Homework #1 Solutions

Reading:

- Define deterministic and probabilistic mathematical models. Give an example of each.

In deterministic models, the value Y is entirely determined by the value of X. In probabilistic models, there is assumed to be another random variable that contributes to the observed value of Y; Y is not solely determined by X.

Example of determ.: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

Example of prob.: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \epsilon_i$

- Write the general equation for a simple linear regression model.

$$E[Y] = \beta_0 + \beta_1 X \quad \text{or, equivalently,}$$

$$Y = \beta_0 + \beta_1 X + \epsilon$$

- Describe, in your own words, the overall concept of the method of least squares.

Answers may vary. Essentially, a method for estimating the intercept and slope values of a line that has minimized the sum of square distances between points and the line.

- State the least-squares estimators for the simple linear regression model.

$$\hat{\beta}_0 : \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 : \frac{\sum(x_i - \bar{x}) \sum(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{\text{cov}(x,y)}{\text{var}(x)}$$

- State the means and variances of the least-squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ in simple linear regression.

$$E[\hat{\beta}_0] = \beta_0 \; ; \; E[\hat{\beta}_1] = \beta_1$$

$$\text{var}(\hat{\beta}_0) = \frac{\sum x_i^2 \sigma^2}{n S_{xx}}$$

$$\text{var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$$

- State a pair of null and alternative hypotheses for making inferences about single regression parameters & linear functions of the parameters

$$H_0: \beta_i = \beta_{i0} \text{ \& } H_a: \beta_i \neq \beta_{i0}$$

Let $\theta = a_0\beta_0 + a_1\beta_1$.

$$H_0: \theta = \theta_0 \text{ ; } H_a: \theta \neq \theta_0 .$$

Practice:

1. a. Fit the model to these data, using least squares.

$$\bar{X} = \frac{\sum X_i}{10} = \frac{720}{10} = 72$$

$$\bar{Y} = \frac{\sum Y_i}{10} = \frac{721}{10} = 72.1$$

$X_i - \bar{X}$	
-62	-36
-60	169
-63	-13
-45	95
-25	
40	

$Y_i - \bar{Y}$	
-63.1	36.9
-58.1	-32.1
-65.1	165.9
-43.1	-12.1
-27.1	97.9

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = 54243$$

$$\sum (X_i - \bar{X})^2 = 54714$$

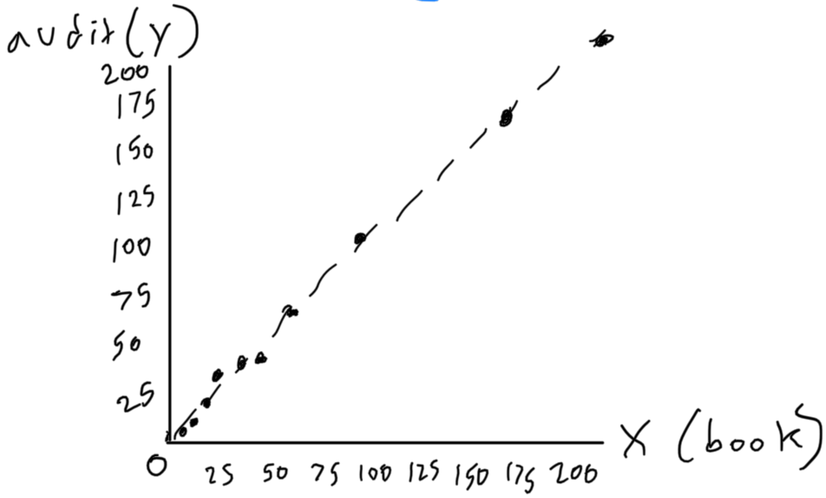
$$\hat{\beta}_1 = \frac{54243}{54714} = \boxed{0.99}$$

$$\hat{\beta}_0 = 72.1 - 0.99(72)$$

$$\hat{\beta}_0 = \boxed{0.72}$$

$$\boxed{Y = 0.72 + 0.99X + \epsilon}$$

b. Plot the 10 data points and graph the line representing the model.



c. Calculate SSE and s^2.

$$SSE = S_{yy} - \hat{\beta}_1 S_{xy} = 53832.9 - (0.9913916)(5424$$

$$S_{yy} = \sum (y_i - \bar{y})^2 = 53832.9$$

$$\boxed{= 56.84544}$$

$$s^2 = \frac{SSE}{n-2} = \frac{56.84544}{8} = \boxed{7.10568}$$

d. Conduct a hypothesis test at the 5% significance level.

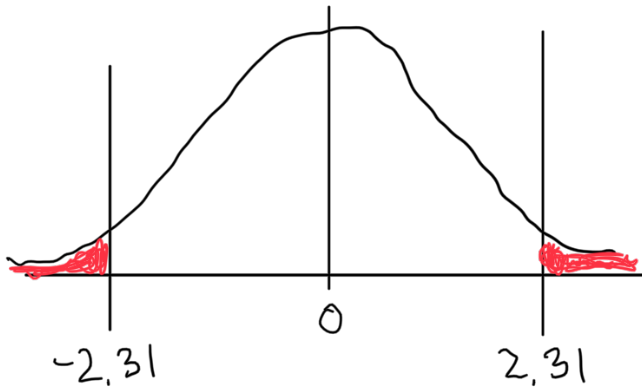
$$\boxed{H_0 : \beta_1 = 0 \quad H_a : \beta_1 \neq 0}$$

test statistic: $t = \dfrac{\beta_1 - \beta_{10}}{s\sqrt{1/S_{xx}}} \approx \dfrac{0.99 - 0}{2.665648(0.0043)}$

$s = \sqrt{7.10568} = 2.665648$

$\boxed{t = 86.9945}$

$\sqrt{\dfrac{1}{S_{xx}}} \approx \sqrt{\dfrac{1}{54714}} \approx \sqrt{1.83 \times 10^{-5}} = 0.0043$



$t_{\alpha/2} = t_{0.025}$ for 8df

$|86.9945| \in (2.31, \infty) \therefore 2\left(1 - P(t \le |86.9945|)\right) \approx$

$\boxed{\text{p-value} \approx 3.40 \times 10^{-13}}$

$p < \alpha$, so we <u>reject</u> the null hypothesis and conclude that the slope is not zero. This means there is a statistically significant relationship between book price & audit price.

e. What is your model's estimate for the expected change in audit value per one-unit change in book value?

About \$0.99.

1. Predict audited value for a book value of $100.

$$\hat{y} = 0.72 + 0.99(100)$$

$$\boxed{\hat{y} = \$99.72}$$

2. Show that $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ will always go through the point $(\bar{x}, \bar{y})$.

If $x = \bar{x}$, $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$

then, given that $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$:

$$\hat{y} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x}$$

$$\hat{y} = \bar{y}$$

3. What value for x is length of prediction interval minimized?

The prediction interval formula is:

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2} s \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}$$

when $x^* = \bar{x}$, it becomes

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2} s \sqrt{1 + \frac{1}{n}}$$

and is, therefore, minimized.