

STAT 213 (SUMMER 2021): HW4 (SLOS A1, B1, B2, C1)

DUE ELECTRONICALLY BY THE CLASS TIME, FRIDAY 6/25/21

INSTRUCTIONS AND TECHNICAL TIPS

Write up your solutions and save/upload the file(s) to the RStudioPro server in the folder `stat213/turnin/hw4/` inside your Home directory. Include `hw3` (all lowercase) in your filenames.

Using an RMarkdown document is recommended, but not required.

RMarkdown Format. If you do use RMarkdown, R code should go in code chunks, and verbal commentary (and any math equations) should go outside code chunks.

Use section headings (a line starting with one or more `#` symbols) to demarcate the start of a problem. You can nest headings by using an additional `#` symbol for each level of nesting: one for a top-level heading, two for the next level within that, etc.

Periodically “Knit” your file to verify that it is working correctly.

“Knitting” Troubleshooting. If your code runs chunk by chunk but won’t Knit, try clearing your environment (broom icon in the upper right) and running chunk by chunk from the start again. The most common cause is an undefined variable. This can happen if you change your variable names some places but not others, but a very common reason is that you read in the data from a file using a menu button instead of using the `read.file()` command, which means the “reading in the data” step is not recorded in your document. If you can’t Knit, I won’t be able to run your code either!

If you aren’t able to Knit directly to `.pdf`, it may be because you are using special characters (such as `≠`) in your `.Rmd`. Replace these with plain text and try again.

If you have done this and are still unable to Knit directly to `.pdf`, convert your Knitted `.html` or `.docx` into a `.pdf` (possibly by “printing” the `html` from your browser to a file) and save that there. A uniform file format across students will make grading 64 problem sets much more streamlined. Thanks!

What to turn in. If using Markdown, turn in the source file (ending in `.Rmd`) and the Knitted output file (preferably as `.pdf`). Otherwise, just turn in a `.pdf`.

Date: June 20, 2021.

Verifying receipt of your work. When I run the script that collects your work after the due date, you will see a “receipt” file in the `~/stat213/receipts/hw4/` directory. If this does not appear within 24 hours or so of you submitting your assignment, let me know.

PROBLEMS

1. **Textbook prices revisited.** We again examine the dataset `TextPrices`. As a reminder, the dataset consists of a random sample of 30 textbooks from the CalPoly campus bookstore, collected in the fall of 2006. The variables of interest are `Price`, the price of each textbook, and `Pages`, the number of pages in the book. The students were interested in the question of whether the number of pages can be used to predict prices.
 - (a) Perform a significance test to address the students’ question of whether the number of pages is a useful predictor of a textbook’s price. Report and interpret the hypotheses, test statistic, and P -value, along with your conclusion.
 - (b) Determine a 95% confidence interval for the population slope coefficient. Also explain what the slope coefficient means in the context of this data.
2. **Metabolic rate of caterpillars.** Marisa Stearns collected and analyzed body size and metabolic rates for *Manduca Sexta* caterpillars. The data is in the set `MetabolicRate` and the variables are:

| Variable | Description |
|-----------------------|--|
| <code>Computer</code> | Number of the computer used to obtain metabolic rate |
| <code>BodySize</code> | Size of the animal in kilograms (Note: the built in dataset documentation incorrectly says that size is measured in grams!) |
| <code>C02</code> | Carbon dioxide concentration in parts per million |
| <code>Instar</code> | Number from 1 (smallest) to 5 (largest) indicating stage of the caterpillar’s life |
| <code>Mrate</code> | Metabolic rate |

The dataset also has variables `LogBodySize` and `LogMrate` containing the logs (base 10) of the size and metabolic rate variables. The researchers would like to build a linear model to predict metabolic rate (either `Mrate` directly or on a log scale with `LogMrate`) using a measure of body size for the caterpillars (either `BodySize` directly or on a log scale with `LogBodySize`).

- (a) Which combination of explanatory and response variable transformation produce a model that best satisfies the regression condition? Support your choice with appropriate plots, and give the prediction equation for your chosen model.

- (b) What metabolic rate does your fitted model from (a) predict for a caterpillar that has a body size of 1 gram? (Note the unit error in the built-in documentation for the data: body size is measured in kilograms in the data, not grams!)
3. Use the dataset `USstamps` to model postal rates over time from 1958 to 2012. Before beginning, use the following code to create a copy of the data set which is filtered to include only observations from years within this range:

```
library(mosaic)
library(Stat2Data)
data(USstamps)
USstamps1958to2012 <- USstamps %>%
  filter(Year >= 1958 & Year <= 2012)
```

- (a) Fit a simple linear regression model to predict postal rates (`Price`) using `Year`. Give the regression equation and interpret the coefficients in context.
- (b) What percentage of the variation in rates in this dataset is predictable using the year?
- (c) Is there significant evidence of a linear association between postal rates and year? Support your answer with a hypothesis test, stating the hypotheses, test statistic and P -value, and interpreting the results in context.
- (d) Give and interpret a 95% confidence interval for the slope of this model.
- (e) Find the ANOVA table that partitions the variability in `Price` for a linear model based on `Year`, and interpret the F -test it contains **as a comparison between two models**. Write out the prediction equations for each model.