**Statistical Analysis Plan (SAP)**

**Categorical outcome**

**Inference on “one proportion or proportion difference (risk difference, RD) or odds ratio (OR)or relative risk (RR)”**

**Magnitude of effect:** …… = xx (95%CI: xx. to xxx; p-value = 0.xxx)

1. **One group -> Proportion**

**RQ: What is the prevalence (proportion) of health status among head of household in SV?**

**Magnitude = xx.x (95%CI: xx.x to xx.x; P-value = 0.xxx)**

. ci status

 Variable Obs Mean Std. Err. [95% Conf. Interval]

 status 300 .55 .0287708 .4933811 .6066189

**Results:** Proportion of bad health status among heads of HH was 55.0% (95%CI: 49.3% to 60.7%)

1. **Two Independent groups -> the difference of two proportion (RD, RR, OR)**

**RQs:**

**Absolute difference: 40 – 20 = 20**

 **4 – 2 = 2**

**Relative difference: 40/20 = 2 times**

 **4/2 = 2 times**

* **What is the difference of health status between male and female among heads of HH? [gender is the comparison group]**
* **Whether gender associate with health status … or not?**
* **What is the magnitude of association between gender and health status …?**
* **Does gender effect on health status …? [Cohort study]**
* **Can gender predict health status …? [Cohort study]**

**Magnitude = RR, RD, OR [xx.x (95%CI: xx.x to xx.x; P-value = 0.xxx)]**

. prtest status, by(v2)

Two-sample test of proportions 1:Number of obs = 157

 2: Number of obs = 143

 Variable Mean Std. Err. z P>z [95% Conf. Interval]

1 .5159236 .0398841 .4377522 .594095

2 .5874126 .0411682 .5067245 .6681007

diff -.071489 .0573198 -.1838338 .0408557

under Ho: .0575083 -1.24 0.214

diff = prop(1) - prop(2) z = -1.2431

Ho: diff = 0

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0

Pr(Z < z) = 0.1069 Pr(Z < z) = 0.2138 Pr(Z > z) = 0.8931

**Reporting:**

Among of 300 HHHs in SV, prevalence of bad health status in male was 51.6% (n = xxx) and that of female was 58.7% (n = xxx). Female had prevalence of bad health status greater than Male (Proportion diff = 7%; 95%CI: 4% to 18%; p-value = 0.214).

**Table 2.** Mean difference of body weight between male and female based on

[two sample t-test]//[simple linear regression]

| **Factor** | **Number** | **Mean** | **(SD)** | **Mean difference** | **95%CI** | **p-value** |
| --- | --- | --- | --- | --- | --- | --- |
| **Gender**  |  |  |  |  |  |  |
| Male |  |  |  |  |  |  |
| Female  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

. regress v3a v2

 Source | SS df MS Number of obs = 300

-------------+------------------------------ F( 1, 298) = 0.59

 Model | 62.2333301 1 62.2333301 Prob > F = 0.4425

 Residual | 31367.1991 298 105.259057 R-squared = 0.0020

-------------+------------------------------ Adj R-squared = -0.0014

 Total | 31429.4324 299 105.115159 Root MSE = 10.26

------------------------------------------------------------------------------

 v3a | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 v2 | .9119149 1.185967 0.77 0.443 -1.422016 3.245846

 \_cons | 63.57407 1.848739 34.39 0.000 59.93583 67.21231

------------------------------------------------------------------------------

1. **Three or more independent means -> the difference of two means (Mean difference)**

**RQs:**

* **What is the difference of BW compared health payments among heads of HH in SV? [health payments is the comparison group]**
* **Whether health payments associate with BW … or not?**
* **What is the magnitude of association between health payments and BW…?**
* **Does health payments effect on BW…? [Cohort study]**
* **Can health payments predict BW…? [Cohort study]**

**Magnitude = mean difference [xx.x (95%CI: xx.x to xx.x; P-value = 0.xxx)]**

. regress v3a i.v9

 Source | SS df MS Number of obs = 300

-------------+------------------------------ F( 2, 297) = 0.78

 Model | 163.366333 2 81.6831663 Prob > F = 0.4612

 Residual | 31266.0661 297 105.27295 R-squared = 0.0052

-------------+------------------------------ Adj R-squared = -0.0015

 Total | 31429.4324 299 105.115159 Root MSE = 10.26

------------------------------------------------------------------------------

 v3a | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 v9 |

 2 | -1.316155 1.510487 -0.87 0.384 -4.288769 1.656459

 3 | .4586378 1.414463 0.32 0.746 -2.325002 3.242277

 |

 \_cons | 65.13093 1.041772 62.52 0.000 63.08074 67.18112

------------------------------------------------------------------------------

Outcome = body weight (kg), compared groups (health payment)- UC is the reference group

. tabstat v3a, stat(n mean sd) by(v9)

Summary for variables: v3a

by categories of: v9

v9 N mean sd

1 97 65.13093 10.94066

2 88 63.81477 10.04103

3 115 65.58957 9.82457

Total 300 64.92067 10.25257

**Reporting:**

| **Factor** | **number** | **Mean** | **(SD)** | **Mean difference** | **95%CI** | **p-value** |
| --- | --- | --- | --- | --- | --- | --- |
| Health payment |  |  |  |  |  | 0.461 |
| UC | 97 |  |  | 0 |  |  |
| CSMBS | 88 |  |  | (CSMBS-UC)-1.32 | -4.29 to 1.66  |  |
| Others | 115 |  |  | (Others-UC).45 |  -2.33 to 3.24 |  |

. oneway v3a v9 /\*ANOVA -> P-value only, not very useful

 Analysis of Variance

 Source SS df MS F Prob > F

------------------------------------------------------------------------

Between groups 163.366333 2 81.6831663 0.78 0.4612

 Within groups 31266.0661 297 105.27295

------------------------------------------------------------------------

 Total 31429.4324 299 105.115159

Bartlett's test for equal variances: chi2(2) = 1.3177 Prob>chi2 = 0.517

There was insufficient evidence (not enough information) to conclude that there was different among health payment groups.

\* Correlation

. scatter v3a v3b, ms(O) c(.) /ms= marker-options, c=connect-options/



There was a “linear relationship” between body weight and height. => “Correlation coefficient” can be estimated

\*\* Pearson correlation \*\*

. corr v3a v3b

(obs=300)

 | v3a v3b

-------------+------------------

 v3a | 1.0000

 v3b | 0.0492 1.0000

. pwcorr v3a v3b, sig /\*P-value only, not very useful\*/

 | v3a v3b

-------------+------------------

 v3a | 1.0000

 |

 |

 v3b | 0.0492 1.0000

 | 0.3962

 |

= Pearson correlation coefficient (r) = 0.05

= p-value = 0.396

How to interpret r value, please click http://www.dummies.com/how-to/content/how-to-interpret-a-correlation-coefficient-r.html

= None of any magnitude of association -> not useful

**Modelling**

. regress v3a v2 i.v9 i.ageg

 Source | SS df MS Number of obs = 299

-------------+------------------------------ F( 5, 293) = 0.45

 Model | 240.716722 5 48.1433443 Prob > F = 0.8097

 Residual | 31018.611 293 105.865566 R-squared = 0.0077

-------------+------------------------------ Adj R-squared = -0.0092

 Total | 31259.3277 298 104.897073 Root MSE = 10.289

------------------------------------------------------------------------------

 v3a | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

 v2 | .8932161 1.199423 0.74 0.457 -1.467361 3.253793

 |

 v9 |

 1 | 1.420142 1.522362 0.93 0.352 -1.576008 4.416293

 3 | 1.715926 1.461738 1.17 0.241 -1.160911 4.592764

 |

 ageg |

 2 | -.1829241 1.687006 -0.11 0.914 -3.503109 3.13726

 3 | -.5114967 1.97926 -0.26 0.796 -4.406865 3.383872

 |

 \_cons | 62.75646 2.523223 24.87 0.000 57.79052 67.7224

------------------------------------------------------------------------------

**Reporting:**

Every one year of age increment, mean BW increased by x.x kgs.

Please added more numbers in the table

| **Factors** | **Number** | **Mean (SD)** | **Mean difference (crude)** | **Mean difference (adjusted)** | **95%CI** | **p-value** |
| --- | --- | --- | --- | --- | --- | --- |
| Gender |  |  |  |  |  | 0.457  |
| Female  | xx | xx.x (x.x) | 0 | 0 |  |  |
| Male | xx | xx.x (x.x) | x.x | 0.89 | x.x - x.x |  |
| Health payment |  |  |  |  |  | 0.816 |
|  UC | xx | xx.x (x.x) | reference | 0 |  |  |
|  CSMBS | xx | xx.x (x.x) | x.x | x.x | x.x - x.x |  |
|  Others | xx | xx.x (x.x) | x.x | x.x | x.x - x.x |  |
| Age group(year) |  |  |  |  |  | 0.963 |
| Less than 30 | xx | xx.x (x.x) | reference | reference |  |  |
| 31 – 50 | xx | xx.x (x.x) | x.x | x.x | x.x - x.x |  |
| 50 or greater | xx | xx.x (x.x) | x.x | x.x | x.x - x.x |  |
| -option of analysis for age- |  |  |  |  |
| *Age (year)* | *xx* | *xx.x (x.x)* | *x.x* | *x.x* | x.x - x.x | *0.xxx* |
| *Age(every 5 years)* | *xx* | *xx.x (x.x)* | *x.x* | *x.x* | x.x - x.x | *0.xxx* |
| *Age(every 10 years)* | *xx* | *xx.x (x.x)* | *x.x* | *x.x* | x.x - x.x | *0.xxx* |

1. regress v3a v2 i.v9 i.ageg
2. regress v3a v2 i.v9 age

gen age5 = age/5

1. regress v3a v2 i.v9 age5

gen age10 = age/10

1. regress v3a v2 i.v9 age10

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Good Luck \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*