Stats Review
Chapter 14

Revised 8/16

Note:

This review is meant to highlight basic concepts from the course. It does not cover all concepts presented by your instructor. Refer back to your notes, unit objectives, handouts, etc. to further prepare for your exam.

The questions are displayed on one slide followed by the answers are displayed in red on the next.

This review is available in alternate formats upon request.

Standard Error of Estimate

Below is a table of the population of St. Cloud from 1970-2010. Find the standard error of estimate of the following data where the years represent the years after 1970, that is 1970=0.

Year	0	10	20	30	40
Population	39691	42566	48812	59107	65842

Standard Error of Estimate

Below is a table of the population of St. Cloud from 1970-2010. Find the standard error of estimate of the following data where the years represent the years after 1970, that is 1970=0.

Year	0	10	20	30	40
Population	39691	42566	48812	59107	65842

We need to use the formula $s_e = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{\sum residuals^2}{n-2}}$

Step 1: Find the least-squares regression line (see chapter 4 for help). For this data the least-squares regression line is $\hat{y} = 37435 + 688.4x$

Step 2: Obtain the predicted value for each year ($\hat{y}$) using the least squares- regression line

Step 3i: Calculate the residuals $(y_i - \hat{y}_i)$

Step 3ii: Calculate the residuals squared, $(y_i - \hat{y}_i)^2$

Year	Population (y_i)	Step 2: $\hat{y}$	Step 3i: $(y_i - \hat{y}_i)$	Step 3ii: $(y_i - \hat{y}_i)^2$
0	39691	$37435 + 688.4(0) = 37435$	(39691-37435)=2256	$(2256)^2$ =5089536
10	42566	44319	-1753	3073009
20	48812	51203	-2391	5716881
30	59107	58087	1020	1040400
40	65842	64971	871	758641

Step 4: Add up the residuals squares, $\sum residuals^2$= 15678467

Step 5: Put it into the formula $s_e = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{15678467}{5-2}}$=**2286.08**

Sample Standard Deviation of $\boldsymbol{\beta_1}$, $\boldsymbol{s_{\beta_1}}$

Find the sample standard deviation of β_1 from the previous problem's data.

Sample Standard Deviation of $\boldsymbol{\beta_1}$, $\boldsymbol{s_{\beta_1}}$

Find the sample standard deviation of β_1 from the previous problem's data.

Option 1:

Step 1: Find the sample standard deviation of the x values (years), s_x.

Using technology (see chapter 3 on doing it by hand) we get s_x=15.811

Step 2: Find s_{β_1} using the formula $s_{\beta_1} = \frac{s_e}{\sqrt{n-1} \cdot s_x}$

s_e was found in the last problem

$s_{\beta_1} = \frac{s_e}{\sqrt{n-1} \cdot s_x} = \frac{2286.08}{\sqrt{5-1} \cdot 15.811} = \mathbf{72.29}$

Option 2:

Step 1: fill out the table below

$\bar{x}\ (mean)$=20

Step 2: Find the sum of the final column and take the square root.

400+100+0+100+400=1000, $\sqrt{1000} = 31.623$. This means $\sqrt{\sum(x_i - \bar{x})^2}$=31.623

Step 3: Put into the formula $s_{\beta_1} = \frac{s_e}{\sqrt{\sum(x_i - \bar{x})^2}}$

s_e was found in the last problem

$s_{\beta_1} = \frac{s_e}{\sqrt{\sum(x_i - \bar{x})^2}} = \frac{2286.08}{31.623} = \mathbf{72.79}$

Year, $\boldsymbol{x_i}$	$\boldsymbol{\bar{x}}$	$(\boldsymbol{x_i - \bar{x}_i})$	$(\boldsymbol{x_i - \bar{x}_i})^2$
0	20	0-20=-20	$(-20)^2$=400
10	20	-10	100
20	20	0	0
30	20	10	100
40	20	20	400

Testing the Significance of the Least-Squares Regression Model

Below are is a table of the population of St. Cloud from 1970-2010. Find the standard error of estimate of the following data where the years represent the years after 1970, that is 1970=0. Assuming the residuals are normally distributed and the residuals are normally distributed with constant error variance, test whether a linear relationship exists between year and population with α=.05 significance level. Also find the confidence interval.

Year	0	10	20	30	40
Population	39691	42566	48812	59107	65842

Testing the Significance of the Least-Squares Regression Model

Below are is a table of the population of St. Cloud from 1970-2010. Find the standard error of estimate of the following data where the years represent the years after 1970, that is 1970=0. Assuming the residuals are normally distributed and the residuals are normally distributed with constant error variance, test whether a linear relationship exists between year and population with α=.05 significance level. Also find the confidence interval.

Year	0	10	20	30	40
Population	39691	42566	48812	59107	65842

Step 1: Determine the null and alternative hypotheses.

$H_0: \beta_1 = 0$ (no linear relationship)

$H_1: \beta_1 \neq 0$ (linear relationship)

It is a two-tailed test because we are seeing if there is a difference.

Step 2: Select the level of significance.

α=.05

Step 3: Compute the test statistic $t_0 = \frac{b_1}{s_{b_1}}$

From the least-squares regression line we have b_1=688.43 (the slope of the regression line) and from slide 21, s_{β_1}= 55.998.

$t_0 = \frac{b_1}{s_{b_1}} = \frac{688.43}{72.29} = 9.52$ with n-2=5-2=3 degrees of freedom

Step 4: Using the p-value approach, reject H_0 if the p-value<α

Using technology, we have a p-value of 0.0025.

.0025<.05, so reject H_0

Therefore there is a linear relationship between year and population exists.

Confidence Intervals for the Slope of the Regression Line

Using the same data as before, find the 95% confidence interval.

Confidence Intervals for the Slope of the Regression Line

Using the same data as before, find the 95% confidence interval.

Step 1: Find b_1

From before, $\hat{y} = 37435 + 688.4x$

Step 2: Verify the conditions (see page 687)

Conditions are met

Step 3: Determine the critical value, $t_{\alpha/2}$ See chapter 9

$t_{.05/2} = t_{.025}$ =3.18 with n-2=5-2=3 degrees of freedom

Step 4: Compute the confidence interval using $b_1 \pm t_{\alpha/2} \cdot \frac{s_e}{\sqrt{\sum(x_i - \bar{x})^2}}$

$688.43 \pm 3.18 \cdot \frac{2286.08}{31.623}$

688.43 ± 229.888

(458.36,918.5)

Confidence Intervals for a Mean Response

Researchers examined the mean surface temperature of a body of salt water and compared it to the mean coral growth (in mm/year). The data is listed below. Construct a 99% confidence interval for the predicted mean of coral reef growth whose temperature is 29.9. The least squares regression line is $\hat{y}$= 11.7159 - 0.3036x.

Surface Temp	29.7	29.9	30.2	30.2	30.5	30.7	30.9
Growth	2.63	2.58	2.68	2.6	2.48	2.38	2.26

Confidence Intervals for a Mean Response

Researchers examined the mean surface temperature of a body of salt water and compared it to the mean coral growth (in mm/year). The data is listed below. Construct a 99% confidence interval for the predicted mean of coral reef growth whose temperature is 29.9. The least squares regression line is $\hat{y}$= 11.7159 - 0.3036x.

Surface Temp	29.7	29.9	30.2	30.2	30.5	30.7	30.9
Growth	2.63	2.58	2.68	2.6	2.48	2.38	2.26

Step 1: Find s_e and $\sum(x_i - \bar{x})^2$ using technology or step shown in the previous slides. Also find $\hat{y}$ for the given value.

s_e=0.0836

$\sum(x_i - \bar{x})^2$=1.1 (using 30.3 as $\bar{x}$)

$\hat{y}$ =11.7159-.3036(29.9)=2.638

Step 2: Determine the critical value, $t_{\alpha/2}$ See chapter 9

$t_{.01/2} = t_{.005}$ =4.032 with n-2=7-2=5 degrees of freedom

Step 3: Put into formula $\hat{y} \pm t_{\frac{\alpha}{2}} \cdot s_e \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}}$ where x^* is the given value, 29.9

$$2.638 \pm 4.032 \cdot .083 \sqrt{\frac{1}{7} + \frac{(29.9 - 30.3)^2}{1.1}}$$

$$2.638 \pm .1797$$

$$\mathbf{(2.4583, 2.8177)}$$

With Technology the answer is (2.456139, 2.8181986)

Prediction Interval for an Individual Response about $\hat{y}$

Using the coral growth data from the previous slides, construct a 99% prediction interval for the predicted growth at a temperature of 29.9.

Prediction Interval for an Individual Response about $\hat{y}$

Using the coral growth data from the previous slides, construct a 99% prediction interval for the predicted growth at a temperature of 29.9.

From the previous problem we know

s_e=0.0836

$\sum(x_i - \bar{x})^2$=1.1

$\hat{y}$ =11.7159-.3036(29.9)=2.638

$t_{\alpha/2} = t_{.01/2} = t_{.005}$ =4.032 with n-2=7-2=5 degrees of freedom

Putting this into the formula $\hat{y} \pm t_{\frac{\alpha}{2}} \cdot s_e\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}}$

$2.638 \pm 4.032 \cdot .083\sqrt{1 + \frac{1}{7} + \frac{(29.9-30.3)^2}{1.1}}$

$2.638 \pm .3798$

(2.2582,3.0178)

With Technology the answer is (2.2544944, 3.0198433)

Confidence and Prediction Intervals

What is the difference between the prediction made in slide 12 and slide 14?

Confidence and Prediction Intervals

What is the difference between the prediction made in slide 12 and slide 14?

The confidence interval on slide 27 is an estimate on the mean coral growth for all bodies of salt water that have a temperature of 29.9. The prediction interval made on slide 29 is an estimate of the coral growth of one body of water whose temperature is 29.9.

Remember confidence interval is for all while a prediction interval is for an individual.