

Problem Set 10

Shane T. Mueller shanem@mtu.edu

November 27, 2018

This problem set covers simple categorical predictors, the link between regression and ANOVA, and post-hoc tests, and multi-way ANOVA.

Categorical Predictors

On each of day of one week, we sampled 100 random company stocks and examined their trading price. Each day a different set of stocks was sampled at random from the NYSE and NASDAQ published prices.

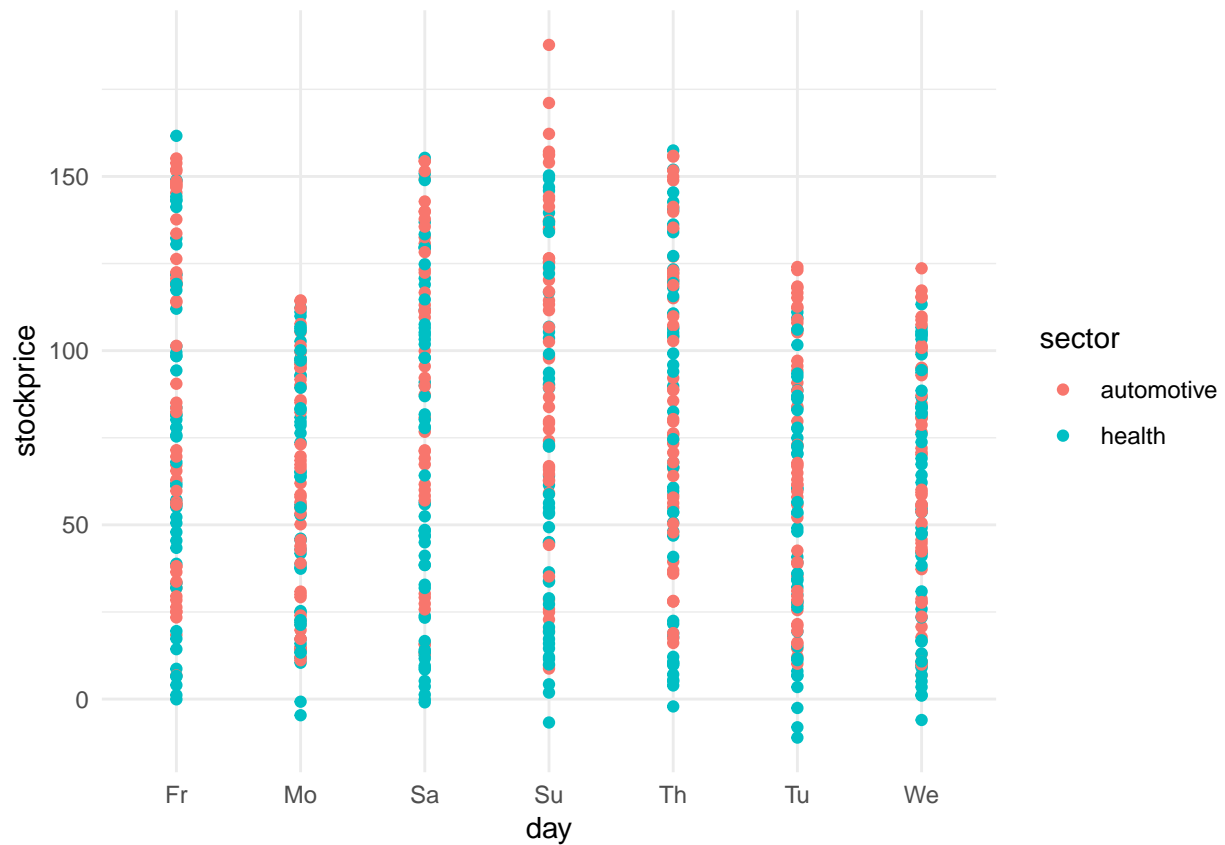
```
library(ggplot2)
data <- read.csv("ps10data.csv")
```

```
head(data)
```

```
##   day      sector stockprice
## 1  Mo automotive    55.11
## 2  Mo automotive    85.10
## 3  Mo    health    99.67
## 4  Mo automotive    19.79
## 5  Mo automotive    69.68
## 6  Mo automotive    61.97
```

This is stored in a matrix. For a regression or ANOVA, we really need each one

```
ggplot(data,aes(x=day,y=stockprice)) + geom_point(aes(color=sector)) + theme_minimal()
```



For this problem, we want to determine, using a number of methods, which days differed from which other days. In each case, run the test, and answer the question in 1-2 sentences describing what you found. Use a $p=.05$ as a criterion for determining whether an effect is statistically significant.

1. First use a contrast that will compare each day to Monday, and report which of the days had prices significantly higher than Monday (report the test obtained directly from the coefficients of `lm` by doing `summary()` on the results of `lm()`).
2. Then, use successive difference coding of the day variable to determine which days of the week differed significantly from the previous day.
3. Use `pairwise.t.test` function to compute all pairwise t-tests and the holm correction between days of the week. Describe concisely which days differed from which other days.
4. Use an `aov()` model to predict stock price by day, and then compute Tukey HSD test on all pairwise comparisons using the Tukey test. Do the result differ from part 3?
5. Compute a Kruskal-Wallis test to see if the non-parametric test shows stock price depended on day-of-week.
6. Compute a one-way BayesFactor ANOVA and report the Bayes factor score determining if day-of-week impacted stock price.

2. Multi-way ANOVA and regression.

The stocks were sampled from two different sectors (health and automotive). Was there a difference in outcome based on sector? What about when day-of-week is considered. Report a standard (Type-I) ANOVA F-test for:

1. the effect of sector on its own (a one-way test), and
2. whether sector has an effect *after* day-of-week is considered: `lm(stockprice~day+sector)`
3. whether the results differ if sector is included in the model first (`lm(stockprice~sector+day)`)

Then compare results of the three tests, including the sum-squared deviations and the results of the F test. Are the results of the tests identical or do they differ? Why? Pick which one you would prefer to use to test the effect, and describe why you feel it is better than the others.