Homework 9

Arun Mahadevan Sathia Narayanan (as235872)

2025-04-21

GitHub Link:
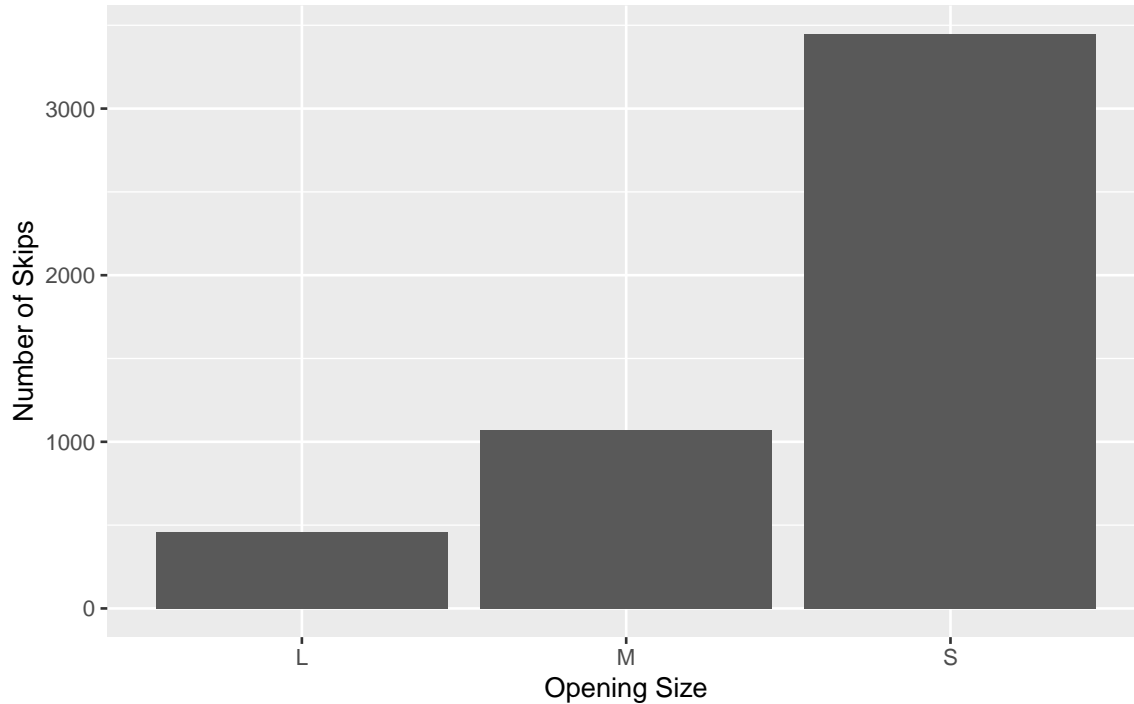To GitHub

GitHub Link (Text Format):
https://github.com/arunmsn/SDS315/tree/main/Week_14/Homework_9
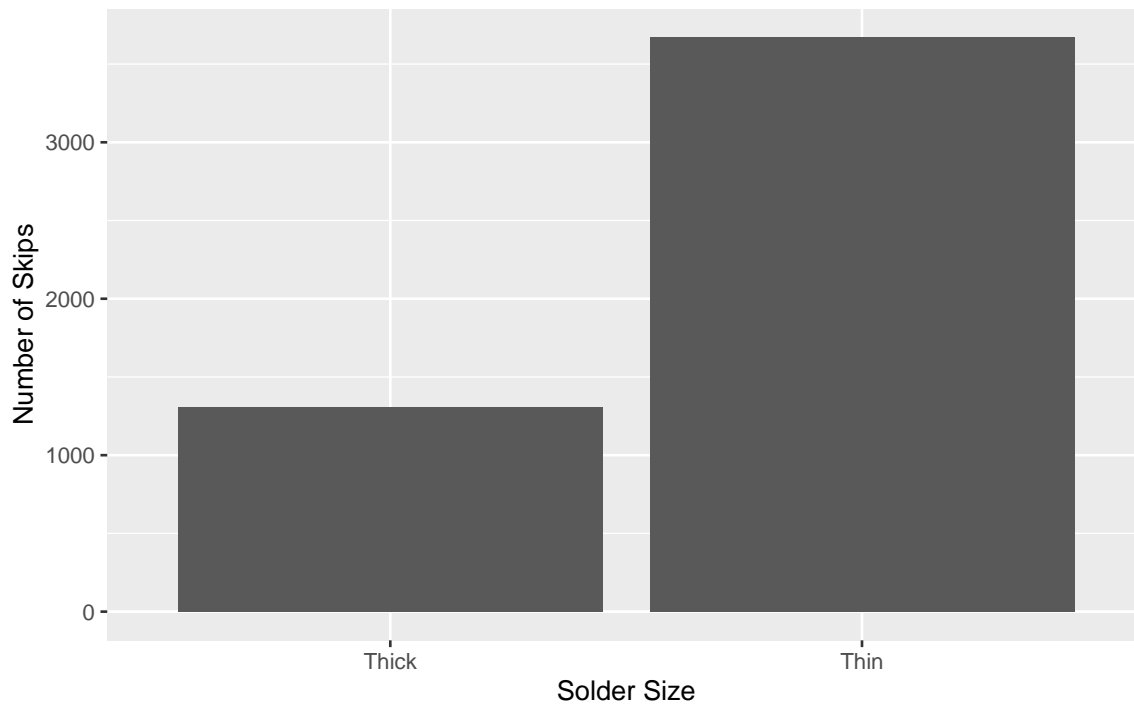
Problem 1 - Manufacturing Flaws in Circuit Boards

1 - Part A

Relationship between Opening size and Number of Skips

Above, we can see that the larger the smaller the opening size, the more number of skips observed.

Relationship between Solder size and Number of Skips

Above, we can see that the thinner solder sizes have a higher number of skips compared to the thicker solder sizes.

1 - Part B

```
##                     name   estimate      lower       upper level
## 1              Intercept  0.3933333  0.2391982   0.6026513  0.95
## 2               OpeningM  2.4066667  1.7279020   3.1051850  0.95
## 3               OpeningS  5.1266667  4.2536780   6.1701060  0.95
## 4              SolderThin 2.2800000  1.6590894   2.8991832  0.95
## 5 OpeningM.SolderThin    -0.7400000 -1.9967406   0.4463558  0.95
## 6 OpeningS.SolderThin     9.6533333  7.3813605  11.9961120  0.95
```

1 - Part C

- The Intercept is the baseline effect of OpeningL and SolderThick, and is 0.39 skips.

- The main, offset effect for the OpeningM variable is 2.41 skips.

- The main, offset effect for the OpeningS variable is 5.13 skips.

- The main, offset effect for the SolderThin variable is 2.28 skips.

- The interaction effect for OpeningM and SolderThin is -0.74 skips.

- The interaction effect for OpeningS and SolderThin is 9.65 skips.
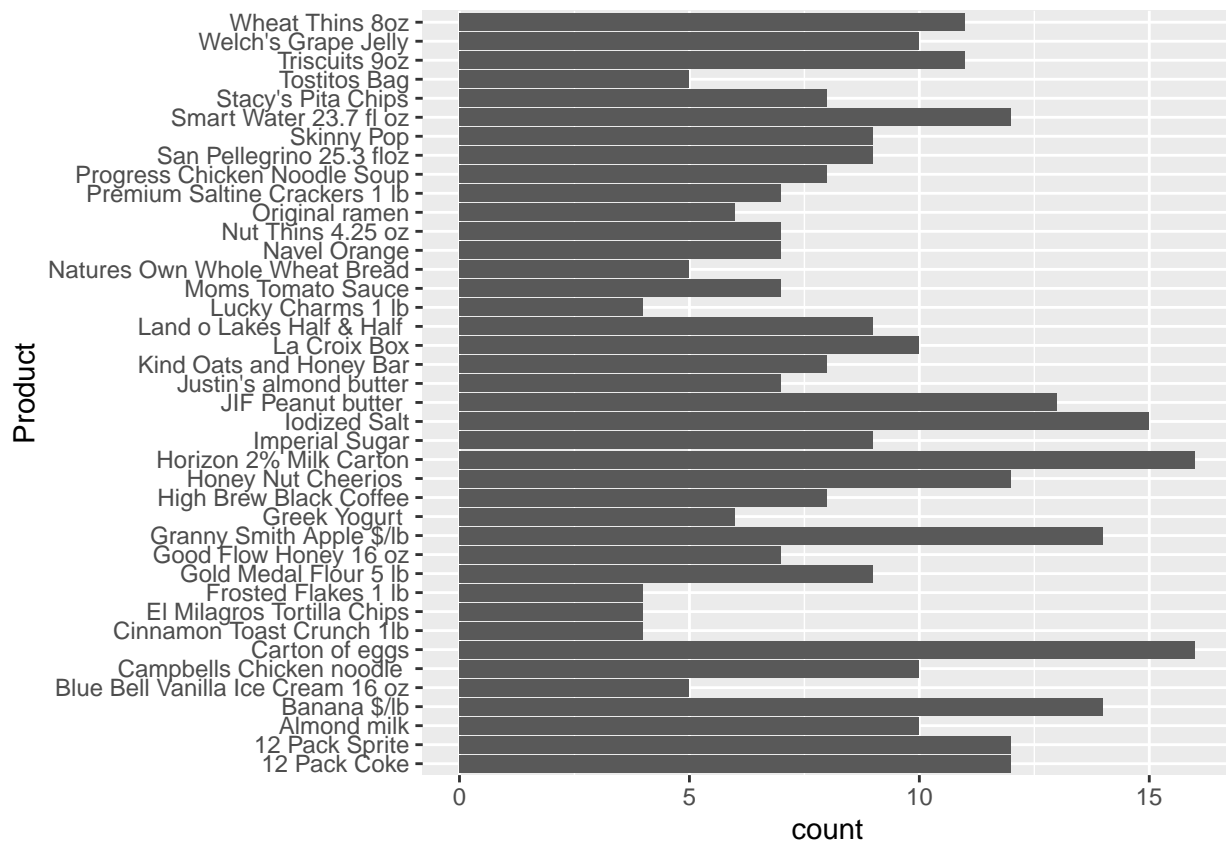
1 - Part D

To minimize the number of skips, I would use a Medium Opening (OpeningM) and Thin Solder size (SolderThin). This is because the minimum number of skips occur when having a medium opening and a thin solder, getting an estimated number of skips of -0.74.

Problem 2 - Grocery Store Prices

2 - Part A

Average Price per Store

A comparison of the average Price of all products, grouped by Store

2 - Part B

2 - Part C

Compared with ordinary grocery stores (like Albertsons, HEB, or Krogers), convenience stores charge somewhere between 5.47 and 6.38 dollars more for the same product.

2 - Part D

The two stores that charge the lowest price when comparing the same product is Walmart and Kroger Fresh Fare. The two stores that charge the highest price when comparing the same product are Wheatsville Food Co-Op and Whole Foods.

2 - Part E

I am using the confidence intervals for the comparisons here. Comparing the values between H-E-B and Central Market (as seen above), the 2.5% values have a difference of 2 cents and the 97.5% values have a difference of 12 cents. While the lower edges (2.5%) are closer together the higher edges (97.5%) are further apart, indicating that **Central Market charges more than H-E-B for the same product**. But let's not immediately determine that answer – we should compare with other stores:

Comparing Kroger Fresh Fare and H-E-B , we get a difference of -42 for the 2.5% and a difference of -10 for the 97.5%.
Comparing Kroger and H-E-B , we get a difference of -22 for the 2.5% and a difference of 10 for the 97.5%.
Comparing Fiesta and H-E-B , we get a difference of -29 for the 2.5% and a difference of 17 for the 97.5%.
Comparing Target and H-E-B , we get a difference of 20 for the 2.5% and a difference of 35 for the 97.5%.
Performing the comparisons, having the other stores being around H-E-B and Central Market in terms of the percentages, we can tell that the range of the difference between Central Market and H-E-B is not as large, giving us the result that **Central Market charges a similar amount to HEB for the same product**.

2 - Part F

- The poorer ZIP codes pay more for the same product, on average, compared to the richer ZIP codes. Given the slope of -0.01 for the Income10K variable (from the list of coefficients) and the smaller number of richer ZIP codes in the dataset (through manual analysis, though can be done using a filter according to a threshold), the slope is still negative even when the higher incomes will disrupt the slope, which indicates the trend follows that prices are higher in the lower income ZIP codes and prices are lower in the higher income ZIP codes.

- The estimated size of the effect of Income10K on Price: 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 (taken from using standardize_parameters).

Problem 3 - Redlining

3 - Statement A - ZIP codes with a higher percentage of minority residents tend to have more FAIR policies per 100 housing units.

This is True; using Figure A1 and the simple linear regression model fitted to the points, we can see that the slope of the regression line is positive (0.51), indicating that with a higher % minority, the FAIR policies per 100 housing units increases.

3 - Statement B - The evidence suggests an interaction effect between minority percentage and the age of the housing stock in the way that these two variables are related to the number of FAIR policies in a ZIP code.

This is Undecidable; even when using the simple linear regression fit to minority and age, while we can see that their relationship is not very strong, there is no information provided that can prove there is an interaction, as there is no explicit interaction terms included in the models.

3 - Statement C - The relationship between minority percentage and number of FAIR policies per 100 housing units is stronger in high-fire-risk ZIP codes than in low-fire-risk ZIP codes.

This is True; looking at Figure C1, we can see that the slope of the line for High Fire Risk is slightly larger than the slope of the line for the Low Fire Risk, and even if that is only barely visible, when using the linear model fit with minority, fire_risk, and their interaction (minority:fire_risk), the R-squared coefficient is positive and more than 0.5 (giving us 0.59), showing that there is a stronger relationship between minority percentage and the number of FAIR policies for the high-fire-rizk ZIP codes compared to the low-fire-risk ZIP codes, even if barely.

3 - Statement D - Even without controlling for any other variables, income "explains away" all the association between minority percentage and FAIR policy uptake.

This is False; comparing the models model_D1 and model_D2, we can see how the coefficient for minority changes; when using the model and not controlling for any of the variables, minority has a coefficient of 0.014 and a p-value of 0 (which makes it statistically significant). However, when using the model and controlling for income, minority still has a coefficient of 0.01 and a p-value of 0.002 (which means it is still statistically significant). This means, even after controlling for income (and no other variable), it cannot "explain away" all the association between minority percentage and FAIR policy uptake, though some of it can be using the fact that the coefficient for minority dropped a bit (partial confounding).

3 - Statement E - Minority percentage and number of FAIR policies are still associated at the ZIP code level, even after controlling for income, fire risk, and housing age.

This is True; using model_E (with multiple predictors), we can see that the R-squared coefficient for the relationship between the number of FAIR policies and minority percentage even after controlling for income, fire-risk, and housing age is still 0.66, which means the two variables are still associated after the controlling of other variables.