Homework 09

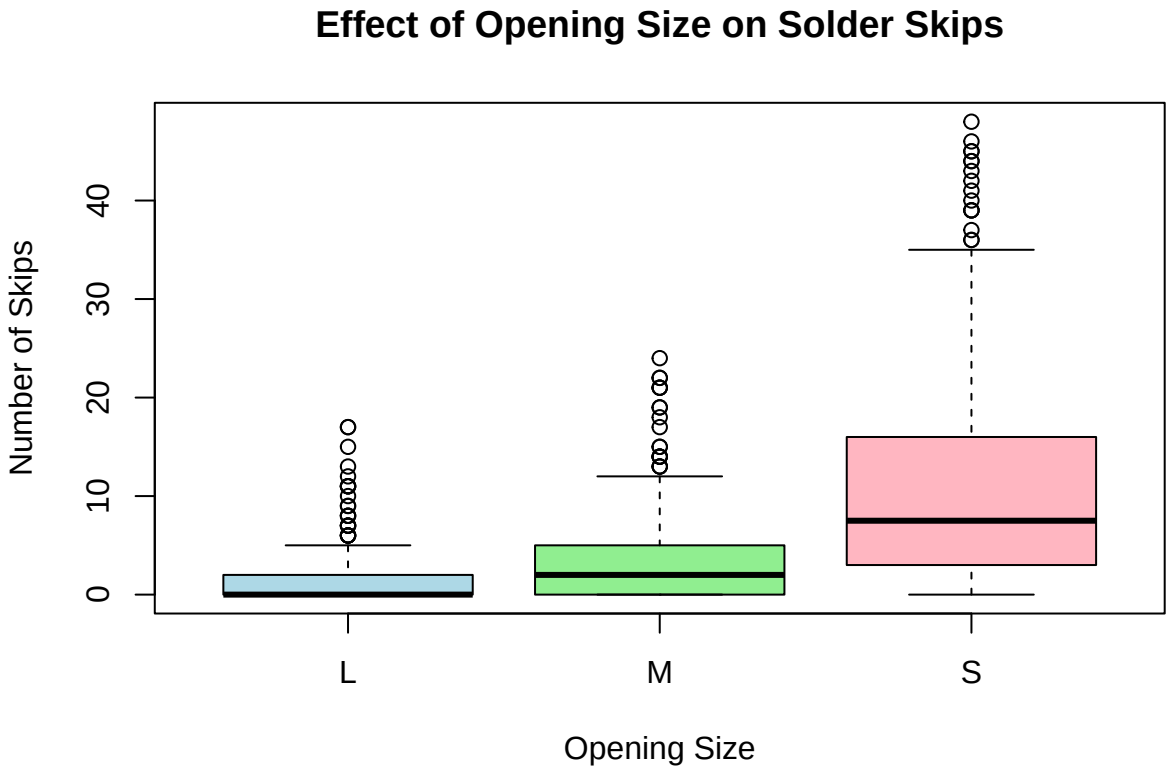
Suresh Bhavsar

2025-04-20

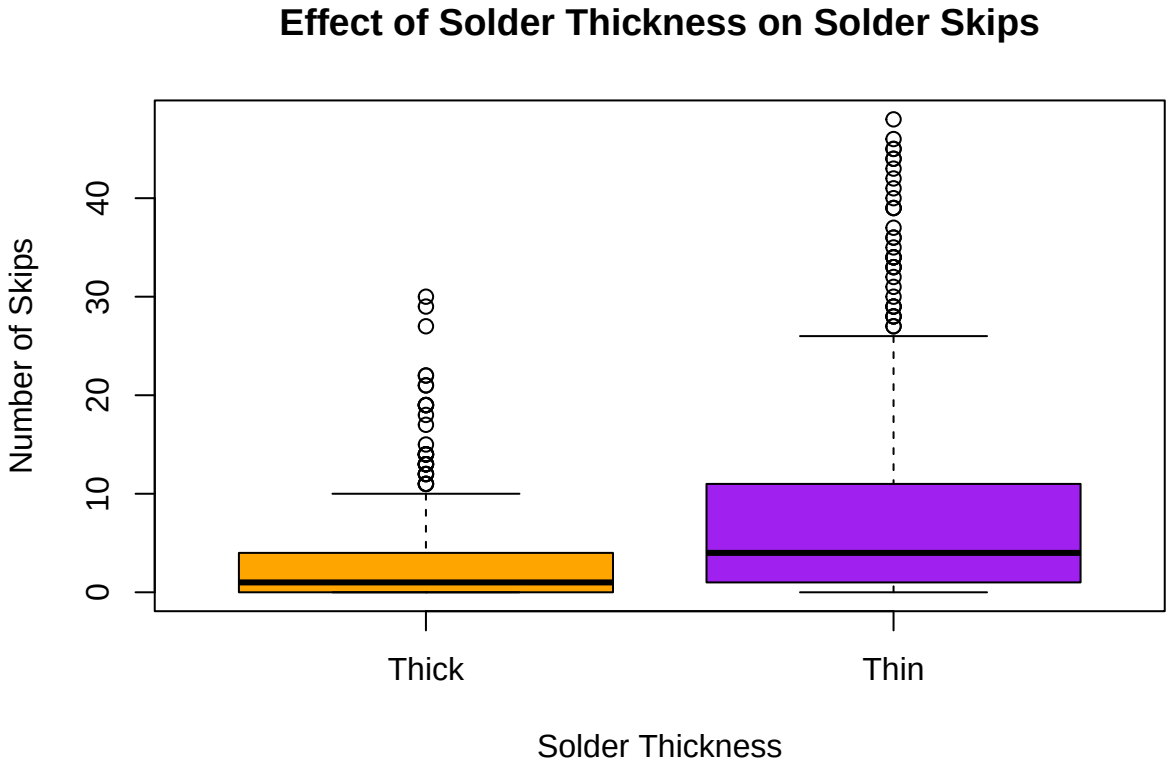
Name: Suresh Bhavsar UT EID: sb65245 GitHub repo: https://github.com/Surbeast/sds-315-hw-9

Problem 1: Manufacturing Flaws in Circuit Boards

Part A



These boxplots show the difference in the distribution of the number of skips between the different opening sizes. The blue boxplot represents the large opening size, the green boxplot represents the medium opening size, and the pink boxplot represents the small opening size.



These boxplots show the difference in the distribution of the number of steps between the different thicknesses of the alloy used for soldering.

Part B

```
## # A tibble: 6 x 4
##   term                  estimate lower_ci upper_ci
##   <chr>                    <dbl>    <dbl>    <dbl>
## 1 intercept                 0.39    -0.63     1.41
## 2 Opening: M                2.41     0.96     3.85
## 3 Opening: S                5.13     3.68     6.57
## 4 Solder: Thin              2.28     0.84     3.72
## 5 Opening: M:SolderThin    -0.74    -2.78     1.3
## 6 Opening: S:SolderThin     9.65     7.61    11.7
```

Part C

The base number of skips for a circuit board that was made with a large opening and thick solder alloy is 0.39. The main effect of a medium sized opening is 2.41 skips. This is the effect of Opening: M in isolation. The main effect of a small sized opening is 5.13 skips. This is the effect of Opening: S in isolation. The main effect of a thing solder alloy is 2.28 skips. This is the effect of Solder: Thin in isolation. The interaction effect for a medium sized opening and a thin solder alloy is -0.74 skips. In other words, circuit boards made with medium openings and a thin solder alloy have 0.74 less skips than what you would expect from summing the individual "isolated" effects of the two variables. The interaction effect for a small sized opening and a thin solder alloy is 9.65 skips. In other words, circuit boards made with small openings and a thin solder alloy have 9.65 more skips than what you would expect from summing the individual "isolated" effects of the two variables.

Part D

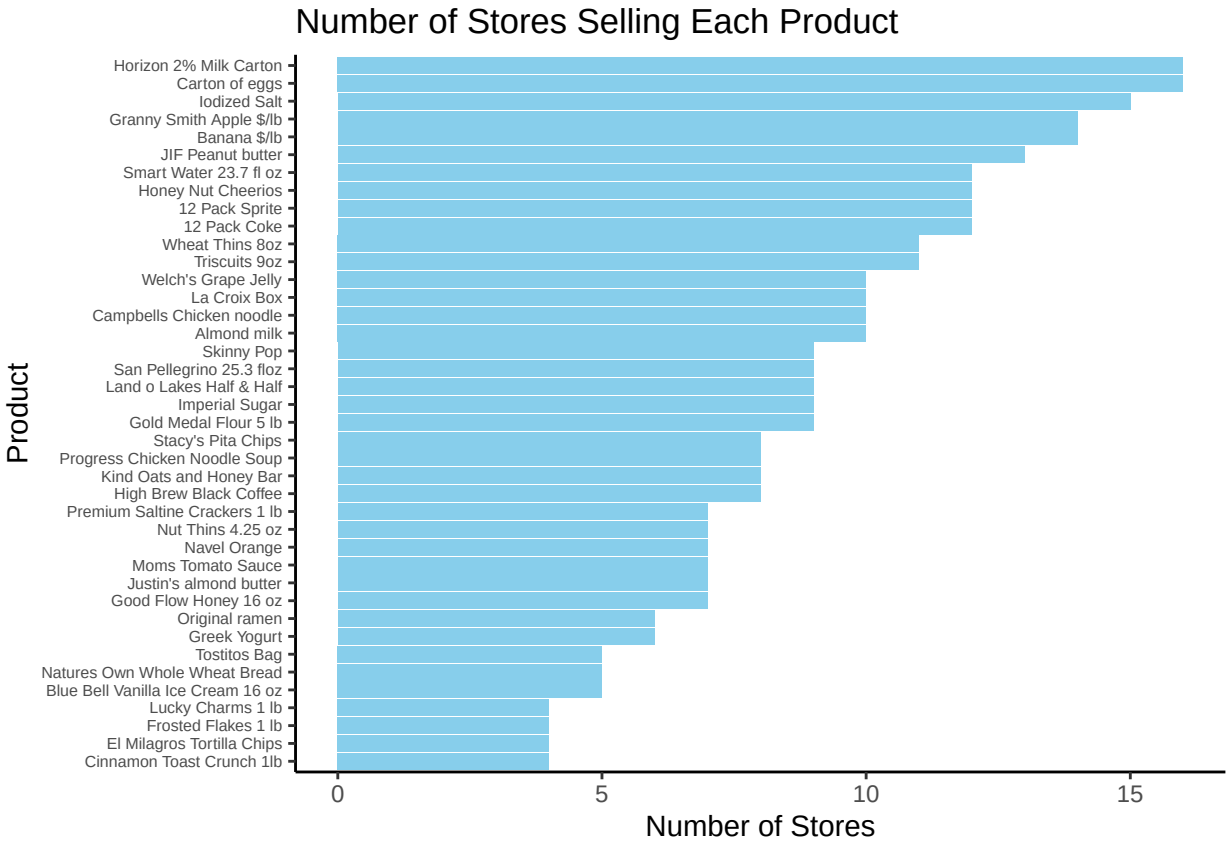
I would recommend a large opening size and a thick solder alloy because it would result in the minimum number of solder skips (0.39).

Problem 2:Grocery Store Prices

Part A

This bar graph shows the relationship of the average price of products at different stores, sorted from lowest to highest.

Part B



This bar graph shows the number of stores selling each product, with the products sold in the most stores at the top, such as milk and eggs, and the products sold in the least stores at the bottom, such as Cinnamon Toast Crunch.

Part C

Compared with ordinary grocery stores (like Albertsons, HEB, or Krogers), convenience stores charge somewhere between 0.41 and 0.92 dollars more for the same product.

Part D

The two stores that seem to charge the lowest prices when comparing the same products is Walmart and Kroger Fresh Fare. Wheatsville Food Co-Op and Whole Foods seem to charge the highest prices when comparing the same products.

Part E

Based on the evidence, it seems more plausible that Central Market charges a similar amount to HEB for the same product. The coefficient for Central Market is -0.57 with a 95% confidence interval of (-0.92, -0.23). The coefficient for HEB is -0.65 with a 95% confidence interval of (-0.95, -0.35). the difference in the estimated coefficients is only 0.08, and there is major overlap between the two confidence intervals. Additionally, the difference between these two grocery stores is much smaller than another pair of stores such as Walmart and Whole Foods. Walmart has an estimated coefficient of -0.99 while Whole Foods has an estimated coefficient of 0.36, making a difference of 1.35. Therefore it seems more likely that the two stores charge similar prices for the same products.

Part F

There seems to be no difference between consumer in rich and poor ZIP codes. the coefficient is only -0.01 and 0 is included in the confidence interval. This suggests that prices are not different based on the income of the ZIP code, and if there is, then it is extremely minor, being an increase of one cent per $10000. A one-standard deviation increase in the income of a ZIP code seems to be associated with a -0.03 standard-deviation change in the price that consumers in that ZIP code expect to pay for the same product.

Problem 3

A. True, figure A1 shows a line of best fit which shows that the number of FAIR policies increases when the % minority increases. Additionally, the slope is positive when a linear regression model is ran for it with a 95% confidence interval of (0.009, 0.018). Since 0 is not in this interval, we can conclude that ZIP codes with a higher percentage of minority residents tend to have more FAIR policies per 100 housing units.

B. Undecidable. While figure B1 does show a correlation between housing built before WWII and the minority percentage, there is no way to know how this affects FAIR policies. The regression model also doesn't take FAIR policies into consideration. We would need a regression model that takes minority:age into consideration to determine whether minority percentage and age of the housing stock would relate to the number of FAIR policies.

C. False. I would say that the relationship between minority percentage and number of FAIR policies per 100 housing units is the same in high fire risk ZIP codes and low fire risk ZIP codes. This is because in the regression model for figure C1, we can see that the estimate for the interaction variable between minority and low fire risk ZIP codes is -0.001 with a 95% confidence interval that includes 0, (-0.012, 0.01). This means that we would not be able to conclude that there is a significant difference between minority percentage and the number of FAIR housing policies in high fire risk ZIP codes.

D. False. I would say that income does not "explain away: all the association between minority percentage and FAIR policy uptake. This is because when we compare the confidence intervals of the minority coefficient between the regression models that takes income into account and the regression model that doesn't take income into account, the 95% confidence intervals overlap heavily. For income not being taken into account it is (0.009, 0.018) and when income is taken into account it is (0.004, 0.015). While the coefficient is lower when income is taken into account, it would need to go to 0 for income to"explain away" everything.

E. True. Regression model E still has a positive coefficient for minority when every other variable is also taken into account. The 95% confidence interval for minority percentage is (0.003, 0.014), which does not include 0, so there is still a relationship between minority percentage and FAIR policies.