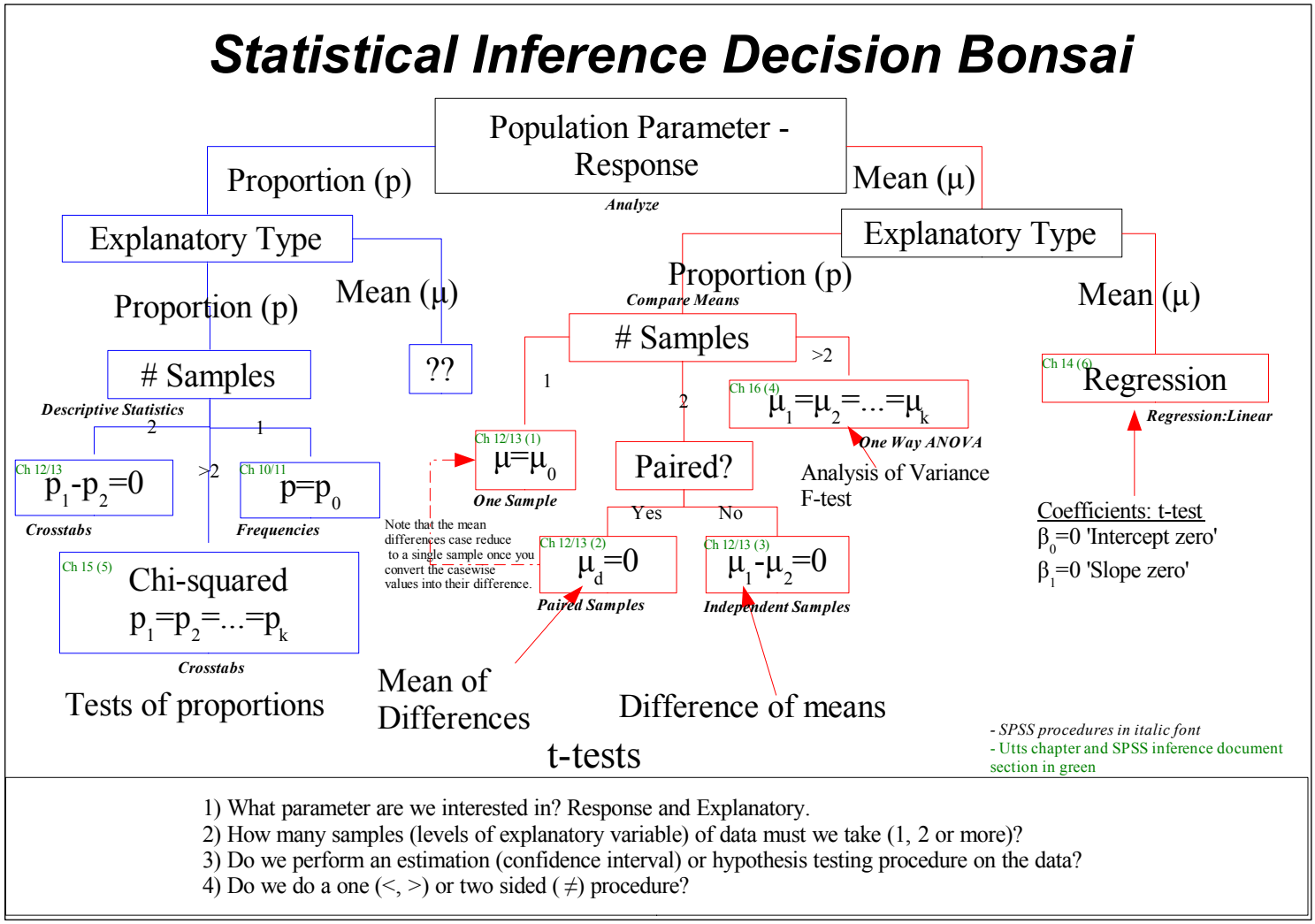
# Examples for use of the Inference Bonsai
# Selecting the correct population parameters.




1) What parameter are we interested in? Response and Explanatory.
2) How many samples (levels of explanatory variable) of data must we take (1, 2 or more)?
3) Do we perform an estimation (confidence interval) or hypothesis testing procedure on the data?
4) Do we do a one (<, >) or two sided ( ≠) procedure?

**For the following questions, state the appropriate type of analysis (confidence interval or hypothesis test), the population parameter(s) of interest and the appropriate statistic(s) or estimate(s). State $H_0$ &$H_a$ in BOTH words and symbols for hypothesis tests. If you feel that a confidence interval is the correct analysis, determine the appropriate multiplier. For one or two proportions or means analysis, setup *but don't calculate* the appropriate standard error, otherwise identify the SPSS procedure you would use for the analysis (such as Analyze:Compare Means:Paired-Samples T Test).**

Setup: We will work with a data set which contains the following variables.

| *Name* | *Type* | *Observed Values* | *Description* |
|---|---|---|---|
| gender | Categorical | {Male, Female} | |
| status | Ordinal | {Fresh, Soph, Junior, Senior, Senior+, Staff, Faculty, Volunteer} | Status in University |
| age | Quantitative | 12...110 (integer) | Age in years |
| hours | Quantitative | 0...168 | Hours on campus/week |
| travel | Quantitative | 0...100 | Miles campus<->home |
| commute | Categorical | Recode{Travel>2 = Yes, Else = No} | Commuter or not |
| ethnicity | Categorical | {Caucasion, Hispanic, AfrAm, Asian, Other} | Primary ethnic background |
| jobmiles | Quantitative | 0..100 | Miles to job |

1) *First we wish to know if the mean number of hours spent on campus is significantly different for different ages.*

Hypothesis test
Response: hours – quantitative
Explanatory: age – quantitative
Bonsai: quantitative (mean)->quantitative(mean)->regression
Population Parameter(s): $\beta_1$ Since we are asking if there is a difference in response for different values of the explanatory variable. I.E. Is the slope parameter not equal to zero. Use the t-test for slope. So our hypotheses are:
$H_0$: ($\beta_1=0$) There is no difference in hours based on age.
$H_a$: ($\beta_1\neq0$) There is a linear relationship between hours and age.
SPSS: Analyze:Regression:Linear, examine t-test for age.

2) *Next we would like to know if the proportion of men depends on ethnic background.*

Hypothesis test
Response: gender – categorical 2 possible values
Explanatory: ethnicity – categorical 5 possible values
Bonsai: Categorical (proportion)->Categorical (proportion) -> (>2) Chi-Squared
Population Parameter(s): $p_1,p_2,p_3,p_4,p_5$ (the 5 subpopulation proportions men)
Here we look at the number of both men and women in each ethnic category and determine the proportion of that group which is men.
$H_0$: ($p_1=p_2=p_3=p_4=p_5$) The proportion of men is the same for all ethnic backgrounds.
$H_a$: ($p_1,p_2,p_3,p_4,p_5$ are not all the same) The proportion for some ethnic background(s) differs.
SPSS: Analyze:Description Statistics:Crosstabs (turn on Chi-Squared test).

3) *We find that for the 55 men and 63 women surveyed, the averages for travel distance for men and women are 15.36 and 10.27 respectively. The respective standard deviations are 5.8 and 4.7. Determine if there is a difference in the mean travel distance for men and women.*

Hypothesis test
Response: travel – quantititative
Explanatory: gender – 2 values (levels)
Bonsai: Quantitative->Categorical->2 samples->Not paired->Independent Means.
(note different sample sizes makes it impossible to be paired)
Population Parameter(s): $\mu_1-\mu_2$ (difference between the means of men and women)
$H_0$: ($\mu_1-\mu_2=0$ ) There is no difference in mean travel miles based on gender.
$H_a$: ($\mu_1-\mu_2\neq0$) There is a difference in mean travel miles based on gender.

$$S.E.=\sqrt{\frac{5.8^2}{55}+\frac{4.7^2}{63}}$$

4) *The survey asked both travel miles to campus and travel miles to work (for those that have and outside job). For the 74 people who have jobs, we find that the difference between the two is 4.2 miles with a standard deviation of 6.5 miles. Is there a significant difference between their miles to campus and their miles to work?*

Hypothesis test
Response: miles traveled (contained in two quantitative variables)
Explanatory: Work miles (jobmiles) versus campus miles (travel)
Bonsai: Quantitative->Categorical->2 samples->Paired t-test
Population Parameter(s): $\mu_d$ (difference between travel and jobmiles)
$H_0$: ($\mu_d=0$ ) The mean difference in job and campus miles is zero.
$H_a$: ($\mu_d\neq0$) The mean difference in job and campus miles is not zero.

$$S.E.=6.5\Big/\sqrt{74}$$

5) *We find that there are 55 men and 63 women in the survey. Do a larger proportion of men commute than women?*

Hypothesis test
Response: commute – categorical
Explanatory: gender – 2 values
Bonsai: Categorical->Categorical->2 levels->difference in proportions ($p_1$-$p_2$=0)
Population Parameter(s): $p_1$ & $p_2$
The definition of $p_1$ & $p_2$ are critical here since we are looking at a one sided alternative.
$p_1$: Proportion of men who commute.
$P_2$: Proportion of women who commute.
$H_0$: ($p_1$-$p_2$=0) The proportion of men and women commuting are the same.
$H_a$: ($p_1$-$p_2$>0) A higher proportion of men commute than women.

$$S.E. = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} : \hat{p} = \frac{\left(n_1\hat{p}_1 + n_2\hat{p}_2\right)}{\left(n_1+n_2\right)}$$

6) *We find that the average miles traveled to campus is 12.56 with a standard deviation of 5.2 and that there were a total of 118 responses. What is the mean miles traveled to campus?*

Confidence Interval
Response: travel – quantitiative
Explanatory: none (we are looking at a single value across the entire population)
Bonsai: Quantitative->Categorical->1 sample
Population Parameter(s): μ(mean miles traveled)
We would use a t* multiplier here with df=117. From Table A.2 our appropriate multiplier is then t*=1.98 (the entry for .95 at df=100).

$$S.E. = 5.2 \Big/ \sqrt{117}$$

7) *Finally, we wish to determine if there is a relationship between hours spent on campus and miles traveled to campus.*

Hypothesis test
Response: hours – quantitative
Explanatory: travel – quantitative
Bonsai: Quantitative->Quantitative->Regression (F-test)
Population Parameter(s): ρ(regression correlation)
This is an F-test instead of a t-test because we are not asking specifically about the significance of the linear slope, but simple if there is a relationship.
$H_0$: (ρ=0 ) There is a correlation between time on campus and miles traveled.
$H_a$: (ρ≠0) There is a correlation between time on campus and miles traveled.
SPSS: Analyze:Regression:Linear, examine F-test in the ANOVA table.