Stat 300: Unit 3 Study Guide [Section 9-2] Fall 2014

Problems involving differences between 2 population proportions (percentages, rates, fractions, etc.): $(p_1 - p_2)$.

$\hat{p}_1$ & $\hat{p}_2$ are sample proportions: $\hat{p}_1 = \dfrac{X_1 \text{ successes}}{N_1 \text{ trials}}$ & $\hat{p}_2 = \dfrac{X_2 \text{ successes}}{N_2 \text{ trials}}$. These may be given in the problem or they may need to be calculated using data provided.

$$CI(p_1 - p_2) = (\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

H.T. $(p_1 - p_2)$

Test Statistic

If difference in H_0: and H_1: is Not zero

value from hypotheses

$$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_0}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$$

Test Statistic

If difference in H_0: and H_1: is zero

$$\frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\bar{p}\,\bar{q}}{n_1} + \frac{\bar{p}\,\bar{q}}{n_2}}}$$

where

$$\bar{p} = \frac{X_1 + X_2}{N_1 + N_2} = \frac{\text{all successes}}{\text{all trials}}$$

Critical Region Picture

Based on Z (standard Normal)

Stat 300: Uni 3 Study Guide [section 9-3] Fall 2014

Problems involving differences between 2 population means (averages, typical values, central values, expected values): $(\mu_1 - \mu_2)$ { **NOT** Matched Pairs }

Issue: Are the variances (variation, standard deviations) different or similar?

- If different, then use each sample standard deviation separately and use the smaller of df_1 and df_2 as the degrees of freedom
- If simiar (same or =) then ~~pool~~ pool the variation into S_p^2 and use $df_1 + df_2$ as deg. of freedom

$$S_p^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{(n_1 - 1) + (n_2 - 1)}$$

Use S_p^2 in place of S_1^2 and S_2^2.

$$CI(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

H. T. $(\mu_1 - \mu_2)$:

Test Statistic

value from hypotheses

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Critical region picture

Based on t with appropriate degrees of freedom

Stat 300 : Unit 3 Study Guide [section 9-4] Fall 2014

Problems involving ~~the~~ differences between 2 population means where the samples are matched pairs:

$$(\mu_1 - \mu_2) = \mu_d$$

$\bar{d}$ = mean of the paired differences

S_d = standard deviation of the paired differences

n = number of pairs.

oberved values in the data are paired by the design of the experiment.

$$CI(\mu_1 - \mu_2) = CI(\mu_d) = \bar{d} \pm t_{\substack{\alpha/2 \\ df=(n-1)}} \left(\frac{S_d}{\sqrt{n}} \right)$$

H.T. $(\mu_1 - \mu_2)$ or H.T. (μ_d)

Test statistic

$$\frac{\bar{d} - \mu_d}{S_d / \sqrt{n}}$$

from hypotheses

critical region picture

based on t distribution with $(n-1)$ deg. of freedom.

Stat 300: Unit 3 Study Guide [section 10-2] Fall 2014

Correlation [linear correlation or Pearson correlation]

Population correlation is "rho" "ρ".
Sample correlation is "r"

To calculate r:

[2nd] [data] (Stat) 2-VAR [ENTER]

[2nd] [data] (STAT) CLRDATA [enter]

[data] x_1 = (value) ~~enter~~ [↓]

y_1 = (value) [↓]

x_2 = (value) [↓]

⋮

y_n = (value) [↓]

[STATVAR]

n $\bar{x}_1$ ......

[◄] [◄] [◄] and "r" is underlined: value is displayed

H. T. (ρ)

Test statistic

$$\frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

critical region picture

based on t
with $(n-2)$ d.f.

Principles:

$-1 \le r \le 1$ if "slope" is negative, then r is negative.

$|r_1| > |r_2|$ means the points for sample #1 appear to be closer to a straight line compared to points for sample #2

Stat 300: Unit 3 Study Guide [Section 10-3] Fall 2014

Linear Regression: Part 1

- Enter data as for correlation
- After pressing $\boxed{\text{STATVAR}}$, slope = $\underline{\underline{a}}$ & intercept = $\underline{\underline{b}}$

To plot the graph of the "regression" line:

(1) pick a low value for X and calculate the $\hat{y}$ value as: $\hat{y} = \underline{\underline{b}} + \underline{\underline{a}} X_1$ or use $y'(X)$ = from calculator

(2) pick a high value for X and calculate the $\hat{y}$ value as in (#1). $\hat{y} = b + a X_2$

(3) Plot the 2 $(X_1, \hat{y}_1)$ and $(X_2, \hat{y}_2)$

(4) Draw a <u>straight</u> line through the two points

The equation of the regression line (best-fit line, least squares line, etc.) is

$$\hat{y} = b + a X = \text{intercept} + \text{slope}(X)$$

values for these

To "predict y" for a known value of $X = X_0$,

(predicted y) = $\hat{y} = b + a(X_0)$

Stat 300: Unit 3 Study Guide [Section 10-~~4~~] Fall 2014

Regression Part 2

Data have been entered as for correlation

"Total" Variation = "Explained" Variation + "Unexplained" Variation

$$\sum(y-\bar{y})^2 = \sum(\hat{y}-\bar{y})^2 + \sum(y-\hat{y})^2$$

value = $S_y^2(n-1)$

value = ("Total")(r^2)

value = "Total" − "Explained" OR ("Total")($1-r^2$)

r^2 = the proportion of the total variation in y that is explained by the regression line

$$= \frac{\text{Explained Variation}}{\text{Total Variation}} = \frac{\sum(\hat{y}-\bar{y})^2}{\sum(y-\bar{y})^2}$$

$$S_e = \text{standard error of Estimate} = \sqrt{\frac{\text{unexplained}}{n-2}} = \sqrt{\frac{\sum(y-\hat{y})^2}{n-2}}$$

CI ($y \mid x_0$ is known) =

$$\hat{y} \pm t_{\alpha/2} \cdot S_e \sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x-\bar{x})^2}}$$

(df = n−2)

see above

value = $S_x^2(n-1)$

Stat 300: Unit 3 Study Guide [section 11-2] Fall 2014

Multi-Nomial Problems:
"Counts in Categories" where each count belongs to only one category

Data:

category	Observed count	Hypothetical proportions	Expected count	$\left[\frac{(O-E)^2}{E}\right]$
1	X_1	p_1	$N \times p_1$	calculate this for each category
2	X_2	p_2	$N \times p_2$	
⋮	⋮	⋮	⋮	
K	X_K	p_K	$N \times p_K$	
	$\Sigma = N$		$\Sigma = N$ $\Sigma p_i = 1$ or 100%	sum of these is the value of the test statistic

Test Statistic

$$\sum_{i=1}^{K} \left[\frac{(O_i - E_i)^2}{E_i}\right]$$

critical value picture

based on χ^2 with K-1 degrees of freedom.

Always a right-tail test.

State hypotheses

H_0: ~~__~~ in english as the problem indicates hypothetical proportions (=)

H_1: Not H_0:

Stat300 : Unit 3 Study Guide [Section 11-3] Fall 2014

Contingency Tables:

"Counts in Categories" where each count belongs to <u>two</u> categories at the same time

Data are explicitly or implicitly arranged in "rows" that represent one "factor" and "columns" that represent another "factor".

Observed counts

Factor #1 \ Factor #2	1	2	...	c	~~cota~~ Row Total
1	O_{11}	O_{12}	...	O_{1c}	
2	O_{21}	O_{22}	...	O_{2c}	
⋮	⋮	⋮	⋱	⋮	
r	O_{r1}	O_{r2}	...	O_{rc}	
col total					Grand total

Expected Counts

$$E = \frac{(\text{row total})(\text{col total})}{\text{grand total}}$$

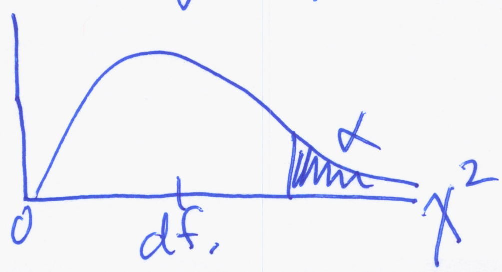
E_{11}	E_{12}	...	E_{1c}
E_{21}	E_{22}	...	E_{2c}
⋮	⋮	⋱	⋮
E_{r1}	E_{r2}	...	E_{rc}

$\boxed{O_{ij} \text{ goes with } E_{ij}}$

Test Statistic

$$\sum \left[\frac{(O-E)^2}{E} \right]$$

critical region picture



$d.f. = (\text{rows}-1)(\text{columns}-1)$

[If there are 3 rows and 4 columns, then $df = (3-1)(4-1) = (2)(3) = 6$]

STAT 300: Unit 3 Study Guide [section 12-2] Fall 2014

Analysis of Variance:

Population means for all treatments or groups are equal (the same, not different, etc.), so

$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ k = # of treatments or groups

H_1: Not H_0:

A.O.V. Table

source	sum of squares (SS)	deg. of Freedom	mean square	F
Treatments or Between Groups	SS(Treat)	K-1	ms(Treat)	ms(Treat) ÷ ms(Error)
Error or Within Groups	SS(Error)	N-k	ms(Error)	The F value is the Test statistic
Total	SS(Total)	N-1		

N: total # of observed values

Sums of squares and degrees of freedom add up in each column to the total at the bottom of the column.

critical region picture

$\text{Total SS} = S_x^2(n-1)$ for all the data together

$MS[\text{Error}]$ = pooled variance for all the Treatment groups

$$S_p^2 = MSE = \frac{S_1^2(n_1-1) + S_2^2(n_2-1) + \cdots + S_k^2(n_k-1)}{(n_1-1)+(n_2-1)+\cdots+(n_k-1)}$$

the critical value is for (k-1) numerator d.f. and (N-k) denominator d.f.