Test assignment

SLR: Conditions + Prediction

INSERT YOUR NAME HERE

2021-08-10

Price of Textbooks

I have made a change to the file. I made another change! Last change to check my style!

```
library(tidyverse)
library(broom)
#library(patchwork)
library(knitr)
```

In this AE, we will look at the price of textbooks and how it varies based on the number of pages. The data contains the price and number of pages for a random sample of 30 college textbooks from the Cal Poly-San Luis Obispo bookstore in Fall 2006.

```
textbooks <- read_csv("data/textbooks.csv")
```

We will use the following variables:

- `Pages`: Number of pages in the textbook
- `Price`: Price of the textbook in US dollars

Visualize distributions

```
p1 <- ggplot(data = textbooks, aes(x = Price)) +
  geom_histogram() +
  labs(title = "Price of Textbooks",
       subtitle = "in 2006")

p2 <- ggplot(data = textbooks, aes(x = Pages)) +
  geom_histogram(binwidth = 100) +
  labs(title = "Pages in Textbooks",
       subtitle = "in 2006")

p3 <- ggplot(data = textbooks, aes(x = Pages, y = Price)) +
  geom_point() +
  labs(y = "Price in US Dollars",
       title = "Price vs. Pages in Textbooks",
       subtitle = "in 2006")

(p1 + p2) / p3
```

Exercise 1: Linear model

Fill in the code to display the model showing 3 digits for numerical values. Then, write the regression equation using mathematical notation.

```
textbook_model <- lm(Price ~ Pages, data = textbooks)
```

```
# code to display model
```

Exercise 2: Conditions for SLR

We use the residuals to check the model conditions for SLR. We can calculate the residuals and fitted (predicted) values using the `augment` function.

Fill in the code below to make a histogram of the residuals, then use functions from the`patchwork` package to arrange the 3 plots in a grid.

```
textbook_aug <- augment(textbook_model)
```

```
resid_fitted <- ggplot(data = textbook_aug, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0, color = "red") +
  labs(x = "Predicted values",
       y = "Residual",
       title = "Residuals vs. Predicted")

resid_qq <- ggplot(data = textbook_aug, aes(sample = .resid)) +
  stat_qq() +
  stat_qq_line() +
  labs(title = "Normal QQ-plot of residuals")

## add code to make a histogram of residuals. Then use the patchwork package to display the plots.
```

Are the conditions satisfied? Briefly explain.

- **Linearity**:
- **Constant variance**:
- **Normality**:
- **Independence**:

Note: You can make a plot of the residuals vs. fitted and the Normal QQ-plot (using standardized residuals) using the ***`autoplot`*** *function in ggfortify package. You still need to create the histogram of residuals.*

Read more about ***`ggfortify`****: https://cran.r-project.org/web/packages/ggfortify/vignettes/plot_lm.html*

```
#install.packages("ggfortify")

#library(ggfortify)
#autoplot(textbook_model, which = 1:2)
```

Exercise 3: Prediction

Below are two prediction tasks:

1. Calculate the predicted price and associated 90% interval for a textbook with 500 pages.
2. Estimate the mean price and associated 90% interval for textbooks with 500 pages.

Which interval will we use to complete each task? How do the intervals compare?

```
x0 <- tibble(Pages = 500)
```

Interval A

```
textbook_model %>%
  predict(x0, interval = "confidence", level = .90) %>%
  kable(digits = 3)
```

fit	lwr	upr
70.242	60.926	79.558

Interval B

```
textbook_model %>%
  predict(x0, interval = "prediction", level = 0.90) %>%
  kable(digits = 3)
```

fit	lwr	upr
70.242	18.766	121.718

Knit your Rmd file to view the updated output. Commit your changes with an informative commit message, and push the updated files to GitHub.