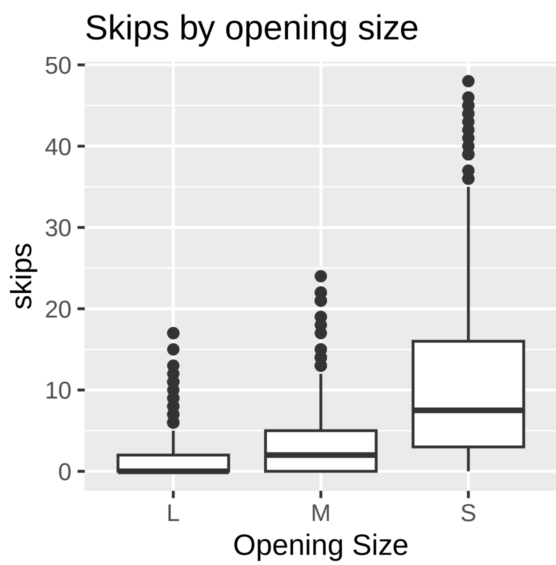
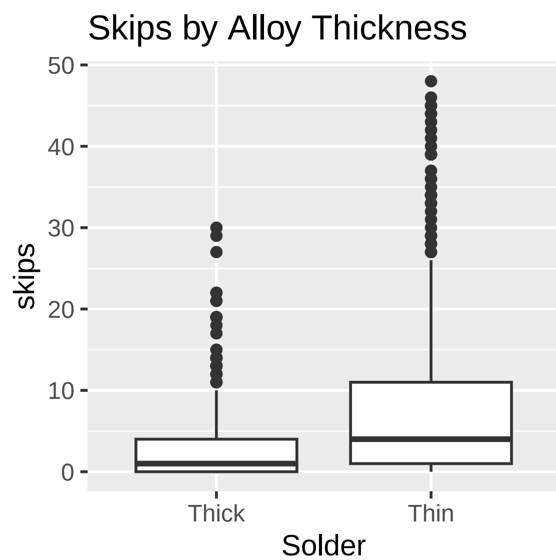
https://github.com/Panda-nny/SDS_315_HW9

Danny Pan

2025-04-20

Problem 1 - Regression Warm Up





According to the two plots above, a greater size of the opening on a solder gun is correlated to a smaller number of skips on circuit boards. Likewise, a thicker alloy used for soldering correlates to a smaller number of skips.

Table 1: Regression Model, Outcome = Skips, Baseline = Opening: L

term	estimate	lower_ci	upper_ci
intercept	0.393	-0.628	1.415
Opening: M	2.407	0.962	3.851
Opening: S	5.127	3.682	6.571
Solder: Thin	2.280	0.836	3.724
Opening: M:SolderThin	-0.740	-2.782	1.302
Opening: S:SolderThin	9.653	7.611	11.696

```
## # A tibble: 6 x 4
##   term                  estimate lower_ci upper_ci
##   <chr>                    <dbl>    <dbl>    <dbl>
## 1 intercept                0.393   -0.628     1.42
## 2 Opening: M               2.41     0.962     3.85
## 3 Opening: S               5.13     3.68      6.57
```

```
## 4 Solder: Thin              2.28     0.836     3.72
## 5 Opening: M:SolderThin   -0.74    -2.78      1.30
## 6 Opening: S:SolderThin    9.65     7.61     11.7
```

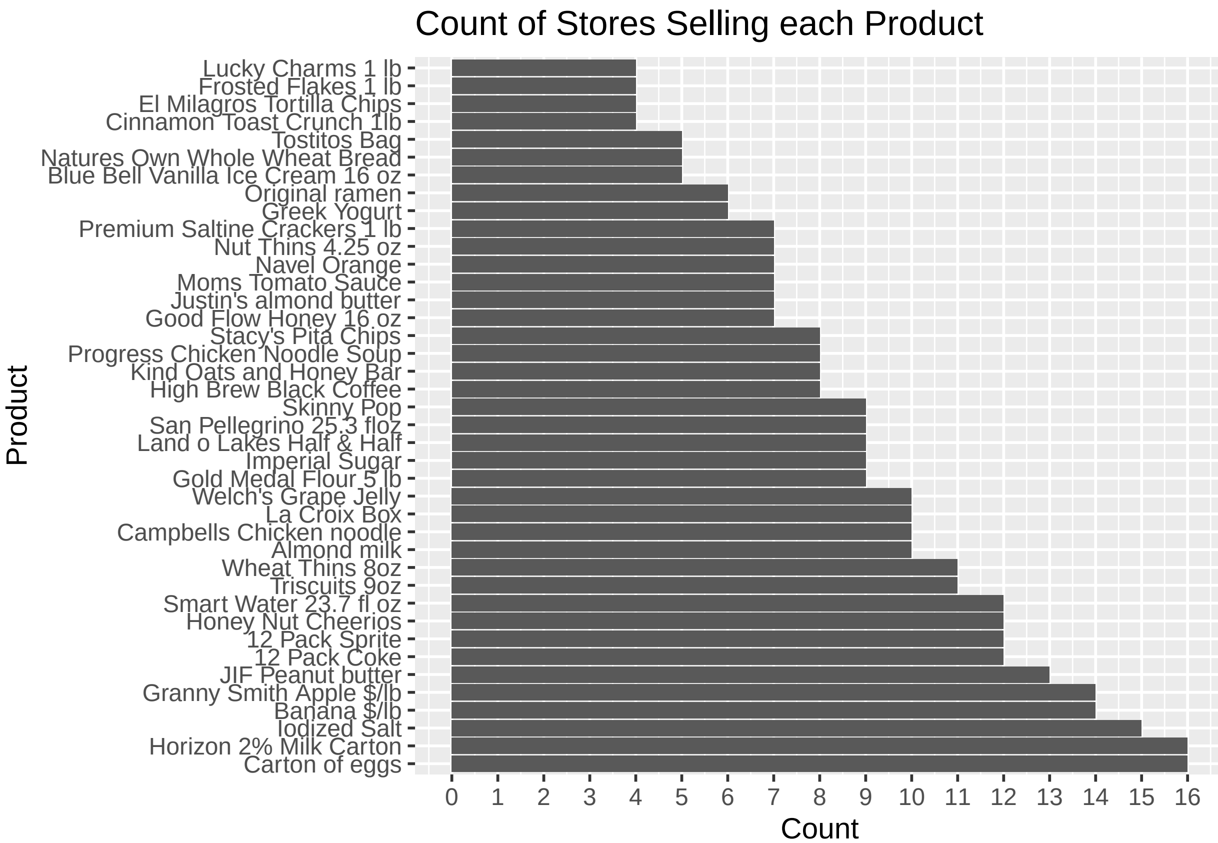
According to the table above, the baseline number of solder skips where a large opening for the solder gun used and a thick alloy used for soldering was about 0.393 with a 95% confidence interval of [-0.628, 1.415]. Comparatively, the effect of a medium and small solder gun opening was used was 2.407 and 5.127 more skips, with a 95% confidence interval of [0.962, 3.851] and [3.682, 6.571], respectively.

The interaction effect of a medium gun opening and thin alloys was estimated to be -0.740 with a 95% confidence interval of [-2.782, 1.302]. The interaction effect was significant for small gun openings and thing allows with a 95% confidence interval of [7.611, 11.696], centered around the estimate of 9.653 more skips.

Using these results, a medium gun opening and thin alloy would be recommended to minimizing the number of skips in the manufacturing process.

Problem 2 - Grocery Store Prices

The average price of products sold at different stores are shown above. Whole Foods appears to sell the most expensive products on average, with Fiesta being the least.



The above barplot displays the number of stores that offer each listed product. Accordingly, every store offered milk and eggs while only a small fraction of stores offered cereals in 1lb boxes such as Lucky Charms and Frosted Flakes.

According to a regression model, 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, with 95% confidence.

The regression model for Price versus Product and Store conveys that Whole Foods and Wheatsville Food Co-Op charge the most for the same product, while Walmart and Kroger Fresh Fare charges the least.

According to the same model, Central Market charges around the same price as HEB for the same product, with a 95% confidence interval of charging [-0.25, 0.40] dollars more. Since the confidence interval contains 0, we conclude that there's no evidence that Central Market charges more than HEB for the same product. Comparatively, Whole Foods, which we concluded earlier charges the most for the same product, has a 95% confidence interval of an offset of [0.686, 1.334] dollars, showing that Whole Foods definitely charges more than HEB for the same product.

The regression model for Price versus Product and Income suggests that consumers in poorer ZIP codes pay more for the same product, as the estimate of the offset is negative. However, the 95% confidence interval of [-0.03, 0.01] for the offset contains 0, suggesting that there may be no difference. The estimated effect size is low, also backed by the fact that a one-standard deviation increase in the income of a ZIP code seems to be associated with a [-0.07, 0.01] standard-deviation change in the price that consumers in that ZIP code expect to pay for the same product, with 95% confidence.

Problem 3 - Redlining

A. True- ZIP codes with a higher percentage of minority residents tend to have more FAIR policies per 100 housing units. This is reflected by the regression line in Figure A1, which has a slope of around 0.014 with a 95% confidence interval of [0.009, 0.018]. An r-squared value of 0.516 for model_A suggests that around 52% of variance in FAIR policies per housing unit can be explained by changes in the minority percentage of a ZIP code.

B. False- There is no evidence that age of housing stock relates to the number of FAIR policies in a ZIP code (model_E includes 0 for the 95% confidence interval of offset by age holding other variables constant), and no evidence is provided that there is an interaction effect between minority percentage and age of housing stock.

C. Uncertain– The offset of the interaction effect of minority percentage with lower fire risks on fair policies per 100 contains the value of 0 in the 95% confidence interval of [-0.012, 0.01], found in model_C. This means that lower or higher fire risk does not necessarily affect the relationship between minority percentage and on fair policies per 100.

D. False- model_D2 demonstrates that even without controlling for any other variables, the association between minority percentage and FAIR policy uptake is significant when income is held constant, with a 95% confidence interval of the effect of [0.004, 0.015].

E. True- 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 supported by the 95% confidence interval of [0.003, 0.014] of the additional effect of minority percentage keeping the other mentioned variables constant in model_E.