STA035B Midterm Practice

Spencer Frei

This provides an idea of what types of questions and material you will have on your midterm exam. In addition to these problems, be sure to review the lecture notes/slides (especially examples of how to calculate things like p values, confidence intervals, etc.), the labs and homework assignments, especially those for the material from the "Visualization" section onwards.

Problem 1

Consider the diamonds dataset:

```
head(diamonds)
```

```
# A tibble: 6 x 10
  carat cut       color clarity depth table price     x     y     z
  <dbl> <ord>     <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
1  0.23 Ideal     E     SI2      61.5    55   326  3.95  3.98  2.43
2  0.21 Premium   E     SI1      59.8    61   326  3.89  3.84  2.31
3  0.23 Good      E     VS1      56.9    65   327  4.05  4.07  2.31
4  0.29 Premium   I     VS2      62.4    58   334  4.2   4.23  2.63
5  0.31 Good      J     SI2      63.3    58   335  4.34  4.35  2.75
6  0.24 Very Good J     VVS2     62.8    57   336  3.94  3.96  2.48
```

```
str(diamonds$cut)
```

```
 Ord.factor w/ 5 levels "Fair"<"Good"<..: 5 4 2 4 2 3 3 3 1 3 ...
```

Which of the following would plot the number of diamonds per cut in a bar plot?

(a)

```
diamonds %>%
  group_by(cut) %>%
  summarise(num = n()) %>%
  ggplot(aes(x = num)) +
  geom_bar()
```

(b)

```
diamonds %>%
  group_by(cut) %>%
  summarise(num = n()) %>%
  ggplot(aes(x = cut)) +
  geom_bar()
```

(c)

```
diamonds %>%
  ggplot(aes(x = cut, y = num)) +
  geom_bar()
```

(d)

```
diamonds %>%
  ggplot(aes(x = cut)) +
  geom_bar()
```

Problem 2

Suppose I want to find the area under the curve for the standard normal that lies to the right of a Z-score of 1.5.

Part (1) Which of the following R code correctly returns this area?

a. `pnorm(1.5)`
b. `1-pnorm(1.5)`
c. `1-pnorm(-1.5)`
d. `qnorm(1.5)`
e. `1-qnorm(1.5)`
f. `qnorm(-1.5)`

Part (2) Which range of values does this value lie in?

a. Between 0.02 and 0.16
b. Between 0.16 and 0.5
c. Between 0.5 and 0.66
d. Between 0.66 and 0.94

Problem 3

Given a linear regression model fit to a dataset df with independent variable x and dependent variable y, the following R code snippet generates a residual plot:

```r
model <- lm(y ~ x, data = df)
df$residuals <- residuals(model)

ggplot(df, aes(x = x, y = residuals)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(title = "Residual Plot", x = "Predictor", y = "Residuals")
```

Explain what the residual plot represents in the context of linear regression. What does the horizontal dashed line at y = 0 signify, and how can this plot be used to evaluate the model's assumptions?

Problem 4

[IMS] 7.19 - Starbucks, calories, and protein (see textbook)

Problem 5

[IMS] 16.3 - Survey on defund the police