Details about hypothesis/significance tests and confidence intervals

A significance test is designed to **DECIDE** the answer to a **YES OR NO** question using **DATA**.

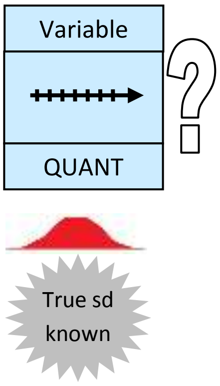
This is the information we need to know about a significance test for Stat Prac I:

- <u>QUESTION:</u>
  Types of yes-or-no questions it can answer, including the types of variables involved.
- <u>PARAMETERS:</u>
  The parameters that need to be defined (if any).
- <u>HYPOTHESES:</u>
  Appropriate null and alternative hypotheses.
- <u>TEST STATISTIC:</u>
  The formula for the test statistic
- <u>DISTRIBUTION:</u>
  The distribution of the test statistic given that the null is true.
- <u>P-VALUE:</u>
  Where the p-value is on a picture (if appropriate)
- <u>COMPUTER:</u>
  Where the test statistic and p-value are in computer output (if appropriate)
- <u>ASSUMPTIONS:</u>
  Assumptions needed, and how to check them.
- <u>CONFIDENCE INTERVAL:</u>
  The confidence interval for the matching “how big is this number” question (if appropriate)

These are the significance tests covered in Stat Prac I:

- Z-test for one mean.
- One-sample t-test for one mean.
- Z-test for one proportion.
- Chi-squared test for goodness of fit.
- Paired t-test.
- Two-sample (unpaired) t-test.
- Wilcoxon rank-sum test.
- ANOVA
- Z-test for two proportions.
- Chi-squared test for homogeneity.
- Chi-squared test for independence.
- T-test for slope in regression
- Confidence and prediction intervals in regression

Z-test for one mean



QUESTION: Is the mean equal to this particular number?

PARAMETERS: Let μ be the true mean of the variable.

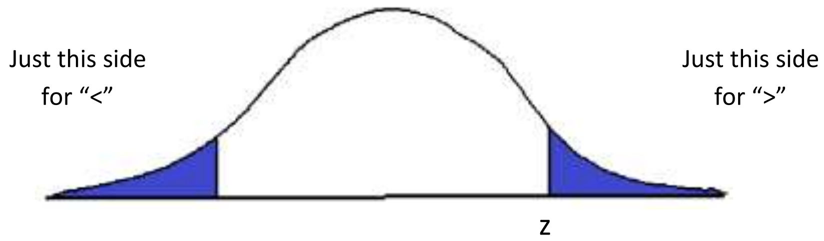
HYPOTHESES: $H_0: \mu = \mu_0$ and $H_a: \mu \neq \mu_0$ (or $\mu < \mu_0$ or $\mu > \mu_0$)

TEST STATISTIC:

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

DISTRIBUTION: $Z \sim N(0,1)$

P-VALUE:



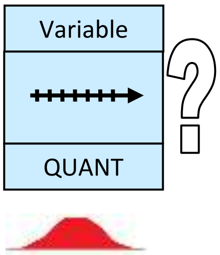
ASSUMPTIONS:

- True standard deviation known – check with story
- Normality – check with histogram
- Independence – check with story

CONFIDENCE INTERVAL:

$$\text{Confidence interval for mean: endpoints} = \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

One-sample T-test for a mean



QUESTION: Is the mean equal to this particular number?

PARAMETERS: Let μ be the true mean of the variable.

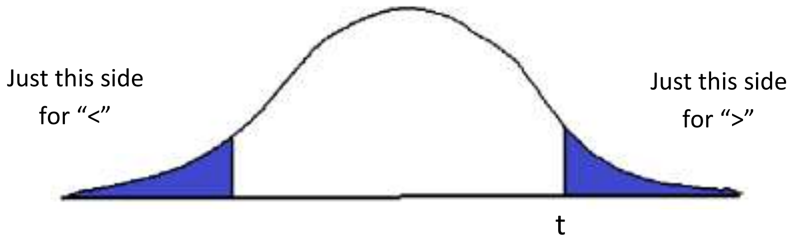
HYPOTHESES: $H_0: \mu = \mu_0$ and $H_a: \mu \neq \mu_0$ (or $\mu < \mu_0$ or $H_0: \mu > \mu_0$)

TEST STATISTIC:

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

DISTRIBUTION: $T \sim t(n-1)$

P-VALUE:



COMPUTER:

One-Sample Test

	Test Value = 5.52					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Density	-1.694	28	.101	-.0690	-.152	.014

Test statistic **P-value**

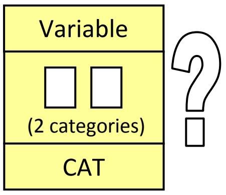
ASSUMPTIONS:

- Normality – check with histogram
- Independence – check with story

CONFIDENCE INTERVAL:

$$\text{Confidence interval for mean: endpoints} = \bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

Z-test for one proportion



QUESTION: Is the proportion in one category equal to this number?

PARAMETERS: Let p be the true proportion in one category.

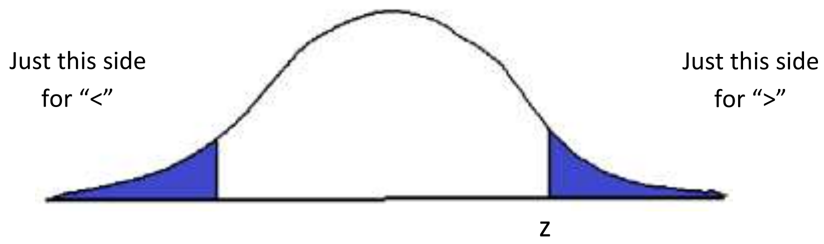
HYPOTHESES: $H_0: p = p_0$ and $H_a: p \neq p_0$ (or $p < p_0$ or $p > p_0$)

TEST STATISTIC:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

DISTRIBUTION: $Z \sim N(0,1)$

P-VALUE:



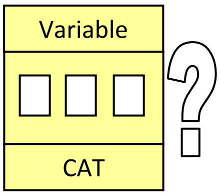
ASSUMPTIONS:

- Normality – $np_0 > 10$ and $n(1-p_0) > 10$
- Independence – check with story

CONFIDENCE INTERVAL:

$$\text{Confidence interval for proportion: endpoints} = \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Chi-squared test for goodness of fit



QUESTION: Are the proportions in the categories distributed according to this list of proportions?

PARAMETERS: Let $p_A,\ p_B,\ p_C, \ldots$ be the true proportions in the categories.

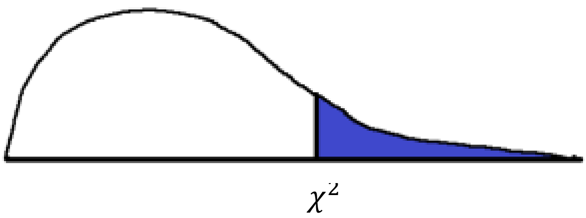
HYPOTHESES: $H_0: p_A = p_{A0},\ p_B = p_{B0},\ p_C = p_{C0}, \ldots$
and H_a: The proportions are not all as stated.

TEST STATISTIC:

$$\chi^2 = \sum \frac{(obs - exp)^2}{exp}, \text{ where } exp \text{ for category X} = n\, p_{X0}$$

DISTRIBUTION: $\chi^2 \sim \chi^2(df)$, where df = (#categories) - 1

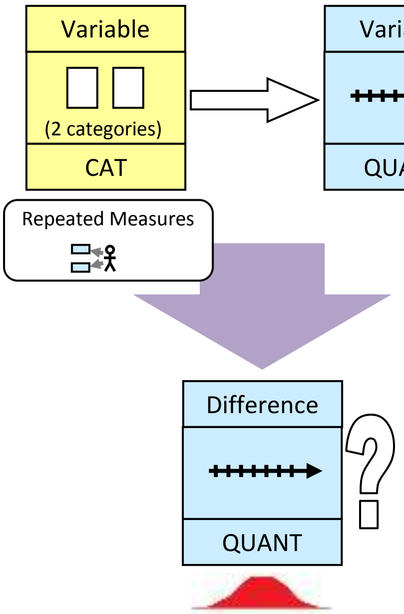
P-VALUE:



ASSUMPTIONS:

- Large sample – most expected counts ($n\, p_{X0}$) more than 5
- Independence – check with story

Paired t-test



<u>QUESTION:</u> Is the mean difference equal to 0?

<u>PARAMETERS:</u> Let μ_D be the true mean difference (calculated as A – B).

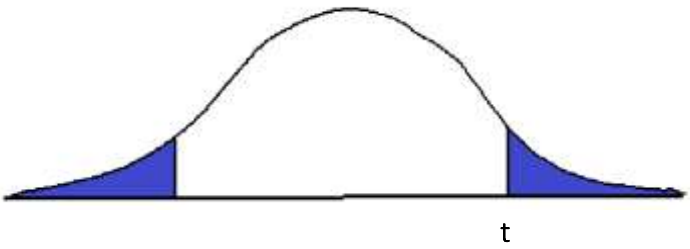
<u>HYPOTHESES:</u> $H_0: \mu_D = 0$ and $H_a: \mu_D \neq 0$

<u>TEST STATISTIC:</u>

$$T = \frac{\bar{d} - 0}{s_d/\sqrt{n}}$$

<u>DISTRIBUTION:</u> $T \sim t(n-1)$

<u>P-VALUE:</u>



<u>COMPUTER:</u>

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1 Dominant - Non Dominant	1.067	2.434	.628	-.281	2.415	1.697	14	.112

Test statistic P-value

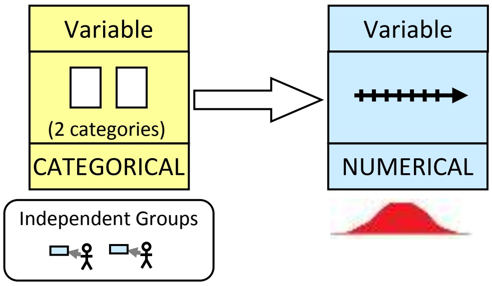
<u>ASSUMPTIONS:</u>

- Normality of differences – check with histogram
- Independence – check with story

<u>CONFIDENCE INTERVAL:</u>

$$\text{Confidence interval for mean difference: endpoints} = \bar{d} \pm t^* \frac{s_d}{\sqrt{n}}$$

Two-sample (unpaired) t-test



QUESTION: Does the category affect the outcome? OR Are the means different under the two categories?

PARAMETERS: Let μ_A be the true mean difference in category A, and μ_A be the true mean difference in category B.

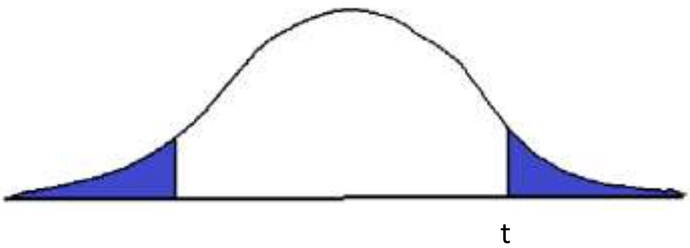
HYPOTHESES: $H_0: \mu_A = \mu_B$ and $H_a: \mu_A \neq \mu_B$

TEST STATISTIC:

$$T = \frac{\overline{x_A} - \overline{x_B}}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

DISTRIBUTION: $T \sim t(df)$ approx. (where df is complicated)

P-VALUE:



COMPUTER:

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Pain	Equal variances assumed	.653	.430	3.991	17	.001	15.533	3.892	7.323	23.744
	Equal variances not assumed			4.047	16.665	.001	15.533	3.838	7.424	23.643

Test statistic   df   P-value

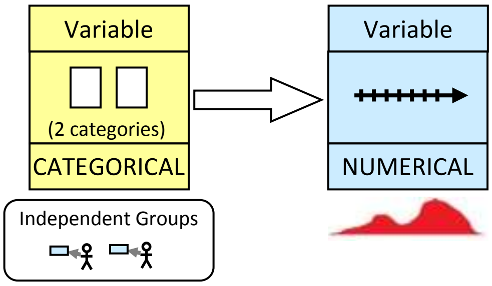
ASSUMPTIONS:

- Normality within each group – check with histograms
- Independence within groups – check with story
- Independence between groups – check with story

CONFIDENCE INTERVAL:

Confidence interval for difference in means: endpoints $= \overline{x_A} - \overline{x_B} \pm t^* \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$

Wilcoxon Rank Sum Test (Mann-Whitney U Test)



QUESTION: Does the category affect the outcome? OR Do the distributions differ under the two categories?

PARAMETERS: None.

HYPOTHESES: H_0: the two distributions are the same and H_a: the two distributions have different locations

TEST STATISTIC: (Exists but difficult to describe.)

DISTRIBUTION: (Difficult to describe.)

P-VALUE: From computer output.

COMPUTER:

P-value

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Pain is the same across categories of Hair Colour.	Independent-Samples Mann-Whitney U Test	.001[1]	Reject the null hypothesis.

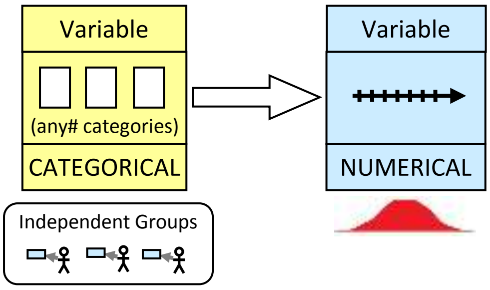
Asymptotic significances are displayed. The significance level is .05.

[1] Exact significance is displayed for this test.

ASSUMPTIONS:

- Independence within groups – check with story
- Independence between groups – check with story

ANOVA (analysis of variance)



QUESTION: Does the category affect the outcome?
OR Are the means different under the different categories?

PARAMETERS: Let μ_A be the true mean difference in category A, and μ_A be the true mean difference in category B, and ....

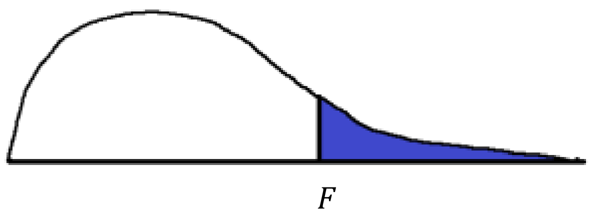
HYPOTHESES: $H_0: \mu_A = \mu_B = \mu_C = \cdots$
and H_a: the means are not all equal.

TEST STATISTIC:

$$F = \frac{MSG}{MSE} \text{ (see computer output)}$$

DISTRIBUTION: $F \sim F(g-1, n-g)$

P-VALUE:



COMPUTER:

Tests of Between-Subjects Effects

Dependent Variable: Bone Density (mg/cm3)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7433.867[a]	2	3716.933	7.978	.002
Intercept	11436717.6	1	11436717.63	24547.190	.000
Group	7433.867	2	3716.933	7.978	.002
Error	12579.500	27	465.907		
Total	11456731.0	30			
Corrected Total	20013.367	29			

a. R Squared = .371 (Adjusted R Squared = .325)

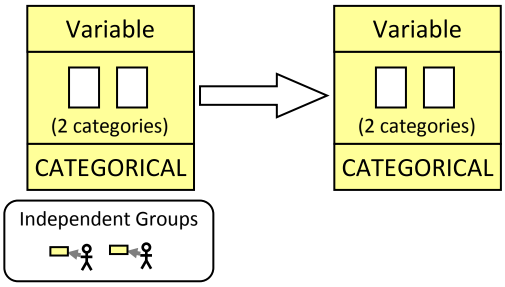
(+ ÷ df Test statistic P-value)

ASSUMPTIONS:

- Normality within each group – check with histograms
- Independence within groups – check with story
- Independence between groups – check with story
- Equal variance in groups – Levene's test has high p-value
  or (biggest stdev) / (smallest stdev) < 2

CONFIDENCE INTERVAL: Computer will calculate a CI for the difference between each pair of means.

Z-test for two proportions



QUESTION: Does the category affect the outcome? OR Is the proportion who are in the outcome category you are interested in different under the two grouping categories?

PARAMETERS: Let p_A and p_B be the true proportions who are in the outcome you are interested in under category A and under category B.

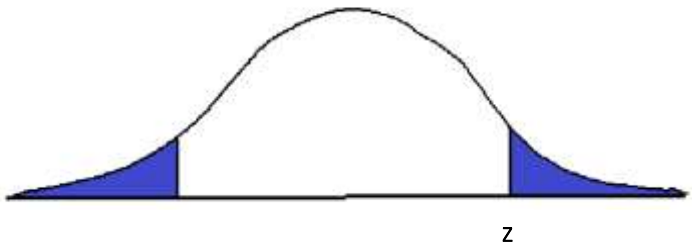
HYPOTHESES: $H_0: p_A = p_B$ and $H_a: p_A \neq p_B$

TEST STATISTIC:

$$Z = \frac{\widehat{p_A} - \widehat{p_B}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$$ where $\hat{p}$ is the pooled proportion.

DISTRIBUTION: $Z \sim N(0,1)$

P-VALUE:



ASSUMPTIONS:

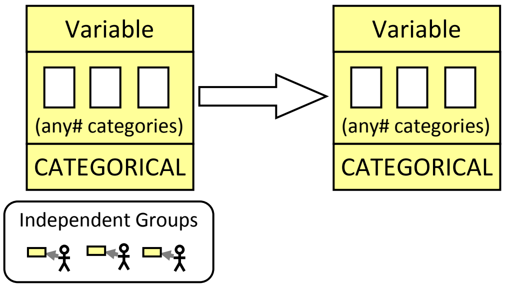
- Normality – number of successes and failures in both grouping categories are all at least 5
- Independence within groups – check with story
- Independence between groups – check with story

CONFIDENCE INTERVAL:

Confidence interval for difference in proportions:

$$\text{endpoints} = \widehat{p_A} - \widehat{p_B} \pm z^* \sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}$$

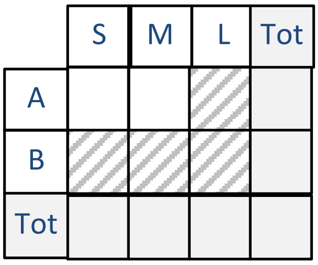
Chi-squared test for homogeneity



QUESTION: Are the proportions of the outcome different under the different grouping categories?

PARAMETERS: The proportions in each grouping category that are in each outcome category.

HYPOTHESES: H_0: The distribution of proportions in each outcome category is the same under each grouping category, and H_a: the distribution is not the same in each grouping category.

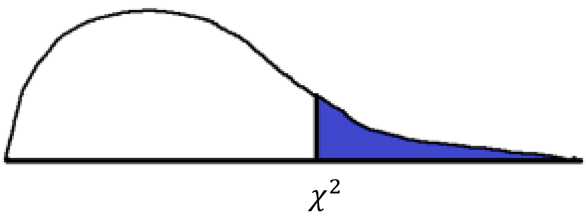


TEST STATISTIC:

$$\chi^2 = \sum \frac{(obs - exp)^2}{exp}, \text{ where } exp = \frac{row\ total \times column\ total}{grand\ total}$$

DISTRIBUTION: $\chi^2 \sim \chi^2(df)$, where df = number of cells left in two-way table after one row and column are removed.

P-VALUE:



COMPUTER:

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.350[a]	2	.309
Likelihood Ratio	2.386	2	.303
N of Valid Cases	35		

a. 1 cells (16.7%) have expected count less than 5. The minimum expected count is 3.86.

Test statistic

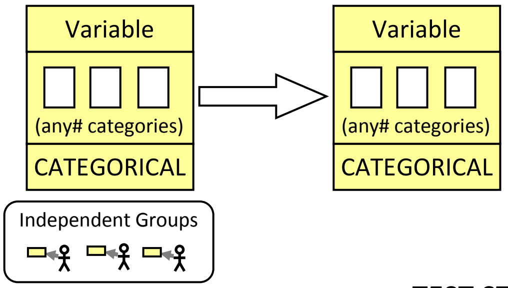
P-value

Assumption check

ASSUMPTIONS:

- Large sample – most expected counts more than 5
- Independence within groups – check with story
- Independence between groups – check with story

Chi-squared test for independence

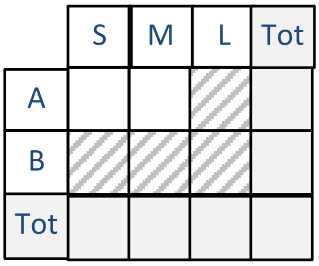


<u>QUESTION:</u> Are the two variables related?

<u>PARAMETERS:</u> None.

<u>HYPOTHESES:</u> H_0: There is no association between the variables, and H_a: There is an association between the variables.

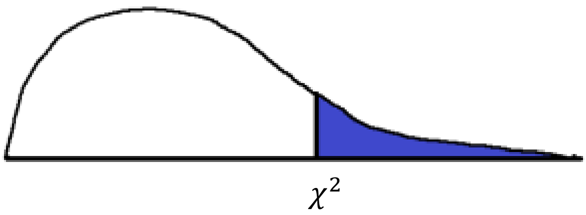
<u>TEST STATISTIC:</u>



$$\chi^2 = \sum \frac{(obs - exp)^2}{exp}, \text{ where } exp = \frac{row\ total \times column\ total}{grand\ total}$$

<u>DISTRIBUTION:</u> $\chi^2 \sim \chi^2(df)$, where df = number of cells left in two-way table after one row and column are removed.

<u>P-VALUE:</u>



<u>COMPUTER:</u>

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.350[a]	2	.309
Likelihood Ratio	2.386	2	.303
N of Valid Cases	35		

a. 1 cells (16.7%) have expected count less than 5. The minimum expected count is 3.86.

Test statistic

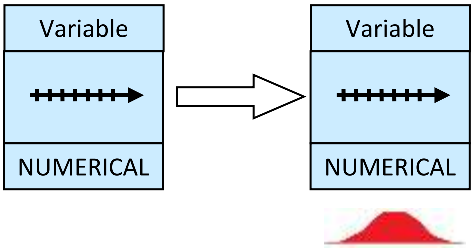
P-value

Assumption check

<u>ASSUMPTIONS:</u>

- Large sample – most expected counts more than 5
- Independence between individuals – check with story

T-test for slope in regression



QUESTION: Is there a linear relationship?

PARAMETERS: The true slope β_1 as per the regression model $y = \beta_0 + \beta_1 x$.

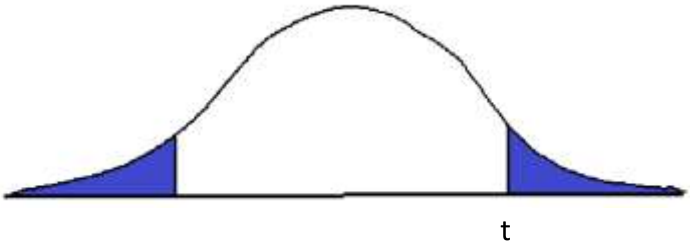
HYPOTHESES: $H_0: \beta_1 = 0$ and $H_a: \beta_1 \neq 0$

TEST STATISTIC:

$$T = \frac{b_1 - 0}{SE(b_1)}$$

DISTRIBUTION: $T \sim t(n-2)$

P-VALUE:



COMPUTER:

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	31.283	1.021		30.639	.000	29.206	33.360
	xvariable	3.444	.053	.996	65.522	.000	3.337	3.551

a. Dependent Variable: yvariable

÷ Test statistic P-value Confidence interval

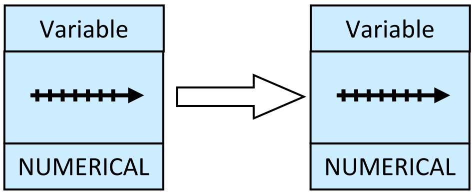
ASSUMPTIONS:

- Normality of errors – check with histogram of residuals
- Linearity – check with residual plot (random scatter)
- Homoskedasticity – check with residual plot (even spread)
- Independence between individuals – check with story

CONFIDENCE INTERVAL:

Confidence interval for slope: endpoints $= b_1 \pm t^* SE(b_1)$

Confidence intervals and prediction intervals in regression



<u>QUESTION:</u> What is the mean y-value at this particular x-value?

<u>CONFIDENCE INTERVAL FORMULA:</u>

$$\text{Confidence interval for mean y at particular } x_0\text{: endpoints} = b_0 + b_1 x_0 \pm t^* s\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}}$$

<u>QUESTION:</u> What could an individual's y-value be at this particular x-value?

<u>CONFIDENCE INTERVAL FORMULA:</u>

$$\text{Prediction interval for individual y at particular } x_0\text{: endpoints} = b_0 + b_1 x_0 \pm t^* s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}}$$

<u>COMPUTER:</u>

M for mean I for individual

	xvariable	yvariable	PRE_1	LMCI_1	UMCI_1	LICI_1	UICI_1
35	19.00	99.50	96.71644	96.23009	97.20280	93.79919	99.63369
36	25.00	.	117.37970	116.56490	118.19449	114.39010	120.36929

Confidence interval for mean Prediction interval for mean

<u>NOTE:</u>

- The prediction interval is wider
- Both intervals are wider the farther x_0 is from $\bar{x}$

INTERPRETING STATISTICS

FILL IN THE BLANK WITH THE WORDS AND NUMBERS IN CONTEXT

SIGNIFICANCE TEST RESULT (REJECT)

Since the p-value is less than [significance level], we reject the null hypothesis. So there is evidence to suggest that [meaning of alternative hypothesis in context].

eg: Since the p-value is less than 0.05, we reject the null hypothesis. So there is evidence to suggest that the mean temperature is different after meals with chilli and without.

SIGNIFICANCE TEST RESULT (DO NOT REJECT)

Since the p-value is greater than [significance level], we do not reject the null hypothesis. So there is not enough evidence to suggest that [meaning of alternative hypothesis in context].

eg: Since the p-value is greater than 0.05, we do not reject the null hypothesis. So there is not enough evidence to suggest that there is an association between gender of author and genre of a novel.

CONFIDENCE INTERVAL

We are [confidence level] % confident that the [parameter] for [population] is between [lower #] [units] and [upper #] [units].

eg: We are 95% confident that the true mean salt content for all chiko rolls is between 20.1 grams and 30.5 grams.

eg: We are 90 % confident that the true proportion of left-handed people for students studying creative arts is between 0.15 and 0.37.

SLOPE IN REGRESSION

An increase of one [unit] in [x variable] corresponds to an [increase/decrease] of [slope] [units] in [y variable] on average.

eg: An increase of one degree Celsius in the internal temperature corresponds to an increase of 1023 mites per square metre in carpet dust-mite concentration on average.

CONFIDENCE INTERVAL FOR SLOPE IN REGRESSION

We are confidence level % confident that an increase of one unit in x variable corresponds to an increase/decrease of between lower # units and upper # units in y variable on average.

eg: We are 95% confident that an increase of one hour of sleep per week corresponds to an increase of between 5.2 percentage points and 10.2 percentage points on the final exam on average.

PREDICTION INTERVAL IN REGRESSION

We predict that prediction level % of all things with an x variable of x_0 units will have a y variable between lower # units and upper # units.

eg: We predict that 90% of all adults on this diet program with a dosage of psyllium husk per day of 75 grams will have a recorded decrease in cholesterol of 3.1 % to 4.8 %.

COEFFICIENT OF DETERMINATION IN REGRESSION

Approximately R^2 % of the variation in y variable is explained by the linear relationship with x variable.

eg: Approximately 26.2% of the variation in IQ test score is be explained by the linear relationship with hours of video game use per week.

USING A GRAPH TO CHECK AN ASSUMPTION

In the name of graph, we see that some aspect appears to some description and so the assumption of name of assumption is reasonable/not reasonable.

eg: In the residual plot, we see that the scatter about the zero line appears to have a strongly curved pattern so the assumption of linearity is not reasonable.

eg: In the residual plot, we see that the spread about the zero line appears to be roughly the same along the whole line so the assumption of homoscedasticity is reasonable.

eg: In the histogram, we see that the shape of the distribution appears to be roughly bell-shaped, so the assumption of normality is reasonable.