

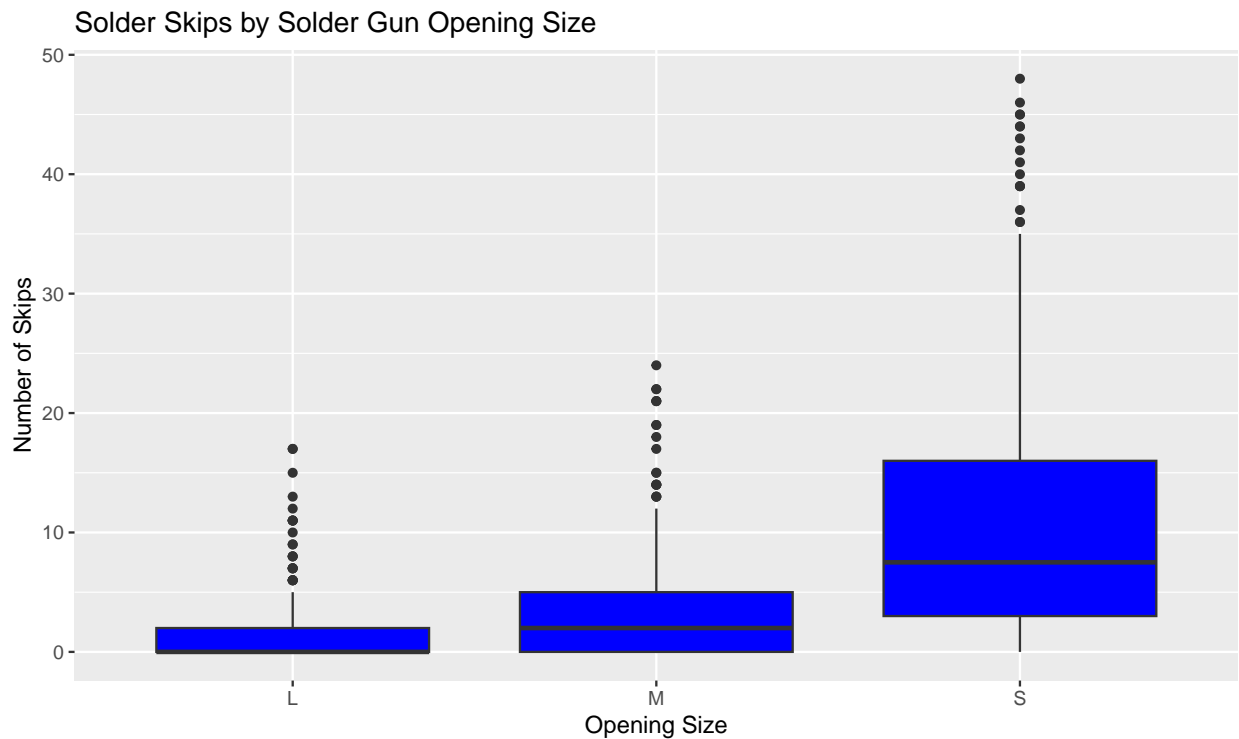
Homework 9

Pranav Belligundu(psb898) - SDS 315 UT Austin

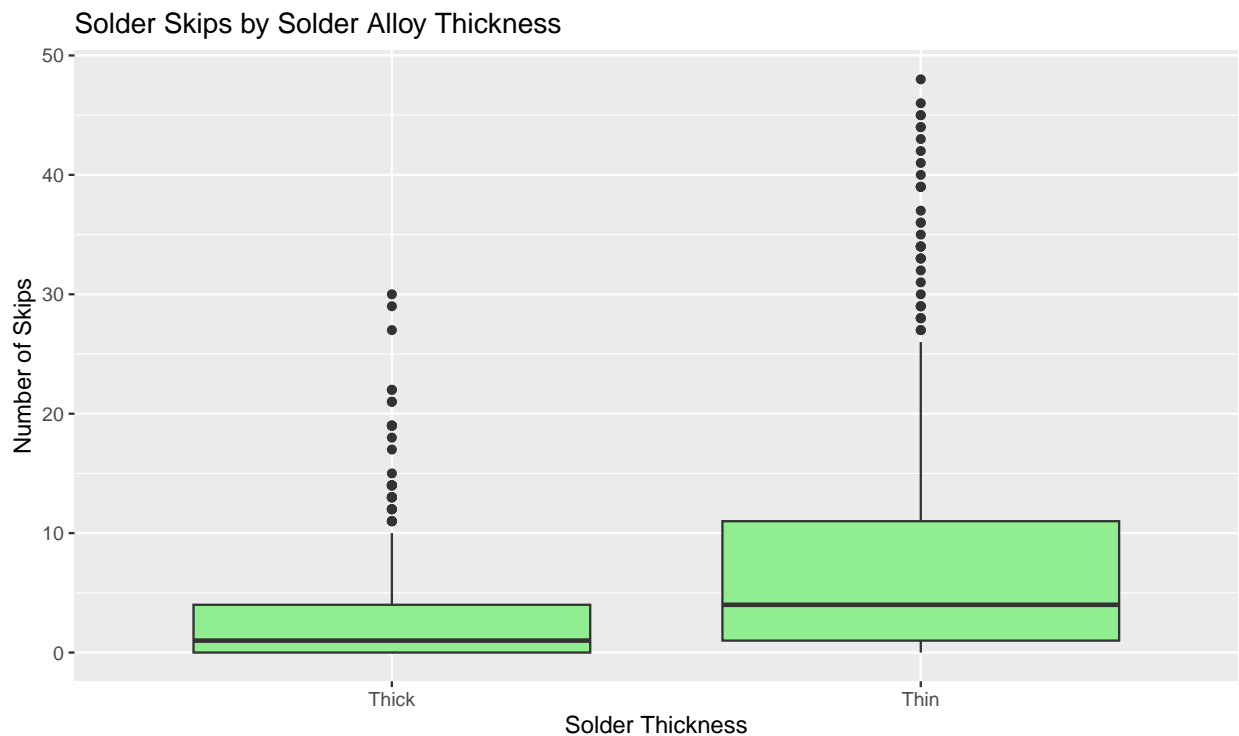
<https://github.com/pranav-B21/SDS-315/tree/main/HW9>

Problem 1: Manufacturing flaws in circuit boards

A. Make 2 plots



This plot shows that smaller openings tend to result in fewer skips, while larger openings have higher median skips and more variability.



This plot indicates that thin solder tends to result in fewer skips on average compared to thick solder.

Part B: Build a regression model with skips

##	Estimate	2.5 %	97.5 %
## (Intercept)	0.393	-0.628	1.415
## OpeningM	2.407	0.962	3.851
## OpeningS	5.127	3.682	6.571
## SolderThin	2.280	0.836	3.724
## OpeningM:SolderThin	-0.740	-2.782	1.302
## OpeningS:SolderThin	9.653	7.611	11.696

Part C: Interpretation

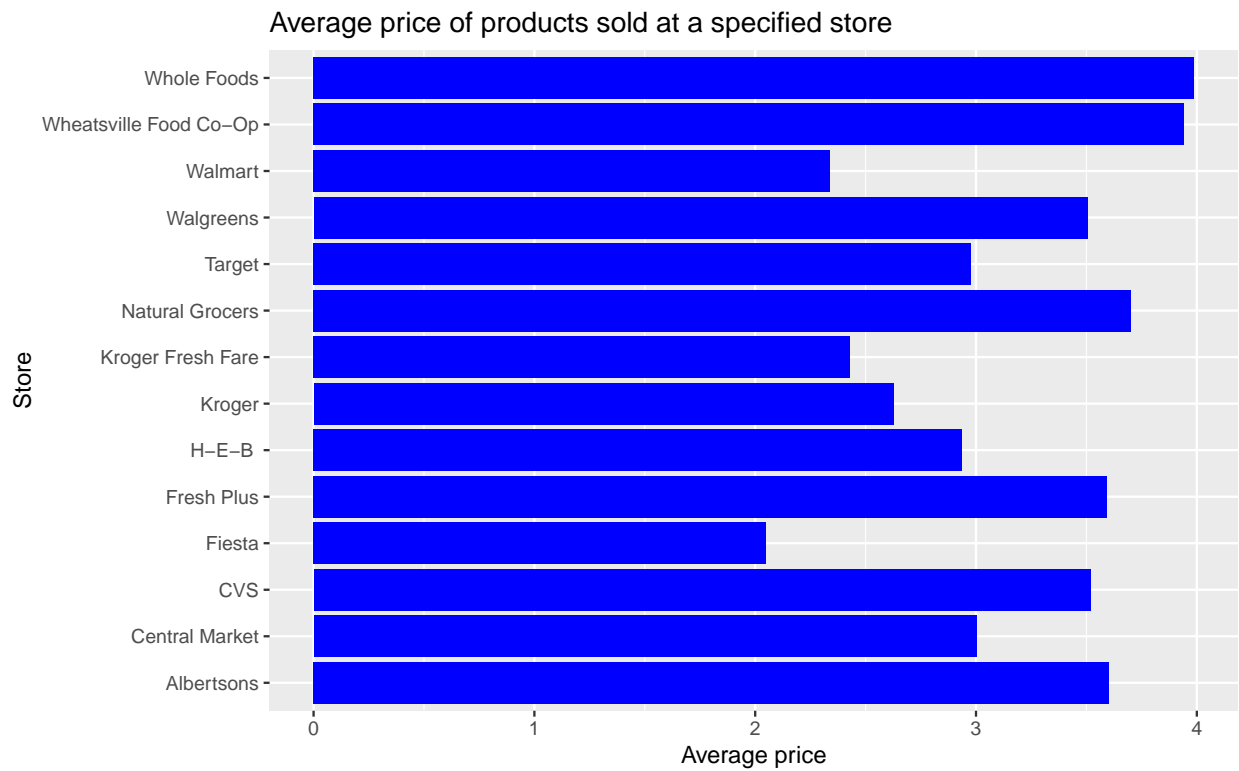
- Intercept: When the Opening is at its baseline level (likely “L”), and the Solder type is “Thick” (also the baseline), the average number of skips is estimated to be 0.393
- OpeningM: for boards made with Opening “M” and Solder “Thick”, the average number of skips is 2.407 higher than those with Opening “L” and Solder “Thick”
- OpeningS: For boards made with Opening “S” and Solder “Thick”, the average number of skips is 5.127 higher than the Opening “L” and Solder “Thick” baseline
- SolderThin: For boards using Solder “Thin”, and Opening “L”, the average number of skips is 2.280 higher
- OpeningM:SolderThin: The interaction between Opening “M” and Solder “Thin” indicates that the effect of using Solder “Thin” is 0.740 fewer skips
- OpeningS:SolderThin: The interaction between Opening “S” and Solder “Thin” suggests that their combined effect produces 9.653 more skips.

Part D: New Combination

Use Opening “L” and Solder “Thick” as the 2 baselines results in the lowest number of skips and is recommended. This baseline setup has the smallest estimated defect count 0.393. To minimize flaws and ensure higher manufacturing quality, Opening “L” with Thick Solder is the most reliable choice.

Problem 2: Manufacturing flaws in circuit boards

A. Make a plot to describe price differences across stores



B. Make a plot to determine number of stores selling each product



C. Model between Price and Product + Type

```
## # 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"           5.92      0.24     24.4      0      5.45      6.4
## 2 "Product: 12 Pack Spr~ -0.02     0.31     -0.06    0.95    -0.63      0.6
## 3 "Product: Almond milk" -2.2       0.33     -6.7      0     -2.85     -1.56
## 4 "Product: Banana $/lb" -4.88      0.3     -16.2     0     -5.48     -4.29
## 5 "Product: Blue Bell V~ -2.95      0.41     -7.2      0     -3.75     -2.14
## 6 "Product: Campbells C~ -3.46      0.33    -10.5     0     -4.11     -2.82
## 7 "Product: Carton of e~ -3         0.29    -10.2     0     -3.58     -2.42
## 8 "Product: Cinnamon To~ -1         0.44     -2.25    0.03    -1.87     -0.12
## 9 "Product: El Milagros~ -2.04      0.44     -4.59     0     -2.91     -1.17
## 10 "Product: Frosted Fla~ -1.25      0.44     -2.82    0.01    -2.12     -0.38
## # i 34 more rows
```

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. Because TypeConvenience is the baseline, the row labeled TypeGrocery tells us the price difference between grocery stores and convenience stores.

D. Model between Price and Product + Store

```
## # A tibble: 53 x 7
##   term                estimate std_error statistic p_value lower_ci upper_ci
```

```
##      <chr>                <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 "intercept"              5.75      0.23     24.6      0        5.29     6.21
## 2 "Product: 12 Pack Spr~   -0.02     0.29    -0.06     0.95    -0.58     0.55
## 3 "Product: Almond milk"   -2.29      0.3     -7.56      0       -2.89    -1.7
## 4 "Product: Banana $/lb"   -4.86      0.28   -17.5       0       -5.4     -4.31
## 5 "Product: Blue Bell V~   -3.08      0.38    -8.17      0       -3.82    -2.34
## 6 "Product: Campbells C~   -3.55      0.3    -11.7       0       -4.15    -2.96
## 7 "Product: Carton of e~   -2.98      0.27   -11.1       0       -3.51    -2.45
## 8 "Product: Cinnamon To~   -1.17      0.41    -2.87      0       -1.98    -0.37
## 9 "Product: El Milagros~   -1.91      0.41    -4.64      0       -2.72    -1.1
## 10 "Product: Frosted Fla~  -1.43      0.41    -3.49      0       -2.23    -0.62
## # i 43 more rows
```

```
## # A tibble: 2 x 5
##   term                estimate std.error statistic  p.value
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>
## 1 StoreWalmart      -0.993    0.233    -4.25 0.0000279
## 2 StoreKroger Fresh Fare -0.902    0.233    -3.87 0.000136
```

```
## # A tibble: 2 x 5
##   term                estimate std.error statistic p.value
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>
## 1 StoreWheatsville Food Co-Op 0.290    0.179     1.62 0.105
## 2 StoreWhole Foods            0.364    0.177     2.06 0.0400
```

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

E. Price Discrimination

Central Market charges a similar amount to HEB for the same product.

The model shows that Central Market charges only about 7 cents more per item than HEB on average. This conclusion comes from comparing their store coefficients: HEB's is -0.65 and Central Market's is -0.57 , resulting in a modest price difference of \$0.08.

When placed in context, this gap is relatively minor compared to other store comparisons—for instance, Walmart prices are around 99 cents lower, while Whole Foods charges approximately 36 cents more than the baseline. These findings suggest that Central Market's higher-end reputation likely stems from the types of products it offers, not significantly higher prices for the same items.

F. Income

```
## # 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      0.25     22.6      0        5.13     6.11
## 2 "Product: 12 Pack Spr~ -0.02     0.33    -0.06     0.96    -0.66     0.63
## 3 "Product: Almond milk" -2.11      0.34    -6.13      0       -2.79    -1.44
## 4 "Product: Banana $/lb" -4.91      0.32   -15.5       0       -5.53    -4.29
## 5 "Product: Blue Bell V~ -2.91      0.43    -6.78      0       -3.75    -2.06
## 6 "Product: Campbells C~ -3.37      0.34    -9.79      0       -4.05    -2.69
```

```
## 7 "Product: Carton of e~ -2.97 0.31 -9.69 0 -3.58 -2.37
## 8 "Product: Cinnamon To~ -1.2 0.46 -2.57 0.01 -2.11 -0.28
## 9 "Product: El Milagros~ -2 0.46 -4.31 0 -2.91 -1.09
## 10 "Product: Frosted Fla~ -1.45 0.46 -3.12 0 -2.36 -0.54
## # i 31 more rows
```

```
## # Standardization method: refit
```

```
##
```

## Parameter	Std. Coef.	95% CI
## (Intercept)	1.08	[0.86, 1.31]
## Product [12 Pack Sprite]	-9.03e-03	[-0.33, 0.31]
## Product [Almond milk]	-1.04	[-1.37, -0.71]
## Product [Banana \$/lb]	-2.42	[-2.72, -2.11]
## Product [Blue Bell Vanilla Ice Cream 16 oz]	-1.43	[-1.85, -1.02]
## ProductCampbells Chicken noodle	-1.66	[-1.99, -1.33]
## Product [Carton of eggs]	-1.46	[-1.76, -1.17]
## Product [Cinnamon Toast Crunch 1lb]	-0.59	[-1.04, -0.14]
## Product [El Milagros Tortilla Chips]	-0.98	[-1.43, -0.53]
## Product [Frosted Flakes 1 lb]	-0.71	[-1.16, -0.26]
## Product [Gold Medal Flour 5 lb]	-1.03	[-1.38, -0.69]
## Product [Good Flow Honey 16 oz]	0.52	[0.15, 0.89]
## Product [Granny Smith Apple \$/lb]	-1.85	[-2.15, -1.54]
## ProductGreek Yogurt	-1.93	[-2.32, -1.54]
## Product [High Brew Black Coffee]	-1.39	[-1.75, -1.03]
## ProductHoney Nut Cheerios	-0.83	[-1.15, -0.52]
## Product [Horizon 2% Milk Carton]	-0.53	[-0.83, -0.23]
## Product [Imperial Sugar]	-1.19	[-1.53, -0.85]
## Product [Iodized Salt]	-1.89	[-2.19, -1.59]
## ProductJIF Peanut butter	-1.35	[-1.67, -1.04]
## Product [Justin's almond butter]	3.38	[3.01, 3.75]
## Product [Kind Oats and Honey Bar]	-0.83	[-1.19, -0.47]
## Product [La Croix Box]	-0.48	[-0.82, -0.15]
## ProductLand o Lakes Half & Half	-1.56	[-1.91, -1.22]
## Product [Lucky Charms 1 lb]	-0.84	[-1.29, -0.39]
## Product [Moms Tomato Sauce]	0.74	[0.37, 1.11]
## Product [Natures Own Whole Wheat Bread]	-1.22	[-1.63, -0.80]
## Product [Navel Orange]	-1.92	[-2.29, -1.55]
## Product [Nut Thins 4.25 oz]	-1.19	[-1.56, -0.82]
## Product [Original ramen]	-2.45	[-2.84, -2.06]
## Product [Premium Saltine Crackers 1 lb]	-1.11	[-1.48, -0.74]
## Product [Progress Chicken Noodle Soup]	-1.49	[-1.85, -1.13]
## Product [San Pellegrino 25.3 floz]	-1.74	[-2.08, -1.39]
## Product [Skinny Pop]	-0.96	[-1.31, -0.62]
## Product [Smart Water 23.7 fl oz]	-1.84	[-2.16, -1.52]
## Product [Stacy's Pita Chips]	-0.83	[-1.18, -0.47]
## Product [Tostitos Bag]	-0.81	[-1.22, -0.39]
## Product [Triscuits 9oz]	-1.13	[-1.45, -0.80]
## Product [Welch's Grape Jelly]	-1.48	[-1.82, -1.15]
## Product [Wheat Thins 8oz]	-1.13	[-1.46, -0.81]
## Income10K	-0.03	[-0.07, 0.01]

Based on the sign of the Income10K coefficient consumers in poorer ZIP codes seem to pay slightly more money for the same products. The coefficient for Income10K is **-0.014**. A negative coefficient means that

as ZIP code income increases, prices decrease for the same product. Thus, lower-income ZIP codes are associated with higher prices.

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: Redlining

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

Verdict: TRUE Justification: Figure A1 shows a strong positive relationship between % minority and FAIR policies. Model A reports an R-squared of 0.516 and a minority coefficient of 0.014 ($p < 0.001$), indicating that FAIR policy rates increase by 0.014 for every 1% increase in minority population.

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.

Verdict: UNDECIDABLE Justification: No regression model includes an interaction term for minority \times age. Figure B1 and Model B only show a weak, non-significant relationship ($p = 0.125$) between age and minority. Therefore, an interaction cannot be confirmed or ruled out.

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.

Verdict: TRUE Justification: Model C includes an interaction between minority and fire_risk. In high fire risk areas (baseline), the minority coefficient is 0.01 ($p = 0.015$). In low fire risk areas, the interaction term is -0.001 ($p = 0.839$), suggesting a flatter slope. The stronger relationship exists in high-fire-risk ZIP codes.

D. Even without controlling for any other variables, income “explains away” all the association between minority percentage and FAIR policy uptake.

Verdict: FALSE Justification: In Model D2, the minority coefficient remains positive (0.01) and significant ($p = 0.002$), even after adding income as a control. The effect is reduced but not eliminated.

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.

Verdict: TRUE Justification: Model E includes all four predictors, and the minority coefficient is still significant: 0.008 ($p = 0.006$). Thus, the association holds after adjusting for other factors.