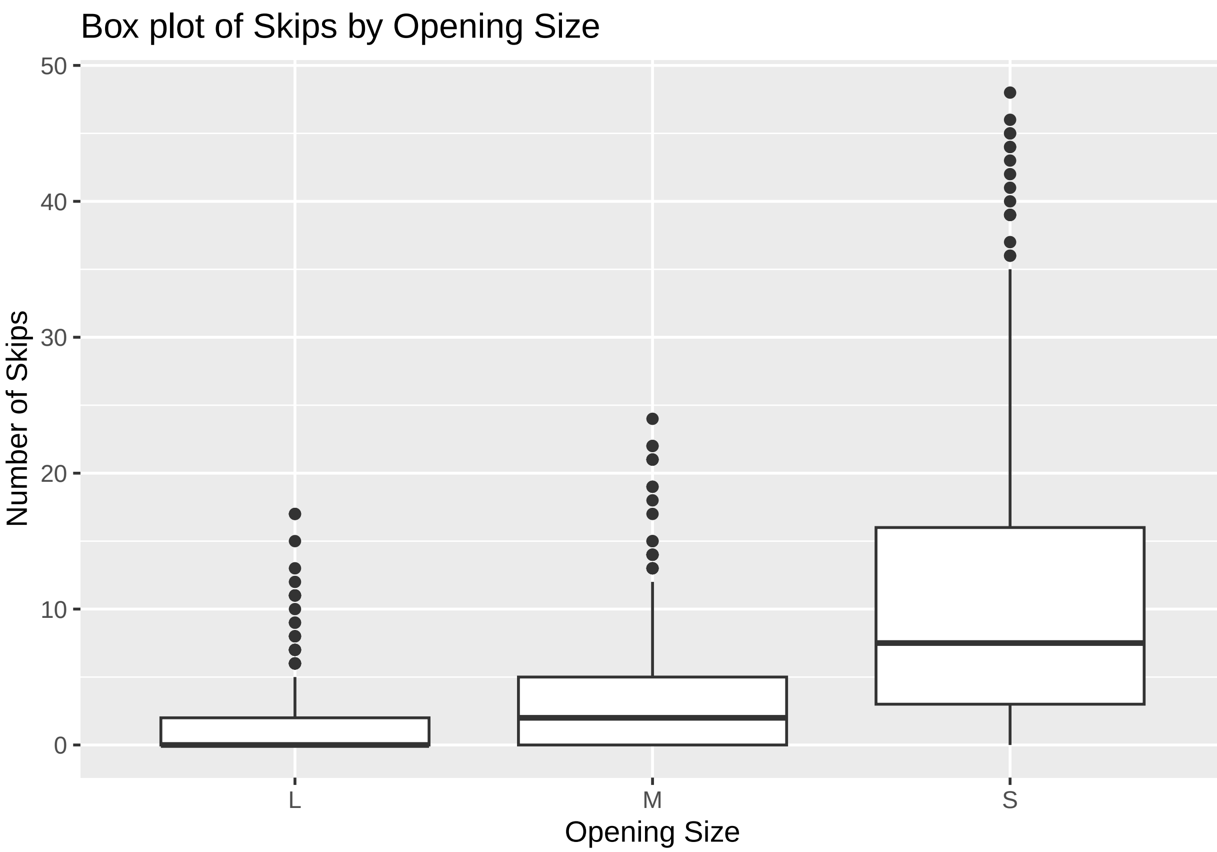
SDS 315 Homework 9

Elizabeth 'Betsy' Sherhart UT EID: eas5778
Click here for link to GitHub repository

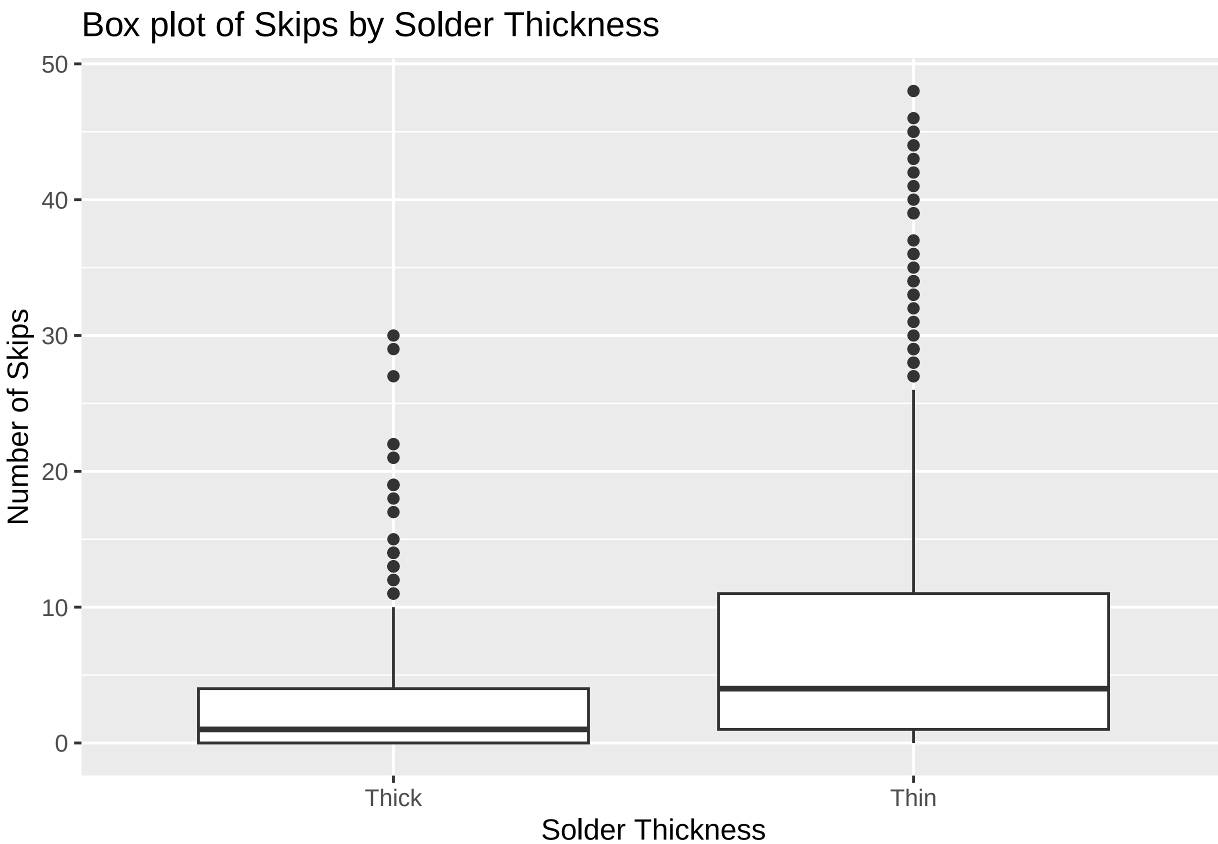
April 21st, 2025

Problem 1: Manufacturing flaws in circuit boards

Part A: Plots



Box plot showing the number of solder skips by solder gun opening size. Smaller openings are associated with a higher number of skips, while larger openings tend to result in fewer skips. This suggests that opening size plays a role in soldering quality and that increasing the opening size may help reduce soldering defects (skips).



Box plot showing the number of solder skips by solder thickness. Thin solder is associated with a higher number of skips, while thick solder tends to result in fewer skips. This suggests that solder thickness plays a role in soldering quality and that using thicker solder may help reduce soldering defects (skips).

Part B: Regression model

```
## # 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
```

Part C: Interpretation of model coefficients

Baseline (Intercept):

The baseline number of skips—when the solder gun has a **large opening (Opening = L)** and the **solder is thick (Solder = Thick)**—is **0.393 skips**. This means the model predicts that a board manufactured with a large opening and thick solder will have an average of 0.393 skips.

Main effect for Opening: M (Medium opening size):

When **Opening = M** and **Solder = Thick**, the main effect for the medium opening is **2.41 skips**. In isolation, switching from a large to a medium opening increases the average number of skips by 2.41, resulting in a **total of 2.80 skips (0.393 + 2.41)**.

Main effect for Opening: S (Small opening size):

When **Opening = S and Solder = Thick**, the main effect for the small opening is **5.13 skips**. This indicates that in isolation switching from a large to a small opening increases the average number of skips by 5.13, leading to a **total of 5.52 skips (0.393 + 5.13)**.

Main effect for Solder: Thin:

When ***Opening = L and Solder = Thin**, the main effect for thin solder is **2.28 skips**. This means that, in isolation, using thin solder instead of thick increases the average number of skips by 2.28, for a **total of 2.67 skips (0.393 + 2.28)**.

Interaction effect for Opening: M and Solder: Thin:

When **Opening = M and Solder = Thin** (i.e., the opening is medium and the solder is thin), the interaction effect is **-0.74 skips**. In other words, boards made with both a medium opening and thin solder result in 0.74 fewer skips than expected based on the sum of their individual (isolated) effects, for a **total of 4.343 skips (0.393 + 2.41 + 2.28 - 0.74)**.

Interaction effect for Opening: S and Solder: Thin:

When **Opening = S and Solder = Thin** (i.e., the opening is small and the solder is thin), the interaction effect is **9.65 skips**. In other words, boards manufactured with both a small opening and thin solder experience 9.65 more skips than would be expected based on the sum of their individual (isolated) effects, for a **total of 17.453 skips (0.393 + 5.13 + 2.28 + 9.65)**.

Part D: Recommendation for AT&T

Based on the regression analysis, if I had to recommend a combination of Opening size and Solder thickness to AT&T to minimize the number of skips (and improve the reliability of the manufacturing process), I would recommend the following:

Opening: Large

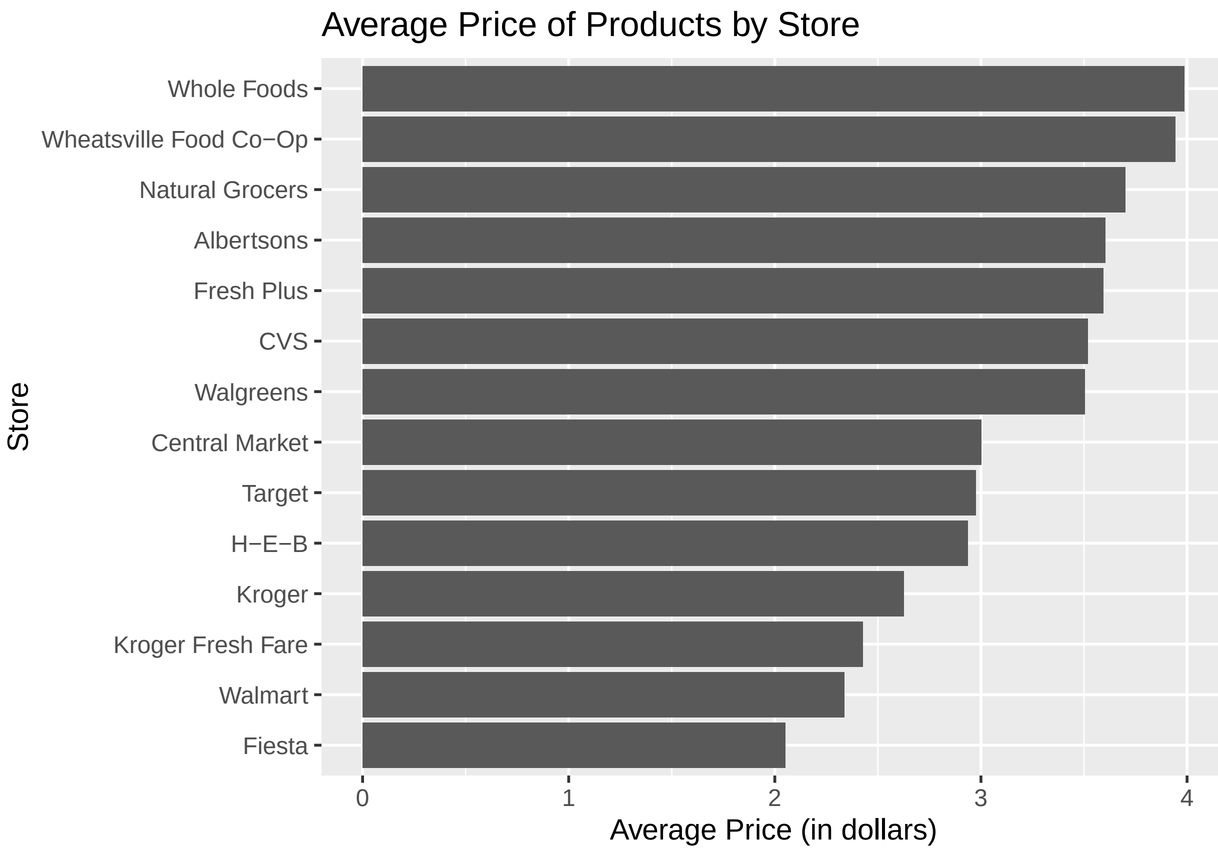
Solder: Thick

Reasoning:

The Large Opening and Thick Solder combination serves as the baseline, with an average of 0.393 skips. From the regression model, we can see that switching to a Medium or Small opening size increases the number of skips. Specifically, a Small Opening increases the skips by 5.13, resulting in a total of 5.52 skips (0.393 + 5.13) compared to the Large Opening. A Medium Opening increases skips by 2.41, leading to 2.80 skips (0.393 + 2.41) compared to the Large Opening. Additionally, using Thin Solder increases skips by 2.28, which results in 2.67 skips (0.393 + 2.28) compared to the Thick Solder. The interaction effects also modify the number of skips: when both Medium Opening and Thin Solder are used, the skips are reduced by 0.74, leading to a total of 4.34 skips (0.393 + 2.41 + 2.28 - 0.74). On the other hand, using both Small Opening and Thin Solder increases the number of skips by 9.65, resulting in 17.45 skips (0.393 + 5.13 + 2.28 + 9.65). These results demonstrate that sticking with the Large Opening and Thick Solder combination minimizes the risk of manufacturing flaws (skips). Therefore, based on the regression model, Large Opening and Thick Solder would be the most reliable combination for minimizing skips and ensuring a high-quality manufacturing process at AT&T.

Problem 2: Grocery store prices

Part A



This bar plot displays the average prices of products across various stores. The x-axis represents the average price in dollars, while the y-axis lists the stores. The plot highlights variations in price between stores, showing that Whole Foods and Wheatsville Food Co-Op have significantly higher prices compared to others, while Walmart and Fiesta have notably lower prices.

Part B

This bar plot shows the number of store entries per product across all 16 stores. Common products, like milk and eggs, are available in every store, while other items appear less frequently. This variation likely reflects product specialization or store format (e.g., natural or convenience stores may not carry certain brands). This plot helps explain why the average price comparisons from the previous plot may not fully represent price differences across stores, as not all stores carry the same products.

Part C

```
##                          Estimate   Lower CI      Upper CI
## TypeGrocery           -0.66425649 -0.9154098 -0.4131031458
## TypeHigh-end Grocery  -0.29769012 -0.5944834 -0.0008968566
## TypeNatural           -0.09223341 -0.4037250  0.2192581912
## TypeSmall Format      -0.44415560 -0.7454724 -0.1428388392
```

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.

Part D

```
##                                Estimate    Lower CI     Upper CI
## StoreWhole Foods             0.36415850  0.01672253  0.711594471
## StoreWheatsville Food Co-Op  0.29029303 -0.06146201  0.642048073
## StoreWalgreens               0.21548328 -0.14031581  0.571282360
## StoreCVS                     0.19311241 -0.16684400  0.553068808
## StoreFresh Plus             -0.03616437 -0.35487988  0.282551136
## StoreNatural Grocers        -0.08117985 -0.46989988  0.307540177
## StoreTarget                 -0.37336967 -0.74706831  0.000328972
```

```
## StoreCentral Market         -0.57338651 -0.92164719 -0.225125837
## StoreH-E-B                  -0.64595932 -0.94513262 -0.346786025
## StoreFiesta                 -0.70323198 -1.23351226 -0.172951704
## StoreKroger                 -0.70331815 -1.16242453 -0.244211765
## StoreKroger Fresh Fare      -0.90177969 -1.36088607 -0.442673303
## StoreWalmart                -0.99254892 -1.45165530 -0.533442534
```

The two stores that seem to charge the lowest prices for the same product when compared to others are Walmart and Kroger Fresh Fare. On the other hand, the stores with the highest prices are Whole Foods and Wheatsville Food Co-Op.

Walmart's prices are, on average, **$0.99** lower than the baseline store (Albertsons), with a 95% confidence interval of [**-$1.45, -$0.53**], which indicates that this difference is statistically significant. Kroger Fresh Fare's prices are, on average, **$0.90** lower than Albertsons, with a 95% confidence interval of [**-$1.36, -$0.44**], which is also statistically significant.

Whole Foods has prices that are, on average, **$0.36** higher than Albertsons, with a 95% confidence interval of [**$0.02, $0.71**], which is statistically significant. However, Wheatsville Food Co-Op has prices that are, on average, **$0.29** higher than Albertsons, with a 95% confidence interval of [**-$0.06, $0.64**]. Since the 95% confidence interval interval includes zero, the difference is not statistically significant.

Part E

```
## H-E-B Estimate and 95% Confidence Interval:

##          Estimate Lower_CI Upper_CI
## Estimate    -0.65    -0.95    -0.35

## Central Market Estimate and 95% Confidence Interval:

##          Estimate Lower_CI Upper_CI
## Estimate    -0.57    -0.92    -0.23

## Price Difference Between Central Market and H-E-B:

##          Estimate_Difference Lower_CI_Difference Upper_CI_Difference
## Estimate                0.07                0.02                0.12

## Total price spread across stores:

## [1] 1.356707
```

The analysis of Central Market and H-E-B reveals that the price difference between the two stores is relatively small. From the model, H-E-B has a coefficient of **-$0.65 (95% confidence interval:** [**-$0.95, -$0.35**]), while Central Market has a coefficient of **-$0.57 (95% confidence interval:** [**-$0.92, -0.23**]). This implies that Central Market charges approximately **$0.07** more than H-E-B for the same product (with a **95% confidence interval of** [**0.02, 0.12**]). Since the confidence interval does not include zero, this difference is statistically significant.

However, when considering this price difference in the larger context, such as comparing Whole Foods to Walmart with a **$1.36** price difference, the **$0.07** difference between H-E-B and Central Market seems relatively minor. This suggests that while Central Market charges slightly more than H-E-B for the same product, the difference is not as pronounced as the variation seen between other stores, like Whole Foods and Walmart.

Therefore, based on the evidence from the model, it seems that Central Market charges slightly more than H-E-B for the same product, but the price difference is small in comparison to other stores.

Part F

```
##     Estimate   Std. Error      t value     Pr(>|t|)
```

```
## -0.014089731  0.009625589 -1.463778628  0.144239123
```

```
## # Standardization method: refit
##
## Parameter | Std. Coef. |        95% CI
## -------------------------------------
## Income10K |      -0.03 | [-0.07, 0.01]
```

Based on the coefficient for **Income10K**, which is **-0.014**, consumers in poorer ZIP codes tend to pay slightly more for the same product on average. The negative sign suggests an inverse relationship between income and price, meaning that as income decreases, the price consumers pay increases. However, the p-value for this coefficient is **0.144**, indicating that this effect is not statistically significant. Therefore, while the sign implies that poorer ZIP codes may pay more, we cannot confidently make this claim based on the model.

In the standardized model, the coefficient is **-0.03**. This suggests that a one-standard deviation increase in income is associated with a **0.03** standard-deviation decrease in the price consumers expect to pay. Again, this reflects a very small, negative relationship between income and price. But since the 95% confidence interval [**-0.07, 0.01**] includes zero, the result is not statistically significant. Thus, while the direction of the effect is consistent, there is insufficient evidence to assert that income significantly impacts price.

Problem 3: redlining

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

This statement is **true**. **Figure A1** shows a positive, linear relationship between the percentage of minority residents in a ZIP code (x) and the number of FAIR policies per 100 housing units (y). According to **regression model_A**, the coefficient for the percentage of minority residents is **0.014**, indicating that for every one percentage point increase in minority residents, the number of FAIR policies per 100 housing units increases by an average of 0.014. This association is statistically significant, as evidenced by a 95% confidence interval of [**0.009, 0.018**], which does not include zero. The model's **R-squared** value is **0.516**, meaning that **51.6%** of the variation in FAIR policy rates across ZIP codes is explained by the percentage of minority residents. Together, these results provide strong evidence of a positive association, supporting the 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 statement is **undecidable** based on the available evidence. While **regression model_B** includes both minority percentage and housing age as predictors, it does not test for an interaction (i.e., a product term between the two variables). Additionally, the **R-squared** value of **0.06** and the 95% confidence interval of [**-0.116, 0.912**], which includes zero, suggest that the relationship between these variables is not statistically significant. **Figure B1** shows a slight upward trend, but it does not indicate that this trend is different for ZIP codes with younger versus older housing stock.

To properly evaluate if there is an interaction, we would need a model that includes an interaction term (minority:age) alongside the main effects of minority percentage and housing age. Additionally, a plot showing the relationship between minority percentage and the number of FAIR policies across different age groups of housing stock would help visualize this potential interaction.

Therefore, the evidence provided does not definitively support or refute the claim of an interaction effect, and further analysis is needed to make this determination.

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 statement is **undecidable**. While **Figure C1** visually suggests a steeper slope for high-fire-risk ZIP codes, implying a stronger relationship between minority percentage and FAIR policy uptake, the statistical evidence is mixed. In **regression model_C**, the coefficient for minority percentage in high-fire-risk ZIP codes is **0.01** and statistically significant (**95% CI: [0.002, 0.017]**), but the interaction term for minority:low-fire-risk ZIPs is **-0.001**, with a wide **95% confidence interval [-0.012, 0.01]** and a **p-value of 0.839**, indicating non-significance. Since the interaction term is not significant, we cannot confidently conclude that the relationship between minority percentage and FAIR policy uptake differs across fire risk levels. While the visual evidence hints at a stronger trend in high-risk areas, the statistical model does not provide definitive evidence of a meaningful difference.

Therefore, more rigorous testing—such as running separate regressions for high- and low-fire-risk ZIP codes and comparing the minority percentage coefficients—would be needed to assess whether fire risk truly moderates the relationship. Without a significant interaction term or clearer statistical evidence, the statement remains undecidable based on the current model.

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

This statement is **false**. While income reduces the strength of the association between minority percentage and FAIR policy uptake, it does not eliminate it. In **regression model_D1** (without any controls), the coefficient for minority percentage is **0.014** and statistically significant (**95 % CI: [0.009, 0.018]**). In **regression model_D2** (controlling for income), the coefficient for minority percentage decreases to **0.01** but remains statistically significant (**95 % CI: [0.004, 0.015]**). Although income itself is a significant predictor (**coefficient: -0.074, 95% CI: [-0.144, -0.003]**), it does not fully account for the relationship between minority percentage and FAIR policy rates. This indicates that minority percentage has an independent effect on FAIR policy uptake, even after controlling for income. The reduction in the coefficient size—from **0.014 to 0.01**—shows that income **partially** "explains away" the association between minority percentage and FAIR policy uptake but does not "explain away" **all** the association between minority percentage and FAIR policy uptake.

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 statement is **true**. In the fully adjusted **regression model_E**, the coefficient for minority percentage is **0.008** with a 95% confidence interval of [**0.003, 0.014**], indicating a statistically significant positive relationship with FAIR policy uptake. Model_E includes controls for income, fire risk, and housing age. Notably, only the coefficients for minority percentage (**95% CI: [0.003, 0.014]**) and fire risk (**95% CI: [0.004, 0.04]**) remain statistically significant, as their intervals do not include zero. The model's **R-squared** value is **0.662**, meaning that model_E explains **66.2%** of the variation in FAIR policy rates across ZIP codes. This provides strong evidence that the positive association between minority percentage and FAIR policy use is not fully explained by other ZIP code characteristics, such as income, fire risk, or housing age. This means that 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.