HW8

Luke Geel

3/31/2021

2.10. For each of the following questions, explain whether a confidence interval for a mean response or a prediction interval for a new observation is appropriate. a. What will be the humidity level in this greenhouse tomorrow when we set the temperature level at 31°C? Prediction interval for new observation because we are trying to find the interval for a future time. b. How much do families whose disposable income is $23,500 spend, on the average, for meals away from home? Confidence interval for mean because we are trying to find the interval for the mean of existing observations. c. How many kilowatt-hours of electricity will be consumed next month by commercial and industrial users in the Twin Cities service area, given that the index of business activity for the area remains at its present level? Prediction interval for new observation because we are trying to find the interval for a future time.

2.13. to Grade point average Problem 1.19. a. Obtain a 95 percent interval estimate of the mean freshman GPA for students whose ACT test score is 28. Interpret your confidence interval. Confidence interval: (3.069, 3.458) We can be 95% confident that the true mean GPA for students whose ACT test score is 28 is between 3.069 and 3.458. b. Mary Jones obtained a score of 28 on the entrance test. Predict her freshman GPA-using a 95 percent prediction interval. Interpret your prediction interval. Prediction interval: (1.536,4.991) We can be 95% confident that Mary Jones' GPA will be between 1.536 and 4.991. c. Is the prediction interval in part (b) wider than the confidence interval in part (a)? Should it be? Yes, the prediction interval is wider than the confidence interval. The prediction interval is wider because it describes the value for a random variable and must have a wider interval to allow for non-parameterized variables to impact the predicted value.

```r
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------------

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
loadRData <- function(fileName){
  load(fileName)
  get(ls()[ls() != "fileName"])
}
gpa <- loadRData("/Users/lukegeel/Downloads/gpa_spring2021.RData")
GPA <- gpa$Y
ACT <- gpa$X
```

```r
GPA.lm <- lm(GPA~ACT)
freshman.gpa <- data.frame(ACT=28)
gpa.confidence.int <- predict(GPA.lm, freshman.gpa, interval = "confidence", level = 0.95, se.fit = TRUE
gpa.confidence.int
```

```
## $fit
##        fit     lwr      upr
## 1 3.263842 3.06938 3.458304
##
## $se.fit
## [1] 0.09819948
##
## $df
## [1] 118
##
## $residual.scale
## [1] 0.8666146
```

```r
#B
library(tidyverse)
loadRData <- function(fileName){
  load(fileName)
  get(ls()[ls() != "fileName"])
}
gpa <- loadRData("/Users/lukegeel/Downloads/gpa_spring2021.RData")
GPA <- gpa$Y
ACT <- gpa$X
GPA.lm <- lm(GPA~ACT)
freshman.gpa <- data.frame(ACT=28)
gpa.prediction.int <- predict(GPA.lm, freshman.gpa, interval = "prediction", level = 0.95, se.fit = TRUE
gpa.prediction.int
```

```
## $fit
##        fit      lwr      upr
## 1 3.263842 1.536727 4.990957
##
## $se.fit
## [1] 0.09819948
##
## $df
## [1] 118
##
## $residual.scale
## [1] 0.8666146
```

*2.14. Refer to Copier maintenance Problem 1.20. a. Obtain a 90 percent confidence interval for the mean gervice time on calls in which six copiers are serviced. Interpret your confidence interval. Confidence interval: (86.472,92.427) We can be 90% confident that the mean service time on calls in which six copiers are serviced is between 86.472 and 92.427 minutes. b. Obtain a 90 percent prediction interval for the service time on the next call in which six copiers are serviced. Is your prediction interval wider than the corresponding confidence interval in part (a)? Should it be? Prediction Interval: (70.213,108.686) Yes, the prediction interval is wider than the corresponding confidence interval which is what it should be. c. Management wishes to estimate the expected service time per copier on calls in which six copiers are serviced. Obtain an appropriate 90 percent confidence interval by converting the interval obtained in part (a). Interpret the converted confidence interval. Converted confidence interval: (14.41204, 15.40449) We can be 90% confident that the true service time per copier on calls in which six copiers are serviced is between 14.4 and 15.4 minutes.

```r
#A
library(tidyverse)
loadRData <- function(fileName){
  load(fileName)
  get(ls()[ls() != "fileName"])
}
copier <- loadRData("/Users/lukegeel/Downloads/copier_spring2021.RData")
Time <- copier$Y
Number <- copier$X
copier.lm <- lm(Time~Number)
six.copier <- data.frame(Number=6)
copier.confidence.int <- predict(copier.lm, six.copier, interval = "confidence", level = 0.90, se.fit =
copier.confidence.int
```

```
## $fit
##        fit      lwr      upr
## 1 89.44961 86.47226 92.42695
##
## $se.fit
## [1] 1.771101
##
## $df
## [1] 43
##
## $residual.scale
## [1] 11.30522
```

```r
#B
library(tidyverse)
loadRData <- function(fileName){
  load(fileName)
  get(ls()[ls() != "fileName"])
}
copier <- loadRData("/Users/lukegeel/Downloads/copier_spring2021.RData")
Time <- copier$Y
Number <- copier$X
copier.lm <- lm(Time~Number)
six.copier <- data.frame(Number=6)
copier.confidence.int <- predict(copier.lm, six.copier, interval = "prediction", level = 0.90, se.fit =
copier.confidence.int
```

```
## $fit
##        fit      lwr      upr
## 1 89.44961 70.21294 108.6863
##
## $se.fit
## [1] 1.771101
##
```

```
## $df
## [1] 43
##
## $residual.scale
## [1] 11.30522
```

```
#C
library(tidyverse)
loadRData <- function(fileName){
  load(fileName)
  get(ls()[ls() != "fileName"])
}
copier <- loadRData("/Users/lukegeel/Downloads/copier_spring2021.RData")
Time <- copier$Y
Number <- copier$X
copier.lm <- lm(Time~Number)
six.copier <- data.frame(Number=6)
copier.confidence.int <- predict(copier.lm, six.copier, interval = "confidence", level = 0.90, se.fit =
copier.confidence.int
```

```
## $fit
##        fit      lwr      upr
## 1 89.44961 86.47226 92.42695
##
## $se.fit
## [1] 1.771101
##
## $df
## [1] 43
##
## $residual.scale
## [1] 11.30522
```

```
lower <-  86.47226
upper <-  92.42695
lower/6
```

```
## [1] 14.41204
```

```
upper/6
```

```
## [1] 15.40449
```

*2.27. Refer to Muscle mass Problem 1.27. a. Conduct a test to decide whether or not there is a negative linear association between amount of muscle mass and age. Control the risk of Type I error at .05. State the alternatives, decision rule, and conclusion. What is the P-value of the test? The hypotheses are: H0: b1 = 0 vs. Ha: b1 < 0. Decision rule: Reject null hypothesis if t-statistic< t(0.95,58df) Fail to reject null hypothesis if t-statistic > t(0.95,58df) Conclusion: There is enough evidence to suggest that there is a negative linear association between amount of muscle mass and age. P-value: < 2.2e-16 (basically 0) b. The two-sided P-value for the test whether B0 = 0 is 0+. Can it now be concluded that b0 provides relevant information on the amount of muscle mass at birth for a female child? No. Even though the test of b0 is significant, b0 does not provide relevant information on the amount of muscle mass at birth for a female child because data was not collected in that region and comparing muscle mass of adults compared to children won't produce relevant information. c. Estimate with a 95 percent confidence interval the difference in expected muscle mass for women whose ages differ by one year. Why is it not necessary to know the specific ages to make this estimate? Confidence interval:(-0.696,-0.508) It is not necessary to know the specific ages because the confidence interval depends on the estimated slope of the regression equation, its standard error, and a t multiplier. All of these values don't change as x changes.

```
library(tidyverse)
loadRData <- function(fileName){
  load(fileName)
  get(ls()[ls() != "fileName"])
}
muscle <- loadRData("/Users/lukegeel/Downloads/muscle_spring2021.RData")
Age <- muscle$X
Mass <- muscle$Y
MuscMass.lm <- lm(Mass~Age)
summary(MuscMass.lm)
```

```
##
## Call:
## lm(formula = Mass ~ Age)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.5897  -3.0549  -0.2494   4.4592  27.2637
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 158.78889    5.85158   27.14   <2e-16 ***
## Age          -1.22932    0.09575  -12.84   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.676 on 58 degrees of freedom
## Multiple R-squared:  0.7397, Adjusted R-squared:  0.7352
## F-statistic: 164.8 on 1 and 58 DF,  p-value: < 2.2e-16
```

```
#C
library(tidyverse)
loadRData <- function(fileName){
  load(fileName)
  get(ls()[ls() != "fileName"])
}
muscle <- loadRData("/Users/lukegeel/Downloads/muscle_spring2021.RData")
Age <- muscle$X
Mass <- muscle$Y
n <- nrow(muscle)
linmod <- lm(Age~Mass)
b1 <- linmod$coef[2]
s.b1 <- summary(linmod)$coef[2, 2]
alpha <- 0.05
qt <- qt(alpha/2, n - 2)
pvalue <- pt(-abs(b1/s.b1),n-2)+(1-pt(abs(b1/s.b1),n-2))
```

```r
lower <- b1+s.b1*qt(alpha/2,n-2)
upper <- b1-s.b1*qt(alpha/2,n-2)
lower
```

```
##       Mass
## -0.6955469
```

```r
upper
```

```
##       Mass
## -0.5079154
```

2.28. Refer to Muscle mass Problem 1.27. a. Obtain a 95 percent confidence interval for the mean muscle mass for women of age 60. Interpret your confidence interval. Confidence interval: (82.787, 87.272) We can be 95% confidentthat the mean muscle mass for women of age 60 is between 82.787 and 87.272. b. Obtain a 95 percent prediction interval for the muscle mass of a woman whose age is 60. Is the prediction interval relatively precise? Prediction interval: (67.518,102.541) This interval is not very precise.

```r
#A
library(tidyverse)
loadRData <- function(fileName){
  load(fileName)
  get(ls()[ls() != "fileName"])
}
muscle <- loadRData("/Users/lukegeel/Downloads/muscle_spring2021.RData")
Age <- muscle$X
Mass <- muscle$Y
MuscMass.lm <- lm(Mass~Age)
womens.age <- data.frame(Age=60)
muscmass.confidence.int <- predict(MuscMass.lm, womens.age, interval = "confidence", level = 0.95, se.fi
muscmass.confidence.int
```

```
## $fit
##        fit      lwr      upr
## 1 85.02951 82.78738 87.27165
##
## $se.fit
## [1] 1.120106
##
## $df
## [1] 58
##
## $residual.scale
## [1] 8.676291
```

```r
#B
library(tidyverse)
loadRData <- function(fileName){
  load(fileName)
  get(ls()[ls() != "fileName"])
}
muscle <- loadRData("/Users/lukegeel/Downloads/muscle_spring2021.RData")
```

```r
Age <- muscle$X
Mass <- muscle$Y
MuscMass.lm <- lm(Mass~Age)
womens.age <- data.frame(Age=60)
muscmass.prediction.int <- predict(MuscMass.lm, womens.age, interval = "prediction", level = 0.95, se.fi
muscmass.prediction.int
```

```
## $fit
##        fit     lwr      upr
## 1 85.02951 67.5179 102.5411
##
## $se.fit
## [1] 1.120106
##
## $df
## [1] 58
##
## $residual.scale
## [1] 8.676291
```

2.66. Five observations on Y are to be taken when X = 4, 8, 12, 16, and 20, respectively. The true regression function is E(y} = 20 + 4X, and the Bi are independent N(O, 25). a. Generate five normal random numbers, with mean 0 and variance 25. Consider these random numbers as the errorterms for the five Y observations at X = 4,8, 12, 16, and 20 and calculate Y1, Y2, Y3, Y4 , and Y5. Obtain the least squares estimates b0 and b1, when fitting a straight line to the five cases. Also calculate Yh when Xh = 10 and obtain a 95 percent confidence interval for E(Yh) when Xh = 10. b0 = 17.264 b1 = 4.081 yh =58.076 Confidence interval: (164.5145, 359.753) b. Repeat part (a) 200 times, generating new random numbers each time. c. Make a frequency distribution of the 200 estimates hI. Calculate the mean and standard deviation of the 200 estimates hI. Are the results consistent with theoretical expectations? Mean of b1s = 3.99 which is consistent because b1 = 4 Standard deviation of b1s = 0.399 which is consistent because sd(b1) = 0.395. d. What proportion of the 200 confidence intervals for E(Yh) when Xh = 10 include E(Yh)? Is this result consistent with theoretical expectations? The proportion of the 200 confidence intervals for E(Yh) when Xh = 10 include E(Yh) is 1 which means that it is consistent with theoretical expectations.

```r
#A
set.seed(1234)
n <- 5
sig <- 5
X <- c(4, 8, 12, 16, 20)
epsilon <- rnorm(n, mean = 0, sd = sig)
Y <- 20 + 4*X + epsilon
linmod <- lm(Y~X)
b0 <- linmod$coef[1]
b1 <- linmod$coef[2]
Y.hat.h <- b0 + 10*b1
b0
```

```
## (Intercept)
##    17.26435
```

```r
b1
```

```
##        X
## 4.081157
```

```
Y.hat.h
```

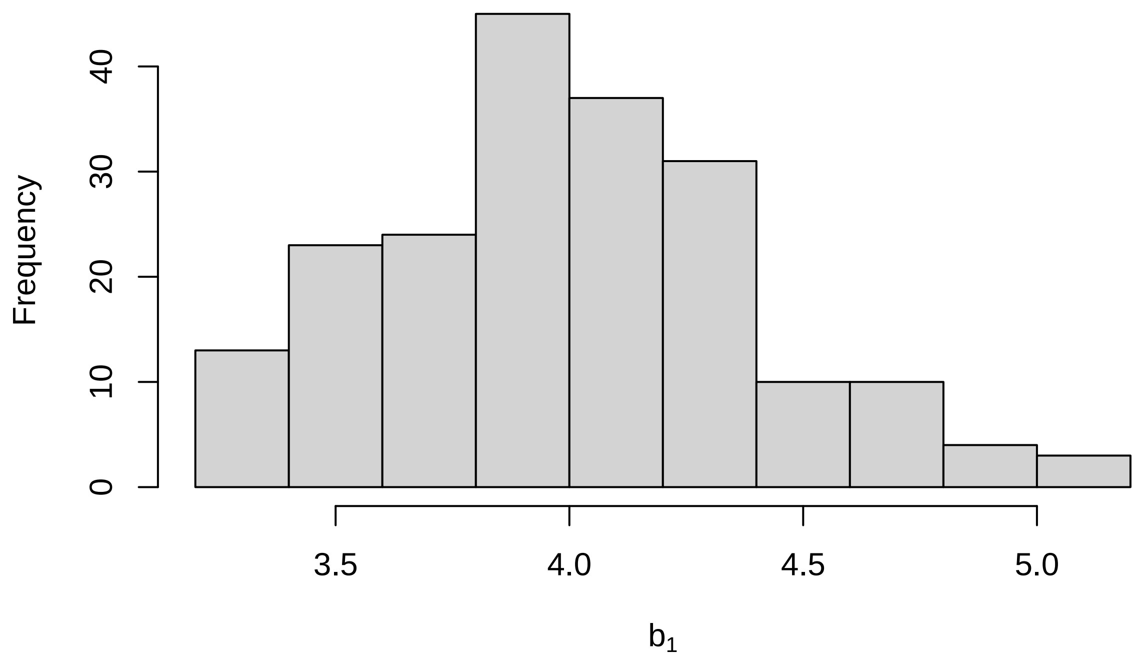
```
## (Intercept)
##    58.07592
```

```
xhat.age <- data.frame(X=60)
muscmass.prediction.int <- predict(linmod, xhat.age, interval = "confidence", level = 0.95, se.fit = TRU
muscmass.prediction.int
```

```
## $fit
##        fit      lwr     upr
## 1 262.1338 164.5145 359.753
##
## $se.fit
## [1] 30.67427
##
## $df
## [1] 3
##
## $residual.scale
## [1] 8.027824
```

```
#B
nsim <- 200
b1s <- numeric(nsim)
for (i in 1:nsim) {
epsilon <- rnorm(n, mean = 0, sd = sig)
Y <- 20 + 4*X + epsilon
linmod <- lm(Y~X)
b1s[i] <- linmod$coef[2]
}
```

```
#C
hist(b1s, xlab = expression(b[1]), main = "")
```




```
mean(b1s)
```

```
## [1] 3.994485
```

```
sd(b1s)
```

```
## [1] 0.3998482
```

6.14. Refel' to Grocery retailer Problem 6.9. Assume that regression model (6.5) for three predictor variables with independent normal error terms is appropriate. Three new shipments are to be received, each with Xh1 = 282,000, Xhb = 7.10, and Xh3 = O. a. Obtain a 95 percent prediction interval for the mean handling time for these shipments. Prediction interval: (3968.297, 4646.054) b. Convert the interval obtained in part (a) into a 95 percent prediction interval for the total labor hours for the three shipments. Prediction interval: (11904.89, 13938.16)

```
#A
library(tidyverse)
loadRData <- function(fileName){
  load(fileName)
  get(ls()[ls() != "fileName"])
}
grocery <- loadRData("/Users/lukegeel/Downloads/grocery_spring2021.RData")
Y <- grocery$Y
X1 <- grocery$X1
```

```
X2 <- grocery$X2
X3 <- grocery$X3
linmod <- lm(Y~X1+X2+X3)
xhat.age <- data.frame(X1=282000,X2=7.1,X3=0)

grocery.confidence.int <- predict(linmod, xhat.age, interval = "prediction", level = 0.95, se.fit = TRU
grocery.confidence.int
```

```
## $fit
##        fit      lwr      upr
## 1 4307.175 3968.297 4646.054
##
## $se.fit
## [1] 26.52769
##
## $df
## [1] 48
##
## $residual.scale
## [1] 166.4423
```

```
#B
library(tidyverse)
loadRData <- function(fileName){
  load(fileName)
  get(ls()[ls() != "fileName"])
}
grocery <- loadRData("/Users/lukegeel/Downloads/grocery_spring2021.RData")
Y <- grocery$Y
X1 <- grocery$X1
X2 <- grocery$X2
X3 <- grocery$X3
linmod <- lm(Y~X1+X2+X3)
xhat.age <- data.frame(X1=282000,X2=7.1,X3=0)

grocery.confidence.int <- predict(linmod, xhat.age, interval = "prediction", level = 0.95, se.fit = TRU
lower <- 3968.297
upper <- 4646.054
lower*3
```

```
## [1] 11904.89
```

```
upper*3
```

```
## [1] 13938.16
```

*6.17. Refer to Patient satisfaction Problem 6.15. Assume that regression model (6.5) for three predictor variables with independent normal error terms is appropriate. a. Obtain an interval estimate of the mean satisfaction when Xh1 = 35, Xh2 = 45, and Xh3 = 2.2. Use a 90 percent confidence coefficient. Interpret your confidence interval. Confidence interval: (63.91482, 72.35177) We can be 90% confident that the true mean satisfaction when Xh1 = 35, Xh2 = 45, and Xh3 = 2.2 is between 63.91482 and 72.35177 b. Obtain a prediction interval for a new patient's satisfaction when X/l 1 = 35, X/l2 = 45, and X/l3 = 2.2. Use a 90 percent confidence coefficient. Interpret your prediction interval Prediction interval: (51.66071, 84.60588) We can be 90% confident that when Xh1 = 35, Xh2 = 45, and Xh3 = 2.2, the average satisfaction is between 51.66071 and 84.60588.

```
#A
library(tidyverse)
loadRData <- function(fileName){
  load(fileName)
  get(ls()[ls() != "fileName"])
}
patient <- loadRData("/Users/lukegeel/Downloads/patient_satisfaction_spring2021.RData")
Y <- patient$Y
X1 <- patient$X1
X2 <- patient$X2
X3 <- patient$X3
linmod <- lm(Y~X1+X2+X3)
xhat.age <- data.frame(X1=35,X2=45,X3=2.2)

patient.confidence.int <- predict(linmod, xhat.age, interval = "confidence", level = 0.90, se.fit = TRUI
patient.confidence.int
```

```
## $fit
##        fit      lwr      upr
## 1 68.13329 63.91482 72.35177
##
## $se.fit
## [1] 2.508083
##
## $df
## [1] 42
##
## $residual.scale
## [1] 9.467135
```

```
#B
library(tidyverse)
loadRData <- function(fileName){
  load(fileName)
  get(ls()[ls() != "fileName"])
}
patient <- loadRData("/Users/lukegeel/Downloads/patient_satisfaction_spring2021.RData")
Y <- patient$Y
X1 <- patient$X1
X2 <- patient$X2
X3 <- patient$X3
linmod <- lm(Y~X1+X2+X3)
xhat.age <- data.frame(X1=35,X2=45,X3=2.2)

patient.confidence.int <- predict(linmod, xhat.age, interval = "prediction", level = 0.90, se.fit = TRUI
patient.confidence.int
```

```
## $fit
##        fit      lwr      upr
```

```
## 1 68.13329 51.66071 84.60588
##
## $se.fit
## [1] 2.508083
##
## $df
## [1] 42
##
## $residual.scale
## [1] 9.467135
```