8.6$

ดังนั้น ค่า $SE = 8.6 / \sqrt{(30)} = 1.6$

เปรียบเทียบกับค่า SD ที่ได้จาก Repeated samples ในข้อ 5.6 ซึ่งเท่ากับ 1.27

กิจกรรมที่ 5.2 :

5.9

ก. ค่าที่ครอบคลุม 95% ของจำนวนข้อมูลทั้งหมดคือ

$65.0 \pm [1.96(8.6)]$ เท่ากับ 48.1 ถึง 81.9 หรือมีจำนวนนักศึกษาที่มีน้ำหนักระหว่าง 48.1 ถึง 81.9 กิโลกรัมคิดเป็น 95% ของนักศึกษาทั้งหมด

- ข. ถ้าทราบว่านักศึกษาที่มหาวิทยาลัยมหิดลก็มีลักษณะข้อมูลน้ำหนักเช่นเดียวกับที่ท่านสำรวจพบในมหาวิทยาลัยขอนแก่น จากนักศึกษา 500 คน จะมีกี่คนที่มือน้ำหนัก 82 กิโลกรัมขึ้นไป
 $(82.0 - 65.0)/8.6 = 2$ นั่นคือมีความแตกต่างเป็น 2 เท่าของ SD ซึ่งสอดคล้องกับประมาณ 2.5% ของพื้นที่ใต้กราฟ (คือครึ่งของ 5%) ดังนั้น มีนักศึกษาที่มีน้ำหนัก 82 กิโลกรัมขึ้นไปเท่ากับ $500 \times 0.25 = 125$ คน

5.10

- ก. ค่าที่ครอบคลุม 95% ของจำนวนข้อมูลทั้งหมด (ในที่นี้คือจำนวนครั้งของการสำรวจ) คือ $65.0 \pm \{1.96[8.6/\sqrt{(30)}]\}$ เท่ากับ $65.0 \pm [1.96(1.6)]$ เท่ากับ 61.9 ถึง 68.1 หรือ สุ่มสำรวจ 100 ครั้ง มี 95 ครั้งที่ได้ค่าเฉลี่ยน้ำหนักอยู่ระหว่าง 61.9 ถึง 68.1 กิโลกรัม
- ข. ถ้าค่าเฉลี่ยน้ำหนักของประชากรเท่ากับ 60 กิโลกรัม การที่จะสุ่มสำรวจหนึ่งครั้งแล้วได้ค่าเฉลี่ยน้ำหนักของนักศึกษาเท่ากับ “...65...” (ใส่ค่าเฉลี่ยที่ได้จากการสำรวจโดยท่าน) กิโลกรัมหรือ สุดขั้วไปทางที่เป็นไปได้น้อย (More extreme) มีโอกาสเท่ากับ $(65 - 60)/1.6 = 3.1$ นั่นคือมีความแตกต่างเป็น 3.1 เท่าของ SE ซึ่งสอดคล้องกับประมาณ 0.5% ของพื้นที่ใต้กราฟ (คือครึ่งของ 1%) ดังนั้น โอกาสที่จะได้ค่าเฉลี่ย 65 กิโลกรัมหรือมากกว่า เท่ากับ 0.005 หรือ สุ่มสำรวจ 1000 ครั้งจะได้ผลคล้ายกันนี้ 5 ครั้ง

5.11 จากข้อ 5.9 ค่าที่ได้จากกราฟของการแจกแจงข้อมูลดิบ X จะนำไปสู่ค่า \bar{X}
 จากข้อ 5.10 ค่าที่ได้จากกราฟของการแจกแจงค่าสถิติ ~~X~~ จะนำไปสู่ค่า μ

กิจกรรมที่ 5.3 :

5.12 -

5.12.1 -

5.12.1.1 โจทย์กำหนด $\alpha = 0.05$ ดังนั้น $\alpha/2 = 0.025$

. su v5					
Variable	Obs	Mean	Std. Dev.	Min	Max
v5	30	64.99333	8.593419	42.9	84.9

แทนค่า : $64.993 \pm 2.045[8.593/\sqrt{(30)}]$

95% CI. : 61.8 ถึง 68.2

STATA: **ci v5** --> เพื่อได้ช่วงเชื่อมั่น 95% ของค่าเฉลี่ยน้ำหนัก

. ci v5					
Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
v5	30	64.99333	1.568936	61.7845	68.20217

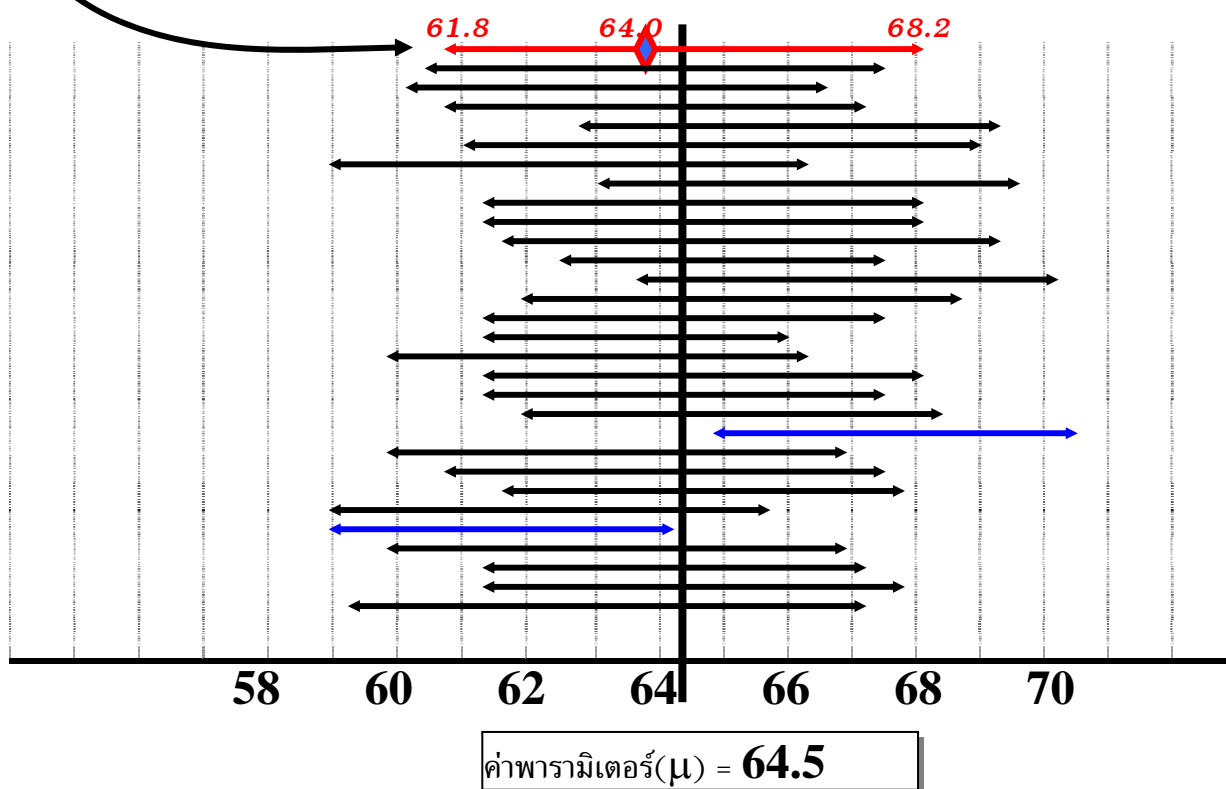
แปลความหมาย : **เชื่อมั่น 95% ว่า ค่าน้ำหนักเฉลี่ยในประชากรอยู่ระหว่าง 61.8 ถึง 68.2 กิโลกรัม**

5.12.1.2

61.8- 68.2	60.7-67.7	60.2-66.8	58.9-66.1	61.3-67.9	61.3-69.1
59.1-66.4	63.1-69.9	62.3-68.1	61.5-68.2	61.9-69.7	62.6-67.9
63.8-70.2	62.1-68.8	61.5-67.6	61.8-66.1	59.9-66.5	61.6-68.1
61.4-67.8	62.3-68.5	65.0-70.6	59.8-67.1	61.1-66.6	61.8-68.3
59.2-66.3	58.9-64.4	60.4-67.0	61.4-67.2	61.6-68.0	58.3-67.1

หมายเหตุ : ในที่นี้ใช้วิธีการสุ่มโดยคอมพิวเตอร์ จำลองเหตุการณ์ที่มีผู้สุ่มหลายคนที่ย่างคนต่างทำ โดยให้ $n = 30$ (ดูวิธีการในภาคผนวก 2)

5.12.1.3 สร้างแผนภูมิแสดงช่วงเชื่อมั่นจากตารางข้างต้น



จากช่วงเชื่อมั่นทั้งหมด 30 ช่วง คร่อมค่าพารามิเตอร์ 28 ช่วง คิดเป็น 93.3 %

5.12.1.4 จากการทดลองข้างต้น เป็นการพิสูจน์ให้เห็นว่า ถ้าสุ่มสำรวจเรื่องเดียวกัน ในประชากรเดียวกัน ด้วยขนาดตัวอย่างเท่ากัน จำนวน 100 ครั้ง จะได้ช่วงเชื่อมั่น 100 ช่วง ในนั้นจะมี 5 ช่วงที่ไม่ครอบคลุมค่าพารามิเตอร์ ช่วงเชื่อมั่นที่ท่านได้มาจึงอาจจะเป็นช่วงที่ครอบคลุมหรือไม่ครอบคลุมค่าพารามิเตอร์ก็ได้ แต่โอกาสที่จะตกเป็นผู้โชคร้ายได้ช่วงไม่ครอบคลุมค่าพารามิเตอร์นั้น มีน้อย (5 ใน 100) ท่านจึงบอกได้ว่าช่วงนั้นครอบคลุมค่าพารามิเตอร์ ด้วยความเชื่อมั่น 95%

5.12.2 ก่อนอื่นเราต้องบอกได้ว่า นี่คือการวิจัยที่เกี่ยวข้องกับการประมาณค่าเฉลี่ยในประชากรสองกลุ่มที่เป็นอิสระต่อกัน ค่าสถิติที่สนใจไม่ใช่ค่าเฉลี่ยแต่ละกลุ่ม (ชาย และ หญิง) แต่เป็นผลต่างค่าเฉลี่ยของสองกลุ่มนั้น กล่าวอีกนัยหนึ่ง นี่เป็นการประมาณค่าผลต่างของสองค่าเฉลี่ย (Differences of the two means)

ท่านสามารถคำนวณด้วยมือ หรือใช้ STATA ตามที่แสดงข้างล่าง เพื่อได้มาซึ่งข้อมูลสรุปนี้ ในทางปฏิบัติ สิ่งนี้คือการสรุปข้อมูลในสองกลุ่มที่เปรียบเทียบ เป็นสิ่งที่พึงต้องทำ จากค่าเฉลี่ย จะเห็นว่าโดยเฉลี่ยชายหนักกว่าหญิง ประมาณ 12 กิโลกรัม (75-63) นี่เป็นค่าสถิติที่สนใจ

	ชาย	หญิง
\bar{X}	74.86	63.02
<i>S.D.</i>	6.36	7.63
<i>n</i>	5	25

. su v5 if v2 == 1

Variable	Obs	Mean	Std. Dev.	Min	Max
v5	5	74.86	6.360269	67.2	84.9

. su v5 if v2 == 2

Variable	Obs	Mean	Std. Dev.	Min	Max
v5	25	63.02	7.625723	42.9	78.7

. ttest v5, by(v2)

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	5	74.86	2.844399	6.360269	66.96268	82.75732
2	25	63.02	1.525145	7.625723	59.87226	66.16774
combined	30	64.99333	1.568936	8.593419	61.7845	68.20217
diff		11.84	3.653709		4.355717	19.32428

Degrees of freedom: 28

Ho: mean(1) - mean(2) = diff = 0

Ha: diff < 0	Ha: diff ~= 0	Ha: diff > 0
t = 3.2405	t = 3.2405	t = 3.2405
P < t = 0.9985	P > t = 0.0031	P > t = 0.0015

95% CI. : 4.35 ถึง 19.32

แปลความหมาย : นักศึกษาชายมีน้ำหนักเฉลี่ยมากกว่านักศึกษาหญิง 11.8 กิโลกรัม เราเชื่อมั่น 95%
ว่าในประชากร ค่าความแตกต่างดังกล่าวอยู่ระหว่าง 4.35 ถึง 19.32 กิโลกรัม

5.12.3 ก่อนอื่นเราต้องบอกได้ว่า นี่คือการวิจัยที่เกี่ยวข้องกับการประมาณค่าเฉลี่ยในประชากรสองกลุ่มที่ไม่เป็นอิสระต่อกัน ค่าสถิติที่สนใจไม่ใช่ค่าเฉลี่ยแต่ละกลุ่ม (ก่อนวิ่ง และ หลังวิ่ง) แต่เป็นค่าเฉลี่ยของผลต่างระหว่างก่อนและหลังวิ่ง กล่าวอีกนัยหนึ่ง นี่เป็นการประมาณค่าเฉลี่ยของผลต่าง (Mean of the differences)

ขั้นตอนการคำนวณจะเริ่มจากหาผลต่างอัตราชีพจรหลังลบก่อนวิ่งของแต่ละคน จากนั้นหาค่าเฉลี่ยของผลต่างดังกล่าว แล้วประมาณค่า

STATA : `ttest v11 = v10`

. ttest v11 = v10						
Paired t test						
Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
v11	30	111.8667	3.197245	17.51203	105.3276	118.4058
v10	30	74.13333	1.497687	8.203167	71.07022	77.19645
diff	30	37.73333	2.845417	15.58499	31.9138	43.55286
Ho: mean(v11 - v10) = mean(diff) = 0						
Ha: mean(diff) < 0		Ha: mean(diff) ~= 0		Ha: mean(diff) > 0		
t = 13.2611		t = 13.2611		t = 13.2611		
P < t = 1.0000		P > t = 0.0000		P > t = 0.0000		

แปลความหมาย : ค่าเฉลี่ยความแตกต่างอัตราชีพจรก่อนและหลังวิ่งกับที่หนึ่งนาทีเท่ากับ 38 ครั้งต่อนาที เราเชื่อมั่น 95% ว่า ในประชากร ค่าความแตกต่างดังกล่าวอยู่ระหว่าง 32 ถึง 44 ครั้งต่อนาที

5.13 -

5.13.1

ก. ตั้งสมมติฐาน

$$H_0: \mu = 55$$

$$H_A: \mu \neq 55$$

ข. กำหนดระดับนัยสำคัญ $\alpha = 0.05$

ค. -

ง. หาค่า p-value ขั้นตอนคือแทนค่าสูตรแล้วคำนวณได้ค่า t จากนั้นนำไปเปิดตาราง t-distribution ที่ให้ไว้ในบทที่ 6 หรือที่ทำตารางสถิติหลายเล่มมีตารางค่าสถิติอยู่ เปิดหาค่า p-value ที่

Degree of freedom = $n - 1$ หรือที่ $Df = 30 - 1 = 29$ นี่คือการเรียกกันว่า One sample t-test ในทางปฏิบัติเราใช้คอมพิวเตอร์ เราใช้ STATA ทำการคำนวณให้ ตามที่แสดงข้างล่างนี้

. ttest v5 = 55						
One-sample t test						
Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
v5	30	64.99333	1.568936	8.593419	61.7845	68.20217
Degrees of freedom: 29						
Ho: mean(v5) = 55						
Ha: mean < 55	Ha: mean ~= 55		Ha: mean > 55			
t = 6.3695	t = 6.3695		t = 6.3695			
P < t = 1.0000	P > t = 0.0000		P > t = 0.0000			

ผลที่ได้คือ p-value = 0.0000 แต่ถ้ารายงานตามนี้ ความหมายคือ "เป็นไปได้" ซึ่งเป็นไปได้น้อยในความเป็นจริง แต่เพราะคอมพิวเตอร์ไม่แสดงผลค่าที่น้อยมาก ๆ ดังนั้น จึงรายงานเป็น p-value < 0.001

- จ. ตัดสินใจและสรุปผล เนื่องจาก p-value น้อยกว่า α ที่กำหนด คือน้อยกว่า 0.05 จึงปฏิเสธสมมติฐาน ($H_0: \mu = 55$) และสรุปว่า น้ำหนักเฉลี่ยของนักศึกษาสูงกว่า 55 กิโลกรัม อย่างมีนัยสำคัญทางสถิติ (p-value < 0.001) กล่าวอีกนัยหนึ่งคือ ถ้าน้ำหนักเฉลี่ยของนักศึกษาทั้งหมด (ประชากรนักศึกษา) เท่ากับ 55 กิโลกรัม โอกาสที่สุ่มตัวอย่างศึกษาครั้งหนึ่ง แล้วได้ค่าเฉลี่ยเท่ากับ 65 กิโลกรัมหรือมากกว่านั้นก็มีโอกาสเป็นไปได้ แต่โอกาสดังกล่าวมีน้อยมาก ๆ

5.13.2

- ก. ตั้งสมมติฐาน $H_0: \mu_{ชาย} = \mu_{หญิง}$ หรือ $H_0: \mu_{ชาย} - \mu_{หญิง} = 0$
 $H_A: \mu_{ชาย} \neq \mu_{หญิง}$ หรือ $H_A: \mu_{ชาย} - \mu_{หญิง} \neq 0$
- ข. กำหนดระดับนัยสำคัญ $\alpha = 0.05$
- ค. -

ง. หาค่า p-value ขั้นตอนคือแทนค่าสูตรแล้วคำนวณได้ค่า t จากนั้นนำไปเปิดตาราง t-distribution ที่ให้ไว้ในบทที่ 6 หรือที่ท้ายตำราสถิติหลายเล่มมีตารางค่าสถิติอยู่ เปิดหาค่า p-value ที่ Degree of freedom = $n_1 + n_2 - 2$ หรือที่ $Df = 5 + 25 - 2 = 28$ นี่คือการเรียกกันว่า Two samples t-test ในทางปฏิบัติ เราใช้คอมพิวเตอร์ เราใช้ STATA ทำการคำนวณให้ ดังผลที่ปรากฏในข้อ 5.12.2 คำสั่ง `ttest v5, by(v2)` ได้ $p\text{-value} = 0.003$

จ. ตัดสินใจและสรุปผล : เนื่องจาก p-value น้อยกว่า α ที่กำหนด จึงปฏิเสธสมมติฐาน $(H_0: \mu_{ชาย} = \mu_{หญิง})$ และสรุปว่า น้ำหนักเฉลี่ยของนักศึกษาชายสูงกว่านักศึกษาหญิงอย่างมีนัยสำคัญทางสถิติ ($p\text{-value} = 0.003$) กล่าวอีกนัยหนึ่งคือ ถ้าผลต่างระหว่างค่าเฉลี่ยน้ำหนักของนักศึกษาชายกับนักศึกษาหญิงในประชากรของนักศึกษา เท่ากับ 0 กิโลกรัม (คือไม่ต่างกัน) แล้ว โอกาสที่สุ่มตัวอย่างมาศึกษาครั้งหนึ่ง แล้วได้ผลต่างค่าเฉลี่ยดังกล่าวเท่ากับ 12 กิโลกรัมหรือมากกว่านั้น มีน้อยมาก คืออาจเกิดขึ้นได้ประมาณ 3 ใน 1000 ดังนั้นการตัดสินใจปฏิเสธสมมติฐานที่ว่าน้ำหนักเฉลี่ยเท่ากันนั้น จึงมีโอกาสผิดพลาดน้อยมากตามที่ระบุ

5.13.3 การทดสอบสมมติฐานค่าเฉลี่ยในประชากรสองกลุ่มที่ไม่อิสระต่อกัน ตัวอย่างคำถามคือ ค่าเฉลี่ยอัตราการเต้นชีพจร (ครั้งต่อนาที) ก่อนและหลังการวิ่งแตกต่างกันหรือไม่ จงแสดงแนวทางที่ท่านวิเคราะห์ข้อมูลเพื่อตอบคำถามวิจัยนี้

ดำเนินการวิเคราะห์คล้ายกับข้อ 5.13.2 ทุกขั้นตอน แต่สูตรต่างกัน ตามที่ทราบมาแล้วว่า จุดที่ต่างกันระหว่างกรณีสองกลุ่มอิสระต่อกันกับไม่อิสระต่อกันนั้น อยู่ที่ค่าสถิติที่สนใจ กล่าวคือกรณีสองกลุ่มอิสระต่อกันสนใจค่า "ผลต่างค่าเฉลี่ย" (Different of the two means) ระหว่างสองกลุ่ม แทนด้วย $\mu_{ชาย} - \mu_{หญิง}$ ในขณะที่กรณีสองกลุ่มไม่อิสระต่อกันนั้น สถิติที่สนใจคือ "ค่าเฉลี่ยของผลต่าง" (Mean of the differences) คือเอาผลต่างระหว่างอัตราการเต้นชีพจรก่อนวิ่งลบอัตราการเต้นชีพจรหลังวิ่งของแต่ละคน แล้วจึงหาค่าเฉลี่ยของผลต่างดังกล่าว แทนด้วย μ_{diff} ดังนั้นสมมติฐานความไม่แตกต่างกรณีสองกลุ่มอิสระต่อกันจึงเขียนเป็น $H_0: \mu_{ชาย} - \mu_{หญิง} = 0$ ในขณะที่กรณีสองกลุ่มไม่อิสระต่อกันเขียนเป็น $H_0: \mu_{diff} = 0$ สถิติที่ใช้ในการทดสอบสมมติฐานคือ Paired t-test ท่านสามารถหาสูตรได้จากตำราพื้นฐานทางสถิติที่ให้ไว้ในบรรณานุกรม ในทางปฏิบัติ เราใช้คอมพิวเตอร์ เราใช้ STATA ทำการคำนวณให้ ดังผลที่ปรากฏในข้อ 5.12.3 คำสั่ง `ttest v11 = v10` ได้ $p\text{-value} < 0.001$ แนวทางการแปลความหมายเหมือนกันกับที่กล่าวในข้อ 5.13.2

หมายเหตุ: สถิติที่ใช้ในการวิเคราะห์ตามโจทย์ข้อ 5.12 ถึง 5.13 ล้วนอยู่ภายใต้ข้อกำหนดเบื้องต้น (Assumption) ว่าข้อมูลมีการแจกแจงแบบ t อย่างไรก็ตามในทางปฏิบัติ ผู้วิเคราะห์ควรตรวจสอบการแจกแจงของข้อมูลก่อนเสมอ ภาคผนวก 2 บทที่ 6 กิจกรรมที่ 6.2 ข้อ 1 ได้แสดงขั้นตอนและวิธีการที่พึงกระทำตามที่กล่าวนี้ กรณีขนาดตัวอย่างเล็กมาก หรือข้อมูลไม่แจกแจงแบบปกติ แม้จะพยายามเปลี่ยนแปลงค่าให้เป็นสเกลอื่นแล้วก็ตาม จะใช้สถิติที่ไม่อาศัยข้อกำหนดเบื้องต้นเกี่ยวกับการแจกแจง *Non-parametric statistics* แทน

กิจกรรมที่ 5.4 :

5.14 องค์ประกอบสิ่งที่ควรนำเสนอในรายงานผลการวิเคราะห์ ได้แก่

1. ขนาดตัวอย่าง ($n_1 = 5$ และ $n_2 = 25$)
2. ค่าสถิติเพื่อพรรณาลักษณะตัวอย่าง ($\bar{X}_1 = 74.8$ กับ $SD_1 = 6.4$ และ $\bar{X}_2 = 63.0$ กับ $SD_2 = 7.6$)
3. ค่าสถิติที่สนใจ (ผลต่างระหว่างสองค่าเฉลี่ย = 11.8)
4. ผลการอนุมานทางสถิติ
 - 4.1 การประมาณค่า (4.35 ถึง 19.32)
 - 4.2 การทดสอบสมมติฐาน ($p\text{-value} = 0.003$)

ลีลาการนำเสนอ :

จากการสำรวจนักศึกษาชายจำนวน 5 คน มีน้ำหนักเฉลี่ย 74.8 กับ (ส่วนเบี่ยงเบนมาตรฐาน = 6.4) และนักศึกษาหญิงจำนวน 25 คน มีน้ำหนักเฉลี่ย 63.0 กับ (ส่วนเบี่ยงเบนมาตรฐาน = 7.6) พบว่าน้ำหนักเฉลี่ยนักศึกษาชายต่างจากนักศึกษาหญิงอย่างมีนัยสำคัญทางสถิติ ($p\text{-value} = 0.003$) กล่าวคือนักศึกษาชายมีน้ำหนักเฉลี่ยมากกว่านักศึกษาหญิง 11.8 กิโลกรัม (95%CI: 4.35 ถึง 19.32 กิโลกรัม)

5.15 ข้อความสรุปผลในข้อ 5.14 อยู่ในส่วนของรายงานวิจัย

ในรายงานวิจัยโดยทั่วไป ประกอบด้วยหัวข้อต่อไปนี้

1. บทคัดย่อ (Abstract)
2. บทนำ (Introduction) หรือที่มา (Background) หรือหลักการและเหตุผล (Rationale)
3. ระเบียบวิธีวิจัย (Methodology)
4. ผลการศึกษา (Results)
5. สรุป อภิปรายผล และข้อเสนอแนะ

สรุปผลการวิเคราะห์ตามข้อ 5.14 พึงมีไว้สำหรับลงในบทคัดย่อและบทสรุปและอภิปรายผล

5.16 จงแสดงรูปแบบการนำเสนอผลการวิเคราะห์ในรายงานวิจัยโดยใช้กรณีเปรียบเทียบค่าน้ำหนักเฉลี่ยระหว่างชายกับหญิงเป็นคำถามวิจัย

ในรายงานวิจัย ภายใต้หัวข้อ "ผลการศึกษา" ควรมีอย่างน้อย 2 ส่วน ส่วนแรกสำหรับอธิบายลักษณะตัวอย่างที่ศึกษา และส่วนที่สองสำหรับตอบคำถามวิจัยโดยจำแนกตอบแต่ละวัตถุประสงค์การวิจัย

ส่วนแรกสำหรับอธิบายลักษณะตัวอย่างที่ศึกษานั้น ควรเลือกนำเสนอเฉพาะตัวแปรต้นที่เกี่ยวข้องโดยตรงกับตัวแปรตาม หรือที่จะเอื้อให้ผู้อ่านใช้วิจารณ์ว่าผลศึกษาน่าเชื่อถือหรือมีข้อจำกัดอย่างไร และสามารถนำองค์ความรู้ที่ได้ไปใช้ในสถานการณ์ของเขาได้หรือไม่ (ลักษณะทางประชากรของกลุ่มตัวอย่าง กับของกลุ่มที่เขาจะนำองค์ความรู้ไปประยุกต์ใช้) ส่วนแรกนี้ไม่จำเป็นต้องมีการอนุมานทางสถิติ กรณีมีเป้าหมายคือการเปรียบเทียบระหว่างกลุ่ม นิยมนำเสนอผลส่วนนี้เป็นตาราง ดังตัวอย่างที่ให้ไว้ท้ายข้อนี้ (ยกมาจากบทที่ 4) โปรดสังเกตว่าจะไม่มีการนำเสนอผลของตัวแปรที่เป็นตัวแปรตาม เพราะเราจะนำไปอนุมานเป็นค่าของประชากร

ส่วนที่สองสำหรับตอบคำถามวิจัยพึงต้องมีการอนุมานทางสถิติ ที่ควรมีทุกกรณีคือการประมาณเสนอเป็นช่วงเชื่อมั่น (95%CI) และการทดสอบสมมติฐาน เสนอเป็นค่า p-value

ตัวอย่าง :

...

ผลการศึกษา

จากการสำรวจนักศึกษาชายจำนวน 5 คน และนักศึกษาหญิงจำนวน 25 คน ทั้งสองกลุ่มมีอายุโดยเฉลี่ยใกล้เคียงกัน คือประมาณ 21 ปี (ตารางที่ 1) ในขณะที่ชายมีความสูงเฉลี่ยมากกว่าหญิง คือ 169 และ 159 เซนติเมตรตามลำดับ

ตารางที่ 1. ลักษณะทางประชากรของกลุ่มที่ศึกษา

ลักษณะประชากร	ชาย (n = 5)	หญิง (n = 25)
1. อายุ		
20 ปี หรือน้อยกว่า	20%	16%
21 - 22 ปี	80%	80%
23 ปีขึ้นไป	0%	4%
รวม	100%	100%
ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน)	20.8 (0.5)	21.1 (0.7)
ค่ามัธยฐาน (ค่าต่ำสุด:ค่าสูงสุด)	21 (20:21)	21 (20:23)
2. ความสูง		
น้อยกว่า 150 ซม.	0%	8%
150 - 159 ซม.	0%	44%
160 ซม. ขึ้นไป	100%	48%
รวม	100%	100%
ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน)	169.7 (2.9)	159.0 (7.0)
ค่ามัธยฐาน (ค่าต่ำสุด:ค่าสูงสุด)	169.5 (166:173)	157.5 (147:170)
3. ... ฯลฯ...		

พบว่านักศึกษาชายมีน้ำหนักเฉลี่ย 74.8 กับ (ส่วนเบี่ยงเบนมาตรฐาน = 6.4) และนักศึกษาหญิงจำนวน 25 คน มีน้ำหนักเฉลี่ย 63.0 กับ (ส่วนเบี่ยงเบนมาตรฐาน = 7.6) น้ำหนักเฉลี่ยนักศึกษาชายต่างจากนักศึกษาหญิงอย่างมีนัยสำคัญทางสถิติ ($p\text{-value} = 0.003$) กล่าวคือนักศึกษาชายมีน้ำหนักเฉลี่ยมากกว่านักศึกษาหญิง 11.8 กิโลกรัม (95%CI: 4.35 ถึง 19.32 กิโลกรัม)

...

****โปรดดูคำสั่ง STATA ในภาคผนวก 2 สำหรับวิเคราะห์ผลลงตารางข้างต้น****

หมายเหตุ: การวิเคราะห์ตามที่กล่าวจะเหมาะสมก็ต่อเมื่อการศึกษานี้เริ่มจากเลือกนักศึกษาชายมา 5 คน และนักศึกษาหญิงมา 25 คน จากนั้นสอบถามและเก็บข้อมูลอื่น ๆ แต่ข้อมูลชุดนี้ไม่มีการแยกกลุ่มก่อน การสรุปผลจึงต้องระมัดระวัง

คำเฉลยบทที่ 6 การประยุกต์แนวคิดสถิติในการวิจัย

กิจกรรมที่ 6.1 :

6.1 -

6.2 -

6.3 -

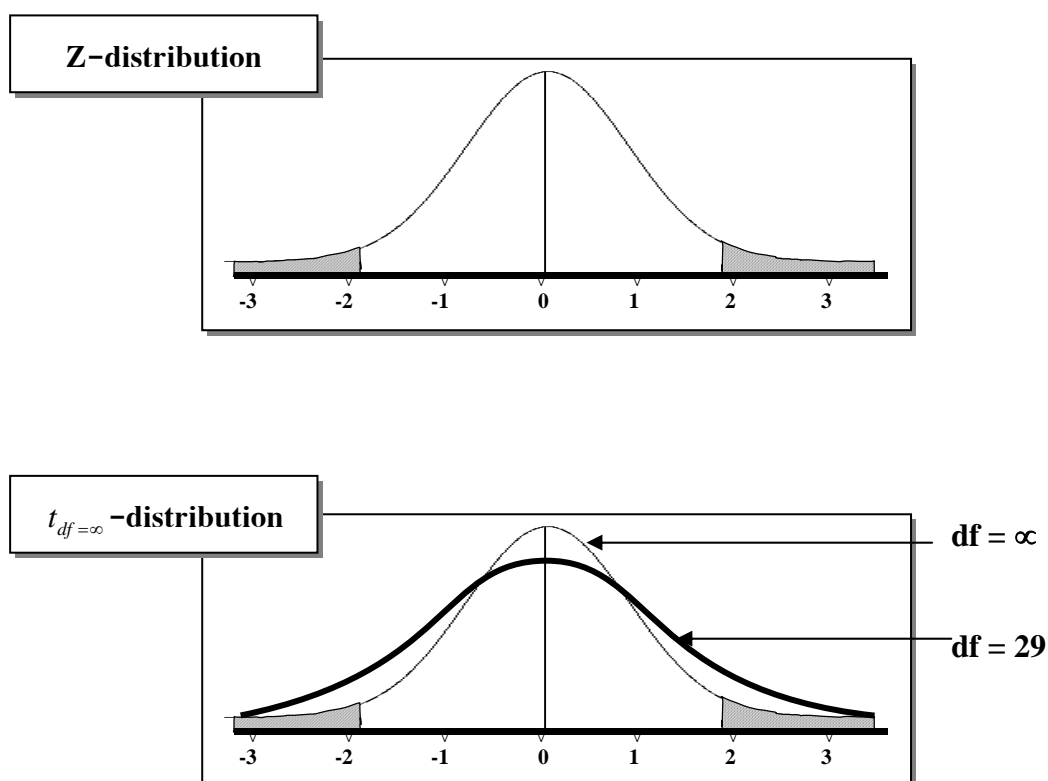
6.4

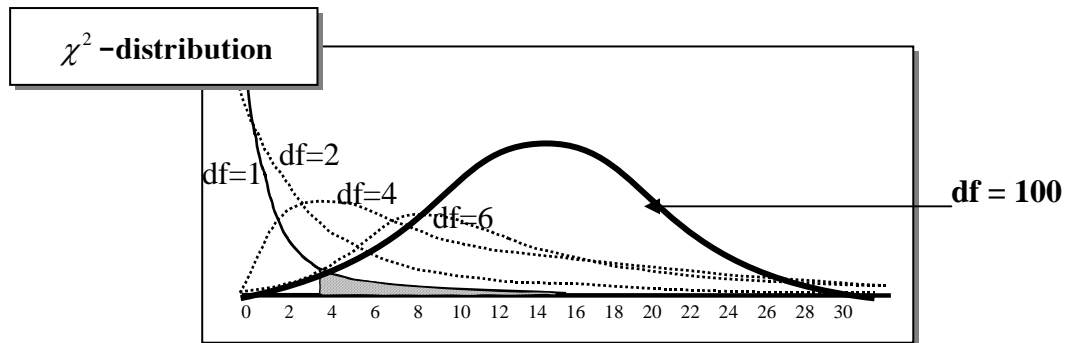
คำถาม : กราฟทั้งสามข้างล่างนี้ เป็นการแจกแจงความน่าจะเป็นของสามค่าสถิติ พื้นที่แรเงาของแต่ละกราฟมีสัดส่วนเท่ากับ 5% ของพื้นที่ทั้งหมดของกราฟนั้น

แกนตั้งแทน ความน่าจะเป็น แกนนอนแทน ค่าสัมประสิทธิ์

จงเขียนกราฟทับ t-distribution โดยให้ $df = 29$

จงเขียนกราฟทับ χ^2 -distribution โดยให้ $df = 100$





6.5

อภิปราย: ผล Significant อาจมีขนาดความแตกต่างที่มีหรือไม่มีค่าสำคัญทางการแพทย์และสาธารณสุขก็ได้ การศึกษาที่ได้ผลแบบ a และ b นั้น สรุปได้แล้วว่ามีความสำคัญค่าเพราะต่างกันมากกว่าระดับที่กำหนด ส่วน c นั้นยังไม่แน่แม้อาจมีโอกาสมีความสำคัญได้สูง แต่ d นั้นสรุปได้แล้วว่าไม่มีความสำคัญ ส่วนอีก 4 การศึกษาที่ผล Non-significant นั้น เฉพาะ b เท่านั้นที่อาจสรุปได้ว่าเท่ากัน ที่เหลือจะสรุปได้เพียงว่า "เราไม่มีเหตุผลเพียงพอที่จะสรุปว่าต่างกัน" แต่ความจริงอาจต่างกันหรือเท่ากันก็ได้ ช่วงเชื่อมั่นเท่านั้นที่เป็นข้อมูลให้ตัดสินใจเรื่องดังกล่าว ค่า p-value ไม่สามารถบอกได้ ตรงข้าม ถ้าช่วงเชื่อมั่น 95% คร่อมค่าที่บอกถึงความไม่แตกต่าง (Null value) ซึ่งในที่นี้คือ 0 สามารถบอกได้ว่า p-value > 0.05 หรือ Non-significant แต่ถ้าไม่คร่อม บอกได้ว่า p-value < 0.05 หรือ Significant

6.6

อภิปราย:

- ข้อ 1 ผิดเพราะค่า p-value ไม่ใช่สิ่งที่พิสูจน์อะไรได้ บอกแต่เพียงโอกาสความผิดพลาดในการตัดสินใจปฏิเสธสมมติฐานที่ถูกต้อง
- ข้อ 2 เป็นคำตอบที่ใกล้เคียงที่สุด และจะถูกต้องเมื่อเพิ่มที่ขีดเส้นใต้เข้าไปในประโยคที่เขียนใหม่ดังนี้ "ถ้าความเป็นจริงคือยาใหม่ไม่มีประสิทธิผลในการรักษาโรคแล้ว โอกาสที่จะได้ผลการรักษาตามที่กล่าวข้างต้นหรือมากกว่านั้น มีน้อยกว่า 5%" ทั้งนี้เพราะ p-value คือค่าความน่าจะเป็น หาได้จากพื้นที่ใต้กราฟ แต่ผลการศึกษารั้งเดียวจะได้สถิติค่าเดียว จึงเป็นเพียงจุดเดียวใต้กราฟไม่มีพื้นที่ของจุด มีแต่พื้นที่ของช่วง ข้อความ กล่าวอีกนัยหนึ่ง ค่าความน่าจะเป็นของเหตุการณ์เดียวนั้นเท่ากับศูนย์ หรือพื้นที่ใต้กราฟเป็นศูนย์นั่นเอง
- ข้อ 3 ผิดเพราะค่า p-value ไม่สามารถบอกขนาดของผลการรักษา (Magnitude of effect) ซึ่งในที่นี้คือผลต่างอัตราการหายจากการใช้ยาสองขนาน ในการศึกษาที่ขนาดตัวอย่างใหญ่มากๆ ขนาดความแตกต่างเพียงเล็กน้อยก็ยังผลให้ได้ค่า p-value น้อยมากได้ คือให้ผล Statistically significant แต่ไม่มีความสำคัญในทางการแพทย์และสาธารณสุข (Clinically unimportant)
- ข้อ 4 โดยทั่วไปยังพบว่ามีคนเลือกตอบข้อนี้มาก นับว่าเป็นเรื่องปกติมาก ขอเป็นกำลังใจให้ศึกษาต่อไป หรืออาจลองเริ่มทำตามคู่มือนี้อย่างจริงจังอีกครั้ง จนกว่า p-value จะไม่ใช่ Magic number สำหรับท่านอีกต่อไป และใช้ประโยชน์อย่างรู้เท่าทัน

ท่านได้รับการร้องขอให้ช่วยเขียนสรุปผลดังกล่าว โปรดเขียนให้อยู่ในรูปที่ผู้วิจัยสามารถตัดแปะ
ในสรุปผลการศึกษาของเขาได้ โดยสมมติตัวเลขใส่ลงไปในงานด้วย

การศึกษาเชิงทดลองทางการแพทย์โดยมีกลุ่มควบคุมเพื่อทราบว่ายาใหม่ขนานหนึ่งมีประสิทธิผลในการ
รักษาโรคหรือไม่ จากกลุ่มทดลอง 200 ราย อัตราการหายจากโรคคิดเป็นร้อยละ 80 และจากกลุ่มทดลอง 200 ราย
อัตราการหายจากโรคคิดเป็นร้อยละ 70 อัตราการหายนี้แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($p\text{-value} = 0.021$) โดย
ยาใหม่มีอัตราการหายสูงกว่ายาหลอกเท่ากับร้อยละ 10 (ช่วงเชื่อมั่น 95% อยู่ระหว่าง 1.6% ถึง 18.4%)

คำอธิบายเพิ่มเติม

ผลจากการศึกษาที่สมมติขึ้นนี้ สอดคล้องกับการศึกษา “d” ในแผนภูมิข้อ 4 ถ้าหากเป็นที่เห็น
ร่วมกันว่าจะตัดสินใจใช้ยาใหม่เมื่อมีอัตราการหายจากโรคสูงกว่าร้อยละ 25 เมื่อเปรียบเทียบกับยาอีก
ขนานหนึ่ง โรคนี้หายเองได้ (กลุ่มที่ใช้ยาหลอก) ถึงร้อยละ 70 แต่การใช้ยาใหม่ทำให้หายเพิ่มเพียงร้อยละ 10
ค่านี้ได้จากการศึกษาครั้งเดียว จากค่าช่วงเชื่อมั่นเราทราบว่าในประชากรความเป็นไปได้สูงสุดที่อัตราการ
หายจากยาใหม่ก็ไม่เกินร้อยละ 20 เมื่อเทียบกับยาหลอก (Upper limit ของช่วงเชื่อมั่น = 18.4%) ซึ่งต่ำ
กว่า 25% ถึงจะถือว่าดีกว่าจริง จึงได้ข้อสรุปว่า ไม่ใช้ยาใหม่ แม้ผลการทดลองจะ Significant ก็ตาม

ดังนั้นการนำเสนอค่าช่วงเชื่อมั่นในงานวิจัยจึงมีความจำเป็น และการแปลความหมายโดยใช้
ค่า $p\text{-value}$ อย่างเดียวจึงมีโอกาสผิดพลาดสูง ในงานวิจัยจึงควรนำเสนอทั้งช่วงเชื่อมั่น และ $p\text{-value}$
แต่ให้แปลความหมายบนช่วงเชื่อมั่นเป็นหลัก แล้วเอา $p\text{-value}$ มาช่วยเสริมว่าพยานหลักฐานเพียงพอที่จะ
ตัดสินใจตามนั้นได้หรือไม่เพียงใด

กิจกรรมที่ 6.2 :

คำเฉลยต่อไปนี มุ่งให้ผู้เรียนเกิดการผูกโยงแนวคิดทางสถิติเข้ากับการวิจัย โดยคาดหวังให้เห็น
ความสัมพันธ์ระหว่างวัตถุประสงค์การวิจัย ตัวแปร และการรูปแบบการวิจัย ว่าเป็นประเด็นพื้นฐานใน
การตัดสินใจเลือกวิธีการทางสถิติในการวิเคราะห์อย่างถูกต้องเหมาะสม

การอ่านวัตถุประสงค์วิจัยแล้วสามารถระบุตัวแปรตามและประเภทตัวแปรตาม และสามารถระบุตัวแปรต้นและประเภทตัวแปรต้นได้นั้น เป็นทักษะที่สำคัญที่ควรฝึก ส่วนประเด็นเกี่ยวกับรูปแบบการวิจัยในที่นี้ให้คุ้นเคยเฉพาะกับกลุ่มอิสระหรือไม่อิสระต่อกัน ในความเป็นจริงยังมีอีกมากมายเกินกว่าที่จะครอบคลุมในที่นี้ คำถามสุดท้ายเกี่ยวกับแนวทางการวิเคราะห์โดยสังเขปนั้น ไม่ได้คาดหวังผู้เรียนจะตอบได้ตามที่เฉลย แต่คาดหวังเพียงให้สามารถมองเป็นองค์รวม ไม่แยกส่วน กล่าวคือรู้ว่าการจะวิเคราะห์ข้อมูลเพื่อตอบคำถามวิจัยนั้น รวมขั้นตอนตั้งแต่การพรรณนาลักษณะกลุ่มที่ศึกษาจนถึงการอนุมานค่าจากตัวอย่างเป็นค่าของประชากร การอนุมานต้องมีทั้งช่วงเชื่อมั่นและการทดสอบสมมติฐาน ต้องมีการทดสอบข้อกำหนดเบื้องต้นในการใช้วิธีการทางสถิติที่เลือกก่อนการวิเคราะห์ซึ่งเป็น

Parametric statistics และทราบว่าต้องเลือกใช้ Non-parametric statistics กรณีที่ไม่ได้ตามข้อกำหนดเบื้องต้น ส่วนที่นอกเหนือจากนี้แม้กล่าวถึงแต่มีรายละเอียดเกินกว่าที่จะครอบคลุมในคู่มือนี้ จึงใส่ไว้เพียงเพื่อเป็นแนวทางสำหรับผู้สนใจเท่านั้น ศัพท์เฉพาะบางคำจึงไม่มีการขยายความแต่อย่างใด อย่างไรก็ตาม ผู้อ่านสามารถค้นคว้าตำราเล่มอื่นที่ให้ไว้ท้ายแผนภูมิที่ 2 กับ 3 ในบทที่ 1 เพื่อทำความเข้าใจต่อไปได้ ตัวอย่างการวิเคราะห์และการแปลความหมายแสดงไว้พอสังเขปที่ภาคผนวก 2

1. เพื่อประมาณค่าเกรดเฉลี่ยสะสมของนักศึกษาในมหาวิทยาลัย

➤ ตัวแปรตามคือ: เกรดเฉลี่ยสะสม (V7) เป็นตัวแปร ☒ ต่อเนื่อง ☐ แฉกนับ

ตัวแปรตามนี้ถูกวัด ☒ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน

➤ ตัวแปรต้นคือ: ไม่มี เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉกนับ

ถ้าเป็นตัวแปรแฉกนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)

➤ แนวทางการวิเคราะห์โดยสังเขป:

เป็นการประมาณค่าเฉลี่ยในกลุ่มตัวอย่างเดียว คือคำนวณค่าเกรดเฉลี่ย และคำนวณค่าช่วงเชื่อมั่น ถ้าข้อมูลแจกแจงแบบปกติ แต่ถ้ากรณีข้อมูลแจกแจงไม่เป็นแบบปกติ ให้พยายามทำ Data transformation หรือจัดการกับ Outliers ให้เหมาะสม จนกว่าจะได้แจกแจงแบบปกติ แล้ววิเคราะห์ด้วยวิธีตามที่กล่าว แต่ถ้ายังไม่แจกแจงแบบปกติ ให้รายงานค่ามัธยฐาน พร้อมกับ Interquartile range หรือคำนวณช่วงเชื่อมั่นโดยวิธี Bootstrapping methods

2. เพื่อศึกษาว่าเกรดเฉลี่ยสะสมในมหาวิทยาลัยของนักศึกษามากกว่า 2.0 หรือไม่

➤ ตัวแปรตามคือ: เกรดเฉลี่ยสะสม (V7) เป็นตัวแปร ☒ ต่อเนื่อง ☐ แฉงนับ

ตัวแปรตามนี้ถูกวัด ☒ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน

➤ ตัวแปรต้นคือ: ไม่มี เป็นตัวแปร ☐ ต่อเนื่อง ☐ แฉงนับ

ถ้าเป็นตัวแปรแฉงนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)

➤ แนวทางการวิเคราะห์โดยสังเขป:

เป็นการทดสอบค่าเฉลี่ยในกลุ่มตัวอย่างเดียว แต่ควรมีการประมาณค่าเฉลี่ยในกลุ่มเดียว คือคำนวณค่าเกรดเฉลี่ย และคำนวณค่าช่วงเชื่อมั่น ร่วมกับหาค่า p-value โดยใช้ One sample t-test ถ้าข้อมูลแฉงแบบปกติ แต่ถ้ากรณีข้อมูลแฉงไม่เป็นแบบปกติ ให้พยายามทำ Data transformation หรือจัดการกับ Outliers ให้เหมาะสม จนกว่าจะได้แฉงแบบปกติ แล้ววิเคราะห์ด้วยวิธีตามที่กล่าว แต่ยังไม่แฉงแบบปกติ ให้รายงานค่ามัธยฐาน พร้อมกับ Interquartile range หรือคำนวณช่วงเชื่อมั่นโดยวิธี Bootstrapping methods และทดสอบด้วย Wilcoxon Signed Rank test

3. เพื่อศึกษาว่าเกรดเฉลี่ยสะสมในมหาวิทยาลัยของนักศึกษาชายและหญิงแตกต่างกันหรือไม่

➤ ตัวแปรตามคือ: เกรดเฉลี่ยสะสม (V7) เป็นตัวแปร ☒ ต่อเนื่อง ☐ แฉงนับ

ตัวแปรตามนี้ถูกวัด ☒ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน

➤ ตัวแปรต้นคือ: เพศ (V2) เป็นตัวแปร ☐ ต่อเนื่อง ☒ แฉงนับ

ถ้าเป็นตัวแปรแฉงนับ ตัวแปรต้นนี้มี 2 ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)

➤ แนวทางการวิเคราะห์โดยสังเขป:

เป็นการอนุมานค่าเฉลี่ยสองกลุ่มตัวอย่างที่เป็นอิสระต่อกัน ขั้นแรกตรวจสอบการแฉงข้อมูลก่อน แล้วดำเนินการ Transform ข้อมูลหรือจัดการกับ Outlier ให้เหมาะสมถ้าจำเป็น เมื่อได้ข้อมูลแฉงแบบปกติ ให้คำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของแต่ละกลุ่ม เพื่อการพรรณนาข้อมูล จากนั้นหาผลต่างค่าเฉลี่ย ซึ่งเป็นค่าสถิติที่น่าสนใจ แล้วประมาณค่าช่วงเชื่อมั่น ตามด้วยการทดสอบสมมติฐานว่าค่าเฉลี่ยสองกลุ่มแตกต่างกันหรือไม่ โดยใช้ Two samples t-test กรณีข้อมูลแฉงไม่เป็นแบบปกติ ให้หาค่ามัธยฐานของแต่ละกลุ่ม และทดสอบทดสอบสมมติฐานโดย Mann-Whitney U test

4. เพื่อศึกษาว่าอัตราการเต้นของชีพจรของนักศึกษาก่อนและหลังวิ่งกับที่หนึ่งนาทีแตกต่างกันหรือไม่

- ตัวแปรตามคือ: อัตราชีพจร (V10 และ V11) เป็นตัวแปร ☒ ต่อเนื่อง ☐ เจนนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☒ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน
- ตัวแปรต้นคือ: ครึ่งที่วัด เป็นตัวแปร ☐ ต่อเนื่อง ☒ เจนนับ
ถ้าเป็นตัวแปรเจนนับ ตัวแปรต้นนี้มี 2 ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)
- แนวทางการวิเคราะห์โดยสังเขป:

เป็นการอนุมานค่าเฉลี่ยสองกลุ่มตัวอย่างที่ไม่เป็นอิสระต่อกัน ขั้นแรกหาผลต่างอัตราชีพจรหลังลบด้วยก่อนวิ่ง แต่ละรายจะได้ค่าผลต่างมาหนึ่งค่า จากนั้นหาค่าเฉลี่ยของผลต่าง แล้วประมาณค่าช่วงเชื่อมั่น ตามด้วยการทดสอบสมมติฐานว่าค่าเฉลี่ยของผลต่างดังกล่าวต่างจากศูนย์หรือไม่โดยใช้ Paired t-test กรณีที่ข้อมูลผลต่างชีพจรแจกแจงแบบไม่ปกติ ไม่ว่าจะ Transformation หรือจัดการ Outlier แล้วก็ไม่สำเร็จ จะใช้ Wilcoxon Matched pairs Signed Ranks Test เป็นสถิติทดสอบ

5. เพื่อศึกษาว่านักศึกษามีภูมิลำเนาต่างกันจะมีกรดเฉลี่ยสะสมในมหาวิทยาลัยแตกต่างกันหรือไม่

- ตัวแปรตามคือ: กรดเฉลี่ยสะสม (V7) เป็นตัวแปร ☒ ต่อเนื่อง ☐ เจนนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน
- ตัวแปรต้นคือ: ภูมิลำเนา (V9) เป็นตัวแปร ☐ ต่อเนื่อง ☒ เจนนับ
ถ้าเป็นตัวแปรเจนนับ ตัวแปรต้นนี้มี 4 ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)
- แนวทางการวิเคราะห์โดยสังเขป:

เป็นการอนุมานค่าเฉลี่ยสี่กลุ่มตัวอย่างที่เป็นอิสระต่อกัน ขั้นแรกตรวจสอบการแจกแจงข้อมูลก่อนก่อน แล้วดำเนินการ Transform ข้อมูลหรือจัดการกับ Outlier ให้เหมาะสมถ้าจำเป็น เมื่อได้ข้อมูลแจกแจงแบบปกติ ให้คำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของแต่ละกลุ่ม เพื่อการพรรณนาข้อมูล จากนั้นหาผลต่างค่าเฉลี่ยระหว่างคู่ที่สนใจจะเปรียบเทียบ แล้วประมาณค่าช่วงเชื่อมั่น (ดูวิธีการใน Wonnacot and Wonnacot, 1990 หน้า 344) ตามด้วยการทดสอบสมมติฐานว่ามีค่าเฉลี่ยอย่างน้อยหนึ่งคู่ที่ต่างกันหรือไม่ โดยใช้ Analysis of Variance (ANOVA) แต่สถิตินี้มีข้อตกลงเบื้องต้นที่สำคัญมากคือค่า Variance ในแต่ละกลุ่มต้องเท่ากัน ดังนั้นต้องทดสอบ Assumption นี้ก่อนก่อนวิเคราะห์ผลการวิเคราะห์โดย ANOVA กรณีผล p-value < 0.05 ให้ทดสอบเพื่อหาว่าคู่ใดที่แตกต่างกัน (ทำ Multiple comparison) กรณีข้อมูลแจกแจงไม่เป็นแบบปกติและความแปรปรวนระหว่างกลุ่มไม่เท่ากัน ให้หาค่ามัธยฐานของแต่ละกลุ่ม และทดสอบทดสอบสมมติฐานโดย Kruskal Wallis test

6. เพื่อศึกษาความสัมพันธ์ระหว่างน้ำหนักกับอัตราชีพจรขณะพัก

- ตัวแปรตามคือ: น้ำหนัก(V5) และอัตราชีพจร(V10) เป็นตัวแปร ☒ ต่อเนื่อง ☐ เจนนับ
ตัวแปรตามนี้ถูกวัด ☒ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน
- ตัวแปรต้นคือ: ไม่มี เป็นตัวแปร ☐ ต่อเนื่อง ☐ เจนนับ
ถ้าเป็นตัวแปรเจนนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)
- แนวทางการวิเคราะห์โดยสังเขป:

เป็นการวิเคราะห์หาระดับความสัมพันธ์ระหว่างสองตัวแปรที่ทั้งคู่เป็นตัวแปรต่อเนื่องโดยใช้สหสัมพันธ์ (Pearson Correlation) ขั้นแรกตรวจสอบการแจกแจงข้อมูลก่อนก่อน แล้วดำเนินการ Transform ข้อมูลหรือจัดการกับ Outlier ให้เหมาะสมถ้าจำเป็น เมื่อได้ข้อมูลแจกแจงแบบปกติ ขั้นที่สองสร้าง Scatter plot เพื่อดูว่าตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นต่อกันหรือไม่ ขั้นที่สามคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ (r) และประมาณค่าช่วงเชื่อมั่น ขั้นสุดท้ายคือทดสอบสมมติฐานว่าค่า r ต่างจากศูนย์อย่างมีนัยสำคัญหรือไม่ กรณีการแจกแจงไม่เป็นแบบปกติ ใช้ Spearman Rank Correlation กรณีความสัมพันธ์ที่เห็นจาก Scatter plot ไม่เป็นเชิงเส้น ต้องตีความขนาดของ r อย่างระมัดระวัง และอาจต้องวิเคราะห์จำแนกตามตัวแปรที่สาม ถ้าพบว่าทำเช่นนั้นแล้วมีความสัมพันธ์เป็นเชิงเส้น

7. เพื่อศึกษาว่าน้ำหนักสามารถใช้ทำนายอัตราชีพจรขณะพักได้หรือไม่ อย่างไร

- ตัวแปรตามคือ: อัตราชีพจร (V10) เป็นตัวแปร ☒ ต่อเนื่อง ☐ เจนนับ
ตัวแปรตามนี้ถูกวัด ☒ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน
- ตัวแปรต้นคือ: น้ำหนัก (V5) เป็นตัวแปร ☒ ต่อเนื่อง ☐ เจนนับ
ถ้าเป็นตัวแปรเจนนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)
- แนวทางการวิเคราะห์โดยสังเขป:

เป็นการวิเคราะห์ความสัมพันธ์ในเชิงคาดคะเนที่ตัวแปรตามเป็นตัวแปรต่อเนื่องโดยที่ตัวแปรต้นเป็นตัวแปรประเภทใดก็ได้ แต่มีเพียงหนึ่งตัว กรณีมีตัวแปรต้นตัวเดียวนี้ สถิติที่ใช้คือการถดถอยเชิงเส้นอย่างง่าย (Simple linear regression) ค่าสถิติที่ต้องคำนวณคือสัมประสิทธิ์การถดถอย (β) ซึ่งเป็นค่าที่ใช้อธิบายผลของตัวแปรต้นต่อตัวแปรตาม จากนั้นประมาณค่าช่วงเชื่อมั่น และการทดสอบสมมติฐานว่าค่า β ต่างจากศูนย์อย่างมีนัยสำคัญหรือไม่ สุดท้ายคือค่า Coefficient of determination หรือ R^2 เพื่อบอกสมการการถดถอยที่ได้นั้นสามารถอธิบายความแปรปรวนของข้อมูลได้ร้อยละเท่าใด บางกรณียังอาจต้องคำนวณค่าตัวแปรตามที่ได้จากการคาดคะเนโดยตัวแปรต้น พร้อมช่วงเชื่อมั่น ทั้งนี้ขึ้นอยู่กับผู้วิจัยว่าต้องการใช้ผลการศึกษานี้เพียงเพื่ออธิบายความสัมพันธ์หรือนำไปใช้คาดเป็นคะแนนสำหรับการคาดคะเนตัวแปรตามในการปฏิบัติงานจริง การทดสอบข้อกำหนดเบื้องต้นสำหรับ Regression จะกระทำหลังจากได้สมการการถดถอยเรียบร้อยแล้ว

8. ผู้วิจัยสร้างตัวแปรขึ้นมาใหม่ชื่อ "การเจ็บป่วย" ให้มีค่าเป็น *ไม่ป่วย* กับ *ป่วย* โดยถือว่านักศึกษา *ไม่ป่วย* เมื่อตัวแปรในข้อ 14 ของแบบสอบถาม คือสถานสุขภาพมีค่า 1 (เยี่ยม) และ 2 (ดี) และ *ป่วย* คือสถานสุขภาพมีค่า 3 (ปานกลาง) และ 4 (แย่) วัตถุประสงค์การวิจัยคือ เพื่อประมาณค่าความชุก (Prevalence) ของการป่วย (ค่าร้อยละ)

➤ ตัวแปรตามคือ: การเจ็บป่วย เป็นตัวแปร [] ต่อเนื่อง [☒] แยกนับ
ตัวแปรตามนี้ถูกวัด ☒ ครั้งเดียว [] มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน

➤ ตัวแปรต้นคือ: ไม่มี เป็นตัวแปร [] ต่อเนื่อง [] แยกนับ
ถ้าเป็นตัวแปรแยกนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)

➤ แนวทางการวิเคราะห์โดยสังเขป:

เป็นการประมาณค่าสัดส่วน (P) ในกลุ่มตัวอย่างเดียว ขั้นแรกคำนวณค่าความชุกการเจ็บป่วย นี้คือค่าสถิติที่สนใจ จากนั้นให้ทดสอบว่าขนาดตัวอย่าง (n) ใหญ่พอจะใช้การแจกแจงปกติประมาณค่าได้หรือไม่โดยทั้ง $n \times P$ และ $n \times (1-P)$ ต้องมากกว่า 5 ถ้าได้ ก็คำนวณช่วงเชื่อมั่นโดย Normal approximation แต่ถ้าไม่ได้ต้องคำนวณ Binomial exact confidence intervals มีสูตรและวิธีการใน Rosner (1990) หน้า 172

9. เพื่อศึกษาว่าความชุกการป่วยของนักศึกษาแตกต่างจากของประชากรทั่วไป ซึ่งเท่ากับ 5% หรือไม่

➤ ตัวแปรตามคือ: การเจ็บป่วย เป็นตัวแปร [] ต่อเนื่อง [☒] แยกนับ
ตัวแปรตามนี้ถูกวัด ☒ ครั้งเดียว [] มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน

➤ ตัวแปรต้นคือ: ไม่มี เป็นตัวแปร [] ต่อเนื่อง [] แยกนับ
ถ้าเป็นตัวแปรแยกนับ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)

แนวทางวิเคราะห์โดยสังเขป:

เป็นการทดสอบค่าสัดส่วน (P) ในกลุ่มตัวอย่างเดียว อย่างไรก็ตามควรมีการประมาณค่าด้วยซึ่งได้กล่าวแล้วในข้อ 8 ส่วนการทดสอบสมมติฐานว่า P ต่างจาก 0.05 (จากโจทย์) หรือไม่นั้น ขั้นแรกให้ทดสอบว่าขนาดตัวอย่าง (n) ใหญ่พอจะใช้การแจกแจงปกติประมาณค่าได้หรือไม่โดยทั้ง $n \times P$ และ $n \times (1-P)$ ต้องมากกว่า 5 ถ้าได้ ก็ใช้สถิติทดสอบ Z-test แต่ถ้าไม่ได้ต้องคำนวณ Binomial exact probability test ซึ่งได้ค่า Exact p-value โดยตรง

10. เพื่อศึกษาว่าความชุกการเจ็บป่วยในกลุ่มนักศึกษาชายและกลุ่มนักศึกษาหญิงแตกต่างกันหรือไม่

➤ ตัวแปรตามคือ: การเจ็บป่วย เป็นตัวแปร ☐ ต่อเนื่อง ☒ แง่

ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน

➤ ตัวแปรต้นคือ: เพศ (V2) เป็นตัวแปร ☐ ต่อเนื่อง ☒ แ่ง

ถ้าเป็นตัวแปรแง่ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)

➤ แนวทางการวิเคราะห์โดยสังเขป:

เป็นการอนุมานค่าสัดส่วนในสองกลุ่มตัวอย่างที่เป็นอิสระต่อกัน ขั้นแรกให้คำนวณค่าความชุกการเจ็บป่วยจำแนกตามเพศ จากนั้นหาผลต่างของค่าความชุกดังกล่าว แล้วประมาณค่าช่วงเชื่อมั่นของผลต่างดังกล่าว จากนั้นทดสอบสมมติฐานว่าค่าความชุกแตกต่างกันหรือไม่ โดยสามารถใช้ได้ทั้ง Z-test หรือ Chi-square test กรณีตัวอย่างขนาดเล็ก ใช้ Fisher's exact test

11. เพื่อศึกษาความสัมพันธ์ระหว่างการเจ็บป่วยกับเพศ

➤ ตัวแปรตามคือ: การเจ็บป่วย เป็นตัวแปร ☐ ต่อเนื่อง ☒ แ่ง

ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน

➤ ตัวแปรต้นคือ: เพศ (V2) เป็นตัวแปร ☐ ต่อเนื่อง ☒ แ่ง

ถ้าเป็นตัวแปรแง่ ตัวแปรต้นนี้มี ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)

➤ แนวทางการวิเคราะห์โดยสังเขป:

เป็นการวิเคราะห์ระดับความสัมพันธ์ในสองกลุ่มตัวอย่างที่เป็นอิสระต่อกัน ขั้นแรกให้คำนวณค่าความชุกการเจ็บป่วยจำแนกตามเพศ จากนั้นคำนวณมาตรวัดความสัมพันธ์ซึ่งใช้ Relative risk (RR) กรณี Cohort study และใช้ Odds ratio (OR) กรณี Cross-sectional study หรือ Case-control study จากนั้นประมาณค่าช่วงเชื่อมั่นของค่าดังกล่าว ตามด้วยการทดสอบสมมติฐานว่าค่า RR หรือ OR นั้นแตกต่างจาก 1 หรือไม่ โดยใช้ Chi-square test กรณีตัวอย่างขนาดเล็ก ใช้ Fisher's exact test

12. เพื่อศึกษาว่าสุขภาพจิต (เกิดอาการเครียด) ของนักศึกษา ก่อนและหลังเข้าเรียนมหาวิทยาลัย แตกต่างกันหรือไม่

➤ ตัวแปรตามคือ: สุขภาพจิต (V12 และ V13) เป็นตัวแปร ☐ ต่อเนื่อง ☒ แยกนับ
ตัวแปรตามนี้ถูกวัด ☐ ครั้งเดียว ☒ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน

➤ ตัวแปรต้นคือ: ครั้งของการวัด เป็นตัวแปร ☐ ต่อเนื่อง ☒ แยกนับ
ถ้าเป็นตัวแปรแยกนับ ตัวแปรต้นนี้มี 2 ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)

➤ แนวทางการวิเคราะห์โดยสังเขป:
เป็นการอนุมานค่าสัดส่วนในสองกลุ่มตัวอย่างที่ไม่เป็นอิสระต่อกัน ขั้นแรกให้คำนวณค่าความชุกก่อนและหลังเข้ามหาวิทยาลัย จากนั้นคำนวณผลต่างของความชุกจากข้อมูลที่แจกแจงลงในตาราง 2×2 ที่แสดงความสัมพันธ์ระหว่างการเจ็บป่วยก่อนและหลังเข้ามหาวิทยาลัย แล้วประมาณค่าช่วงเชื่อมั่นของผลต่างดังกล่าว จากนั้นทดสอบสมมติฐานว่าค่าความแตกต่างความชุกแตกต่างจากศูนย์หรือไม่ โดยสามารถใช้ได้ทั้ง Z-test หรือ McNemar Chi-square test กรณีตัวอย่างขนาดเล็ก ใช้ Binomial exact probability test

13. เพื่อศึกษาว่าภูมิฉันทามีความสัมพันธ์กับความชุกการเจ็บป่วยหรือไม่

➤ ตัวแปรตามคือ: การเจ็บป่วย เป็นตัวแปร ☐ ต่อเนื่อง ☒ แยกนับ
ตัวแปรตามนี้ถูกวัด ☒ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน

➤ ตัวแปรต้นคือ: ภูมิฉันทา (V9) เป็นตัวแปร ☐ ต่อเนื่อง ☒ แยกนับ
ถ้าเป็นตัวแปรแยกนับ ตัวแปรต้นนี้มี 4 ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)

➤ แนวทางการวิเคราะห์โดยสังเขป:
เป็นการวิเคราะห์ระดับความสัมพันธ์ในสี่กลุ่มตัวอย่างที่เป็นอิสระต่อกัน ขั้นแรกให้คำนวณค่าความชุกการเจ็บป่วยจำแนกตามภาค จากนั้นคำนวณมาตรวัดความสัมพันธ์ซึ่งใช้ Odds ratio โดยให้ภาคใดภาคหนึ่งเป็นฐานเปรียบเทียบ เรียกว่า Local odds ratio จากนั้นประมาณค่าช่วงเชื่อมั่นของค่าดังกล่าวซึ่งตามตัวอย่างนี้มี 3 ค่า ตามด้วยการทดสอบความสัมพันธ์โดยใช้ Chi-square test กรณีตัวอย่างขนาดเล็ก ใช้ Fisher's exact test

14. เพื่อศึกษาปัจจัยที่มีผลต่อเกรดเฉลี่ยสะสมในมหาวิทยาลัย

➤ ตัวแปรตามคือ: เกรดเฉลี่ย (V7) เป็นตัวแปร ☒ ต่อเนื่อง ☐ แยกนับ

ตัวแปรตามนี้ถูกวัด ☒ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน

➤ ตัวแปรต้นคือ: ทุกตัวแปรที่อาจมีผลต่อเกรดเฉลี่ยสะสม เช่นอายุ เพศ ความเครียด ฯลฯ

➤ แนวทางการวิเคราะห์โดยสังเขป:

นี่คือคำถามวิจัยที่ต้องวิเคราะห์ข้อมูลแบบ Multivariable analysis คือวิเคราะห์ตัวแปรตามหนึ่งตัวร่วมกับตัวแปรต้นหลายตัวในคราวเดียวกัน กรณีนี้ตัวแปรตามเป็นตัวแปรต่อเนื่อง สถิติที่ใช้คือ Multiple regression มีวิธีการคล้ายกับที่กล่าวในข้อ 7 คือคำนวณ β ประมาณค่าช่วงเชื่อมั่น ทดสอบสมมติฐานเกี่ยวกับค่า β คำนวณค่า R^2 บางกรณีอาจต้องคำนวณค่าเกรดเฉลี่ยสะสมที่คาดคะเนโดยตัวแปรต้นทั้งหมดในคราวเดียวกัน พร้อมช่วงเชื่อมั่น และการทดสอบข้อกำหนดเบื้องต้นสำหรับ Regression จะกระทำหลังจากได้ผลการถดถอยเรียบร้อยแล้ว ที่ต่างไปจากข้อ 7 คือการได้มาซึ่งสมการ (Model fitting strategies) คือจะนำตัวแปรต้นใดเข้าใน Model นำตัวแปรใดออก เป็นต้น ซึ่งมีความสลับซับซ้อนกว่ามาก

15. เพื่อศึกษาปัจจัยที่มีผลต่อ การเจ็บป่วย

➤ ตัวแปรตามคือ: การเจ็บป่วย เป็นตัวแปร ☐ ต่อเนื่อง ☒ แยกนับ

ตัวแปรตามนี้ถูกวัด ☒ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน

➤ ตัวแปรต้นคือ: ทุกตัวแปรที่อาจมีผลต่อการเจ็บป่วย เช่นอายุ เพศ ความเครียด ฯลฯ

➤ แนวทางการวิเคราะห์โดยสังเขป:

คล้ายกับข้อ 14 เพียงแต่กรณีนี้ตัวแปรตามเป็นตัวแปรแยกนับ ในที่นี้เป็น Dichotomous variable เพราะมีค่าเป็นค่าใดค่าหนึ่งจากทั้งหมดสองค่า สถิติที่ใช้คือ Multiple Logistic regression การวิเคราะห์จะได้ค่า β ซึ่งอธิบายโอกาสที่จะป่วยอันเนื่องมาจากผลของตัวแปรต้นนั้น ถ้าเป้าหมายเพื่อหาปัจจัยเสี่ยงตามที่โจทย์กำหนดนี้ เป้าหมายของค่าสถิติที่ต้องคำนวณออกมาคือนำโอกาสที่จะป่วยที่ได้จากค่า β ไปหาค่าความเสี่ยงที่เรียกว่า Odds ratio (OR) คล้ายกับผลที่ได้ตามข้อ 13 เพียงแต่ในข้อ 13 มีตัวแปรต้นตัวเดียวจึงเป็น Crude OR แต่ที่ได้จาก Multiple Logistic regression จะเรียก Adjusted OR คือได้ควบคุมผลกระทบตัวแปรอื่นๆ ด้วยแล้ว จากนั้นคำนวณช่วงเชื่อมั่นของ OR ตามด้วยทดสอบสมมติฐานว่าค่า Adjusted OR นั้นต่างจาก 1 หรือไม่ การได้มาซึ่งสมการมีความสลับซับซ้อน ได้สมการแล้วก็ต้องมีการตรวจสอบความถูกต้องเหมาะสมของสมการด้วย ซึ่งรายละเอียดเหล่านี้ ผู้อ่านสามารถศึกษาได้ในหนังสือ "การวิเคราะห์ข้อมูลการวิจัยทางวิทยาศาสตร์สุขภาพโดยใช้การถดถอยโลจิสติก" โดยผู้แต่งคนเดียวกับผู้แต่งคู่มือเล่มนี้

16. ประยุกต์แนวคิดทางสถิติกับการคำนวณขนาดตัวอย่าง

โจทย์: ผู้เข้ารับการอบรมได้รับมอบหมายให้ทำการศึกษาวิจัยเพื่อเปรียบเทียบน้ำหนักของนักศึกษา


ชายกับนักศึกษาหญิงที่มหาวิทยาลัยเชียงใหม่ จงคำนวณขนาดตัวอย่างสำหรับการศึกษาวิจัยดังกล่าว

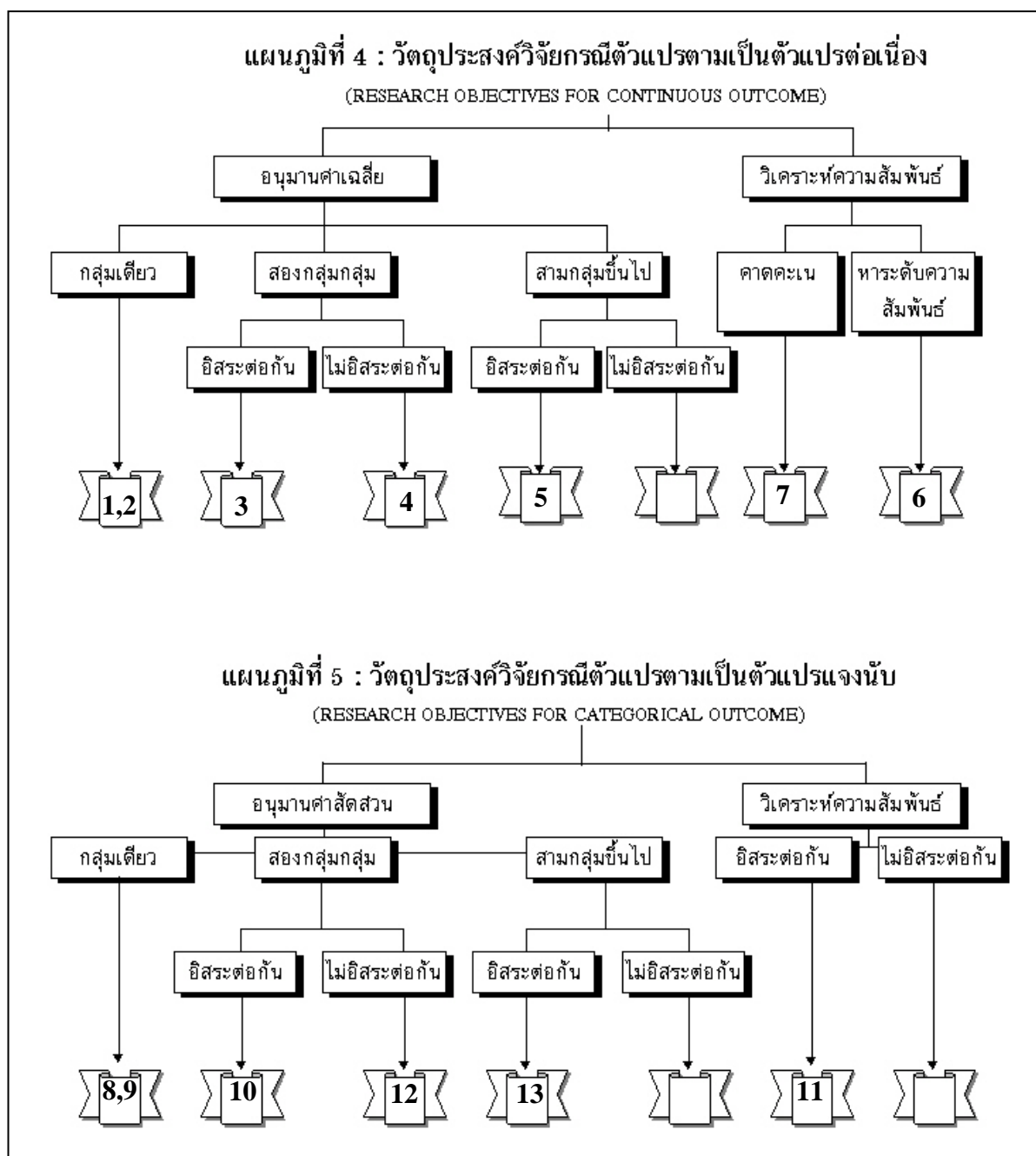
- ตัวแปรตามคือ: น้ำหนัก (V5) เป็นตัวแปร ☒ ต่อเนื่อง ☐ แยกนับ
ตัวแปรตามนี้ถูกวัด ☒ ครั้งเดียว ☐ มากกว่าหนึ่งครั้ง ในนักศึกษาแต่ละคน
- ตัวแปรต้นคือ: เพศ (V2) เป็นตัวแปร ☐ ต่อเนื่อง ☒ แยกนับ
ถ้าเป็นตัวแปรแยกนับ ตัวแปรต้นนี้มี 2 ค่า (ซึ่งหมายถึงจำนวนกลุ่มนั่นเอง)
- แนวทางการคำนวณขนาดตัวอย่างโดยสังเขป:

ค้นตารางสถิติ หาสูตรการคำนวณขนาดตัวอย่างสำหรับค่าเฉลี่ยในตัวอย่างสองกลุ่มที่เป็นอิสระต่อกัน แล้วคำนวณตามแนวทางที่ให้ไว้ในตารางเล่มนั้น

หมายเหตุ: ข้อสุดท้ายนี้ คำถามเกี่ยวกับ แนวทางการคำนวณขนาดตัวอย่างโดยสังเขปนั้น โปรดบอกเพียงแนวทางการค้นหาสูตรการคำนวณขนาดตัวอย่างที่เหมาะสม

กิจกรรมที่ 6.3 :

จากตัวอย่างวัตถุประสงค์การวิจัยข้อ 1 ถึง 15 ในกิจกรรมที่ 6.2 จงนำหมายเลขหน้าวัตถุประสงค์การวิจัยดังกล่าว ใส่ในกรอบ  สำหรับแต่ละกรณีในแผนภูมิที่ 4 และ 5 ต่อไปนี้



14 = Multivariable analysis for continuous outcome

15 = Multivariable analysis for categorical outcome

คำอธิบายเพิ่มเติมเนื้อหาที่ยังไม่ครอบคลุม

ในแผนภูมิที่ 4 ช่องที่เว้นว่างคือกรณีตัวแปรต่อเนื่องสามกลุ่มขึ้นไปที่ไม่เป็นอิสระต่อกันนั้น ตัวอย่างวัตถุประสงค์การวิจัยเช่น เพื่อศึกษาว่าเกรดเฉลี่ยสะสมในมหาวิทยาลัย ปีที่ 1 ปีที่ 2 ปีที่ 3 และปีที่ 4 โดยเฉลี่ย มีแนวโน้มอย่างไร ซึ่งไม่มีการเก็บข้อมูลนี้ในแบบสอบถามการสำรวจนักศึกษา นี่เป็นการวัดตัวแปรเดียวซ้ำหน่วยสังเกต (นักศึกษา) เรียกว่า Repeated measures data ต้องวิเคราะห์โดยคำนึงถึงความไม่เป็นอิสระของแต่ละค่าข้อมูลด้วย เช่นการสรุปค่าเกรดเฉลี่ยแต่ละคนให้เป็นค่าเดียวก่อน แล้ววิเคราะห์ที่ค่าสรุปนั้นโดยใช้สถิติอย่างง่ายตามที่กล่าวข้างต้น ค่าที่สรุป Repeated measures data ในแต่ละหน่วยสังเกตนี้เรียกว่า Summary measure ซึ่งอาจเป็นค่าเฉลี่ย ค่าต่ำสุด ค่าสูงสุด หรือค่า Slope เป็นต้น วิธีนี้นิยมกันเพราะง่ายต่อการทำการวิเคราะห์และแปลความหมาย แต่ที่ตรงและมีประสิทธิภาพกว่าคือ Repeated measure ANOVA ซึ่งสลับซับซ้อน อาจต้องปรึกษานักสถิติที่มีประสบการณ์ ข้อมูลที่ไม่เป็นอิสระต่อกันอาจเป็นรูปแบบวัดซ้ำตามเวลาดังตัวอย่างนี้ เรียกว่า Longitudinal data ยังมีรูปแบบอื่นเช่น น้ำหนักแรกเกิดของลูกสุกร เพราะแม่สุกรหนึ่งมีลูกหลายตัว แล้วเราเก็บข้อมูลจากหลายแม่สุกร เช่นนี้เรียก Clustered data เหล่านี้ล้วนต้องวิเคราะห์โดยคำนึงถึงความไม่เป็นอิสระของแต่ละค่าข้อมูลเสมอ

ในแผนภูมิที่ 5 ช่องที่เว้นว่าง 2 ช่อง ช่องแรกคือกรณีตัวแปรแจกแจงสามกลุ่มขึ้นไปที่ไม่เป็นอิสระต่อกันนั้น ตัวอย่างเช่น "เพื่อศึกษาว่าสุขภาพจิต (เกิดอาการเครียด) ของนักศึกษา ในแต่ละปีการศึกษาของ 4 ปีแรกที่ศึกษาในมหาวิทยาลัย ว่าแตกต่างกันหรือไม่ระหว่างนักศึกษาชายและหญิง เป็น Repeated measures data เช่นเดียวกันกับกรณีแรก แต่การวิเคราะห์ยากกว่ามาก และไม่มี Summary measure ที่จะเอื้อให้ใช้สถิติอย่างง่ายได้ สถิติที่ใช้ได้แก่ Generalized Linear Model using Generalized Estimating Equations หรือที่เรียกสั้นๆ กันว่า GEE หรือวิธีการที่มีประสิทธิภาพสูงขึ้นได้แก่ Linear mixed model ช่องสุดท้ายของแผนภูมิคือการวิเคราะห์ความสัมพันธ์กลุ่มไม่อิสระต่อกัน จะวิเคราะห์คล้าย Repeated measure data แต่ถ้าเป็นการวิเคราะห์ความสัมพันธ์ระหว่างสองตัวแปร โดยแต่ละตัวแปรเป็น Dichotomous variable จะคล้ายกับวัตถุประสงค์วิจัยในข้อ 12 แต่เขียนใหม่เป็น "เพื่อศึกษาว่าสุขภาพจิต (เกิดอาการเครียด) ของนักศึกษาสัมพันธ์กับการเข้าเรียนมหาวิทยาลัยหรือไม่" การวิเคราะห์ยังมีเป้าหมายเพื่อคำนวณค่า OR แต่มีวิธีการที่คำนึงถึงความไม่เป็นอิสระต่อกันด้วย วิธีการคำนวณจึงต่างจากในข้อ 13

ที่ไม่มีในทั้งสองแผนภูมิคือ กรณีตัวแปรตามเป็นระยะปลอดเหตุการณ์ ศัพท์เฉพาะเรียกข้อมูลประเภทนี้ว่า Censored data การวิเคราะห์จะต้องใช้ Survival analysis ตัวอย่างข้อมูลประเภทนี้ได้แก่ "ผลการรักษา (ตาย, รอด)" ของการทดลองเพื่อศึกษาประสิทธิผลของยาสองขนาน การติดตามผลการรักษาผู้ป่วย 5 ปีหลังการรักษา แล้ววิเคราะห์เป็นอัตราตายเปรียบเทียบกับเสมือนกับที่กล่าวแล้ว

ข้างต้นนั้นไม่ถูกต้องเพราะผู้ป่วยบางรายอาจตายเพราะสาเหตุอื่นที่ไม่ใช่โรคที่ยานันรักษา หรือบางคนยังไม่ตายภายใน 5 ปีหลังการรักษา ผู้ป่วยเหล่านี้เรียก Censored observation จะต้องได้รับการจัดการที่เหมาะสมในการวิเคราะห์ข้อมูล โดยวิธี Survival analysis รายละเอียดสามารถศึกษาในหนังสือ “การวิเคราะห์ระยะปลอดเหตุการณ์ในการวิจัยด้านวิทยาศาสตร์สุขภาพ” โดยผู้แต่งคนเดียวกันกับที่แต่งคู่มือนี้

ทั้งหมดนี้เป็นภาพที่สมบูรณ์ของสถิติสำหรับการวิจัยในทางการแพทย์และสาธารณสุขที่ผู้อ่านพึงรู้เพื่อเป็นฐานสำคัญในการศึกษารายละเอียดในแต่ละเรื่องเมื่อถึงเวลาที่ต้องใช้ต่อไป

ภาคผนวก 2

การวิเคราะห์ข้อมูลตามทีละบุในเฉลยคู่มือโดยใช้ STATA

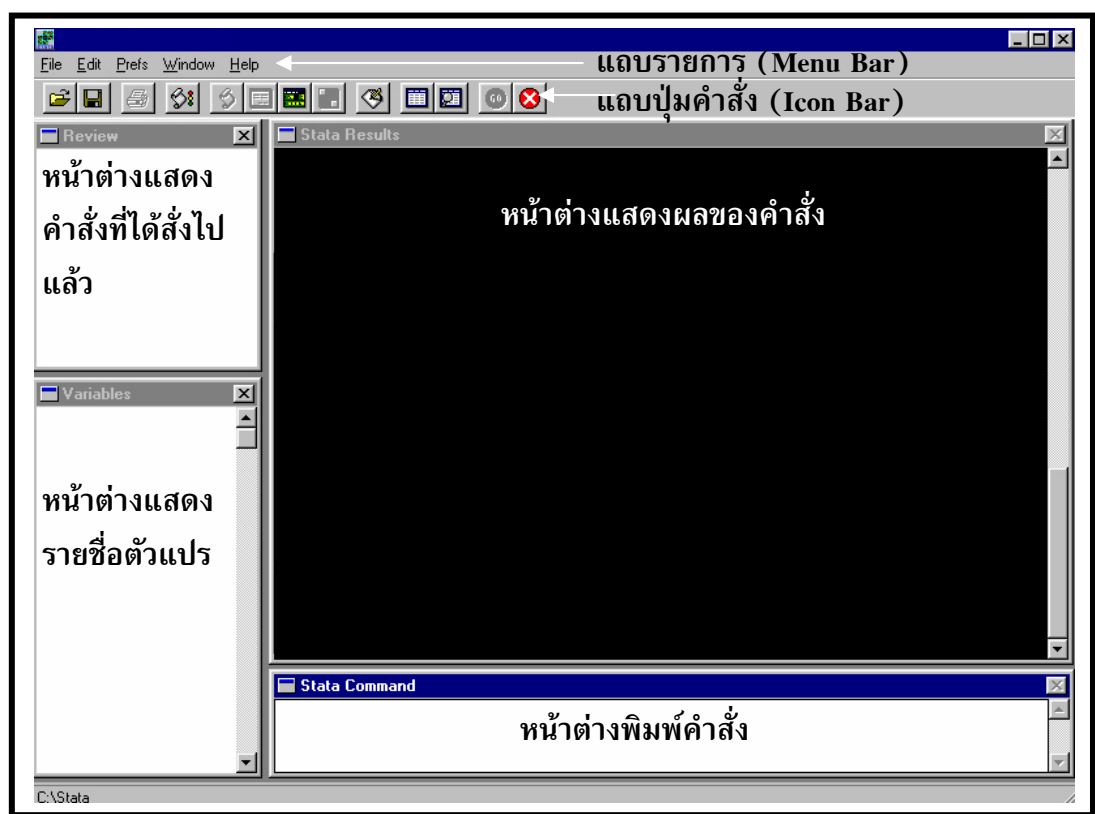
คำสั่ง STATA แสดงด้วยอักษรทึบ ส่วนผลแสดงด้วยอักษรธรรมดา แฟ้มข้อมูลที่ใช้วิเคราะห์ทั้งหมดสามารถ Download ได้จาก <http://web.kku.ac.th/~bandit/data>

บทนำสู่ STATA

ขั้นตอนการป้อนข้อมูลเข้า STATA (ต่อไปนี้จะทุกคำสั่งของ STATA จะแสดงไว้หลังจุด แต่การพิมพ์คำสั่งลงใน STATA ไม่ต้องพิมพ์จุด)

การนำข้อมูลเข้าวิเคราะห์ใน STATA ทำได้ 4 ทางคือ (1)ป้อนเข้าที่ Data Editor โดยตรง (2) แปลงจากแฟ้มข้อมูลอื่น เช่น dBASE หรือ Excel ซึ่งสามารถใช้การคัดลอกแล้วข้ามมาวางที่หน้าต่าง Data Editor ของ STATA ก็ได้ (3)อ่านจากแฟ้มข้อมูลที่เป็น STATA Format คือแฟ้มที่มีสกุลเป็น .dta และ (4)พิมพ์เข้าโดยตรงต่อท้ายคำสั่งประเภท Immediate commands ในเบื้องต้นนี้ แสดงการนำเข้าโดยวิธีแรกคือป้อนข้อมูลเข้าที่ Data Editor ของ STATA โดยตรง

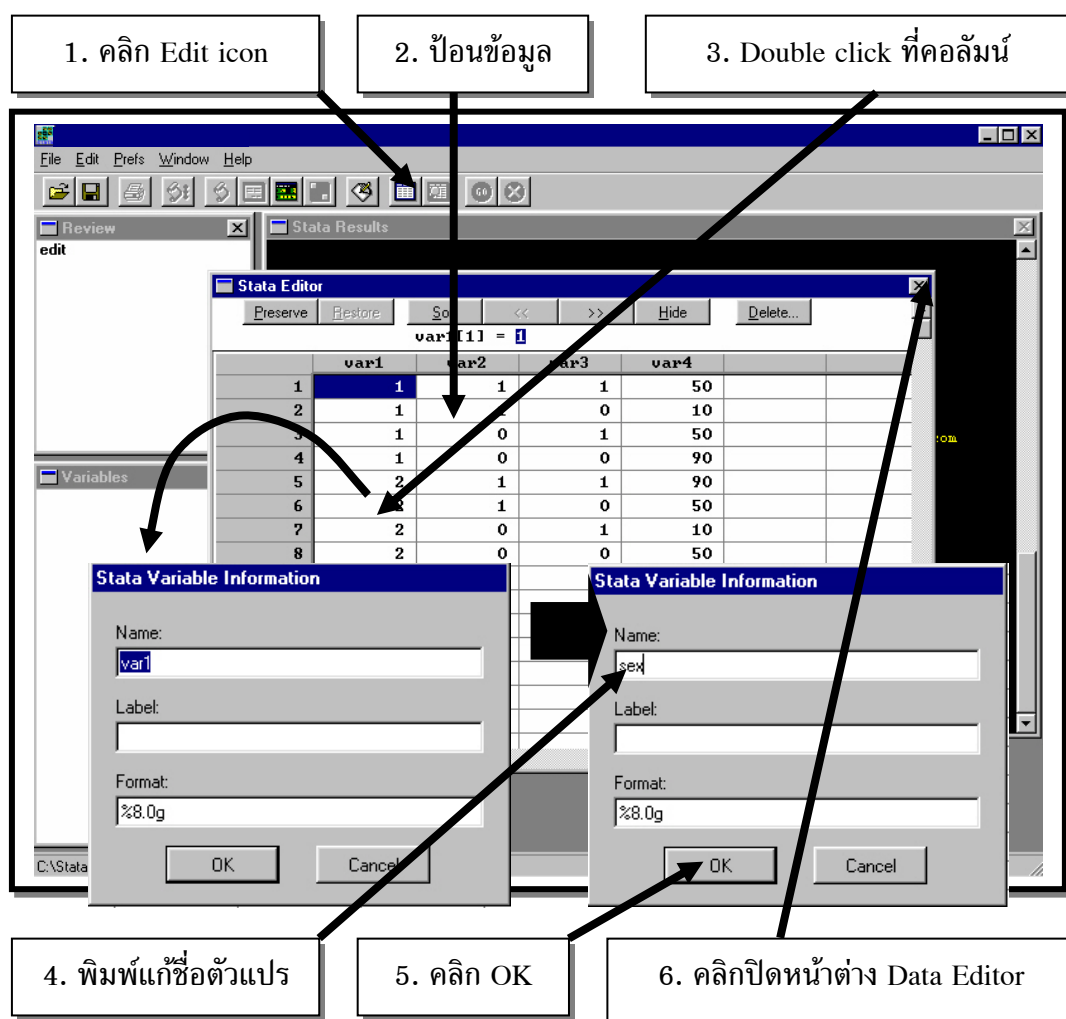
ขั้นแรกเปิดโปรแกรม STATA ได้หน้าจอดังนี้



ขั้นที่สอง ที่หน้าต่างพิมพ์คำสั่ง พิมพ์ edit

. edit <ENTER> จะได้หน้าต่างป้อนข้อมูล STATA Editor

ขั้นที่สามพิมพ์ข้อมูลลงไปจนครบทุกคอลัมน์ แต่ละคอลัมน์ ใช้เมาส์คลิกสองครั้งเพื่อเปลี่ยนชื่อตัวแปรตามที่ต้องการ (ตัวแปรที่กำหนดโดยอัตโนมัติจาก STATA คือ var<X> เมื่อ X คือหมายเลขตั้งแต่ 1 เป็นต้นไป) จากนั้นคลิกมุมขวาบนสุดของหน้าต่าง STATA Editor เพื่อจบการป้อนข้อมูล กลับมาที่หน้าต่างพิมพ์คำสั่ง มีลำดับการทำงานตามรูปข้างล่างนี้



จากนี้ท่านสามารถวิเคราะห์ข้อมูลชุดนี้ได้ และสามารถ SAVE เก็บไว้ใช้ได้ โดยคลิกที่ File menu แล้วคลิก Save ตามด้วยการป้อนชื่อ File แล้วคลิก OK

ลำดับต่อไปมีการสอดแทรกวิธีการนำเข้าข้อมูลครบทั้ง 3 แบบที่เหลือ

บทที่ 3 ข้อ 3.1

```
. use ps30.dta, clear    ---> เปิดแฟ้มข้อมูลที่สุ่มมา 30 คนจากทั้งหมด 143 คน (Sample data)
. list v5                ---> เรียกดูข้อมูล
```

v5

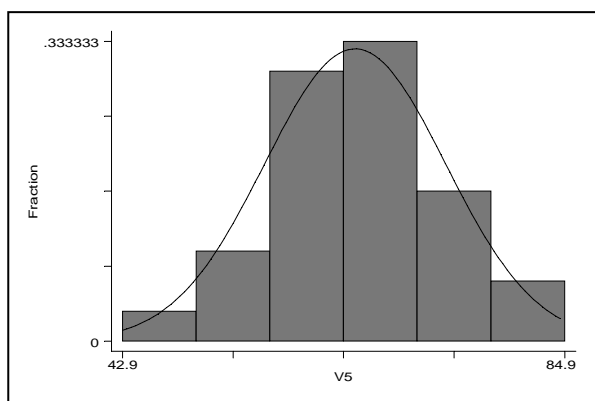
1.	42.9
2.	53.5
3.	54.5

---ข้าม Output บางส่วน ---

28.	76.4
29.	78.7
30.	84.9

บทที่ 3 ข้อ 3.3

```
. use ps30.dta, clear    ---> เปิดแฟ้มข้อมูล
. gr v5, bin(6) norm      ---> สร้าง Histogram ตัวแปร V5 โดยแบ่งข้อมูลออกเป็น
                           6 ช่วงเท่า ๆ กัน พร้อมให้มีเส้นโค้งทาบกับบนกราฟ
```

**บทที่ 4 ข้อ 4.2****ตารางที่ 1 ข้อ 1**

```
. use ps30.dta, clear
. tab v2    ---> สร้างตารางแจกแจงความถี่ตัวแปร V2
```

V2	Freq.	Percent	Cum.
1	5	16.67	16.67
2	25	83.33	100.00
Total	30	100.00	

ตารางที่ 1 ข้อ 2

```
. generate v3g = .
(30 missing values generated)
. replace v3g = 1 if v3 <= 20
(5 real changes made)
. replace v3g = 2 if v3 > 20 & v3 <= 22
(24 real changes made)
. replace v3g = 3 if v3 > 22
(1 real change made)
```

----> สร้างตัวแปรขึ้นมาใหม่ ชื่อ v3g โดยให้ทุก
ค่าเป็น Missing value ไว้ก่อน

แทนค่า Missing value
ด้วยค่าที่ระบุตามเงื่อนไข

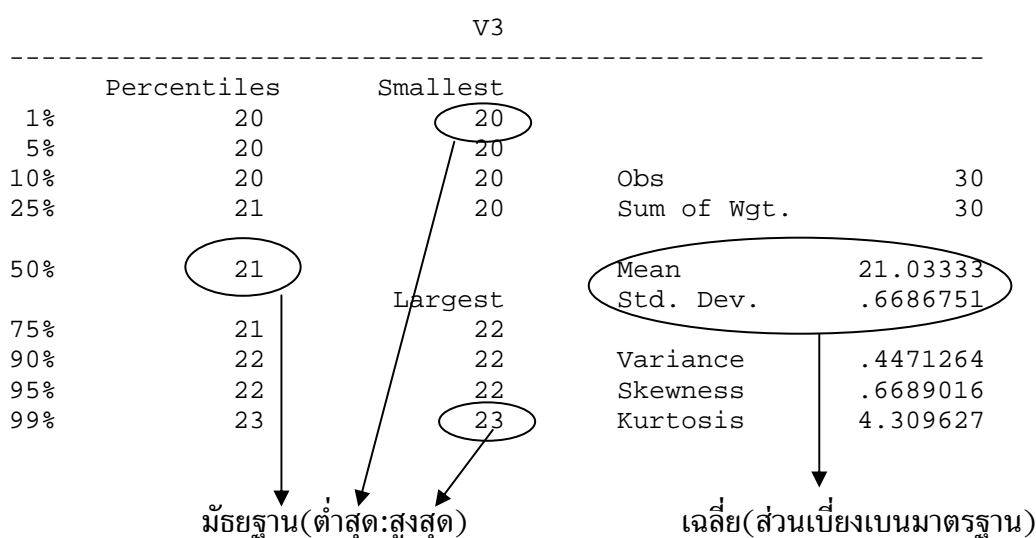
```
. tab v3g
```

----> สร้างตารางแจกแจงความถี่

v3g	Freq.	Percent	Cum.
1	5	16.67	16.67
2	24	80.00	96.67
3	1	3.33	100.00
Total	30	100.00	

```
. summarize v3, detail
```

----> สรุปข้อมูลตัวแปรต่อเนื่อง

**ตารางที่ 1 ข้อ 3**

```
. gen v4g = .
(30 missing values generated)
. replace v4g = 1 if v4 < 150
(2 real changes made)
. replace v4g = 2 if v4 >= 150 & v4 <160
(11 real changes made)
. replace v4g = 3 if v4 >= 160
(17 real changes made)
```

สร้างตัวแปรขึ้นมาใหม่ ชื่อ v4g โดยให้ทุก
ค่าเป็น Missing value ไว้ก่อน

แทนค่า Missing value
ด้วยค่าที่ระบุตามเงื่อนไข

. tab v4g

---> สร้างตารางแจกแจงความถี่

v4g	Freq.	Percent	Cum.
1	2	6.67	6.67
2	11	36.67	43.33
3	17	56.67	100.00
Total	30	100.00	

. su v4, de

---> สรุปข้อมูลตัวแปรต่อเนื่องด้วยคำสั่งแบบสั้น

v4				
Percentiles	Smallest			
1%	147.5	147.5		
5%	149	149		
10%	151.5	150	Obs	30
25%	155	153	Sum of Wgt.	30
50%	160		Mean	160.7967
			Std. Dev.	7.632213
75%	168.4	Largest		
		169.5		
90%	169.75	170	Variance	58.25067
95%	172	172	Skewness	-.0224196
99%	173	173	Kurtosis	1.653667

ตารางที่ 1 ข้อ 4

. gen v5g = .
(30 missing values generated)

} สร้างตัวแปรขึ้นมาใหม่ ชื่อ v5g โดยให้ทุก
ค่าเป็น Missing value ไว้ก่อน

. replace v5g = 1 if v5 < 50
(1 real change made)

. replace v5g = 2 if v5 >= 50 & v5 < 60
(8 real changes made)

} แทนค่า Missing value
ด้วยค่าที่ระบุตามเงื่อนไข

. replace v5g = 3 if v5 >= 60
(21 real changes made)

. tab v5g

v5g	Freq.	Percent	Cum.
1	1	3.33	3.33
2	8	26.67	30.00
3	21	70.00	100.00
Total	30	100.00	

```
. su v5, de
```

V5					
Percentiles			Smallest		
1%	42.9		42.9		
5%	53.5		53.5		
10%	55.25		54.5	Obs	30
25%	58.9		56	Sum of Wgt.	30
50%	64.75			Mean	64.99333
				Std. Dev.	8.593419
75%	69.7		Largest		
			74.6		
90%	75.5		76.4	Variance	73.84684
95%	78.7		78.7	Skewness	-.0532572
99%	84.9		84.9	Kurtosis	3.376095

บทที่ 5 ข้อ 5.3.1

การสุ่มครั้งที่ 1

```
. use PS143.dta, clear ---> เปิดแฟ้มข้อมูลนักศึกษาทั้งหมด 143 คน (Population data)
```

```
. sample 21 ---> สุ่มตัวอย่างมา 21% ของจำนวนข้อมูลที่มีทั้งหมดในขณะนั้น (21% ของ 143 = 30)
(113 observations deleted)
```

```
. su v5
```

Variable	Obs	Mean	Std. Dev.	Min	Max
v5	30	63.68667	8.957013	38.5	77.6

การสุ่มครั้งที่ 2

```
. use PS143.dta, clear
```

```
. sample 21
(113 observations deleted)
```

```
. su v5
```

Variable	Obs	Mean	Std. Dev.	Min	Max
v5	30	63.58667	7.368466	53.3	75.4

การสุ่มครั้งที่ 3

```
. use PS143.dta, clear
```

120

```
. sample 21  
(113 observations deleted)
```

```
. su v5
```

Variable	Obs	Mean	Std. Dev.	Min	Max
v5	30	64.86	9.847794	38.5	84.9

ทำเช่นนี้ไปเรื่อย ๆ ได้ไม่จำกัด แต่ในที่นี้ทำเพียง 40 ครั้ง

บทที่ 5 ข้อ 5.3.2

```
. use MeanWt.dta, clear ---> เปิดแฟ้มข้อมูลค่าเฉลี่ยน้ำหนักจากการสุ่ม 40 ครั้ง (Sample means data)
```

```
. su
```

Variable	Obs	Mean	Std. Dev.	Min	Max
mean	40	64.0225	1.270471	61.3	66.5

บทที่ 5 ข้อ 5.3.6

```
. use PS143.dta, clear
```

```
. su v5
```

Variable	Obs	Mean	Std. Dev.	Min	Max
v5	143	64.52517	8.509402	38.5	84.9

บทที่ 5 ข้อ 5.12.1.2

สัปดาห์ที่ 1

```
. use PS143.dta, clear
```

```
. sample 21  
(113 observations deleted)
```

```
. ci v5 ---> คำนวณช่วงเชื่อมั่น 95% ของตัวแปร V5
```

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
v5	30	64.14333	1.737899	60.58893	67.69773

สัปดาห์ที่ 2

```
. use PS143.dta, clear
```

```
. sample 21
(113 observations deleted)
```

```
. ci v5
```

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
v5	30	63.52667	1.611966	60.22983	66.82351

สัปดาห์ที่ 3

```
. use PS143.dta, clear
```

```
. sample 21
(113 observations deleted)
```

```
. ci v5
```

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
v5	30	62.5	1.755248	58.91011	66.08989

ทำเช่นนี้เรื่อยไปจนครบ 30 ครั้ง (ตามที่เฉลี่ยในคู่มือ)

บทที่ 5 ข้อ 5.16

การวิเคราะห์เปรียบเทียบกลุ่มชาย และหญิงของตัวแปรอายุ

```
. use PS30.DTA, clear
```

```
. generate v3g = .
(30 missing values generated)
```

```
. replace v3g = 1 if v3 <= 20
(5 real changes made)
```

```
. replace v3g = 2 if v3 > 20 & v3 <= 22
(24 real changes made)
```

```
. replace v3g = 3 if v3 > 22
(1 real change made)
```

ให้คำนวณค่าร้อยละโดยให้จำนวนรวมในแนว

คอลัมน์เป็นตัวหาร (Column percent)

```
. tab v3g v2, col ----> สร้างตาราง 2 x 3
```

v3g	v2		Total
	1	2	
1	1 20.00	4 16.00	5 16.67
2	4 80.00	20 80.00	24 80.00
3	0 0.00	1 4.00	1 3.33
Total	5 100.00	25 100.00	30 100.00

สรุปค่าสถิติตัวแปร

```
. table v2, c(mean v3 sd v3 median v3 min v3 max v3 ) ----> แรกจำแนกตาม
```

V2	mean(v3)	sd(v3)	med(v3)	min(v3)	max(v3)
1	20.8	.4472136	21	20	21
2	21.08	.7023769	21	20	23

กลุ่มตัวแปรที่สอง

ต่อไปนี้เป็น การวิเคราะห์ที่เกี่ยวข้องกับบทที่ 6 ข้อ 1 กิจกรรมที่ 6.2 ทั้งหมด โดยใช้แฟ้มข้อมูล PS143.DTA ซึ่งมีข้อมูลที่แสดงในภาคผนวก ต่อไปนี้เป็นขั้นตอนการวิเคราะห์ตามแนวทางที่ระบุในเฉลย อย่างไรก็ตามยังมีรายละเอียดหลายอย่างที่ไม่สามารถใส่ในนี้ได้หมด ดังนั้นจะเสนอทุกขั้นตอนเฉพาะในข้อ 1 เท่านั้น ข้ออื่นๆ แสดงเฉพาะคำสั่งที่จำเป็น พร้อมกับท้าวรื้อล้อมรอบผลที่ต้องใช้ตอบคำถามวิจัย รายละเอียดแต่ละคำสั่ง ท่านสามารถสั่ง help ตามด้วยคำสั่งที่แสดงให้ดูเป็นตัวอย่างนั้น หนึ่งข้อมูลอาจดูไม่สมจริงในบางกรณี ดังนั้นโปรดอย่าได้สนใจการเนื้อหาของเรื่องที่วิจัยและแปลความหมาย แต่ให้ความสำคัญที่วิธีการวิเคราะห์และการนำเสนอผลการวิเคราะห์

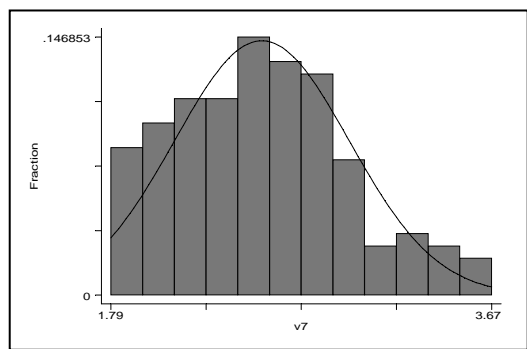
บทที่ 6 กิจกรรมที่ 6.2 ข้อ 1

เพื่อประมาณค่าเกรดเฉลี่ยสะสมของนักศึกษาในมหาวิทยาลัย

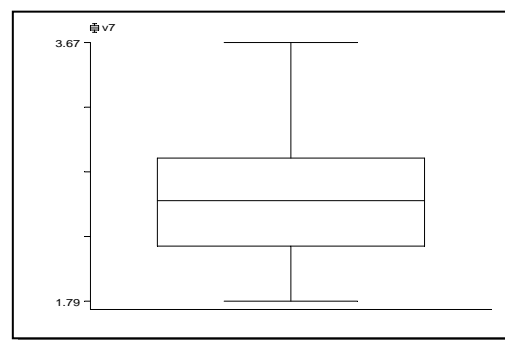
ขั้นตรวจสอบการแจกแจงข้อมูล

. use PS143.DTA, clear

. gr v7, bin(12) norm

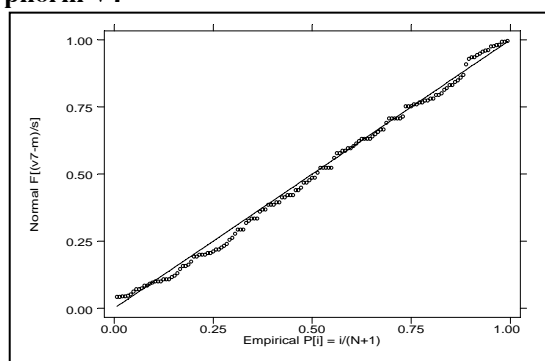


. gr v7, box



จาก Histogram พบว่าข้อมูลเบ้ขวาแต่ไม่มาก และจาก Box-and-Whisker Plot ไม่พบว่ามี Outlier
ดูกราฟ P-P Plot สำหรับ Standard normal probability เพื่อดูว่าข้อมูลมีการแจกแจงว่าเป็นแบบปกติหรือไม่

. pnorm v7



ข้อมูลแจกแจงไม่ปกติ แต่ต่างไปไม่มากนัก เพราะ Plot ส่วนมากอยู่ชิดเส้นทะแยง มีห่างเส้นไปไม่มากนัก

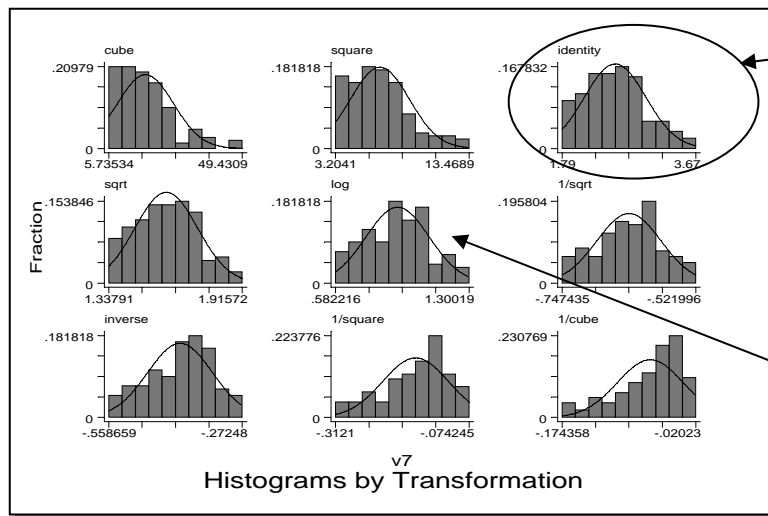
ทดสอบการแจกแจงว่าเป็นแบบปกติหรือไม่

. swilk v7

Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Pr > z
v7	143	0.97883	2.365	1.946	0.02580

พบว่า p-value < 0.05 คือไม่เป็นแบบปกติ ต้อง Transform data ให้เป็น Scale อื่น

อยากทราบว่า Transform เป็น Scale ได้จริงจะมีโอกาสเป็น Normal distribution ได้ ดูจากกราฟ
`. gladder v7`



นี่คือกราฟจาก
ข้อมูลจริง

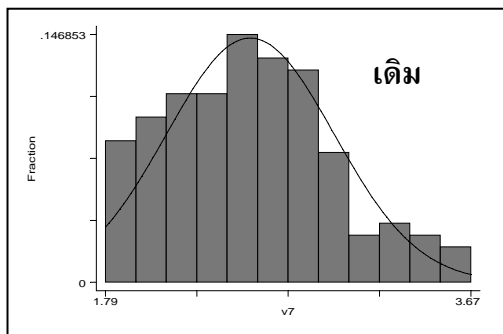
นี่คือกราฟจาก
Logarithm
ของข้อมูลจริง
ซึ่งสมมาตร

ลองทำ Logarithmic Transformation

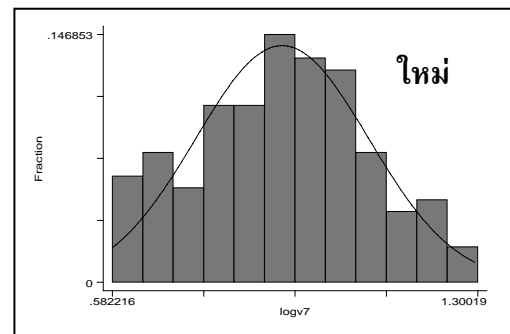
```
. gen logv7 = log(v7)
```

ตรวจสอบอีกครั้ง

```
. gr v7, bin(12) norm
```



```
. gr logv7, bin(12) norm
```



การแจกแจงสมมาตรมากขึ้น แต่ลองทดสอบอีกครั้งเพื่อความมั่นใจ

```
. swilk logv7
```

Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Pr > z
logv7	143	0.98542	1.629	1.103	0.13504

พบว่า p-value > 0.05 หมายถึงมีการแจกแจงแบบปกติ

ต่อไปเป็นขั้นตอนการวิเคราะห์ข้อมูล คือประมาณค่าเฉลี่ย โดยใช้ข้อมูลที่แปลงค่าแล้ว

```
. ci logv7
```

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
logv7	143	.916004	.0142288	.8878764	.9441315


ค่าที่เป็น Logarithmic scale ได้ค่าเฉลี่ยและช่วงเชื่อมั่นเป็น 0.92 (95%CI: 0.89 ถึง 0.94)

ต้องการค่าจริง คือเกรดเฉลี่ยสะสม จึงต้องแปลงค่ากลับเป็น Scale ดั้งเดิม (Back transformation) ในที่นี้ เมื่อ Transform โดย Log ก็ต้องทำ Back transform กลับด้วย Anti-logarithm ดังนี้

```
. display exp(.916004), exp(.8878764), exp(.9441315)
```

2.4992833 2.4299639 2.5705799

→ คำสั่ง disp เหมือน
ใช้ STATA เป็น
เครื่องคิดเลข

ค่าที่ได้คือ 2.50 (95%CI: 2.43 ถึง 2.57)  นี่เป็นคำตอบที่ถูกต้องที่สุด

ถ้าไม่ Transform ข้อมูลก่อนการวิเคราะห์ จะได้ 2.53 (95%CI: 2.46 ถึง 2.61) จากคำสั่งต่อไปนี้

```
. ci v7
```

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
v7	143	2.535385	.0361071	2.464008	2.606761

ซึ่งต่างกันไม่มากนัก ทั้งนี้เพราะการแจกแจงข้อมูลไม่ได้ต่างจาก Normal distribution มากนักนั่นเอง แต่ถ้าข้อมูลไม่เป็นแบบ Normal distribution ชัดเจน (เช่นข้อมูลเบ้มากๆ) ผลที่ได้จะต่างกันมาก และการไม่ Transform ก่อน จะไม่ถูกต้อง

สรุปผล: จากการสำรวจนักศึกษา 143 คน พบว่าเกรดเฉลี่ยสะสมของนักศึกษาในมหาวิทยาลัย
เท่ากับ 2.50 (95%CI: 2.43 ถึง 2.57)

กรณี Transform แล้วไม่สำเร็จ คือยังไม่ได้ Normal distribution ให้คำนวณช่วงเชื่อมั่นโดยไม่อาศัยข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจง แต่จะดีกว่าแบบที่กล่าวข้างต้น ในที่นี้เพียงแสดงให้เห็นตัวอย่างการคำนวณ โดยใช้ Bootstrapping methods ดังต่อไปนี้

คำสั่งที่ใช้

คำนวณ

ค่าสถิติ

ดูในคู่มือ STATA ที่หัวข้อ "Saved Results" ของคำสั่งที่ใช้ เช่น กรณีตัวอย่างนี้ ดู Volume 1 A-G หน้า 199 บรรทัดที่ 4 จากล่าง)

```
. bs "ci v7" "r(mean)", rep(1000)
```

จำนวนครั้งของการทำ Bootstrap (ค่า 1000 สำหรับคำนวณช่วงเชื่อมั่นที่ใช้ Percentile หรือ Bias corrected methods)

```
command:      ci v7
statistic:    r(mean)
(obs=143)
```

Bootstrap statistics

Variable	Reps	Observed	Bias	Std. Err.	[95% Conf. Interval]		
bs1	1000	2.535385	.0012288	.0356045	2.465516	2.605253	(N)
					2.463881	2.605734	(P)
					2.46035	2.603636	(BC)

N = normal, P = percentile, BC = bias-corrected

สรุปค่าที่ได้คือ 2.53 (95%CI: 2.46 ถึง 2.60)

บทที่ 6 กิจกรรมที่ 6.2 ข้อ 2

เพื่ออำนวยความสะดวกในการทำความเข้าใจ จึงใช้ข้อมูลที่ยังไม่แปลงค่า ซึ่งยอมรับได้กรณีนี้ เพราะการแจกแจงไม่ผิดไปจาก Normal distribution มากนัก และ t-test ก็ค่อนข้างทนต่อ Assumption เกี่ยวกับ Normality ดังจะเห็นได้จากค่าช่วงเชื่อมั่นระหว่างที่ได้จาก Log transformation กับข้อมูลดั้งเดิมนั้น ต่างกันน้อยมาก

. ttest v7 = 2.0 ----> คำสั่งวิเคราะห์ One sample t - test

One-sample t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
v7	143	2.535385	.0361071	.4317778	2.464008	2.606761

Degrees of freedom: 142

Ho: mean(v7) = 2.0

Ha: mean < 2.0
t = 14.8277
P < t = 1.0000

Ha: mean ~= 2.0
t = 14.8277
P > |t| = 0.0000

Ha: mean > 2.0
t = 14.8277
P > t = 0.0000

สรุปผล: จากการสำรวจนักศึกษา 143 คน พบว่าเกรดเฉลี่ยสะสมของนักศึกษาในมหาวิทยาลัย โดยเฉลี่ย เท่ากับ 2.53 (95%CI: 2.46 ถึง 2.61) ซึ่งมากกว่า 2.0 อย่างมีนัยสำคัญทางสถิติ ($p\text{-value} < 0.001$)

กรณีข้อมูลแจกแจงไม่เป็นแบบปกติ หรือขนาดตัวอย่างน้อย ใช้ Non-parametric statistics ต่อไปนี้

. **signrank v7 = 2** ----> คำสั่งวิเคราะห์ Wilcoxon Signed Ranks test

Wilcoxon signed-rank test

sign	obs	sum ranks	expected
positive	123	9955.5	5145
negative	17	334.5	5145
zero	3	6	6
all	143	10296	10296

unadjusted variance 246246.00
 adjustment for ties -15.25
 adjustment for zeros -3.50

adjusted variance 246227.25

Ho: v7 = 2

z = 9.694
 Prob > |z| = 0.0000

บทที่ 6 กิจกรรมที่ 6.2 ข้อ 3

. **ttest v7, by(v2)** ----> คำสั่งวิเคราะห์ Two sample t-test

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	30	2.412333	.0704186	.3856985	2.268311	2.556356
2	113	2.568053	.0412973	.4389961	2.486228	2.649878
combined	143	2.535385	.0361071	.4317778	2.464008	2.606761
diff		-0.1557198	.088023		-0.3297351	.0182956

Degrees of freedom: 141

Ho: mean(1) - mean(2) = diff = 0

Ha: diff < 0
 t = -1.7691
 P < t = 0.0395

Ha: diff ~= 0
 t = -1.7691
 P > |t| = 0.0790

Ha: diff > 0
 t = -1.7691
 P > t = 0.9605

สรุปผล: จากการสำรวจนักศึกษาชาย 30 คน และหญิง 113 คน พบว่าเกรดเฉลี่ยสะสมในมหาวิทยาลัยโดยเฉลี่ยเท่ากับ 2.41 (SD = 0.39) และ 2.57 (SD = 0.44) ตามลำดับ นักศึกษาชายมีเกรดเฉลี่ยสะสมน้อยกว่านักศึกษาหญิงเท่ากับโดยเฉลี่ยเท่ากับ 0.16 (95%CI: -0.33 ถึง 0.32) อย่างไรก็ตาม ความแตกต่างดังกล่าวนี้ไม่มีนัยสำคัญทางสถิติ (p-value = 0.079)

กรณีข้อมูลแจกแจงไม่เป็นแบบปกติ หรือขนาดตัวอย่างน้อย ใช้ Non-parametric statistics ต่อไปนี้

. **ranksum v7, by(v2)** ----> คำสั่งวิเคราะห์ Mann-Whitney U test

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

v2	obs	rank sum	expected
1	30	1855.5	2160
2	113	8440.5	8136
combined	143	10296	10296

unadjusted variance 40680.00
 adjustment for ties -8.93

adjusted variance 40671.07

Ho: v7(v2==1) = v7(v2==2)

z = -1.510

Prob > |z| = 0.1311

บทที่ 6 กิจกรรมที่ 6.2 ข้อ 4

. **ttest v11 = v10** ----> คำสั่งวิเคราะห์ Paired t-test

Paired t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
v11	143	113.3776	1.699352	20.32129	110.0183	116.7369
v10	143	76.70629	.7853324	9.39121	75.15384	78.25875
diff	143	36.67133	1.543894	18.46229	33.61934	39.72332

Ho: mean(v11 - v10) = mean(diff) = 0

Ha: mean(diff) < 0

t = 23.7525

P < t = 1.0000

Ha: mean(diff) ~= 0

t = 23.7525

P > |t| = 0.0000

Ha: mean(diff) > 0

t = 23.7525

P > t = 0.0000

หรือวิเคราะห์โดยสร้างตัวแปรใหม่ใส่ผลต่างของสองตัวแปรนั้น แล้วประมาณค่าช่วงเชื่อมั่นดังต่อไปนี้

```
. gen diff = v11 - v10
```

```
. ci diff
```

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
diff	143	36.67133	1.543894	33.61934	39.72332

สรุปผล: จากการสำรวจนักศึกษา 143 คน พบว่าอัตราชีพจรก่อนวิ่งเท่ากับ 77 ครั้งต่อนาที (SD = 9) และหลังวิ่งเท่ากับ 113 ครั้งต่อนาที (SD = 20) มีความแตกต่างเท่ากับ 37 ครั้งต่อนาที (95%CI: 34 ถึง 40) ซึ่งแตกต่างกันอย่างมีนัยสำคัญทางสถิติ (p-value < 0.001)

กรณีข้อมูลแจกแจงไม่เป็นแบบปกติ หรือขนาดตัวอย่างน้อย ใช้ Non-parametric statistics ต่อไปนี้

```
. signrank v11 = v10 ---> คำสั่งวิเคราะห์ Wilcoxon matched-pairs signed-ranks
```

Wilcoxon signed-rank test

sign	obs	sum ranks	expected
positive	143	10296	5148
negative	0	0	5148
zero	0	0	0
all	143	10296	10296

```
unadjusted variance    246246.00
adjustment for ties    -128.88
adjustment for zeros    0.00
-----
adjusted variance      246117.12
```

Ho: v11 = v10

```
z = 10.377
Prob > |z| = 0.0000
```

บทที่ 6 กิจกรรมที่ 6.2 ข้อ 5

One Way ANOVA

Multiple comparison

```
. oneway v7 v9, tabulate bonferroni
```

		Summary of v7		
v9		Mean	Std. Dev.	Freq.
1		2.3528571	.35513916	7
2		2.5543243	.43380059	111
3		2.5495652	.43042767	23
4		1.96	.02828424	2
Total		2.5353846	.4317778	143

p-value > 0.05 หมายถึง
ไม่มีหลักฐานเพียงพอที่จะ
สรุปว่าค่าเฉลี่ยในแต่ละ
กลุ่มต่างกัน

Analysis of Variance						
Source	SS	df	MS	F	Prob > F	
Between groups	.939790944	3	.313263648	1.71	0.1687	
Within groups	25.5335632	139	.183694699			
Total	26.4733542	142	.186432072			

Bartlett's test for equal variances: $\chi^2(3) = 4.2731$ Prob> $\chi^2 = 0.233$

Comparison of v7 by v9
(Bonferroni)

Row Mean- Col Mean	1	2	3
2	.201467 1.000		
3	.196708 1.000	-.004759 1.000	
4	-.392857 1.000	-.594324 0.324	-.589565 0.385

p-value > 0.05 หมายถึงไม่มี
หลักฐานเพียงพอที่จะสรุปว่า
Variance ในแต่ละกลุ่มต่างกัน คือไม่
ละเมิด Assumption ในเรื่องนี้

ค่าความแตกต่างค่าเฉลี่ยระหว่างกลุ่ม
ค่า p-value ทดสอบความแตกต่างค่าเฉลี่ยระหว่างกลุ่ม

สรุปผล: จากการสำรวจเกรดเฉลี่ยสะสมในระดับมหาวิทยาลัยในนักศึกษา 143 คน พบว่าเกรดเฉลี่ยสะสมโดยเฉลี่ยของนักศึกษาจากภาคเหนือ 7 คนเท่ากับ 2.35 (SD=0.36) จากภาคตะวันออกเฉียงเหนือ 111 คนเท่ากับ 2.55 (SD=0.43) จากภาคกลาง 23 คนเท่ากับ 2.55 (SD=0.43) และจากภาคใต้ 2 คนเท่ากับ 1.96 (SD=0.02) เราไม่มีหลักฐานเพียงพอที่จะสรุปว่ามีนักศึกษาน้อยจากภาคใดภาคหนึ่งมีค่าเฉลี่ยเกรดเฉลี่ยสะสมต่างไปจากภาคอื่นๆ (p-value = 0.169)

กรณีข้อมูลแจกแจงไม่เป็นแบบปกติ หรือขนาดตัวอย่างน้อย ใช้ Non-parametric statistics ต่อไปนี้

. kwallis v7, by(v9) ----> คำสั่งวิเคราะห์ Kruskal-Wallis Test

Test: Equality of populations (Kruskal-Wallis Test)

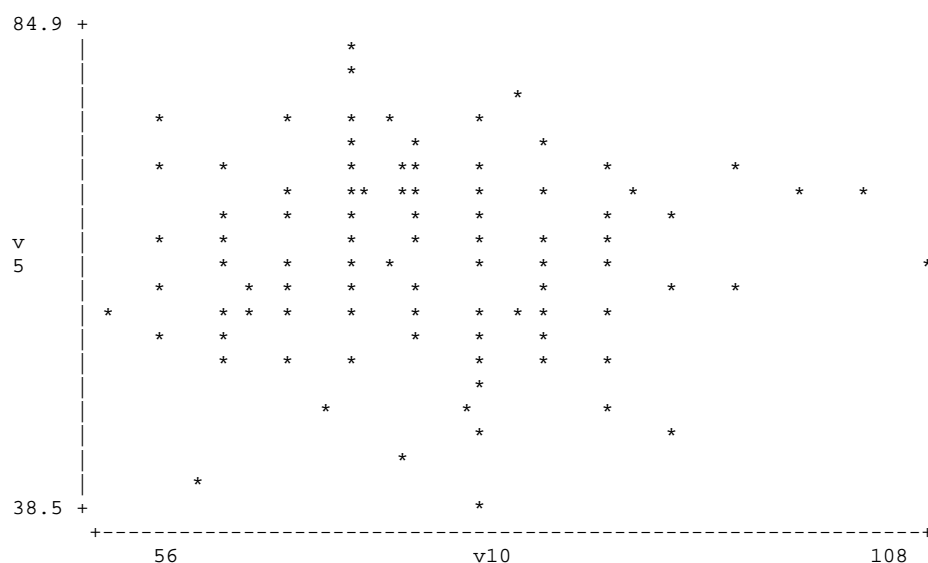
v9	_Obs	_RankSum
1	7	393.50
2	111	8204.50
3	23	1670.50
4	2	27.50

chi-squared = 5.214 with 3 d.f.
probability = 0.1568

บทที่ 6 กิจกรรมที่ 6.2 ข้อ 6

. plot v5 v10

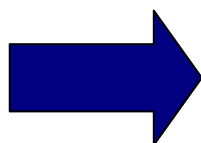
----> Scatter plot เพื่อดูว่ามีความสัมพันธ์รูปแบบอื่นที่ไม่ใช่ Linear relationship หรือไม่
พบว่าไม่มี จึงสามารถหา Correlation coefficient ได้



. pwcorr v5 v10, sig

----> คำสั่งวิเคราะห์เพื่อคำนวณค่า Pearson correlation coefficient (r)

	v5	v10
v5	1.0000	
v10	-0.0653 0.4382	1.0000



r = -0.065
p-value = 0.438

ประมาณค่าช่วงเชื่อมั่น โดยใช้สองคำสั่งต่อไปนี้ ซึ่งไม่มีใน STATA Version 6 แต่ต้อง Download (ฟรี) จาก <http://www.uc.pt/pessoal/ramalheira/stb.htm> โดยเลือก STB32V4 และโปรแกรมนี้อยู่ใน SG51 เป็นโปรแกรมเล็ก ๆ ทำงานเฉพาะอย่าง ที่เรียก Automatic Do File หรือ .ADO File มีอยู่สองไฟล์ชื่อ z_r.ado และ z_rci.ado ส่วนอีกสองไฟล์ซึ่งเป็น Help file เพื่อไว้ดูคำช่วยเหลือการใช้โปรแกรมคือไฟล์ชื่อเดียวกันแต่สกุล .hlp (โปรแกรมนี้เขียนโดย John R. Gleason) จากนั้นก็อปไปใส่ลงในโฟลเดอร์ C:\STATA\ADO\UPDATES\Z

```
. z_r v5 v10
(sample correlations, n=143)
      v5      v10
v5    1.0000
v10   -0.0653   1.0000

(lower\upper 95% confidence limits)
      v5      v10
v5    1.0000   0.0999
v10   -0.2271  1.0000

. z_rci v5 v10
Variables      r      95% confidence limits
v5, v10      0.0653      -0.2271      0.0999
```

สรุปผล: จากการสำรวจนักศึกษา 143 คน พบว่าอัตราการเต้นของชีพจรมีความสัมพันธ์เชิงลบกับน้ำหนัก กล่าวคือเมื่อน้ำหนักมากขึ้น อัตราชีพจรจะต่ำลง แต่ระดับความสัมพันธ์ดังกล่าวนั้นต่ำมาก คือค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันเท่ากับ -0.07 (95%CI: -0.22 ถึง 0.10) ซึ่งต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ (p-value = 0.438)

กรณีข้อมูลแจกแจงไม่เป็นแบบปกติ หรือขนาดตัวอย่างน้อย ใช้ Non-parametric statistics ต่อไปนี้

```
. spearman v5 v10      ---> คำสั่งวิเคราะห์ Spearman rank correlation

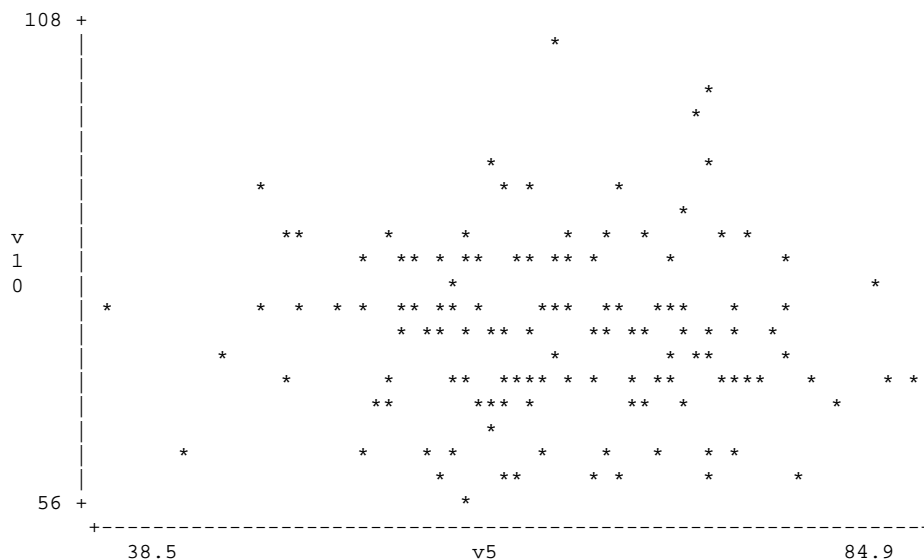
Number of obs =      143
Spearman's rho =      -0.1011

Test of Ho: v5 and v10 independent
Pr > |t| =      0.2294
```

บทที่ 6 กิจกรรมที่ 6.2 ข้อ 7

. plot v10 v5

----> Scatter plot เพื่อดูว่ามีความสัมพันธ์รูปแบบอื่น
ที่ไม่ใช่ Linear relationship หรือไม่ ซึ่งพบว่าไม่มี
จึงสามารถวิเคราะห์ด้วย Linear regression ได้



. regress v10 v5

----> คำสั่งวิเคราะห์ Simple linear regression

Source	SS	df	MS	Number of obs =	143
Model	53.4649184	1	53.4649184	F(1, 141) =	0.60
Residual	12470.1994	141	88.4411306	Prob > F	= 0.4382
Total	12523.6643	142	88.1948193	R-squared	= 0.0043
				Adj R-squared	= -0.0028
				Root MSE	= 9.4043

v10	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
v5	-0.0721093	.0927436	-0.778	0.438	-.255457 .1112384
_cons	81.35916	6.03575	13.480	0.000	69.42689 93.29142

สรุปผล: จากการสำรวจนักศึกษา 143 คน พบว่าความสัมพันธ์ระหว่างอัตราการเต้นของชีพจรและน้ำหนัก สามารถอธิบายด้วยสมการการถดถอยดังต่อไปนี้

$$Y = 81.4 - 0.07X \quad \text{หรือเขียนใหม่เป็น}$$

$$\text{อัตราชีพจร} = 81.4 - 0.07(\text{น้ำหนัก})$$

ค่า R^2 เท่ากับ 0.004 บ่งชี้ว่าสมการอธิบายความแปรปรวนของข้อมูลได้เพียงร้อยละ 0.4 จากสมการ เราสามารถแปลความหมายได้ว่า ทุกๆ 1 กิโลกรัมที่น้ำหนักเพิ่มขึ้น อัตราการเต้นของชีพจรลดลง 0.7 ครั้งต่อนาที อย่างไรก็ตาม ความสัมพันธ์ดังกล่าวนี้ ไม่มีนัยสำคัญทางสถิติ ($p\text{-value} = 0.438$)

บทที่ 6 กิจกรรมที่ 6.2 ข้อ 8

```
. gen disease = .
(143 missing values generated)

. replace disease = 1 if v14 > 2
(51 real changes made)

. replace disease = 0 if v14 < 3
(92 real changes made)
```

สร้างตัวแปรใหม่ชื่อ DISEASE มี
ค่าเป็น 1 เมื่อป่วยและ 0 เมื่อไม่
ป่วย (ดูเฉลยประกอบ)

```
. tab disease
```

disease	Freq.	Percent	Cum.
0	92	64.34	64.34
1	51	35.66	100.00
Total	143	100.00	

```
. disp 143 * 0.3566
50.9938
```

```
. disp 143 * (1-0.3566)
92.0062
```

```
. ci disease
```

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]
disease	143	.3566434	.0401975	.2771805 .4361062

ดูว่าขนาดตัวอย่างใหญ่พอที่จะใช้ Normal approximation หรือต้องใช้ Binomial exact methods พบว่าทั้งคู่ค่าเกิน 5 จึงใช้ Normal approximation ได้

สรุปผล: จากการสำรวจนักศึกษาทั้งหมด 143 คน ป่วย 51 คน คิดเป็นความชุก 35.7%
(95%CI: 27.7 ถึง 43.6%)

กรณีที่ต้องใช้ Binomial exact methods สั่งคำสั่งต่อไปนี้

```
. ci disease, bin
```

Variable	Obs	Mean	Std. Err.	-- Binomial Exact -- [95% Conf. Interval]
disease	143	.3566434	.0400567	.2783773 .4409897

บทที่ 6 กิจกรรมที่ 6.2 ข้อ 9

. **prtest disease = 0.05** ----> คำสั่งเพื่อทดสอบค่าสัดส่วนในกลุ่มตัวอย่างเดียว โดยใช้ Z-test

One-sample test of proportion disease: Number of obs = 143

Variable	Mean	Std. Err.	z	P> z	[95% Conf. Interval]
disease	.3566434	.0400567	8.90347	0.0000	.2781337 .435153

Ho: proportion(disease) = .05

Ha: disease < .05

z = 16.825

P < z = 1.0000

Ha: disease ~= .05

z = 16.825

P > |z| = 0.0000

Ha: disease > .05

z = 16.825

P > z = 0.0000

สรุปผล: จากการสำรวจนักศึกษาทั้งหมด 143 คน ป่วย 51 คน คิดเป็นความชุก 35.7% (95%CI: 27.7 ถึง 43.6%) ซึ่งสูงกว่า 5.0% อย่างมีนัยสำคัญทางสถิติ (p-value < 0.001)

กรณีที่ต้องใช้ Binomial exact methods สั่งคำสั่งต่อไปนี้

. **bitest disease = 0.05** ----> คำสั่งวิเคราะห์ Binomial exact probability test

Variable	N	Observed k	Expected k	Assumed p	Observed p
disease	143	51	7.15	0.05000	0.35664

Pr(k >= 51) = 0.000000 (one-sided test)

Pr(k <= 51) = 1.000000 (one-sided test)

Pr(k >= 51) = 0.000000 (two-sided test)

Note: Lower tail of two-sided p-value is empty.

บทที่ 6 กิจกรรมที่ 6.2 ข้อ 10

. prtest disease, by(v2)

----> คำสั่งวิเคราะห์ Z-test for two proportions

Two-sample test of proportion

1: Number of obs = 30
2: Number of obs = 113

Variable	Mean	Std. Err.	z	P> z	[95% Conf. Interval]	
1	.2333333	.0772202	3.02166	0.0025	.0819845	.3846822
2	.3893805	.0458705	8.4887	0.0000	.2994761	.479285
diff	-.1560472	.0898168			-.332085	.0199906
	under Ho:	.098381	-1.58615	0.1127		

Ho: proportion(1) - proportion(2) = diff = 0

Ha: diff < 0
z = -1.586
P < z = 0.0564

Ha: diff ~= 0
z = -1.586
P > |z| = 0.1127

Ha: diff > 0
z = -1.586
P > z = 0.9436

สรุปผล: จากการสำรวจนักศึกษาชาย 30 คน และหญิง 113 คน พบว่าความชุกของการป่วยเท่ากับ 23.3% และ 38.9% ตามลำดับ นักศึกษาชายมีความชุกการป่วยน้อยกว่านักศึกษาหญิงเท่ากับ 15.6% (95%CI: 33.2% ถึง 2.0%) อย่างไรก็ตาม ความแตกต่างดังกล่าวนี้ไม่มีนัยสำคัญทางสถิติ (p-value = 0.113)

↑ ให้แสดง Row percent

↑ ให้ทดสอบสมมติฐานโดยใช้ Chi-square test

. tab v2 disease, row chi2 exact

v2	disease		Total
	0	1	
1	23 76.67	7 23.33	30 100.00
2	69 61.06	44 38.94	113 100.00
Total	92 64.34	51 35.66	143 100.00

ให้ทดสอบสมมติฐานโดยใช้ Fisher's exact test

Pearson chi2(1) = 2.5159 Pr = 0.113
Fisher's exact = 0.136
1-sided Fisher's exact = 0.083

บทที่ 6 กิจกรรมที่ 6.2 ข้อ 11

. recode v2 1=1 2=0
(113 changes made)

---> คำสั่งแปลงรหัสตัวแปร ให้ค่า 1 คงเดิม ส่วนค่า 2 ให้เปลี่ยนเป็น 0 เป็นสิ่งที่จำเป็นเพราะการวิเคราะห์ความสัมพันธ์กรณีตั้งแปรแจกแจงนับนี้ ต้องให้อยู่ในรูปตัวแปร (1,0) โดยให้ 1 คือกลุ่มที่สนใจและ 0 คือกลุ่มเปรียบเทียบ

. cs disease v2, or

---> คำสั่งวิเคราะห์ความสัมพันธ์ในรูปแบบตาราง 2 x 2

	v2		
	Exposed	Unexposed	Total
Cases	7	44	51
Noncases	23	69	92
Total	30	113	143
Risk	.2333333	.3893805	.3566434
	Point estimate		[95% Conf. Interval]
Risk difference	-.1560472	-.332085	.0199906
Risk ratio	.5992424	.3010147	1.192937
Prev. frac. ex.	.4007576	-.1929365	.6989853
Prev. frac. pop	.084075		
Odds ratio	.4772727	.1934759	1.183432 (Cornfield)
	chi2(1) = 2.52		Pr>chi2 = 0.1127

Relative Risk (RR)
สำหรับกรณี Cohort study

Odds Ratio (OR) สำหรับกรณี Cross-sectional study หรือ Case-control study

สรุปผล: จากการสำรวจนักศึกษาชาย 30 คน และหญิง 113 คน พบว่าความชุกของการป่วยเท่ากับ 23.3% และ 38.9% ตามลำดับ นักศึกษาชายมีความเสี่ยงต่อการป่วยต่ำกว่านักศึกษาหญิง (OR = 0.48 และ 95%CI: 0.19 ถึง 1.18) กล่าวอีกนัยหนึ่ง นักศึกษาหญิงมีความเสี่ยงสูงเป็น 2.1 เท่าของนักศึกษาชาย อย่างไรก็ตามความสัมพันธ์ดังกล่าวนี้ไม่มีนัยสำคัญทางสถิติ (p-value = 0.113)

บทที่ 6 กิจกรรมที่ 6.2 ข้อ 12

. mcc v12 v13

----> คำสั่งวิเคราะห์ค่าสัดส่วนจากสองตัวอย่างที่ไม่เป็นอิสระต่อกัน

(MCC ย่อมาจาก Matched case-control data)

Cases	Controls		Total
	Exposed	Unexposed	
Exposed	24	0	24
Unexposed	119	0	119
Total	143	0	143

McNemar's chi2(1) = 119.00 Pr>chi2 = 0.0000
 Exact McNemar significance probability = 0.0000

Proportion with factor

Cases	.1678322		
Controls	1	[95% conf. interval]	
difference	-.8321678	-.9004132	-.7639224
ratio	.1678322	.1165129	.2417555
rel. diff.	.	.	.
odds ratio	0	0	.0314863 (exact)

Diagram: Cases (circled) points to Exposed (circled) and Unexposed (circled). Exposed points to 'เครียด' (Stressed) and 'ไม่เครียด' (Not stressed). Unexposed points to 'ในมหาวิทยาลัย' (In university) and 'นอกมหาวิทยาลัย' (Out of university).

เนื่องจากข้อมูลชุดนี้ ไม่มีจำนวนเหตุการณ์เพียงพอที่จะวิเคราะห์ได้ จึงขอสมมติข้อมูล เพื่อแสดงวิธีการสรุปผลการวิเคราะห์ดังนี้

. mcci 100 20 4 19 ----> คำสั่ง mcc ที่คำนวณได้ในทันทีเมื่อป้อนข้อมูลลงไปท้ายคำสั่ง

Cases	Controls		Total
	Exposed	Unexposed	
Exposed	100	20	120
Unexposed	4	19	23
Total	104	39	143

McNemar's chi2(1) = 10.67 Pr>chi2 = 0.0011
 Exact McNemar significance probability = 0.0015

Proportion with factor

Cases	.8391608		
Controls	.7272727	[95% conf. interval]	
difference	.1118881	.0403023	.1834739
ratio	1.153846	1.058815	1.257406
rel. diff.	.4102564	.2211872	.5993256
odds ratio	5	1.675012	20.11794 (exact)

Diagram: Cases points to Exposed (circled) and Unexposed (circled). Exposed points to 'เครียด' (Stressed) and 'ไม่เครียด' (Not stressed). Unexposed points to 'ในมหาวิทยาลัย' (In university) and 'นอกมหาวิทยาลัย' (Out of university).

สรุปผล: จากการสำรวจนักศึกษาชาย 143 คน ในภาพรวมแล้วพบว่ากลุ่มผู้ที่มีความเครียดเป็นเมื่ออยู่ในมหาวิทยาลัยมีสัดส่วน 83.9% สำหรับกลุ่มผู้ที่ไม่มีความเครียดเป็นอยู่ในมหาวิทยาลัยมีสัดส่วนเท่ากับ 72.7% พบว่าการอยู่ในมหาวิทยาลัยมีความเสี่ยงต่อการเกิดความเครียดสูงเป็น 5 เท่าของการไม่อยู่ในมหาวิทยาลัย (95%CI: 1.7 ถึง 20.1) ความสัมพันธ์ดังกล่าวนี้มีนัยสำคัญทางสถิติ (p-value = 0.002)

บทที่ 6 กิจกรรมที่ 6.2 ข้อ 13

. tab v9 disease, row chi2 exact

----> คำสั่งวิเคราะห์ข้อมูลตาราง 2 × 4

v9	disease		Total
	0	1	
1	2	5	7
	28.57	71.43	100.00
2	77	34	111
	69.37	30.63	100.00
3	13	10	23
	56.52	43.48	100.00
4	0	2	2
	0.00	100.00	100.00
Total	92	51	143
	64.34	35.66	100.00

Pearson chi2(3) = 9.3479
Fisher's exact =

Pr = 0.025
0.017

Chi-square test

Fisher's exact test

. csi 34 77 5 2, or exact

-----> คำสั่งวิเคราะห์หาค่า OR

	Exposed	Unexposed	Total
Cases	34	77	111
Noncases	5	2	7
Total	39	79	118
Risk	.8717949	.9746835	.940678
	Point estimate		[95% Conf. Interval]
Risk difference	-.1028887	-.2133828	.0076055
Risk ratio	.8944389	.7889516	1.01403
Prev. frac. ex.	.1055611	-.0140305	.2110484
Prev. frac. pop	.0348888		
Odds ratio	.1766234	0	.837579

ข้อสังเกต
ตะวันออกเฉียงเหนือ
โดยใช้ภาคเหนือเป็น

1-sided Fisher's exact P = 0.0388
2-sided Fisher's exact P = 0.0388

. csi 10 13 5 2, or exact

-----> คำสั่งวิเคราะห์หาค่า OR

	Exposed	Unexposed	Total
Cases	10	13	23
Noncases	5	2	7
Total	15	15	30
Risk	.6666667	.8666667	.7666667
	Point estimate		[95% Conf. Interval]
Risk difference	-.2	-.4941156	.0941156
Risk ratio	.7692308	.5109065	1.158169
Prev. frac. ex.	.2307692	-.1581687	.4890935
Prev. frac. pop	.1153846		
Odds ratio	.3076923	.0569277	1.734838

ของภาคกลางโดยใช้
ภาคเหนือเป็นกลุ่ม

1-sided Fisher's exact P = 0.1949
2-sided Fisher's exact P = 0.3898

สรุปผลด้วยตารางดังต่อไปนี้ (โปรดสังเกตการนำข้อมูลจากผลของ STATA มานำเสนอ)

ภูมิสำเนา	จำนวน	เครียด (%)	OR	95%CI	p-value
ภาคเหนือ	7	71.4	1.0		0.017
ภาคตะวันออกเฉียงเหนือ	111	30.6	0.2	(0.0 ถึง 0.8)	
ภาคกลาง	23	43.5	0.3	(0.6 ถึง 1.7)	
ภาคใต้	2	100.0	*	(0.1 ถึง *)	

*ประมาณค่าไม่ได้เนื่องจากจำนวนข้อมูลไม่พอเพียง

จากตาราง 2 × 4 ไม่ใช่จากตารางย่อย 2 × 2 ที่ใช้หา Local OR

บทที่ 6 กิจกรรมที่ 6.2 ข้อ 14

. regress v7 v2 v3 v4 v5 disease ----> คำสั่งวิเคราะห์ Multiple regression

Source	SS	df	MS	Number of obs =	143
Model	.950070282	5	.190014056	F(5, 137) =	1.02
Residual	25.5232839	137	.186301342	Prob > F =	0.4084
Total	26.4733542	142	.186432072	R-squared =	0.0359
				Adj R-squared =	0.0007
				Root MSE =	.43163

v7	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
v2	.1175785	.1222911	0.961	0.338	-.1242437 .3594006
v3	.014352	.0237405	0.605	0.546	-.0325931 .0612972
v4	-.00186	.0074577	-0.249	0.803	-.016607 .012887
v5	-.0036743	.0060483	-0.607	0.545	-.0156345 .0082859
disease	.0739798	.0778351	0.950	0.344	-.0799337 .2278934
_cons	2.508784	1.353555	1.853	0.066	-.1677777 5.185345

สรุปผล: จากการสำรวจนักศึกษา 143 คน พบว่าผลของเพศ อายุ ส่วนสูง น้ำหนัก และการเจ็บป่วยต่อเกรตเฉลี่ยสะสมนั้น สามารถอธิบายด้วยสมการการถดถอยดังต่อไปนี้

$$\text{เกรตเฉลี่ยสะสม} = 2.509 + 0.118(\text{เพศ}) + 0.014(\text{อายุ}) - 0.002(\text{ความสูง}) - 0.004(\text{น้ำหนัก}) + 0.074(\text{การเจ็บป่วย})$$

อย่างไรก็ตาม ทุกตัวแปรมีผลต่อค่าเกรตเฉลี่ยสะสมอย่างไม่มีนัยสำคัญทางสถิติ ($p\text{-value} > 0.05$) ค่า R^2 เท่ากับ 0.036 บ่งชี้ว่าสมการอธิบายความแปรปรวนของข้อมูลได้เพียงร้อยละ 3.6 แสดงให้เห็นว่าตัวแปรที่นำมาหาขนาดของความสัมพันธ์ต่อค่าเกรตเฉลี่ยสะสมร่วมกันในที่นี้ สามารถอธิบายการเปลี่ยนแปลงค่าเกรตเฉลี่ยสะสมได้น้อยมาก ต้องมีตัวแปรที่ไม่ได้นำมาวิเคราะห์ ไม่ได้เก็บข้อมูลมา หรือแม้กระทั่งไม่สามารถวัดได้

สรุปผลด้วยตารางดังต่อไปนี้ แต่เนื่องจากมีข้อพิจารณาอื่นอีกมาก จึงขอเสนอเป็นตัวอย่างเฉพาะผลของ “เพศ” ต่อการเจ็บป่วย เพื่อเป็นข้อสังเกตการนำข้อมูลจากผลของ STATA มานำเสนอ

ภูมิลำเนา	จำนวน	เครียด (%)	OR _{CRU} DE	OR _{ADJUSTE} D	95%CI	p-value
1. เพศ						
ชาย	30	23.3	0.5	0.7	0.2 ถึง 2.6	0.658
หญิง	113	38.9	1.0			
2. อายุ						
. . .						

นำมาจากผลในค่าเฉลี่ยก่อนหน้านี้
(บทที่ 6 ข้อ 1 กิจกรรมที่ 6.2 ข้อ 11)

ผลจาก Logistic regression

การแปลความหมายตัวแปรใดๆ ต่อตัวแปรตาม (การเจ็บป่วย) จะเน้นว่าได้ควบคุมผลกระทบจากตัวแปรอื่นทุกตัวแปรเพศที่อยู่ใน Model เพราะมาตรวัดความสัมพันธ์ที่ได้จาก Multivariable model นี้เป็น Adjusted OR เช่น “เมื่อควบคุมผลกระทบจาก อายุ น้ำหนัก และส่วนสูงแล้ว นักศึกษาชายมีความเสี่ยงต่อการป่วยต่ำกว่านักศึกษาหญิง (OR = 0.7 และ 95%CI: 0.2 ถึง 2.6) อย่างไรก็ตามความสัมพันธ์ดังกล่าวนี้ไม่มีนัยสำคัญทางสถิติ (p-value = 0.658) ”

บทที่ 6 กิจกรรมที่ 6.2 ข้อ 16

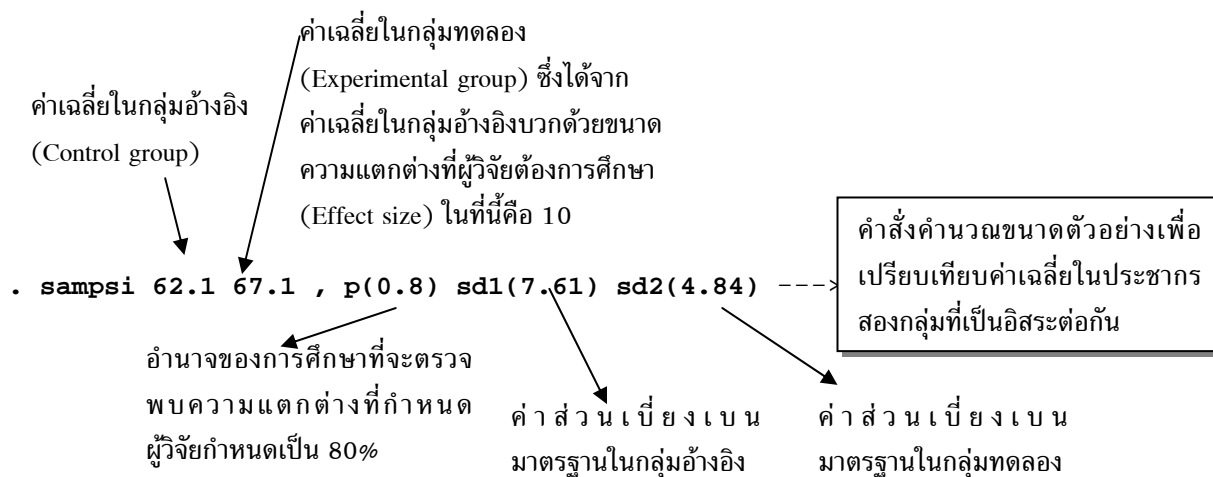
จากเฉลย ทำให้ทราบว่า งานวิจัยที่จะวางแผนทำต่อไปนั้นเป็น Two-sample comparison of mean1 to mean2 ซึ่งจะเริ่มง่าย แต่ต้องหาสูตรที่เหมาะสมแล้วคำนวณด้วยมือ แต่มีคำสั่งใน STATA ที่สามารถคำนวณขนาดตัวอย่างในการศึกษาที่ใช้อยู่ ดังต่อไปนี้

```
. table v2, c(mean v5 sd v5)
```

v2	mean(v5)	sd(v5)
0	62.11947	7.609403
1	73.58666	4.84154

--->

คำสั่งคำนวณค่าสถิติที่จะนำไปใช้ในการคำนวณขนาดตัวอย่าง ปกติค่าเหล่านี้ได้จากการทบทวนวรรณกรรม แต่ถ้าได้ข้อมูลงานวิจัยที่คล้ายกับที่จะทำต่อไป ก็ใช้ได้เสมือนเป็นการศึกษานำร่องให้แล้ว



หมายเหตุ ผู้วิจัย เป็นผู้กำหนดขนาดความแตกต่างที่ต้องการศึกษา (Effect size) และอำนาจของการศึกษา (โดยทั่วไปกำหนดค่าเป็น 80% หรือ 90%) นอกเหนือจากนี้ควรได้จากการมบทวนวรรณกรรมหรือ การศึกษานำร่อง (Pilot study)

ต่อไปนี้เป็นผลของการสั่งคำสั่งข้างต้น

Estimated sample size for two-sample comparison of means

Test Ho: $m1 = m2$, where $m1$ is the mean in population 1
and $m2$ is the mean in population 2

Assumptions:

```
alpha = 0.0500 (two-sided)
power = 0.8000
m1 = 62.1
m2 = 67.1
sd1 = 7.61
sd2 = 4.84
n2/n1 = 1.00
```

Estimated required sample sizes:

```
n1 = 26
n2 = 26
```

สรุปผล: เพื่อให้การศึกษามีอำนาจ 80% ในการตรวจพบความแตกต่างน้ำหนักเฉลี่ยในนักศึกษาชาย 10 กิโลกรัมสูงกว่านักศึกษาหญิงซึ่งมีน้ำหนักเฉลี่ย 62.1 กิโลกรัม เมื่อค่าส่วนเบี่ยงเบน มาตรฐานของน้ำหนักในชายและหญิงเท่ากับ 7.61 และ 4.84 ตามลำดับ โดยให้ระดับ ความมีนัยสำคัญเท่ากับ 0.05 ขนาดตัวอย่างที่ต้องใช้ไม่น้อยกว่ากลุ่มละ 26 คน

ภาคผนวก 3

ข้อมูลการสำรวจนักศึกษา

แถวที่ 0 สำหรับกรอกข้อมูลของท่านจากแบบสอบถามที่แนบท้ายคู่มือนี้ ทั้งนี้ไม่ต้องนับรวมเป็นกรอบ
สำหรับสุ่มตัวอย่าง **[ตัวทึบเอียงคือรายชื่อที่ถูกสุ่มเป็นตัวอย่างในการเจลาคุมือนี่]**

ID	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
1.	2	21	160.5	57.3	2.88	3.42	2	2	76	104	2	2	2
2.	2	20	168.0	65.2	2.98	2.35	2	2	72	140	1	2	2
3.	2	21	169.0	78.7	3.64	2.56	2	3	72	100	2	2	2
4.	2	21	166.0	67.3	3.50	3.58	2	2	88	120	2	1	3
5.	2	21	162.5	71.6	3.41	2.92	2	2	68	85	2	2	2
6.	2	21	156.0	54.8	2.58	2.22	2	3	88	120	2	1	3
7.	2	21	168.4	67.5	2.89	2.21	2	2	80	112	2	2	2
8.	1	21	176.0	72.2	2.94	2.28	2	2	100	142	1	2	2
9.	2	19	164.0	65.2	2.93	3.19	2	3	88	120	2	1	2
10.	2	21	153.0	59.3	3.30	2.44	2	2	76	96	2	1	2
11.	2	21	155.0	42.9	3.45	2.87	2	2	62	120	2	1	2
12.	2	22	160.0	66.1	2.97	2.11	2	2	76	104	2	2	2
13.	2	21	149.0	53.5	2.53	2.83	2	2	64	108	2	1	4
14.	1	24	175.0	69.1	3.09	2.87	2	3	68	82	2	2	2
15.	2	21	170.0	67.7	3.50	2.85	2	2	60	84	1	2	3
16.	1	20	183.0	82.2	1.33	2.68	2	2	82	100	2	2	2
17.	2	21	159.0	68.6	3.55	2.33	2	2	76	100	2	1	3
18.	2	21	150.0	45.5	3.40	2.26	2	2	75	90	2	2	3
19.	2	19	159.5	67.3	3.34	2.86	2	2	76	104	2	2	2
20.	1	21	168.0	84.9	2.33	2.38	2	3	72	128	1	1	2
21.	2	21	168.0	75.4	2.62	2.83	2	2	72	128	2	1	2
22.	1	21	163.5	68.2	3.02	2.19	2	2	80	100	2	2	2
23.	1	22	159.5	73.4	2.84	2.72	2	2	72	162	1	1	2
24.	2	20	159.0	72.6	2.19	1.98	2	4	104	146	2	1	4
25.	2	22	164.0	69.2	1.92	3.22	2	3	88	164	2	2	2
26.	2	22	160.0	62.8	3.05	2.77	2	2	72	100	2	2	2

ID	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
27.	2	20	164.0	62.1	1.51	3.11	2	3	84	108	2	1	2
28.	2	20	155.5	49.3	2.85	2.30	2	2	70	120	1	1	1
29.	1	21	166.0	74.6	2.09	2.45	2	1	80	146	2	2	2
30.	1	27	169.0	70.7	2.75	2.90	2	2	80	92	1	2	1
31.	2	21	168.0	66.9	2.80	2.51	2	3	64	80	2	2	3
32.	1	22	173.0	80.0	3.03	2.68	2	1	68	120	2	1	3
33.	2	20	152.0	47.6	2.80	2.56	2	3	92	120	2	1	2
34.	2	21	155.0	61.5	2.98	2.89	2	2	76	92	2	1	2
35.	2	23	167.0	69.4	2.49	2.97	2	2	76	120	2	1	3
36.	2	21	169.0	64.6	3.40	3.27	2	2	84	109	2	1	2
37.	2	20	170.0	64.6	1.33	3.02	2	3	108	124	2	1	3
38.	2	20	159.0	59.4	2.89	2.41	2	2	88	104	2	1	2
39.	2	20	160.0	67.9	3.45	2.60	2	2	92	124	2	2	2
40.	1	22	162.0	69.6	3.33	2.63	2	3	68	124	1	1	2
41.	1	22	163.0	77.6	3.20	2.72	2	2	80	156	2	1	3
42.	2	20	155.0	61.6	2.85	2.83	2	3	60	71	2	2	2
43.	2	21	161.0	66.4	3.69	2.42	2	3	72	85	1	1	3
44.	2	21	169.0	69.7	3.25	2.68	2	2	80	120	1	1	2
45.	2	22	150.0	60.5	3.45	2.62	2	2	76	117	2	2	1
46.	1	20	172.0	73.7	3.00	3.30	2	2	72	88	2	2	2
47.	1	21	172.0	73.6	2.01	1.87	2	3	88	142	2	1	3
48.	2	21	157.0	57.4	3.33	2.56	1	2	76	92	2	2	3
49.	2	21	155.0	68.2	3.48	2.66	1	2	80	112	1	1	2
50.	1	22	172.0	71.0	3.25	2.68	1	2	75	125	2	1	2
51.	2	22	153.3	49.5	2.87	2.75	1	2	79	102	2	2	3
52.	2	21	167.0	69.9	2.72	2.71	1	2	64	116	2	2	1
53.	2	21	167.0	57.6	2.89	2.54	1	2	84	140	2	2	2
54.	2	21	158.0	55.3	3.78	2.42	1	2	80	155	2	1	2
55.	2	21	160.0	47.6	2.87	3.00	1	2	80	108	1	1	2
56.	2	20	153.8	51.8	3.48	3.39	1	2	80	120	2	2	1
57.	1	22	167.0	70.0	2.12	2.00	1	2	72	100	2	1	4
58.	2	22	155.0	64.8	3.31	3.43	1	2	80	120	2	2	2
59.	2	21	152.0	53.3	3.03	2.69	1	2	80	95	2	1	3
60.	1	22	165.0	72.9	3.21	2.95	1	2	60	74	2	2	2

ID	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
61.	2	21	161.0	68.7	3.20	2.95	1	2	68	116	2	2	2
62.	1	20	180.0	76.2	2.80	2.77	1	2	72	140	2	2	2
63.	2	22	163.0	59.5	2.85	2.44	1	2	56	104	2	2	2
64.	1	21	169.5	67.2	2.99	1.97	1	2	64	132	2	1	2
65.	2	22	166.0	59.7	2.51	2.39	1	2	84	104	1	1	2
66.	2	21	152.0	61.6	3.24	2.84	1	2	72	110	2	1	3
67.	2	21	155.0	50.0	3.19	2.13	1	2	88	104	2	1	3
68.	2	21	155.0	50.0	3.19	2.13	1	2	88	104	2	1	3
69.	2	21	165.0	72.8	3.30	2.45	1	3	64	108	2	1	2
70.	2	21	154.0	48.9	3.69	2.65	1	2	88	99	2	1	2
71.	2	21	154.0	48.9	3.69	2.65	1	2	88	99	2	1	2
72.	2	21	160.0	53.4	2.45	2.16	1	2	84	116	2	2	2
73.	2	21	153.0	55.6	3.25	2.70	1	2	84	120	2	1	3
74.	2	22	158.0	74.7	2.91	2.63	1	2	72	110	2	2	3
75.	2	21	168.0	66.6	3.36	2.89	1	2	60	76	2	1	3
76.	2	21	155.0	57.9	0.41	3.17	1	2	80	120	2	1	3
77.	2	21	161.0	64.9	3.63	3.67	1	2	84	164	2	2	3
78.	1	21	165.5	63.0	2.79	2.48	1	2	68	98	2	1	3
79.	2	22	153.0	76.4	2.50	2.77	1	2	76	100	1	1	2
80.	2	21	154.0	62.2	3.50	2.77	1	2	72	84	2	1	3
81.	2	21	160.0	68.8	3.45	2.85	1	2	68	92	2	2	2
82.	2	21	145.0	38.5	3.53	3.59	1	2	80	120	2	1	2
83.	2	20	155.0	55.8	3.33	2.84	1	2	80	99	1	1	2
84.	2	20	156.5	59.2	3.50	3.29	1	2	88	120	2	1	2
85.	2	20	162.0	59.8	3.18	2.52	1	2	80	120	2	2	2
86.	2	21	161.0	64.6	3.19	2.98	1	2	80	100	1	1	3
87.	2	20	163.0	62.6	2.86	2.86	1	2	76	105	2	2	3
88.	2	21	161.0	58.1	2.80	2.62	1	2	82	139	2	1	3
89.	2	21	160.0	58.2	3.29	2.64	1	2	80	104	2	2	2
90.	2	20	152.0	60.7	3.40	3.24	1	2	65	90	2	1	2
91.	2	28	155.0	63.8	2.58	2.78	1	1	72	105	2	2	2
92.	1	28	159.0	68.4	3.03	2.56	1	2	72	140	2	1	4
93.	1	30	179.0	74.5	3.48	2.20	1	2	64	145	2	2	2
94.	1	27	167.0	72.1	3.00	2.30	1	1	73	115	2	1	3

ID	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
95.	2	27	150.0	54.3	2.92	3.19	1	2	68	85	2	2	2
96.	2	21	153.5	53.4	3.03	2.64	3	2	84	100	2	1	2
97.	2	21	169.0	71.0	2.60	2.05	3	2	80	120	2	2	2
98.	2	21	152.2	56.3	2.50	2.00	3	1	84	104	2	1	3
99.	2	21	156.1	59.2	2.97	2.03	3	2	84	140	1	1	1
100.	2	20	160.0	58.4	3.82	1.98	3	2	72	106	2	1	2
101.	2	21	159.0	66.7	2.88	2.18	3	3	84	116	2	1	2
102.	2	21	158.0	55.8	3.14	1.98	3	2	76	156	2	1	3
103.	2	20	162.5	63.8	2.55	2.34	3	2	80	128	2	1	2
104.	2	21	155.0	77.4	2.95	2.17	3	2	84	148	2	1	3
105.	2	22	150.5	54.5	2.50	2.35	3	2	68	99	2	1	2
106.	2	21	155.0	58.8	3.40	2.08	3	2	64	112	2	2	2
107.	2	23	162.0	71.5	3.25	1.79	3	1	80	112	1	1	3
108.	2	22	160.0	73.2	3.56	1.94	3	4	75	110	2	2	3
109.	2	21	160.0	58.8	2.78	1.80	3	2	72	120	2	2	2
110.	2	20	153.0	60.8	2.68	2.39	3	3	68	120	1	1	3
111.	2	20	158.0	61.7	3.33	2.41	3	3	68	99	2	1	3
112.	2	21	162.0	74.8	3.08	2.16	3	2	88	117	2	2	2
113.	2	21	169.0	67.2	2.79	2.17	3	2	80	136	1	1	2
114.	2	21	147.5	64.2	3.30	2.67	3	2	84	116	1	2	3
115.	1	20	167.0	71.4	2.79	1.83	3	2	76	90	2	1	2
116.	2	21	157.5	58.9	3.20	1.90	3	3	72	91	2	1	3
117.	1	23	173.0	83.0	2.55	2.50	3	2	72	108	2	2	1
118.	1	21	173.0	73.9	2.30	1.80	3	2	72	110	2	1	2
119.	1	20	179.0	72.8	2.80	2.00	3	2	96	104	2	2	2
120.	2	22	160.0	72.9	2.75	1.91	3	2	76	84	2	1	3
121.	1	21	174.0	74.0	2.50	2.10	3	2	72	128	1	1	2
122.	2	24	165.0	71.6	3.20	2.20	3	2	90	110	1	1	3
123.	2	22	157.0	59.9	2.45	2.56	3	2	84	112	2	1	3
124.	1	20	176.5	71.7	2.50	2.23	3	2	80	140	2	2	2
125.	2	20	160.0	62.7	2.75	1.96	3	2	84	156	2	1	3
126.	2	21	157.0	62.7	2.80	2.35	3	2	92	120	2	1	2
127.	2	20	153.0	56.7	2.25	2.02	3	2	64	108	2	1	3
128.	2	21	160.0	62.8	2.52	1.94	3	2	84	121	2	2	2

ID	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
129.	2	21	167.0	70.6	2.10	2.47	3	3	72	148	2	2	2
130.	2	22	164.0	64.3	2.85	2.41	3	2	74	122	2	1	2
131.	2	21	153.0	74.3	3.23	2.45	3	2	76	110	2	1	3
132.	2	20	153.0	60.2	2.85	2.30	3	2	84	124	2	1	2
133.	2	22	155.0	54.5	3.16	2.18	3	2	72	90	1	1	2
134.	2	21	158.0	63.7	3.21	2.77	3	2	64	112	2	2	2
135.	2	21	158.0	60.2	2.30	2.10	3	2	68	96	2	2	2
136.	1	21	183.0	77.6	2.20	1.81	3	2	74	100	2	2	2
137.	2	21	167.0	70.7	2.62	2.25	3	2	84	153	2	1	2
138.	2	20	150.0	56.0	2.69	2.17	3	3	80	112	2	1	3
139.	2	21	160.0	61.7	2.16	2.47	3	1	92	122	2	1	3
140.	2	21	158.5	61.8	2.67	3.38	3	3	60	78	2	2	3
141.	1	21	178.0	78.1	2.82	2.52	3	2	60	76	2	2	2
142.	2	22	155.0	60.4	2.50	1.79	3	2	65	120	2	1	2
143.	2	21	160.0	64.4	2.50	1.90	3	3	80	120	2	1	2

ภาคผนวก 4

แบบสอบถามนักศึกษา

ID.....

เขียนหมายเลขจากกรอบตัวอย่างช่องนี้

แบบสอบถามนักศึกษา

โปรดตอบแบบสอบถาม โดยกาเครื่องหมาย ✓ ลงในช่อง [] ตรงคำตอบของท่าน กรณีคำถามที่เว้นช่องว่างให้กรอกคำตอบ โปรดเขียนคำตอบลงในช่องว่าง

คำถาม	ช่องสำหรับลงรหัส
1. ชื่อ..... รหัส -	V1[][][][][][]-[]
2. เพศ []1. ชาย []2. หญิง	V2[]
3. อายุ (เต็มปี)ปี	V3[][]
4. ความสูง.....เซนติเมตร	V4[][][][]
5. น้ำหนัก.....กิโลกรัม	V5[][][][]
6. เกรดเฉลี่ยสะสมเมื่อจบการศึกษาระดับมัธยมปลาย.....	V6[][][][]
7. เกรดเฉลี่ยสะสมในมหาวิทยาลัยจนถึงเทอมที่ผ่านมา.....	V7[][][][]
8. ท่านสอบเข้ามหาวิทยาลัยโดยระบบใด []1. โควต้า []2. เอนทรานซ์	V8[]
9. ภูมิลำเนาอยู่ภาคใด []1. เหนือ []2. ตะวันออกเฉียงเหนือ []3. กลาง []4. ใต้	V9[]
10. อัตราการเต้นของชีพจร (นับจำนวนครั้งการเต้นของชีพจรของท่านในระยะเวลา 15 วินาที แล้วคูณด้วย 4 และเขียนคำตอบลงในช่องว่าง) อัตราการเต้นของชีพจรก่อนการวิ่งที่ 1 นาที เท่ากับ:.....	V10[][][][]
11. วิ่งกั๊บที่อย่างรวดเร็ว เป็นเวลา 1 นาที แล้ววัดอัตราการเต้นของชีพจรอีกครั้ง อัตราการเต้นของชีพจรหลังการวิ่งกั๊บที่ 1 นาที เท่ากับ:.....	V11[][][][]
12. สุขภาพจิตของท่านก่อนเข้าเรียนมหาวิทยาลัย []1. มีความเครียด []2. ไม่มีความเครียด	V12[]
13. สุขภาพจิตของท่านหลังเข้าเรียนมหาวิทยาลัยแล้ว []1. มีความเครียด []2. ไม่มีความเครียด	V13[]
14. สถานสุขภาพกายของท่านในปัจจุบันนี้อยู่ในระดับใด []1. เยี่ยม []2. ดี []3. ปานกลาง []4. แย่	V14[]

แบบสอบถามนี้ สำหรับใช้อ้างอิงในการปฏิบัติการตลอดหลักสูตร