

Name - vvgupta2

Homework 4

Homework Instructions

For questions that require code, please create a code chunk directly below the question and type your code there. Your knitted pdf will show both your code and your output. You are encouraged to knit your file as you work to check that your coding and formatting is done so appropriately.

For written responses or multiple choice questions, please bold your (selected) answer.

Grading Details

All questions will be graded full credit (1 point), half credit (0.5 point) or no credit (0 points).

Full credit responses should have the correct response and appropriate code (if applicable). Half credit responses will have a reasonable attempt (typically no more than one small error or oversight), and no credit responses will be either non-attempts or attempts with significant errors.

Exercise 1

Consider a random variable X that has a t distribution with 7 degrees of freedom. Calculate $P[X > 1.3]$.

Hint: By default `pt()` considers area under the curve to the left of the given value.

```
#solution  
pt(1.3, df=7, lower.tail = FALSE)
```

```
## [1] 0.1173839
```

Exercise 2

Consider a random variable Y that has a t distribution with 9 degrees of freedom. Find c such that $P[X > c] = 0.025$.

```
#solution  
qt(0.025, df=9, lower.tail = FALSE)
```

```
## [1] 2.262157
```

c = 2.262157

Exercise 3

For this Exercise, use the built-in `trees` dataset in R. Fit a simple linear regression model with `Girth` as the response and `Height` as the predictor. Extract the p-value for testing $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$

Hint: Section 8.6 of the book offers several examples of how to extract different elements from a linear model summary. That might be helpful to reference.

#solution

```
?trees
model_1 = lm(Girth ~ Height, data = trees)
names(model_1)
```

```
## [1] "coefficients" "residuals"      "effects"      "rank"
## [5] "fitted.values"  "assign"       "qr"           "df.residual"
## [9] "xlevels"        "call"        "terms"       "model"
```

```
(summary(model_1))$coefficients[2,4]
```

```
## [1] 0.002757815
```

Exercise 4

Continue using the Tree model you just created. Create a **90%** confidence interval for β_1 ?. Then calculate the total width of that interval.

Hint: Remember to specify the confidence level.

#solution

```
a = confint(model_1, level = 0.90)
a
```

```
##              5 %      95 %
## (Intercept) -16.3155305 3.938741
## Height      0.1229463 0.388548
```

```
total_width = abs(a[2,2])-abs(a[2,1])
total_width
```

```
## [1] 0.2656018
```

Exercise 5

Calculate a **95%** confidence interval for the mean tree girth of a tree that is 79 feet tall.

Report only the lower and upper bound of your interval.

Hint: Use brackets to extract only those elements.

#solution

```
a = predict(model_1, data.frame(Height=79), interval = "confidence", level = 0.95)
bound=c(a[2],a[3])
bound
```

```
## [1] 12.90480 15.12646
```

Exercise 6

For the next several exercises, use the `faithful` dataset, which is built into `R`.

Suppose we would like to predict the duration of an eruption of the Old Faithful geyser in Yellowstone National Park based on the waiting time before an eruption. Fit a simple linear model in `R` that accomplishes this task.

What is the value of $SE[\hat{\beta}_1]$?

Hint: Section 8.6 of the book offers several examples of how to extract different elements from a linear model summary. That might be helpful to reference.

```
#solution
?faithful
model_2 = lm (eruptions ~ waiting, data = faithful)
std_err = (summary(model_2))$coefficients[2,2]
std_err
```

```
## [1] 0.002218541
```

Exercise 7

What is the value of the test statistic for testing $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$?

```
#solution
t = (((summary(model_2))$coefficients[2,1]) - 0 )/std_err
t
```

```
## [1] 34.08904
```

Exercise 8

Take a look at your linear model summary from the `faithful` data.

Test $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$ with $\alpha = 0.01$. What decision do you make?

- Fail to reject H_0
- **Reject H_0**
- Reject H_1
- Not enough information

Exercise 9

Now let's take a look at the intercept. Calculate and report a 90% confidence interval for β_0 .

```
#solution
summary(model_2)
```

```
##
## Call:
## lm(formula = eruptions ~ waiting, data = faithful)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29917 -0.37689  0.03508  0.34909  1.19329
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.874016   0.160143  -11.70  <2e-16 ***
## waiting      0.075628   0.002219   34.09  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4965 on 270 degrees of freedom
## Multiple R-squared:  0.8115, Adjusted R-squared:  0.8108
## F-statistic: 1162 on 1 and 270 DF,  p-value: < 2.2e-16
```

```
beta_0_hat = summary(model_2)$coefficients[1,1]
beta_0_err = summary(model_2)$coefficients[1,2]
t = qt(0.95, 270)
upper_b = beta_0_hat + (t*beta_0_err)
lower_b = beta_0_hat - (t*beta_0_err)
ci= c(lower_b, upper_b)
ci
```

```
## [1] -2.138335 -1.609697
```

Exercise 10

Create a 99% **prediction** interval for a new observation's eruption duration for a waiting time of 72 minutes.

Output only the **lower** and **upper** bound of your interval.

Hint: Use brackets at the end of the argument.

```
#solution
b = predict(model_2, data.frame(waiting=72), interval = "predict", level=0.99)
ans = c(b[2],b[3])
ans
```

```
## [1] 2.280781 4.861612
```

Exercise 11

Consider a 90% confidence interval for the mean response and a 90% prediction interval, both at the same x value. Which interval is narrower?

- **Confidence interval**
- Prediction interval
- No enough information, it depends on the value of x

Exercise 12

Suppose you test $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$ with $\alpha = 0.01$ and fail to reject H_0 . Which statement best describes our finding?

- There is no relationship between the response and the predictor.
- **If β_1 really equals 0, then we would observe a $\hat{\beta}_1$ value as extreme as we did more than 1% of the time.**
- We have strong evidence that β_1 is not equal to 0.
- The probability that $\beta_1 = 0$ is very high.