Quiz 3 — 2/1/2023

Instructions

This take-home quiz is due on **Wednesday, February 1 at 23:59**.

You may use your own course materials, as well as any materials directly linked from the course website. **No collaboration allowed.**

Type your answers **directly in this Jupyter notebook**, and submit this notebook (just the `ipynb` file) using the submission form on the course website.

> **Feedback.** This quiz problem has many similarities to STAT2 Exercises 1.19 and 1.21, assigned for homework.

Problem 1

The data frame `GrinnellHouses` in the `Stat2Data` package contains data from 929 house sales in Grinnell, Iowa, between 2005 and into 2015.

For this quiz, we want to investigate the relationship between the list price of the home (what the seller asks for the home) and the final sale price. One would expect there to be a strong relationship. In most markets, including Grinnell during this period, the list price almost always exceeds the sale price.

(a) Load the data frame `GrinnellHouses` from the `Stat2Data` package. Display the first few rows.

In [ ]:

> **Feedback.** Follow the instructions carefully. Recall that `head()` displays the first few rows of a data frame. This is an important step — you should familiarize yourself with the layout of the data before you analyze it.

(b) Make a scatterplot with `ListPrice` on the horizontal axis and `SalePrice` on the vertical axis.

In [ ]:

(c) Briefly comment on any patterns you observe in your scatterplot from part (b).

Write your answer here. Double-click to edit.

> **Feedback.** Some of you may see in your graded quizzes that this code cell results in an error. This is my fault — I had intended this cell to be a Markdown cell.
>
> In the future, use a Markdown cell for your non-code, written repsonses.

(d) Use R to find the least squares regression line for predicting the sale price of a home based on the list price of that home. Your output should include:

- a table that shows the estimated parameters of the least squares regression line;

- a scatterplot of the two variables with the least squares regression line on the same axes.

In [ ]:

Feedback.

- Be careful when specifying the response variable and the explanatory variable in `lm()`: it should be `lm(response ~ explanatory, ...)`.
- Follow the instructions carefully. Your code for *this part* should output a table that shows the estimated parameters for the least squares regression line.

  In class, we saw one way to output a table with the estimated parameters for the least squares regression line: `summary(fit)` – see Lesson 6.

(e) Interpret the value (not just the sign) of the slope of the fitted model in the context of this setting.

Write your answer here. Double-click to edit.

Feedback. See Example 3b in Lesson 6 and STAT2 Exercise 1.19c for similar examples.

Note that the estimated slope of the least squares regression line gives you information about how changes in the explanatory variable affect the response variable **on average** (or equivalently, **in expectation**). So, your answer to this part needs to indicate how the sales price changes on average, or how the expected sales price changes.

(f) What sales price would the fitted model predict for a house that has a listing price of $99,500? Use the code cell below as a calculator to show your work.

In [ ]:

Feedback. See STAT2 Exercise 1.21a for a similar example.

(g) The house at 1317 Prince Street has a listing price of $99,500 and a sales price of $95,000. Find the residual for this observation. Use the code cell below as a calculator to show your work.

In [ ]:

Feedback. See Example 3d in Lesson 6 and STAT2 Exercise 1.21b for similar examples.

(h) Make a residual vs. fitted values plot.

In [ ]:

(i) Based on your residual vs. fitted values plot in part (h), does the fitted model satisfy the constant variance condition? Briefly explain.

Write your answer here. Double-click to edit.

Feedback. See Lesson 7 Part 2 for a similar example.

Note that when we are looking at a residuals vs. fitted values plot to determine whether constant variance (or uniform

spread) is satisfied, we are looking to see whether the **vertical width** of the band of residuals is relatively constant for different fitted values (across the x-axis).

Grading rubric

Problem	Weight
1a	0.5
1b	0.5
1c	0.5
1d	1.0
1e	1.0
1f	0.5
1g	0.5
1h	0.5
1i	1.0
Max Score	**60**