



SESSION-11: STATA DESCRIPTIVE STATISTICS-STATISTICAL TESTS

Course detail: <http://julhas.com/jsedutech/stata-level-one.html>

Mentor: Julhas Sujan


Recap-Session 10

- Descriptive statistics – cross table


Session-11

Statistical tests:




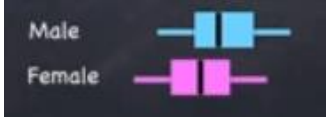

Population



Sample



- one sample proportion
- chi-squared test
- t-test
- ANOVA
- correlation test

What we observe in our sample data	Is it real?	Purpose	
One categorical		One sample proportion test	Comparison
Two categorical		Chi-squared	Comparison
One numeric		t-Test	Comparison
One numeric and one categorical		t-Test or ANNOVA (3 levels of group)	Comparison
Two numeric (Continuous only)		Correlation test	Relationship

Three key areas:

- Question and hypothesis
- Null hypothesis and Alpha value
- Analyse data

Decision: If the P-value is less than (or equal to), **reject** the **null hypothesis** in favor of the alternative hypothesis. **If the P-value is greater than, do not reject the null hypothesis.**

Q1. Is a difference in the number of men and women the in the population?

H0 = There is no difference



Ha = There is difference



Test: One sample proportion test

Command: `prtest gender2 = 0.5, ttest gender2 = .5`

Result:

```
. prtest gender2=0.5
```

```
One-sample test of proportion          Number of obs      =          6
```

Variable	Mean	Std. Err.	[95% Conf. Interval]	
gender2	.5	.2041241	.099924	.900076

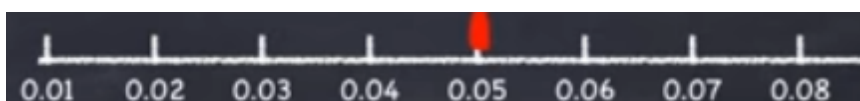
```
    p = proportion(gender2)                z =    0.0000
```

```
Ho: p = 0.5
```

```
    Ha: p < 0.5  
Pr(Z < z) = 0.5000
```

```
    Ha: p != 0.5  
Pr(|Z| > |z|) = 1.0000
```

```
    Ha: p > 0.5  
Pr(Z > z) = 0.5000
```



Reject Null Hypothesis because the p-value is

A hypothesis is an **educated guess** about something in the world around you. It should be testable, either by experiment or observation. For example:

- A new medicine you think might work.
- A way of teaching you think might be better.
- A possible location of new species.
- A fairer way to administer standardized tests.

It can really be *anything at all* as long as you can put it to the test². The hypothesis statement could be: *If I...**(do this to an independent variable)....then (this will happen to the dependent variable)***.

Ways of working on hypothesis:

- Figure out your null hypothesis: always the accepted fact,
- State your null hypothesis,
- Choose what kind of test you need to perform,
- Either support or reject the null hypothesis.

Hypothesis testing can be done using Z-Test, One sample proportion, Mean using TI 83, Bayesian testing,

Q2. Does the male and female differ across the age groups?

H₀ = There is no difference



H_a = There is difference



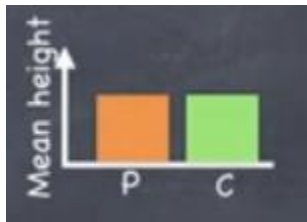
Test: Chi-squared test (Categorical variables)

Command: ***prtest gender2 = 0.5***

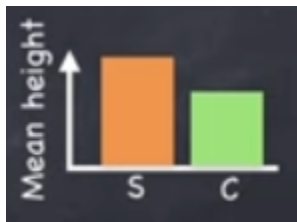
Result: If p value is less the the alpha vlue (0.05) then we will reject the null hypothesis

Q3. Is the average height different from the established previous height (1.4m)?

H0 = There is no difference



Ha = There is difference



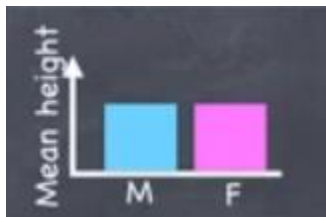
Test: t-Test (Only numeric)

Command: ***prtest gender2 = 0.5***

Result: If p value is less than the alpha value (0.05) then we will reject the null hypothesis

Q4. Is there a difference in height between men and women?

H0 = There is no difference



Ha = There is difference



Test: t-Test (One categorical and one numeric)

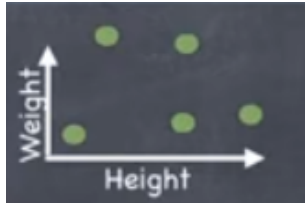
Command: ***prtest gender2 = 0.5***

Result: If p value is less than the alpha value (0.05) then we will reject the null hypothesis

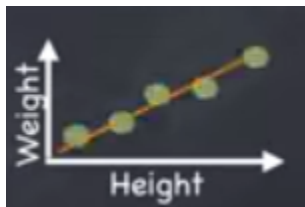
Note: If more than two categories then we will perform ANNOVA.

Q5. Is there a relationship between height and weight?

H0 = There is no difference



Ha = There is difference

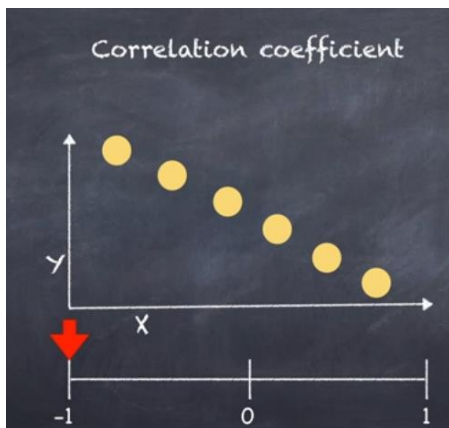


Test: Correlation test (two numeric)

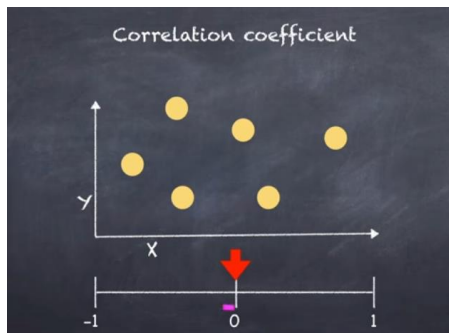
Command: ***prtest gender2 = 0.5***

Result: Correlation coefficient which tells us the measures of association between the two variables. If p value is less than the alpha value (0.05) then we will reject the null hypothesis.

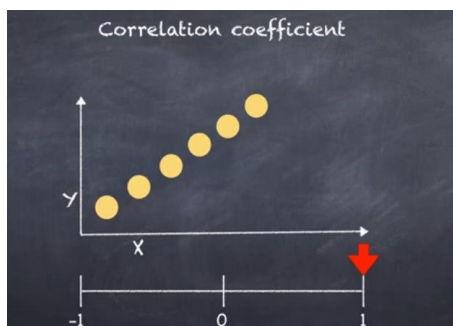
Negative correlation:



No relation:



Positive correlation:



Type of study based tests:

Type of study	Possible tests
Cross sectional	Regression analysis: Linear, Multi linear, Logistics, Multiplte logistics Chi-squared
Case control	Chi-squared, T-test, Z-test
Cohort	Cox proportional regression, sensitivity analysis, survival analysis.

References:

[1] <https://www.youtube.com/watch?v=l10q6fjPxJ0>

[2] Hypothesis testing. Accesible from: <https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/>