Cheat Sheet Final

Quantitative Methods 1 (York University)

Describe 1 Population (1 sample given)

- Data Type:
  - Interval Data (#'s) → Focus of Question:
    - Unknown Population Mean (μ):
      - Population standard deviation (σ) is given → 1
      - Population standard deviation (σ) is unknown → 2
    - Unknown Population Variance (σ^2) → 3
  - Nominal Data (categories) → Focus of Question:
    - Focus on one category → 4
    - Describe entire distribution → 14

Compare 2 Populations (2 samples given)

- Data Type:
  - Interval Data (#'s) → Focus of Question:
    - Difference between Population Means:
      - Independent groups:
        - Population variances given → 5
        - Population variances unknown → First Perform F-test (9) to decide equal or unequal variances:
          - Do Not Reject Ho → 6
          - Reject Ho → 7
      - Matched Pairs → 8
    - Ratio of population variances → 9
  - Nominal Data (categories):
    - CASE1: Is one population proportion... - greater than - less than - different than the other population proportion? → 10
    - CASE 2: Is one population proportion... - greater than - less than - different than the other population proportion by a specified amount? → 11

Compare 3+ Populations (3+ samples given) Interval Data → 12

Are They Related?

- Data Type:
  - 2 Nominal Variables → 13
  - 2 Interval Variables → 15

---

1 z-test: one population mean μ

HYPOTHESES

$H_0: \mu = \#$ ← given in question

$H_1: \mu \underset{\text{choose}}{< \text{ or } \neq \text{ or } >} \#$

TEST STATISTIC

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$ ← μ is the # from the H_0

REQUIRED CONDITIONS

The population is normal, or approximately so.

ESTIMATION

$$LCL = \bar{x} - z_{\alpha/2}(\sigma/\sqrt{n})$$
$$UCL = \bar{x} + z_{\alpha/2}(\sigma/\sqrt{n})$$

2 t-test: one population mean μ

HYPOTHESES

$H_0: \mu = \#$ ← given in question

$H_1: \mu \underset{\text{choose}}{< \text{ or } \neq \text{ or } >} \#$

TEST STATISTIC

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$ ← μ is the # from the H_0

DEGREES FREEDOM

$v = n - 1$

REQUIRED CONDITIONS

Population is normal.

STANDARD DEVIATION

$$s = \sqrt{\left(\frac{1}{n-1}\right)\left(\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right)}$$

ESTIMATION

$$LCL = \bar{x} - t_{\alpha/2,n-1}(s/\sqrt{n})$$
$$UCL = \bar{x} + t_{\alpha/2,n-1}(s/\sqrt{n})$$

3 χ^2-test: one population variance σ^2

HYPOTHESES

$H_0: \sigma^2 = \#$ ← given in question

$H_1: \sigma^2 \underset{\text{choose}}{< \text{ or } \neq \text{ or } >} \#$

TEST STATISTIC

$$X^2 = \frac{(n-1)s^2}{\sigma^2}$$ ← σ^2 is the # from the H_0

DEGREES FREEDOM

$v = n - 1$

REQUIRED CONDITIONS

The population is normal.

ESTIMATION

$$LCL = \frac{(n-1)s^2}{X^2_{\alpha/2,n-1}}$$
$$UCL = \frac{(n-1)s^2}{X^2_{1-\alpha/2,n-1}}$$

4 z-test: one population proportion p

HYPOTHESES

$H_0: p = 0.\#$ ← given in question

$H_1: p \underset{\text{choose}}{< \text{ or } \neq \text{ or } >} 0.\#$

TEST STATISTIC

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$ ← p is the 0.# from the H_0

$$\hat{p} = \frac{x}{n}$$ ← x is the number of successes in the sample

REQUIRED CONDITIONS

$np > 5$, and...
$n(1-p) > 5$

ESTIMATION

$$LCL = \hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$
$$UCL = \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

REQUIRED CONDITIONS

$n\hat{p} > 5$, and...
$n(1-\hat{p}) > 5$

5 z-test: difference between two population means (μ_1-μ_2)

HYPOTHESES

$H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 \underset{\text{choose}}{< \text{ or } \neq \text{ or } >} 0$

[$H_0: \mu_1 - \mu_2 = 0$ unless testing for specified difference given in question]

TEST STATISTIC

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

[$(\mu_1 - \mu_2)$ = value stated in H_0]

REQUIRED CONDITIONS

The population is normal, or approximately so.

ESTIMATION

$$LCL = (\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$
$$UCL = (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

StatsDoesntSuck.com

6 Equal Variances t-test: difference between two population means (μ_1-μ_2)

HYPOTHESES

$H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 \underset{\text{choose}}{< \text{ or } \neq \text{ or } >} 0$

[$H_0: \mu_1 - \mu_2 = 0$ unless testing for a specified difference given in the question]

TEST STATISTIC

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

DEGREES FREEDOM

$v = n_1 + n_2 - 2$

REQUIRED CONDITIONS

The population is normal.

ESTIMATION

$$LCL = (\bar{x}_1 - \bar{x}_2) - t_{\alpha/2,v}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$
$$UCL = (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2,v}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

PRE-MIDTERM STUFF:

SAMPLE VARIANCE

$$s^2 = \left(\frac{1}{n-1}\right)\left(\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right)$$

SAMPLE MEAN

$$\bar{x} = \frac{\sum x}{n}$$

7 Unequal Variances t-test: difference between two population means (μ_1-μ_2)

HYPOTHESES

$H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 \underset{\text{choose}}{< \text{ or } \neq \text{ or } >} 0$

[$H_0: \mu_1 - \mu_2 = 0$ unless testing for a specified difference given in the question]

TEST STATISTIC

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

DEGREES FREEDOM

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

REQUIRED CONDITIONS

The population is normal.

ESTIMATION

$$LCL = (\bar{x}_1 - \bar{x}_2) - t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$
$$UCL = (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

8 Matched pairs t-test

HYPOTHESES

$H_0: \mu_D = 0$

$H_1: \mu_D \underset{\text{choose}}{< \text{ or } \neq \text{ or } >} 0$

TEST STATISTIC

$$t = \frac{\bar{x}_D - \mu_D}{s_D/\sqrt{n_D}}$$ ← use $\mu_D = 0$ from H_0

DEGREES FREEDOM

$v = n_D - 1$

REQUIRED CONDITIONS

The population of difference scores are normal.

STANDARD DEVIATION

$$s = \sqrt{\left(\frac{1}{n-1}\right)\left(\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right)}$$

ESTIMATION

$$LCL = \bar{x}_D - t_{\alpha/2,n-1}(s_D/\sqrt{n_D})$$
$$UCL = \bar{x}_D + t_{\alpha/2,n-1}(s_D/\sqrt{n_D})$$

9 F-test: Ratio of population variances $\frac{\sigma_1^2}{\sigma_2^2}$

HYPOTHESES

$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1 \qquad H_1: \frac{\sigma_1^2}{\sigma_2^2} \underset{\text{choose}}{< \text{ or } \neq \text{ or } >} 1$

TEST STATISTIC

$$F = \frac{s_1^2}{s_2^2}$$

DEGREES FREEDOM

$v_1 = n_1 - 1$
$v_2 = n_2 - 1$

REQUIRED CONDITIONS

The populations are normal and independent.

ESTIMATION

$$LCL = \left(\frac{s_1^2}{s_2^2}\right)\frac{1}{F_{\alpha/2,v_1,v_2}}$$
$$UCL = \left(\frac{s_1^2}{s_2^2}\right)F_{\alpha/2,v_2,v_1}$$

10 z-test: CASE 1 Difference between two population proportions (p_1-p_2)

HYPOTHESES

$H_0: p_1 - p_2 = 0$

$H_1: p_1 - p_2 \underset{\text{choose}}{< \text{ or } \neq \text{ or } >} 0$

TEST STATISTIC

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \hat{p}_1 = \frac{x_1}{n_1} \quad \hat{p}_2 = \frac{x_2}{n_2}$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$ x_1, x_2 = # successes in sample 1,2

REQUIRED CONDITIONS

$n_1\hat{p}_1 > 5$ and... $n_2\hat{p}_2 > 5$
$n_1(1-\hat{p}_1) > 5$ and... $n_2(1-\hat{p}_2) > 5$

10 & 11 z-estimate of the difference between two population proportions (p_1-p_2)

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

REQUIRED CONDITIONS

$n_1\hat{p}_1 > 5$ and $n_2\hat{p}_2 > 5$
$n_1(1-\hat{p}_1) > 5$ and $n_2(1-\hat{p}_2) > 5$

Steps of Hypothesis Test:

1. State the hypotheses: H_0 and H_1
2. Identify the significance level from question (use $\alpha = 0.05$ unless question states otherwise)
3. State rejection region
4. Calculate the test statistic
5. Compare test statistic & rejection region. Conclude.

StatsDoesntSuck.com

11 z-test: CASE 2 Difference between two population proportions (p_1-p_2)

HYPOTHESES

$H_0: p_1 - p_2 = 0.\#$ ← given in question

$H_1: p_1 - p_2 \underset{\text{choose}}{< \text{ or } \neq \text{ or } >} 0.\#$

TEST STATISTIC

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \qquad \hat{p}_1 = \frac{x_1}{n_1} \quad \hat{p}_2 = \frac{x_2}{n_2}$$

x_1, x_2 = # successes in sample 1,2

REQUIRED CONDITIONS

$n_1\hat{p}_1 > 5$ and... $n_2\hat{p}_2 > 5$
$n_1(1-\hat{p}_1) > 5$ and... $n_2(1-\hat{p}_2) > 5$

Rejection Regions (and p-values for z-tests) depending on signs in hypotheses

	$H_0 =$ $H_1 <$	$H_0 =$ $H_1 \neq$	$H_0 =$ $H_1 >$
z	Reject H_0 if $z < -z_\alpha$	Reject H_0 if $z < -z_{\alpha/2}$ or if $z > +z_{\alpha/2}$	Reject H_0 if $z > +z_\alpha$
	$pvalue = P\left(z < \text{test statistic}\right)$	$pvalue = 2 \times P\left(z > \text{positive test statistic}\right)$	$pvalue = P\left(z > \text{test statistic}\right)$
t	Reject H_0 if $t < -t_{\alpha,v}$	Reject H_0 if $t < -t_{\alpha/2,v}$ or if $t > +t_{\alpha/2,v}$	Reject H_0 if $t > +t_{\alpha,v}$
χ^2	Reject H_0 if $X^2 < X^2_{1-\alpha,v}$	Reject H_0 if $X^2 < X^2_{1-(\alpha/2),v}$ …	Reject H_0 if $X^2 > X^2_{\alpha,v}$
F	Reject H_0 if $F < \frac{1}{F_{\alpha,v_2,v_1}}$	… or if $F > F_{(\alpha/2),v_1,v_2}$	* always use for (12)

ALWAYS

Reject the H_0 if p-value < α regardless of sign ($<, \neq, >$) in H_1

Interpreting the p-value

Wrong!	The p-value is the probability that the Ho is true
Correct:	The p-value is the probability of obtaining a test statistic at least as extreme as the one that's actually observed, assuming the Ho is true

As p-value gets closer to 0, test results become more significant

Conclusions for Hypothesis Test

…nce (at the … that the H_1 is true.

…evidence (at the … level) to infer that the H_1 is true.

Minimum Sample Sizes

μ

$$n = \left(\frac{z_{\alpha/2}\sigma}{B}\right)^2$$

↑ B is the given amount the estimation must be within ↓

p

$$n = \left(\frac{z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})}}{B}\right)^2$$

if $\hat{p}$ is not given from a previous study, use $\hat{p}$ = 0.5 or as close to 0.5 as possible

12 ANOVA: One way ANalysis Of VAriance

HYPOTHESES

$H_0: \mu_1 = \mu_2 = \mu_3 = \ldots$ *(keep going for more groups)*

H_1: At least two means differ

REQUIRED CONDITIONS

population data must be normal

population variances must be equal

GRAND MEAN

$$\bar{\bar{x}} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3 + \ldots}{n_1 + n_2 + n_3 \ldots}$$

TEST STATISTIC

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F-STATISTIC
Treatments	$df_T = k - 1$ ← k = # of samples	$SST = n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + n_3(\bar{x}_3 - \bar{\bar{x}})^2 + \ldots$	$MST = \frac{SST}{k-1}$	$F = \frac{MST}{MSE}$
Error	$df_E = (n_1 + n_2 + n_3 + \ldots) - k$	$SSE = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2 + \ldots$	$MSE = \frac{SSE}{n-k}$	
Total	$df_{TOTAL} = (n_1 + n_2 + n_3 + \ldots) - 1$	$SS_{TOTAL} = SST + SSE$	ANOVA SUMMARY TABLE	

SST: Sum of Squares for Treatments
(between treatments variation) Measures how close the sample means are to each other.

SSE: Sum of Squares for Error
(within treatments variation) Measures variation in responses not explained by the different treatments.

Relationship between F and t statistics

$F = t^2$ but only if:

- comparing exactly 2 population means
- $H_1: \mu_1 \neq \mu_2$ in the t-test
- comparing independent groups with equal variances

WARNING: multiple t-tests cannot replace the ANOVA test, otherwise an increased probability of Type I error will result

13 χ^2-test of a Contingency Table

HYPOTHESES

H_0: The two variables are independent

H_1: The two variables are related

DEGREES FREEDOM

$df = (r-1) \times (c-1)$ r = # of rows, c = # of columns in table

EXPECTED FREQUENCIES

$$e_{ij} = \frac{(row\ i\ total) \times (column\ j\ total)}{n}$$

⇐ This box applies to Test of Contingency + Goodness of Fit ⇒
(fill in given frequencies in 1st column, then complete other columns)

TEST STATISTIC

f ← given in question	e	$(f-e)$	$(f-e)^2$	$(f-e)^2/e$
⇓+ (adds to)	⇓+	⇓+		⇓+
n	n	zero		$X^2 = \sum \frac{(f-e)^2}{e}$

REQUIRED

Rule of Five: Each expected frequency (e) must be at least 5

14 χ^2 Goodness of Fit Test

HYPOTHESES

$H_0: p_1 = 0.\#, p_2 = 0.\#, p_3 = 0.\#\ldots$

H_1: At least one p_i is not equal to its specified value

DEGREES FREEDOM

$df = k - 1$ k = # of categories

EXPECTED FREQUENCIES

$e_i = np_i$

Relationship between X^2 and z

$X^2 = (z)^2$ only as follows:

- the Goodness of fit test is equivalent to a two-tailed $(=, \neq)$ z-test of one proportion when there are only two categories.
- the Test of a Contingency Table is equivalent to a two-tailed $(=, \neq)$ z-test of the difference in two proportions (CASE I).

The 'model' decribes the relationship between x,y in the population. The least squares regression line describes the relationship in the sample

MODEL

$y = \beta_0 + \beta_1 x + \epsilon$

LEAST SQUARES REGRESSION LINE

$\hat{y} = b_0 + b_1 x$

- y = dependent variable
- x = independent variable
- β_0 = y-intercept
- β_1 = slope
- ε = error variable
- $\hat{y}$ = predicted y
- b_0 = estimated y-intercept
- b_1 = estimated slope

REGRESSION STEP #1: Set up table to calculate sums

x	y	x^2	y^2	xy
⇓+ (adds to)	⇓+	⇓+	⇓+	⇓+
$\sum x$	$\sum y$	$\sum x^2$	$\sum y^2$	$\sum xy$

STEP #2: Use sums to calculate statistics below.

COVARIANCE

$$s_{xy} = \left(\frac{1}{n-1}\right)\left(\sum xy - \frac{(\sum x)(\sum y)}{n}\right)$$

VARIANCE OF X

$$s_x^2 = \left(\frac{1}{n-1}\right)\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)$$

VARIANCE OF Y

$$s_y^2 = \left(\frac{1}{n-1}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)$$

MEAN X $\bar{x} = \frac{\sum x}{n}$ **MEAN Y** $\bar{y} = \frac{\sum y}{n}$

Y-INTERCEPT	SLOPE	INTERPRETATION OF SLOPE
$b_0 = \bar{y} - b_1\bar{x}$	$b_1 = s_{xy}/s_x^2$	*For every unit increase in x, y increases on average by b_1*

SSE (SUM OF SQUARES ERROR)

$$SSE = (n-1)\left(s_y^2 - \frac{s_{xy}^2}{s_x^2}\right)$$

INTERPRETATION OF S_ε

S_ε measures (on average) how close values predicted by the regression equation are to the actual observed values

STANDARD ERROR OF ESTIMATE

$$s_\varepsilon = \sqrt{\frac{SSE}{n-2}}$$

Compare S_ε to $\bar{y}$ (when available) to determine the relative size of the error. If S_ε is close to 0 the model is a good fit

15 Test of Slope

HYPOTHESES

$H_0: \beta_1 = 0$ ← *no relationship*

$H_1: \beta_1 \underbrace{< \text{ or } \neq \text{ or } >}_{\text{choose}} 0$

< for negative, > for positive relationships

TEST STATISTIC

$t = \frac{b_1 - \beta_1}{s_{b_1}}$ ← *β_1 is 0 from H_0*

STANDARD ERROR OF SLOPE

$$s_{b_1} = \frac{s_\varepsilon}{\sqrt{(n-1)s_x^2}}$$

DEGREES FREEDOM

$v = n - 2$

REQUIRED CONDITIONS

Variables are bivariate normal.

Test of rho

HYPOTHESES

$H_0: \rho = 0$ ← *no relationship*

$H_1: \rho \underbrace{< \text{ or } \neq \text{ or } >}_{\text{choose}} 0$

TEST STATISTIC

$t = r\sqrt{\frac{n-2}{1-r^2}}$ **DEGREES FREEDOM** $v = n - 2$

COEFFICIENT OF CORRELATION

$r = \frac{s_{xy}}{\sqrt{(s_x^2)(s_y^2)}}$ ← *measures the strength and direction of the linear relationship*

REQUIRED

Variables are bivariate normal.

Coefficient of Determination

$$R^2 = \frac{(s_{xy})^2}{(s_x^2)(s_y^2)}$$

$$R^2 = 1 - \frac{SSE}{\underbrace{\sum(y_i - \bar{y})^2}_{\text{Sum of Squares Total}}}$$

INTERPRETATION

R^2 represents the proportion of variation in y that is explained by the variation in x. The remaining proportion $(1 - R^2)$ cannot be explained

If R^2 is close to 1.0 the model is a good fit

Correlation ≠ Causation... If x and y are correlated, this DOES NOT mean that x causes y!

PREDICTION INTERVAL	CONFIDENCE INTERVAL
$\hat{y} \pm t_{\alpha/2,v}\, s_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x_g - \bar{x})^2}{(n-1)s_x^2}}$ for estimating an individual y	$\hat{y} \pm t_{\alpha/2,v}\, s_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_g - \bar{x})^2}{(n-1)s_x^2}}$ for estimating the mean $\bar{y}$
Where x_g is the given value of x and $\hat{y} = b_0 + b_1 x_g$	*Where x_g is the given value of x and $\hat{y} = b_0 + b_1 x_g$*

Qualities of a Good Estimator:

- **Unbiased**: The average (expected) value of the estimator equals the population parameter being estimated.
- **Consistent**: The estimator gets closer to the population parameter as the sample size increases.
- Has **Relative Efficiency**: If two unbiased estimators are available as estimator, choose the one with less variability

Probabilities of Type I and II Errors

Decision	H_0 is TRUE	H_0 is FALSE
Reject the H_0	Type I Error P(Type I) = α	correct decision (power)
Don't Reject H_0	correct decision	Type II Error P(Type II) = β

DISTRIBUTIONS

t distribution

- Symmetrical with 0 at center
- Wider and flatter than the z (normal) distribution
- Becomes wider, flatter as df ⇩

χ^2 and F distributions

- Positively skewed (stretched to the right)
- Starts at 0. Positive values only
- Changes shape as df change

Interpreting a Confidence Interval:

The mean is estimated to fall between (LCL*) and (UCL*). This type of estimation is correct (confidence level*)% of the time.

* Fill in the values

Factors affecting WIDTH of Confidence Intervals

As sample size n⇧...	width ⇩ (good)
As sample size n⇩...	width ⇧ (wide intervals are bad)
As conf. level (1-α) ⇧...	width ⇧ (wide intervals are bad)
As conf. level (1-α) ⇩...	width ⇩ (loss of confidence is bad)
As variance⇧...	width ⇧ (bad: can't adjust variance)
As variance⇩...	width ⇩ (bad: can't adjust variance)
As sample mean⇧...	width is unaffected
As sample mean⇩...	width is unaffected

Estimating the TOTAL in a population	Estimating the TOTAL # SUCCESSES in a population
(when population size N is given)	
$N\left[\bar{x} \pm t_{\alpha/2,v}\left(\frac{s}{\sqrt{n}}\right)\right]$	$N\left[\hat{p} \pm z_{\alpha/2}\left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)\right]$

***Using the results of a Confidence Interval Estimation to test =,≠ Hypotheses**

*only if (conf. level %) + (significance level %) = 100%

Reject H_o	If H_0 is DOES NOT fall between the LCL and UCL
Do not Reject H_o	If H_0 is DOES fall between the LCL and UCL

CHECK OF CONDITIONS: Central Limit Theorem

Sampling Distribution of the Sample Mean

1. $\mu_{\bar{x}} = \mu$
2. $\sigma_{\bar{x}}^2 = \sigma^2/n$ and $\sigma_{\bar{x}} = \sigma/\sqrt{n}$
3. If x is normal, then $\bar{x}$ is normal. If x is nonnormal, then $\bar{x}$ is approximately normal for suficiently large sample sizes. The definition of "sufficiently large" depends on the extent of nonnormality of x.

DEFINITIONS

Analysis of Variance:	compares two or more population means.
Between Treatments Variation:	Variation between the treatment means. Measured with SST.
Confidence Level (1-α):	The probability that your estimation will be correct.
Dependent Variable (y):	The variable that we attempt to forecast / predict in a regression analysis.
Error Variable (ε):	accounts for all the variables, measureable and immeasureable, that are not part of the regression model.
Expected Frequency:	The frequency you expect to observe when the null is true.
Experimental Data:	Subjects are assigned to groups (the independent variable is manipulated) and the values of the dependent variable are observed. Studies using this type of data can show cause-effect relationships as well as correlational relationships.
Interval Estimate:	Range of values that estimate an unknown *population parameter*.
Least Squares Method:	Used in regression to create a best-fit line for the observed data by minimizing the sum of squared differences between the predicted and actual observed y-values.
Model (first-order-linear):	attempts to represent the relationship between the dependent variable (y) when given the value of the independent variable (x). Both variables must be interval.
Observational Data:	The independent AND dependent variables are both observed. Studies using this type of data can ONLY show correlational relationships (not cause-effect like experimenta data can).
P value:	The probability of obtaining a test statistic at least as extreme as the one that's actually observed, assuming the Ho is true
Point Estimate:	Single value that estimates an unknown *population parameter*.
Pooled Variance:	Weighted average of the sample variances (weighted according to df)
Population Parameter:	A characteristic of the population (like the mean, the proportion, etc).
Power of a Test:	The probabilty of rejecting a false null.
Residual (e = y - ŷ):	the difference between the actual (observed) and predicted y.
Rejection Region:	If the test statistic falls into this range of values then the Ho is rejected.
Regression Analysis:	Method for predicting the value of the dependent variable (y) when given the value of the independent variable (x). Both variables must be interval.
Significance Level (α):	The probability of a *Type I error*.
Slope:	The value of the slope shows the average change to the independent variable (y) for every unit increase to the independent variable (x).
Statistically Significant:	A test is statistically significant when the Ho is rejected.
Sum of Squares Total (SStotal):	The total variation in scores. SST+SSE
Type I Error:	REJECTING a TRUE null hypothesis.
Type II Error:	NOT rejecting a FALSE null hypothesis.
Within Treatments Variation:	Variation in the scores that is not caused by the different treatments being compared. Measured with SSE.
y-intercept:	The value of the dependent variable (y) when the independent variable (x) = 0.