315HW9

Elan Satchit

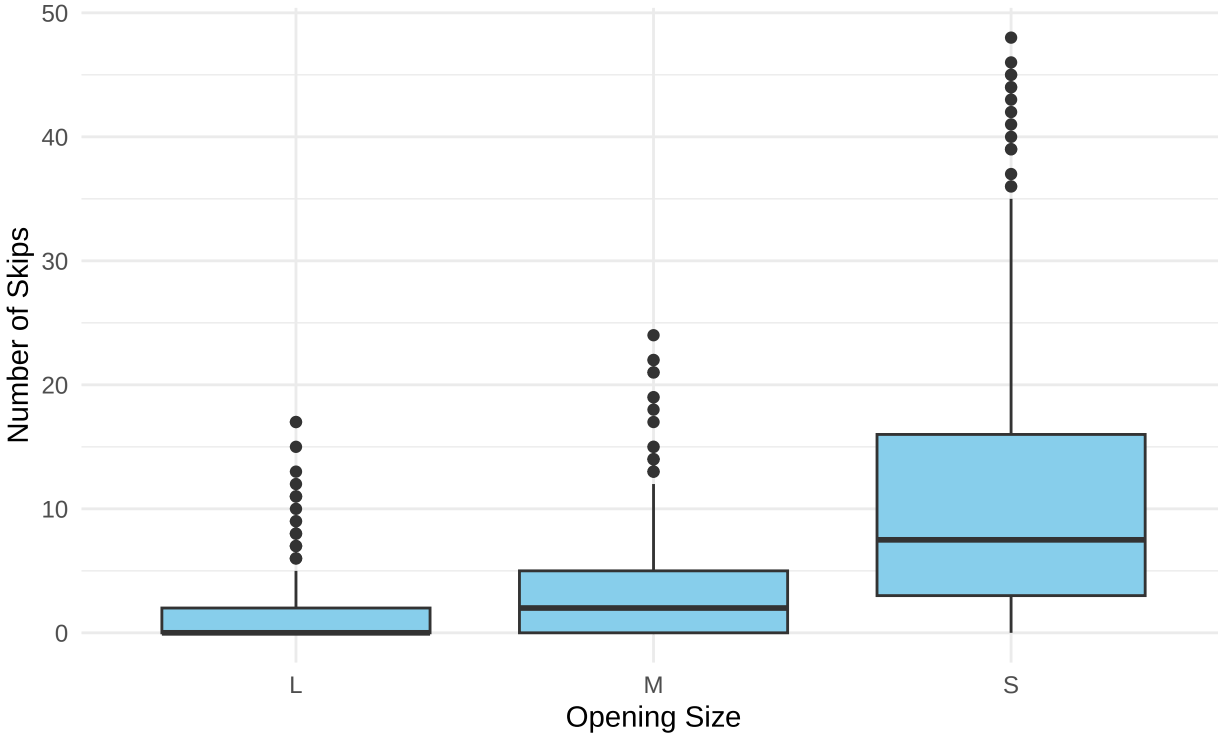
2025-04-21

GITHUB REPO

Problem 1

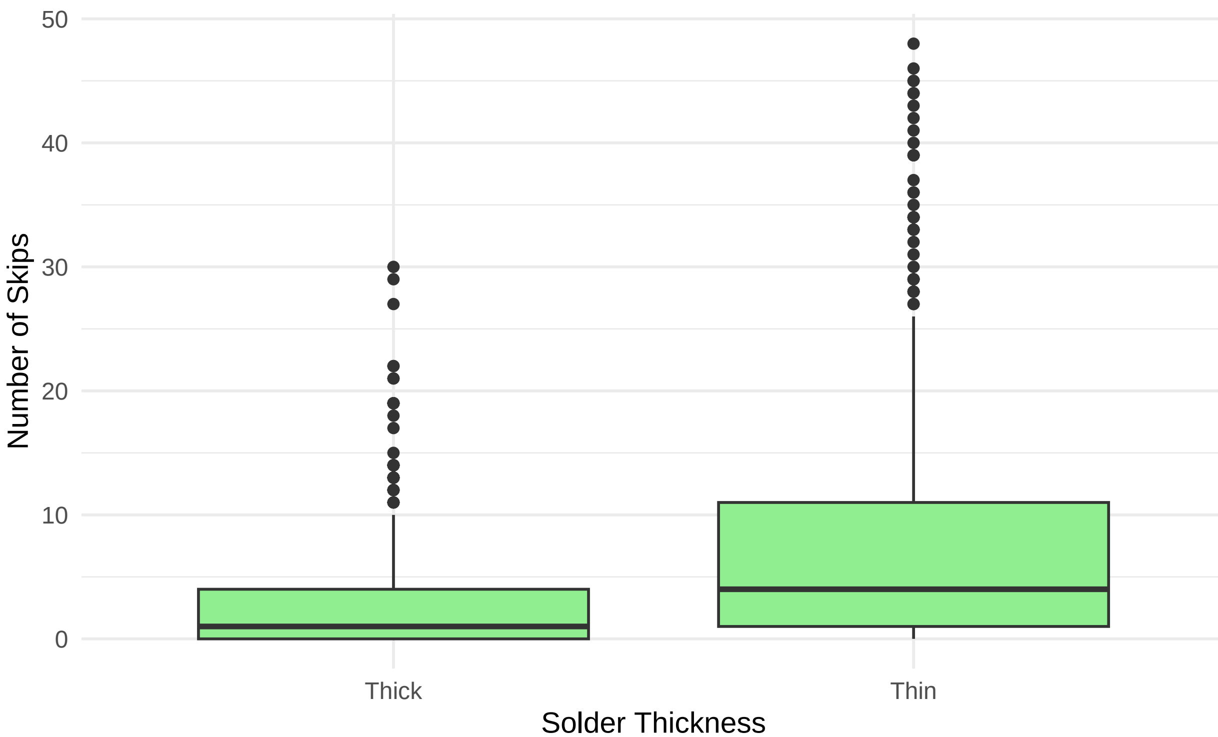
Part A

Number of Solder Skips by Opening Size of Solder Gun



This plot shows the distribution of solder skips between different solder gun openings. Large openings tend to result in fewer skips.

Number of Solder Skips by Solder Alloy Thickness



This boxplot compares solder skips between thick and thin solder alloy. Thicker solder also appears to reduce the number of skips.

Part B

```
## # A tibble: 6 x 7
##   term                   estimate std_error statistic p_value lower_ci upper_ci
##   <chr>                     <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept                 0.393     0.52      0.756   0.45    -0.628     1.42
## 2 Opening: M                2.41      0.736     3.27    0.001    0.962     3.85
## 3 Opening: S                5.13      0.736     6.97    0        3.68      6.57
## 4 Solder: Thin              2.28      0.736     3.10    0.002    0.836     3.72
## 5 Opening: M:SolderThin    -0.74      1.04     -0.711   0.477   -2.78      1.30
## 6 Opening: S:SolderThin     9.65      1.04      9.28    0        7.61     11.7
```

```
## # A tibble: 6 x 4
##   term                 estimate lower_ci upper_ci
##   <chr>                   <dbl>    <dbl>    <dbl>
## 1 (Intercept)             0.393   -0.628     1.41
## 2 OpeningM                2.41     0.962     3.85
## 3 OpeningM:SolderThin    -0.740   -2.78      1.30
## 4 OpeningS                5.13     3.68      6.57
## 5 OpeningS:SolderThin     9.65     7.61     11.7
## 6 SolderThin              2.28     0.836     3.72
```

Part C

(Intercept) = 0.393 For boards with Opening = L and Solder = Thick, the expected number of skips is approximately 0.39.

OpeningM = 2.41 For boards with Opening = M and Solder = Thick, the expected number of skips is 2.41 more than boards with Opening = L and Solder = Thick.

OpeningM:SolderThin = -0.74 The interaction term indicates that for boards with Opening = M and Solder = Thin, the effect of OpeningM is 0.74 skips fewer than expected from simply adding the separate effects of OpeningM and SolderThin.

OpeningS = 5.13 For boards with Opening = S and Solder = Thick, the expected number of skips is 5.13 more than boards with Opening = L and Solder = Thick.

OpeningS:SolderThin = 9.65 For boards with Opening = S and Solder = Thin, the interaction adds 9.65 additional skips, above and beyond the separate effects of OpeningS and SolderThin.

SolderThin = 2.28 For boards with Opening = L and Solder = Thin, the expected number of skips is 2.28 more than with Solder = Thick.

Part D

Based on the regression analysis, I recommend that AT&T use Opening = L and Solder = Thick in their manufacturing process. This combination corresponds to the baseline group in the model and resulted in the lowest number of skips on average. All other combinations of opening size and solder type led to a statistically significant increase in skips. In particular, smaller openings (M and S) and the use of thin solder were associated with higher skip counts, with the combination of Opening S and Thin Solder producing the highest number of skips overall. Therefore, to minimize defects and improve product quality, the optimal choice is to use large openings with thick solder.

Problem 2

Part A

This plot shows that some stores like Whole Foods charge more on average than others like Walmart or Fiesta.

Part B

Some products are sold at all 16 store locations (e.g., milk, eggs), while others appear in fewer.

Part C

```
## # A tibble: 44 x 7
##    term                     estimate std_error statistic p_value lower_ci upper_ci
##    <chr>                       <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
##  1 intercept                   5.26      0.226    23.3     0         4.82     5.70
##  2 Product: 12 Pack Spri~     -0.018     0.313    -0.059   0.953    -0.634    0.597
##  3 Product: Almond milk       -2.20      0.329    -6.70    0        -2.85    -1.56
##  4 Product: Banana $/lb       -4.88      0.303   -16.1     0        -5.48    -4.29
##  5 Product: Blue Bell Va~     -2.95      0.409    -7.20    0        -3.75    -2.14
##  6 Product: Campbells Ch~     -3.46      0.329   -10.5     0        -4.11    -2.82
##  7 Product: Carton of eg~     -3.00      0.293   -10.2     0        -3.58    -2.42
##  8 Product: Cinnamon Toa~     -0.997     0.444    -2.25    0.025    -1.87    -0.124
##  9 Product: El Milagros ~     -2.04      0.444    -4.59    0        -2.91    -1.16
## 10 Product: Frosted Flak~     -1.25      0.444    -2.82    0.005    -2.12    -0.379
## # i 34 more rows
```

```
## # A tibble: 1 x 3
##   term             `2.5 %` `97.5 %`
##   <chr>              <dbl>    <dbl>
## 1 TypeConvenience     0.41     0.92
```

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

Part D

```
## # A tibble: 13 x 7
##    term                    estimate std_error statistic p_value lower_ci upper_ci
##    <chr>                      <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
##  1 Store: Walmart            -0.993     0.233    -4.25    0        -1.45    -0.533
##  2 Store: Kroger Fresh F~    -0.902     0.233    -3.86    0        -1.36    -0.443
##  3 Store: Fiesta             -0.703     0.269    -2.61    0.01     -1.23    -0.173
##  4 Store: Kroger             -0.703     0.233    -3.01    0.003    -1.16    -0.244
##  5 Store: H-E-B              -0.646     0.152    -4.25    0        -0.945   -0.347
##  6 Store: Central Market     -0.573     0.177    -3.24    0.001    -0.922   -0.225
##  7 Store: Target             -0.373     0.19     -1.97    0.05     -0.747    0
##  8 Store: Natural Grocers    -0.081     0.198    -0.411   0.681    -0.47     0.308
##  9 Store: Fresh Plus         -0.036     0.162    -0.223   0.823    -0.355    0.283
## 10 Store: CVS                 0.193     0.183     1.06    0.292    -0.167    0.553
## 11 Store: Walgreens           0.215     0.181     1.19    0.234    -0.14     0.571
## 12 Store: Wheatsville Fo~     0.29      0.179     1.62    0.105    -0.061    0.642
## 13 Store: Whole Foods         0.364     0.177     2.06    0.04      0.017    0.712
```

Whole Foods and Wheatesville Food Co-op seem to charge the highest prices when comparing the same product.

Part E

```
## # A tibble: 2 x 7
##   term                    estimate std_error statistic p_value lower_ci upper_ci
##   <chr>                      <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 Store: Central Market     -0.573     0.177     -3.24   0.001   -0.922   -0.225
## 2 Store: H-E-B              -0.646     0.152     -4.25   0       -0.945   -0.347
```

The data suggest that Central Market charges slightly more than H-E-B for the same product — approximately 7 cents more on average. However, this difference is small in the context of the overall variation in store pricing. For example, the difference between Walmart and Whole Foods exceeds $1.35, so the Central Market vs. H-E-B gap is relatively minor.

Part F

```
## # A tibble: 41 x 7
##    term                    estimate std_error statistic p_value lower_ci upper_ci
##    <chr>                      <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
##  1 intercept                  5.62      0.249    22.6     0         5.13     6.11
##  2 Product: 12 Pack Spri~    -0.018     0.328    -0.056   0.955    -0.664    0.627
##  3 Product: Almond milk      -2.11      0.345    -6.13    0        -2.79    -1.44
##  4 Product: Banana $/lb      -4.91      0.316   -15.5     0        -5.53    -4.29
##  5 Product: Blue Bell Va~    -2.91      0.429    -6.78    0        -3.75    -2.06
##  6 Product: Campbells Ch~    -3.37      0.345    -9.78    0        -4.05    -2.70
##  7 Product: Carton of eg~    -2.97      0.307    -9.68    0        -3.58    -2.37
##  8 Product: Cinnamon Toa~    -1.20      0.465    -2.57    0.011    -2.11    -0.281
##  9 Product: El Milagros ~    -2         0.464    -4.31    0        -2.91    -1.09
## 10 Product: Frosted Flak~    -1.45      0.465    -3.12    0.002    -2.36    -0.536
## # i 31 more rows
```

```
## # A tibble: 1 x 7
##   term             estimate std_error statistic p_value lower_ci upper_ci
##   <chr>               <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 scale(Income10K)   -0.032     0.022     -1.46   0.144   -0.074    0.011
```

The standardized coefficient for Income10k is -0.03. This means a one-standard deviation increase in the income of a ZIP code is associated with a 0.03 standard-deviation decrease in the price that consumers in that ZIP code pay for the same product. Since the coefficient is negative, consumers in poorer ZIP codes pay slightly more on average for the same product. However, the effect size is very small (-0.03 SD), suggesting that while the pattern exists, it's not a strong one.

Problem 3

A. True. Figure A1 shows a clear upward trend: as the percentage of minority residents goes up, so does the number of FAIR policies. In model_A, the coefficient for minority is positive (0.014) and statistically significant ($p < 0.001$), with a 95% confidence interval from 0.009 to 0.018. The R^2 is 0.516, meaning minority percentage alone explains over half the variation in FAIR policy use.

B. Undecidable. None of the models include an interaction between minority and age, so we can't say for sure. Figure B1 shows a slight trend between housing age and minority percentage, but the regression model (model_B) shows the relationship isn't statistically significant ($p = 0.125$) and the R^2 is very low (0.06). We would need a model that directly includes the interaction to know more.

C. True. Figure C1 shows a steeper trend for high-fire-risk ZIP codes than for low-risk ones. In model_C, the main effect of minority is significant (estimate = 0.01, $p = 0.015$). While the interaction term with fire risk (minority:fire_riskLow) isn't significant, the figure does suggest the relationship is stronger in high-risk areas.

D. False. Controlling for income in model_D2 reduces the effect of minority, but doesn't remove it. The minority coefficient is still significant (estimate = 0.010, $p = 0.002$), even after adjusting for income. So, income matters, but it doesn't explain away the relationship.

E. True. In model_E, the coefficient for minority is still positive and statistically significant (estimate = 0.008, $p = 0.006$), even after controlling for income, fire, and age. That means minority percentage is still linked to higher FAIR policy use, even when adjusting for these other factors.