Homework #9

Advay Vyas

4/21/25

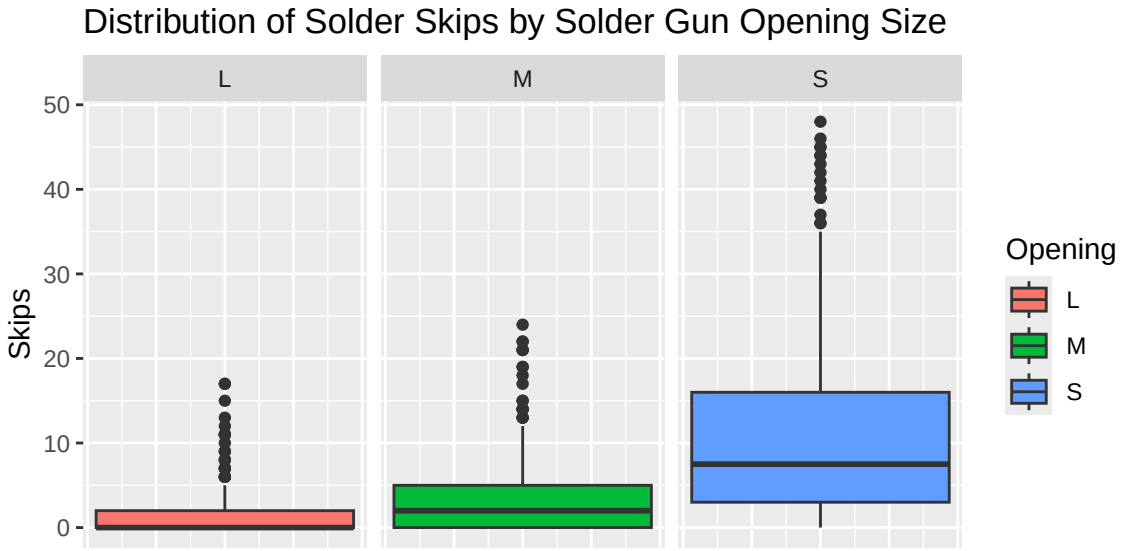
Contents

Introduction

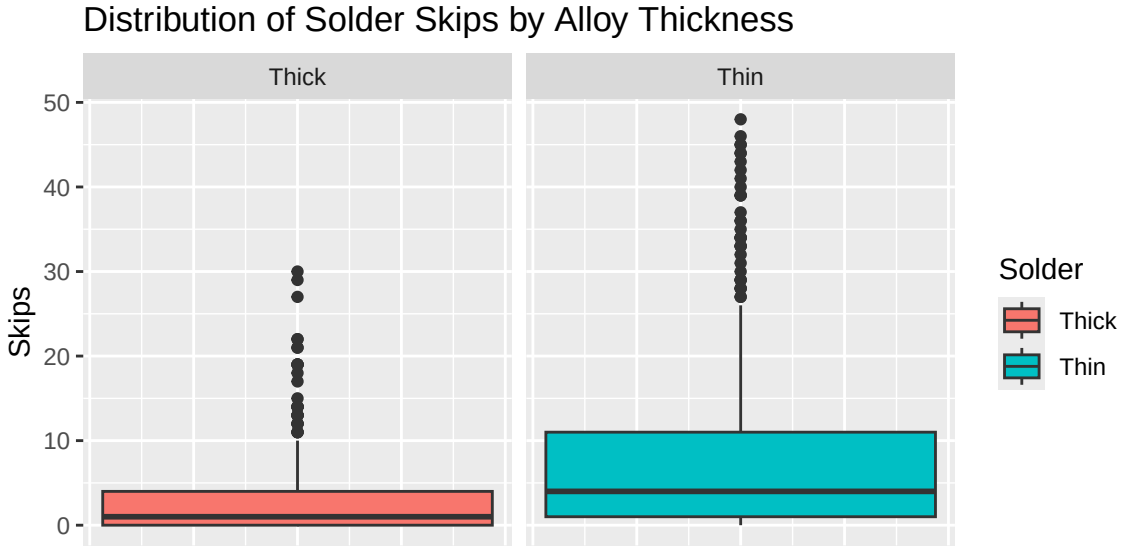
I'm Advay Vyas, EID: av37899, and this is my submission for SDS 315 Statistical Thinking Homework #9. The GitHub repository for my code is at this link.

Problem 1

Part A



From this plot, we can easily see that **the size of the opening positively affects the amount of solder skips** by examining the faceted box plots' vertical placement.



From this plot, we can see that **alloy thickness has an effect on the amount of solder skips**, with thinner alloys contributing to usually higher amounts of solder skips.

Part B

Table 1: Coefficients of the fitted linear regression model

Condition	Estimate	CI_Lower	CI_Upper
Baseline	0.395	0.23077	0.59396
SolderThin	2.281	1.70992	2.90189
OpeningMid	2.404	1.71335	3.12207
OpeningSmall	5.135	4.17674	6.15757
SolderThin:OpeningMid	-0.741	-1.95046	0.47085
SolderThin:OpeningSmall	9.640	7.34981	11.89140

Part C

If we round every coefficient to 0, we now create the equation below representing the amount of skips as $\hat{y}$ and the corresponding effect variables and y-intercept.

$$\begin{aligned}\hat{y} = \; & 0 + 2 \cdot \text{SolderThin} + 2 \cdot \text{OpeningMid} + 5 \cdot \text{OpeningSmall} \\ & - 1 \cdot \text{SolderThin} \cdot \text{OpeningMid} + 10 \cdot \text{SolderThin} \cdot \text{OpeningSmall}\end{aligned}$$

- The y-intercept indicates that the predicted skips for SolderThick & OpeningLarge is 0 (the baseline)
- The effect of the SolderThin variable is 2, so you add two skips when you have SolderThin
- Similarly, this means the effect of having the OpeningMid variable raises the prediction by 2
- Having OpeningSmall raises the predicted amount of skips by 5
- Having both SolderThin and OpeningMid together reduces the predicted amount of skips by 1
- Having both SolderThin and OpeningSmall greatly increases the predicted amount of skips by 10

Part D

We first have to tabulate every possible combination. For SolderThin, we immediately have 2 for the expected amount of skips. Considering small, medium, and large openings, we have that OpeningMid adds 2 and OpeningSmall adds 5 so 7, 4, 2 respectively. Considering the interactions, SolderThin and OpeningMid reduces the predicted skips by 1 and SolderThin and OpeningSmall increases the predicted skips by 10 resulting in 17, 3, 2 for SolderThin and OpeningSmall, OpeningMed, OpeningLarge respectively.

The calculations become much simpler for SolderThick due to no interaction variables. We simply add the effect variable to 0 since SolderThick has "no effect." With OpeningSmall (+5), OpeningMid (+2), and OpeningLarge (+0 b/c no effect variable and it is the baseline) and we get 5, 2, 0 respectively. The results are tabulated in the table below.

Table 2: Combinations of Solder Thicknesses and Solder Gun Openings and Predicted Skips

Solder	Opening	Skips
Thin	S	17
Thin	M	3
Thin	L	2
Thick	S	5
Thick	M	2
Thick	L	0

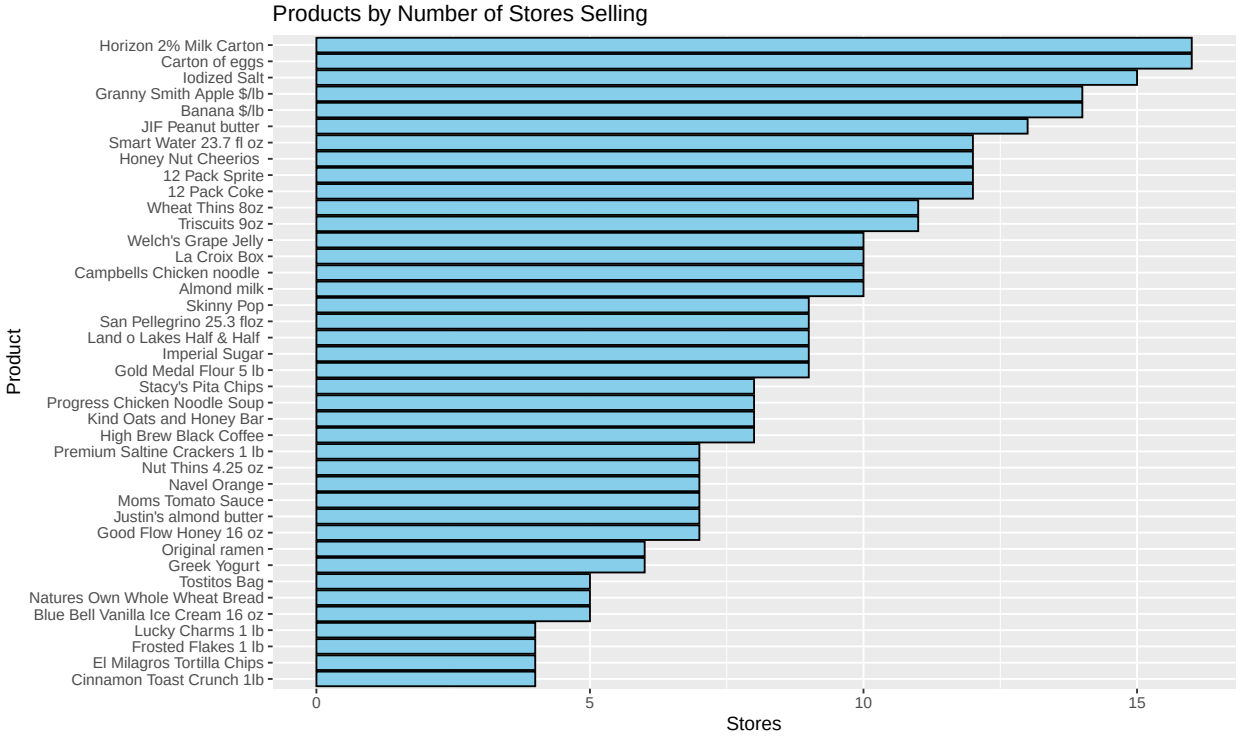
Since soldering skips are manufacturing defects, I'd like to recommend the combination with the least amount of defects which is **SolderThick and OpeningLarge**. This trend is also seen in the earlier plots, as a thick solder and a large solder gun opening each had the least average skips in their categories.

Problem 2

Part A

This graph above shows the average price of a arbitrary product by store.

Part B



This graph above shows the amount of stores that carry/sell each of these products.

Part C

Compared with ordinary grocery stores (like Albertsons, HEB, or Krogers), **convenience stores charge somewhere between $0.45 and $0.88 more for the same product**.

Part D

The two stores that charge the lowest prices for the same product are **Walmart (-0.99) and Kroger Fresh Fare (-0.90)**. The two stores that charge the highest prices for the same product are **Whole Foods (+0.36) and Wheatsville Food Co-Op (+0.29)**.

Part E

The coefficient for H-E-B is -0.65 while the coefficient for Central Market is -0.57, giving a difference of about 8 cents. If we compare this to other stores like the "cheap" Walmart & Kroger, we get about 9 cents. On the other hand, if we compare to the "expensive" Whole Fools & Wheatsville, those two stores have a difference of 7 cents. Even in close categories or "sides", we see that the differences are around 8 cents; from stores like Kroger to Whole Foods, the difference would easily exceed a dollar. Therefore, we can likely say that **Central Market charges similar prices to H-E-B for the same product**.

Part F

Consumers in poorer ZIP codes usually pay slightly more for a product on average because Income10K had a coefficient of -0.014 which indicates that for every $10,000 increase in average income for the ZIP code, the average price of products falls by a cent and a half.

The effect of this variable isn't terribly important as shown in the following sentence. A one-standard deviation increase in the income of a ZIP code seems to be associated with a 0.01 standard-deviation change in the price that consumers in that ZIP code expect to pay for the same product.

Problem 3

Statement A is **true.**

Statement B is **undecidable.**

Statement C is **true.**

Statement D is **false.**

Statement E is **true.**