T TEST

2 POPULATIONS

ARE MEANS DIFFERENT?

NULL HYPOTHESIS: MEANS SAME.

ALTERNATIVE = DIFFERENT.

YOU MIGHT GET PARTICULAR ABOUT HOW DIFFERENT.

E.G., BLUE ≠ RED

Q: IF NO DIFF, HOW LIKELY IS WHAT YOU SAW?
ANOVA
ANALYSIS OF VARIANCE.

DRUG TRIALS:
Define what will be measured and what success will be, before trials.

BAD DATA DREDGING:
Look at results to find interesting things.
LINEAR REGRESSION

STATISTICS KIND
PARAMETERS, TEST
HYPOTHESES, FIND RELATIONS.
PREDICT FUTURE.

Eg SPD00

\[ y = x + w \]

FIT ACROSS TO DATA \( y=ax+b \).

\( (x_1, y_1) \)}
\[ y = ax + b \]

Find \( a, b \) to minimize error

\[ \varepsilon_j = (ax_j + b - y_j)^2 \]

\[ \eta \eta = \sum \varepsilon_j \]

\[ \eta \eta = \sum (a x_j + b - y_j)^2 \]

Pick \( a, b \) to make that

\[ \varepsilon = a \sum x_j^2 + nb^2 + \sum y_j^2 \]

\[ + 2ab \sum x_j - 2a \sum x_j y_j \]

\[ + 2n \sum b y_j \]
\[ \frac{dE}{da} = 2a \sum x_i + 2b \sum x_i \]
- \[ 2 \sum x_i N_i \]
- \[ \sum x_i N_i = 0 \]
- \[ a \sum x_i^2 + b \sum x_i - \sum x_i N_i = 0 \]
- \[ \frac{dE}{da} = 2N M_a + 2a \sum x_i + 2b \sum x_i N_i \]
- \[ dE = 0 \]
- \[ N M_a + a \sum x_i + N \sum x_i N_i = 0 \]

Solve for \( a, b \).

Compute \( E \).
**Multilinear Regression**

\[ Y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \]

**Stepwise Regression**

- Pick most important independent variables
- Add to mix \( \beta \frac{1}{c} \).
- Use correlation coefficient \( p \).
- Compute \( p \) for each independent var + dependent var.
- Pick var with highest \( |p| \).
- Compute residual errors.
- Add new most important var.
NON PARAMETRIC STATS

- DON'T ASSUME ANYTHING ABOUT DISTRIBUTION -
  MORE ROBUST
  BUT WEAKER -

ARE RED OBSERVATIONS SMALLER. DOES RED POPULATION HAVE SMALLER MEAN?
USE ONLY THEIR ORDER.
COUNT TIMES \( R_a < B_j \)
\[
U = 5 + 5 + 4 + 3 + 3 = 20
\]

"MANN-WHITNEY U STAT"

IF NULL \( \bar{U} = 12.5 \)
\[
\text{VAR}(\bar{U}) = \frac{n_1 n_2}{12} \cdot s^2 = 4.8
\]
Null hypothesis that red torque are same mean 5.5 normal FTA took shortcuts. This is unlikely.

ML Machine Learning

Current Big App of Stats