Alexander Martinez

Hw 1 571A Chapter 1

→1.2 $Y = 300 + 2x$. This is a functional relationship b/c the explanatory variables explain exactly the $$ dollar amount you will recieve based on the number of visits you have

→ 1.4 It is b/c Y_i is a random variable. It is the sum of $\beta_0 + \beta_1 X_i$ and the error term. The error term is the reason the response, Y_i, has variation and does not always turn out to be the same everytime

→ 1.9 False, its impossible to find a regression model of any lots w/o having any sort of data.

→ 1.20 a. The Model will be created in R

$\hat{Y} = -.058016 + 15.0352x$

d. 74.5958

→1.21

a. In R,
Model: $\hat{Y} = 10.2 + 4x$
According to the graph in R, it looks like a good fit

b. $\hat{Y} = 10.2 + 4x = 10.2 + 4(1) = 14.2$

c. $\hat{Y}(4) = 18.2 \Rightarrow 18.2 - 14.2 = 4$
The expected # of ampules to break when there is 2 transfers as compared to just 1 transfer on the route is 4 more

d. Again using R,
$\bar{X} = 1$, $\bar{Y} = 14.2$
Putting values in shows it goes through that value
$14.2 = 10.2 + 4(1)$

Scanned with CamScanner

→ 1.24

a. Refer to R code for the work

$\sum_{i=1}^{n} e_i^2 = 3416.377$

Q is just a minimization of the squared residuas.

$Min \sum e_i^2$

b. MSE → Unbiased estimator for σ^2

$MSE = \sum_{i=1}^{n} e^2 / n-2 = 3416.377 / 43 = \boxed{79.45063}$

$\sigma = \sqrt{MSE} = 8.913508$ minutes

→ 1.25

a. Residual: $16 - 14.2 = 1.8$

The Model Error term(ε_i) is Y_i minus the true regression which is unknown.

The Residual is $Y_i - \hat{Y}_i$, which is known by using the Fitted line.

b. In R,

$\sum_{i=1}^{n} e_i^2 = 17.6$, $MSE = \frac{17.6}{8} = 2.2$.

It estimates the average of the squared errors.

Chapter 2

→ 2.5 a. Using R, the $s\{b_1\} = .4831$

$t(.95; 43) = 1.68$

Conf int. : $[15.0352 \pm 1.68(.4831)]$

$\boxed{14.22359 \leq \beta_1 \leq 15.84681}$

I verified this using R, P.val. = .2e-16

b. $H_0: \beta_0 = 0$ $\quad t^* = \left|\frac{b_1 - b_0}{s\{b_1\}}\right| = \left|\frac{15.0352}{.4831}\right| = \boxed{31.12233}$

$H_a: \beta_0 \neq 0$

Reject Null b/c $|t^*| > 1.68$.

c. Yes they are b/c in the CI, Ø is not included, which means w/ 90% significance β_1 is not equal to Ø. The T-test concluded the same thing

d. $H_0: \beta_1 \leq 14$

$H_a: \beta_1 > 14$ $\quad t^* = \frac{15.0352 - 14}{.4831} = \boxed{2.142828}$

$t^* > 1.68$, so this means that a 95% significance level, it takes more than 14 minutes to service a copier.

→2.6 a. $s\{b_1\} = \frac{MSE}{\Sigma(x-\bar{x})^2}$, Using R

$s\{b_1\} = .46904$; $t(9.75; 8) = 2.31$

Conf int: $[4 \pm 2.31(.46904)]$

$\boxed{2.9165 \leq \beta_1 \leq 5.082}$

By 95% Confidence, the mean number of ampules that break increases between 2.917 and 5.082 every transfer of cartons

b. ON OTHER PAGE

b. $t^* = \frac{b_1}{s\{b_1\}}$ $t(.975, 8) = 2.31$

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

$t^* = \frac{4}{.46904} = 8.528$, the P. Val. = Very small

$t^* > 2.31$, Reject Null, Conclude that at 95% Sig level, the mean number of ampleles that break have linear association w/ the transfer of Cartons

c.
* $s\{b_0\} = \sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{\Sigma(x_i - \bar{x}^2)}\right)} = \sqrt{2.2\left(\frac{1}{10} + \frac{1^2}{10}\right)} = .663$

$= \boxed{.663}$

* $t(.975, 8) = 2.31$

$10.2 - 2.306(.663) = 8.6712$
$10.2 + 2.306(.663) = 11.7288$

* At 95% CI is $\boxed{8.6712 \leq \beta_0 \leq 11.7288}$

* At a 95% Conf. the mean # of broken ampules are between 8.6712 and 11.7288 w/ no transfers.

d.

Next PAGE

d. $H_0: \beta_0 \leq 9.0$ $\quad t^* = \dfrac{b_0 - \beta_0}{s\{b_0\}} = \dfrac{10.2 - 9}{.663} = 1.8099$

$H_a: \beta_0 > 9.0$

$t(.975, 8) = 2.31$

$t^* = 1.8099 < 2.306$, Do Not Reject Null

$.05 \leq p.val \leq .10$

This means $P.Val > \alpha$, and Don't Reject Null. This Null means that the mean # of ampules should not exceed 9.0 w/ no transfers at a 97.5% significance level.

e part b. $H_0: \beta_1 = \beta_{10} = 0$

$H_a: \beta_1 \neq \beta_{10} = 0$

$\beta_1 = 2.0$

$\delta = \dfrac{|2.0 - 0|}{.5} = 4$. The power value of $\delta = 4$ + 8 df = .94

* The power of the test that $\beta_1 = 2.0$ is .94

→2.14 a. $s\{\hat{Y}_h\} = \sqrt{MSE\left(\dfrac{1}{n} + \dfrac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)}$

Using R, the CI is:

$87.2838 \leq E\{Y_h\} \leq 91.978$

The Mean Response says that at a 90% Confidence interval, it will take a mean service time between 87.28 and 91.978 minutes total to service 6 copiers

b. $S\{Pred\} = \sqrt{MSE\left(1 + \frac{1}{n} + \frac{(X_{hnew} - \bar{x})^2}{\sum_{i=1}^{n}(X_i - \bar{x})^2}\right)}$

Using R, the prediction Interval is:

$74.46433 \leq \hat{Y}h(new) \leq 104.7983$

* Yes it is wider
* Yes it should be because it is trying to observe the next call, which may be more subject to more variation, while the CI is more centered around the mean.

d.

→2.15 a. Using R at $X = 2$,

$15.976 \leq E\{Yh\} \leq 20.42571$

Using R at $X = 4$,

$21.22316 \leq E\{Yh\} \leq 31.17684$

* The Mean Response says that at a 99% Confidence interval there will be a total of between 15.976 and 20.42571 Broken ampules in a 2 transfer route and between 21.22316 and 31.17684 total broken ampules in a 4 transfer route.

Scanned with CamScanner

b. Using R_1 the prediction intervals are

$12.74814 \leq Y_{h\,new} \leq 23.65186$

Based on the sample data there is a 99% Prediction Interval that if a $Y_{h\,new}$ came to be and X_h had 2 transfers represented, then the total # of broken ampules will be between $12.74814 \leq Y_{h\,new} \leq 23.65186$.

d. $WS = 2F(.99; 2, 8) = (8.649) \cdot 2 = 17.218$

$W = 4.159$

At $X=2$; $18.2 \pm 4.159(.663)$, $15.443 \leq \beta_0 + \beta_1 X \leq 20.957$

At $x=4$; $26.2 \pm 4.159(1.483)$, $20.032 \leq \beta_0 + \beta_1 X \leq 32.3$

→ 2.14.d (Accidentally skipped)

$\hat{Y} \pm W_\alpha s\{\hat{Y}_h\}$ where $W_\alpha = \sqrt{2F(1-\alpha; 2, n-2)}$

$W^2 = 2(F(.90, 2, 43)) = 2(2.4304) = \sqrt{4.8608} = W = 2.2047$

$89.613 \pm 2.2047(1.3964)$

$86.5527 \leq \beta_0 + \beta_1 X_h \leq 92.7099$.

It is wider and should be

2.17

The Test: $H_0: \beta_1 = 0$
$H_a: \beta_1 \neq 0$

If the p value > α, then Do Not Reject. This means the analyst had an α > PValue. If the α = .01, then the analyst would have concluded H_0 instead of H_a. Whether or not this is appropriate depends on the data and strategies used

2.18

The t test is mainly more versatile b/c it can be used for a one sided test.

The F test can test whether $\beta_1 = 0$, $\beta_1 \neq 0$. The t test can do the same and expand beyond that.

2.24

a. Table 2.2

Source of Variation	SS	df	MS	E{MS}
Regression	SSR = 76,960.4	1	MSR = 76,960.4	$\sigma^2 + \beta_1^2(340.44)$
Error	SSE = 3416.37	43	MSE = 79.4506	σ^2
Total	SSTO = 80,376.78.44			

Table 2.3

	SS	df	MS
regression	SSR = 76,940.4	1	76,940.4
Error	SSE = 3416.37	43	79.4506
Total	SSTO = 80376.78	44	
Correction for mean	261,747.2		
Total uncorrected	342,124	45	

Scanned with CamScanner

* The SS part is additive. Also SS divided by df gives the MS.
* The main way the 2 differ is extra information. The 2.3 table provides the total uncorrected and the correction for mean. The 2nd table leaves out the expected value of MS.

b. $H_0: \beta_1 = 0$ $F^* = \frac{MSR}{MSE} = \frac{76,966.4}{79.4406} = 968.77$

$H_a: \beta_1 \neq 0$

* If $F^* \le 2.82$, Do not Reject
$F(.90; 1, 43) = 2.82$ * If $F^* > 2.82$, Reject.

Reject Null, conclude H_a

c. * $\frac{SSR}{SSTO} = \frac{76,960.4}{80,376.78} = \boxed{.957495}$

* This is relatively large, but we must be careful b/c it just measures the % of variation the Y_i's explained by the Model. It is more used as a hint, rather than concrete evidence.
* This is $R^2 \rightarrow$ The coeff of determination

d. $r = \sqrt{R^2} = 97.8516(+)$

*2.25 a.

	SS	df	MS	E{MS}
regression	SSR = 160	1	160	$\sigma^2 + \beta_1^2(10)$
ERROR	SSE = 17.6	8	2.2	σ^2
Total	SSTO = 177.6	9		

Next Page

b. $H_0: \beta_1 = 0$ $\quad F(.95; 1, 8) = 5.317$
$H_a: \beta_1 \neq 0$ $\quad$ Conclude H_0 if $F^* \leq 5.317$
Conclude H_a if $F^* > 5.317$

$$F^* = \frac{MSR}{MSE} = \frac{160}{2.2} = \boxed{72.7272}$$

Conclude Ha b/c $F^* > F(.95; 1, 8)$

c. $t^* = \frac{4-0}{.469} = 8.529$ (From 2.6.b.)

$$\boxed{(t^*)^2 = (8.529)^2 = 72.74 = F^*}$$

d. $R^2 = \frac{SSR}{SSTO} = \frac{160}{177.6} = .9009 \rightarrow$ 90% of variation is accounted for when introducing X

$r = \sqrt{R^2} = .949158(+)$

→ 2.42

a. The plot was made in R, I will upload screenshot
Yes, it does look appropriate

b. Using R, $r_{12} = .95284$, ρ_{12}
ρ_{12} is the coeff of correlation between Y_1 and Y_2. Thus, .95284, indicates strong linear assoc. between Y_1 and Y_2.

c.
$$t^* = \frac{r_{12}\sqrt{n-2}}{\sqrt{1-r_{12}^2}} = \frac{.95284\sqrt{13}}{\sqrt{1-(.95284)^2}} = \frac{3.434}{.3034} = \boxed{11.318}$$

$H_0: \rho_{12} = 0$ $\quad$ if $|t^*| \leq 2.87$
$H_a: \rho_{12} \neq 0$ $\quad$ if $|t^*| > 2.87$

$|t^*| > 2.87$, so Reject Null, conclude H_a that w/ 99% significance, Y_1 and Y_2 are not independent

d. No, b/c you are testing at an exact correlation. This is not used often and there is a specific method for it

2.46

a. Using R, we can find spearman's coeff

$r_s = .9454874$

b. H_0: No Y_1 vs. Y_2 association
H_a: Some Y_1 vs. Y_2 association

Use R for this too,
P-value $= 1.036 \times 10^{-7} < \alpha = .01$, So Reject H_0.

c. They both have shown at a significantly high level (99%) that there is lin. Assoc. between Y_1 and Y_2

Scanned with CamScanner