

Contents

1	Leatherman Homework #2 (2019/09/21)	1
1.1	10 (p91)	1
1.1.1	a	1
1.1.2	b	1
1.1.3	c	2
1.2	12 (p91)	2
1.3	Using the Grade Point Average Dataset	2
1.3.1	4 (p90)	2
1.3.2	13 (p91)	3
1.3.3	23 (p93)	4
1.4	Using the Plastic Hardness Dataset	4
1.4.1	16	4
1.4.2	26	5

1 Leatherman Homework #2 (2019/09/21)

Do the following exercises from the Applied Linear Regression Models textbook. Each individual part (a, b, c, etc.) is worth 6.5 points. p. 91 – #2.10, #2.12 p. 90-93, #2.4, #2.13abc, #2.23abc p. 92-93, #2.16ab, #2.26ab

1.1 10 (p91)

Explain whether a confidence or prediction interval is appropriate

1.1.1 a

What will be the humidity level in this greenhouse tomorrow when we set the temperature at 31° C?

A prediction interval for a new observation is appropriate since the point of interest is a single observation.

1.1.2 b

How much do families whose disposable income is \$23,500 spend, on the average, for meals away from home?

A confidence interval for a mean response is appropriate since the point of interest is an average.

1.1.3 c

/How many kw-hours of electricity will be consumed next month by commercial and industrial users in the Twin Cities service area, give that the index of usiness activity for the area remains at its present level?/

A prediction interval for a new observation is appropriate since the point of interest is a single observation.

1.2 12 (p91)

/Can $\sigma^2(pred)$ in 2.37 be brought increasingly close to 0 as n becomes large? Is this also the case for $\sigma^2(Y_h)$ in 22.9b? What is the implication of this difference?/

As n becomes larger, $Var(pred)$ approaches: $Var(pred) = \sigma^2 + (X_h - \bar{x})^2$
This is not equivalent to 0.

As n increases, $Var(\hat{Y}_h^2) = (X_h - \bar{X})^2$

This means that as sample sizes grow, the variance of a prediction is equal to the squared distance distance of a given predictor from its average. When $X_h = \bar{x}$, the variance of the prediction interval is equal to the variance of Y_i .

1.3 Using the Grade Point Average Dataset

1.3.1 4 (p90)

1. a *Obtain a 99% C.I. for β_1 . Interpret. Does it include zero? Why might the director of admissions be interested in whtether the C.I. includes zero?*

```
gpa <- readxl::read_excel("~/snap/firefox/common/Downloads/GradePointAverage.xlsx")
gpa.model1 <- lm(GPA ~ ACT, data = gpa)
confint(gpa.model1, level = 0.99)
# [0.0054, 0.0723]
```

With 99% confidence, each point scored on the ACT for a class of incoming freshmen is associated with a mean increase in GPA between 0.0054 and 0.0723.

This confidence interval does not include zero which indicates that there is convincing evidence that there is a relationship between a student's GPA and their ACT score.

2. b /Test, using the test statistic t^* , whether or not a linear association exists between student's ACT score and GPA at the end of the freshman year Y. Use a level of significance of 0.01. State the alternatives, decision rule, and conclusion./

```
summary(gpa.model1)
```

```
Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 2.11405
0.32089 6.588 1.3e-09 * ACT 0.03883 0.01277 3.040 0.00292 **
```

$H_0: \beta_1 = 0$

$H_A: \beta_1 \neq 0$

There is convincing evidence that a linear association exists ($\beta_1 \neq 0$) between a student's ACT score and their GPA at the end of their freshman year (p-value = 0.00292, two-sided t-test).

3. c *What is the p-value of your test in part b? How does it support the conclusion reached in part b?*

The p-value is 0.00292. This is less than the significance level which indicates that there is significant evidence that a relationship exists between GPA at the end of freshman year and ACT score.

1.3.2 13 (p91)

1. a *Obtain a 95% C.I. for students whose act test score is 28. Interpret*

```
predict(gpa.model1, data.frame(ACT=c(28)), interval = "confidence")
#           fit           lwr           upr
# 1 3.201209 3.061384 3.341033
```

It is estimated that a student with an ACT score of 28 will have a GPA of 3.2 at the end of their freshman year. With 95% confidence, an ACT score of 28 is associated with an average GPA between 3.0613 and 3.341 at the end of their freshman year.

2. b *Mary Jones obtained a score of 28 on the entrance test. Predict her freshman GPA using a 95% P.I. Interpret.*

```
predict(gpa.model1, data.frame(ACT=c(28)), interval = "prediction")
#           fit           lwr           upr
# 1 3.201209 1.959355 4.443063
```

Given that Mary Jones has an ACT score of 28, it is estimated that Mary Jones will have a 3.2 GPA at the end of her freshman year. With 95% confidence, her GPA at the end of freshman year will be between 1.9593 and 4.443.

3. c As expected, the prediction interval is wider than the confidence interval. This should always be the case.

1.3.3 23 (p93)

1. a

```
anova(gpa.model1)
```

Source	SS	df	MS	F Statistic
Regression	3.588	1	3.5878	9.2402
Error	45.818	118	0.3883	
Total	49.406	119		

2. b The estimated MSR is 3.588. The MSE is 0.3883. When $n = 3$ and $\hat{Y}_i = \bar{y}$, they will estimate the same quantity.

3. c

$$H_0: \beta_1 = 0$$

$$H_A: \beta_1 \neq 0$$

The F statistic is 9.2402 which corresponds to a p-value of 0.002917. There is convincing evidence that there is a relationship between ACT and GPA at the end of freshman year.

1.4 Using the Plastic Hardness Dataset

1.4.1 16

1. a Obtain a 98% C.I. for the mean hardness of molded items with an elapse time of 30 hours. Interpret.

```
plastic <- readxl::read_excel("~/Downloads/PlasticHardness.xlsx")
plastic.model1 <- lm(Hardness ~ Hours, data = plastic)
predict(plastic.model1, newdata = data.frame(Hours=c(30)), interval = "confidence")
```

```
fit lwr upr 1 229.6312 227.4569 231.8056
```

It is estimated that the average hardness for a molded item that has aged for 30 hours is 229.6312. With 98% confidence, a curation time of 30 hours is associated with an average hardness of between 227.4569 and 231.8056.

2. b *Obtain a 98% P.I. for the hardness of a newly molded test item with an elapsed time of 30 hours.*

```
predict(plastic.model1, newdata = data.frame(Hours=c(30)), interval = "prediction")
```

```
fit lwr upr 1 229.6312 220.8695 238.393
```

With 98% confidence, a random mold that has curated for 30 hours will have a hardness between 220.8695 and 238.393.

1.4.2 26

1. a

```
anova(plastic.model1)
```

Source	SS	df	MS	F Statistic
Regression	5297.5	1	5297.5	506.51
Error	146.4	14	10.5	
Total	5443.9	15		

2. b

$$H_0: \beta_1 = 0$$

$$H_A: \beta_1 \neq 0$$

The F statistic is 506.51 which corresponds to a p-value of 2.159e-12. There is convincing evidence that there is a relationship between Hours cured and Hardness.