Comprehension Check: Conditional Probabilities Part 2

Q6

1/1 point (graded)

We are now going to write code to compute conditional probabilities for being male in the `heights` dataset. Round the heights to the closest inch. Plot the estimated conditional probability $P(x) = \Pr(\text{Male}|\text{height} = x)$ for each x.

Part of the code is provided here:

```
library(dslabs)
data("heights")
MISSING CODE
        qplot(height, p, data =.)
```

Which of the following blocks of code can be used to replace **MISSING CODE** to make the correct plot?

-
```
heights %>%
        group_by(height) %>%
        summarize(p = mean(sex == "Male")) %>%
```

-
```
heights %>%
        mutate(height = round(height)) %>%
        group_by(height) %>%
        summarize(p = mean(sex == "Female")) %>%
```

-
```
heights %>%
        mutate(height = round(height)) %>%
        summarize(p = mean(sex == "Male")) %>%
```

```
heights %>%
        mutate(height = round(height)) %>%
        group_by(height) %>%
        summarize(p = mean(sex == "Male")) %>%
```

Submit

You have used 1 of 2 attempts

Answers are displayed within the problem

Q7

1/1 point (graded)

In the plot we just made in Q1 we see high variability for low values of height. This is because we have few data points. This time use the quantile $0.1, 0.2, \ldots, 0.9$ and the `cut` function to assure each group has the same number of points. Note that for any numeric vector `x`, you can create groups based on quantiles like this: `cut(x, quantile(x, seq(0, 1, 0.1)), include.lowest = TRUE)`.

Part of the code is provided here:

```
ps <- seq(0, 1, 0.1)
heights %>%
        MISSING CODE
        group_by(g) %>%
        summarize(p = mean(sex == "Male"), height = mean(height)) %>%
        qplot(height, p, data =.)
```

Which of the following lines of code can be used to replace **MISSING CODE** to make the correct plot?

```
mutate(g = cut(male, quantile(height, ps), include.lowest = TRUE)) %>%
```

```
mutate(g = cut(height, quantile(height, ps), include.lowest = TRUE)) %>%
```

```
mutate(g = cut(female, quantile(height, ps), include.lowest = TRUE)) %>%
```

```
mutate(g = cut(height, quantile(height, ps))) %>%
```

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Q8

1/1 point (graded)

You can generate data from a bivariate normal distrubution using the `MASS` package using the following code:

```
Sigma <- 9*matrix(c(1,0.5,0.5,1), 2, 2)
dat <- MASS::mvrnorm(n = 10000, c(69, 69), Sigma) %>%
        data.frame() %>% setNames(c("x", "y"))
```

And you can make a quick plot using `plot(dat)`.

Using an approach similar to that used in the previous exercise, let's estimate the conditional expectations and make a plot. Part of the code has again been provided for you:

```
ps <- seq(0, 1, 0.1)
dat %>%
        MISSING CODE
        qplot(x, y, data =.)
```

Which of the following blocks of code can be used to replace **MISSING CODE** to make the correct plot?

```
mutate(g = cut(x, quantile(x, ps), include.lowest = TRUE)) %>%
group_by(g) %>%
summarize(y = mean(y), x = mean(x)) %>%
```

```
mutate(g = cut(x, quantile(x, ps))) %>%
group_by(g) %>%
summarize(y = mean(y), x = mean(x)) %>%
```

```
mutate(g = cut(x, quantile(x, ps), include.lowest = TRUE)) %>%
summarize(y = mean(y), x = mean(x)) %>%
```

```
mutate(g = cut(x, quantile(x, ps), include.lowest = TRUE)) %>%
group_by(g) %>%
summarize(y =(y), x =(x)) %>%
```

Submit

You have used 1 of 2 attempts

Answers are displayed within the problem