MINI PROJECT 3

Lori Frager and Sol Graciela Vloebergh

2024-11-05

ANALYSIS 1: E-cigarettes

Question 1: Do the data suggest that the participants' cigarettes per day decreased over the course of the study?

This analysis aims to investigate whether e-cigarette usage reduces the number of cigarettes smoked per day (CPD) over a six-week period. The dataset includes 29 participants, each reporting their CPD at baseline and week 6 of the study, along with additional information such as adherence to the study protocol and age category.

Descriptive Summary

To begin with the analysis, we will calculate the summary statistics for CPD_BL (Baseline Cigarettes per Day) and CPD_W6 (Cigarettes per Day at Week 6) to provide a descriptive summary.

```
## [1] 24.17241
```

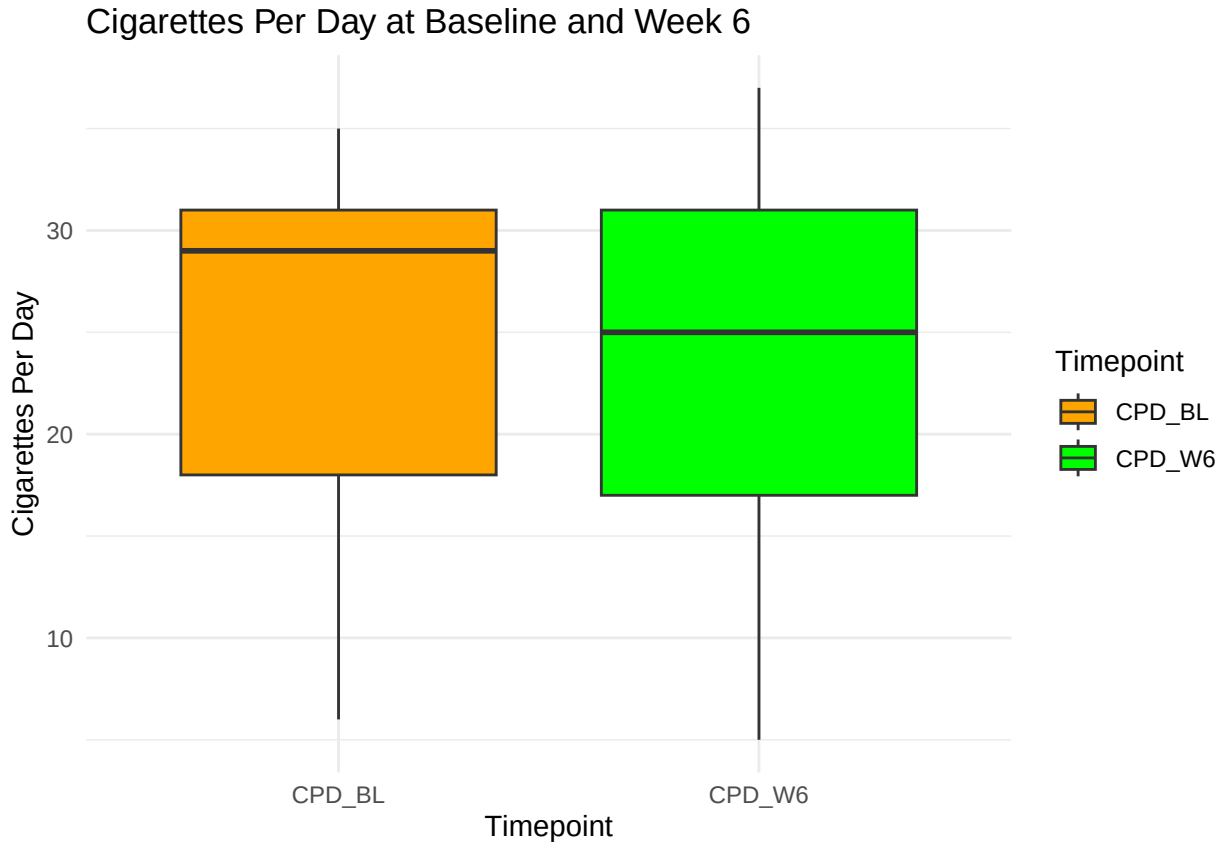
```
## [1] 8.996305
```

```
## [1] 23.17241
```

```
## [1] 9.250749
```

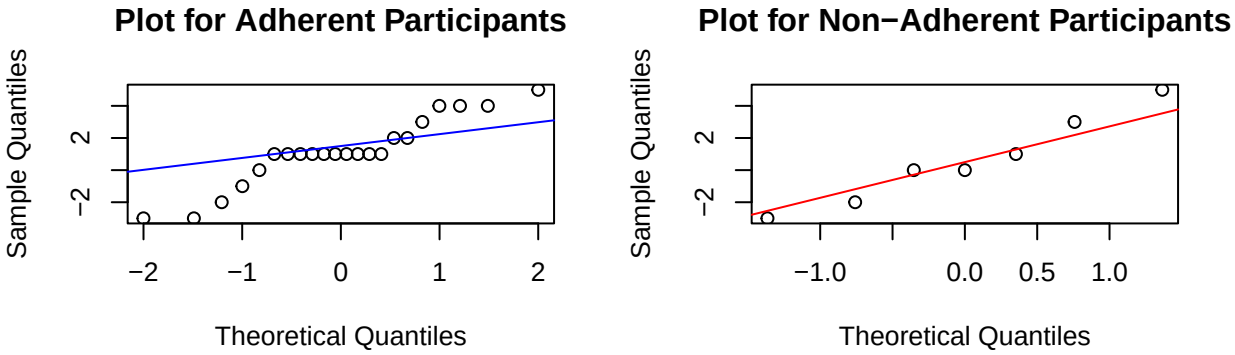
As we observe above, the cigarettes per day (CPD) at baseline was 24.17 (Standard Deviation = 8.99), while at week 6, it was 23.17 (SD = 9.25). There is a slight decrease in smoking behavior over the six-week period.

Visualization: Boxplot to compare CPD_BL and CPD_W6



The plot above shows two boxplots: one for Cigarettes per Day at Baseline (CPD_BL) in orange, and Cigarettes per day at Week 6 (CPD_W6) in green. We can assess there is a potential decrease in the median number of cigarettes smoked per days from the beginning of the study to the end of it.

We noticed that the dataset includes a column called "adherence". The participants who were classified as "adherent" (=1) are those who followed the study protocol. And those classified as "non-adherent" (=0) did not fully comply with it. We will look to see how normal is the distribution for both variables CPD_BL and CPD_W6.



We can observe that the plot for adherent participants shows that most points lie relatively close to the diagonal line, indicating that the differences in CPD (Baseline - Week 6) are approximately normally distributed. Even though there are slight deviations in the tails, the normality assumption is reasonable for this group.

The plot for non-adherent participants exhibits stronger deviations from the diagonal line, particularly in the lower and upper tails. This suggests that the differences in CPD for non-adherent participants may not follow a normal distribution.

Inferential Analysis: Paired t-test

```
## t-statistic: 2.391299
```

```
## p-value: 0.02375017
```

We decided to conduct a paired t-test in order to asses if the observed change in CPD is statistically significant. The results above shows a t-statistic of 2.39 with a p-value of 0.023. Since the conventional significance level is 0.05, we reject the null hypothesis, indicating there is an important decrease in CPD from the beginning of the study until week 6.

Conclusion

Due to the results obtained, we observe there is a significant reduction in cigarette consumption over the six-week period of e-cigarette usage. However, while the decrease in CPD is statistically significant, the

study's design does not fully control for other factors influencing smoking cessation, such as addiction or diseases.

Additionally, the lack of a control group means we cannot confidently attribute the observed reduction solely to e-cigarette use, as well as the relatively small sample size may limit generalization. The differences in adherence and participant behavior could also impact results.

ANALYSIS 2: Heifer Feed

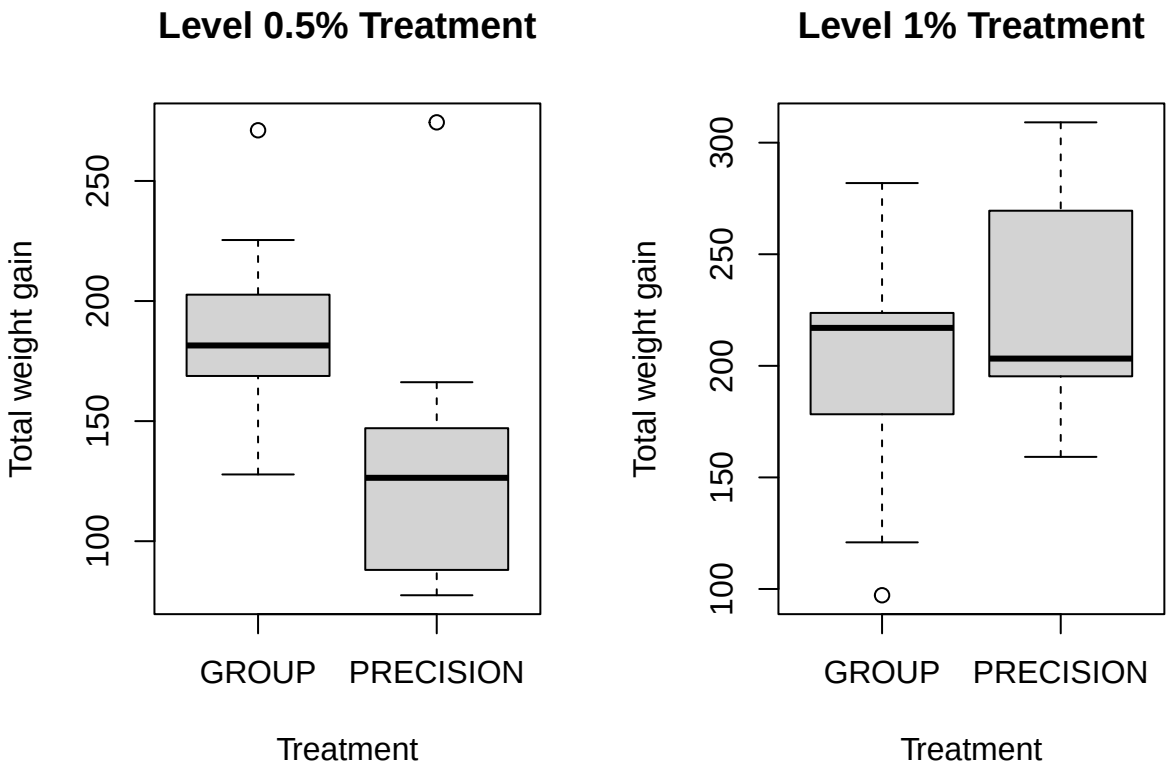
Question 2: On average, did the heifers using GROUP feeding have a different total weight gain than the heifers using the PRECISION feeding at the low feed level, 0.5%? What about in the high feed level, 1%?

A descriptive analysis: We are going to analyze the results on the work the team of animal science researchers did when they gave supplemental feedings to young heifers. They used varied amounts and varied feeding methods. We want to look at the total weight gain of the heifers after 6 months to see if the different amounts and feeding methods made a difference in the weight gain.

```
##
##             0.5  1
##   GROUP      16 16
##   PRECISION  15 16
```

The number of heifers in each set is well balanced. These are two independent samples since the researchers were consistent in the feeding amounts and methods over the 6-month period.

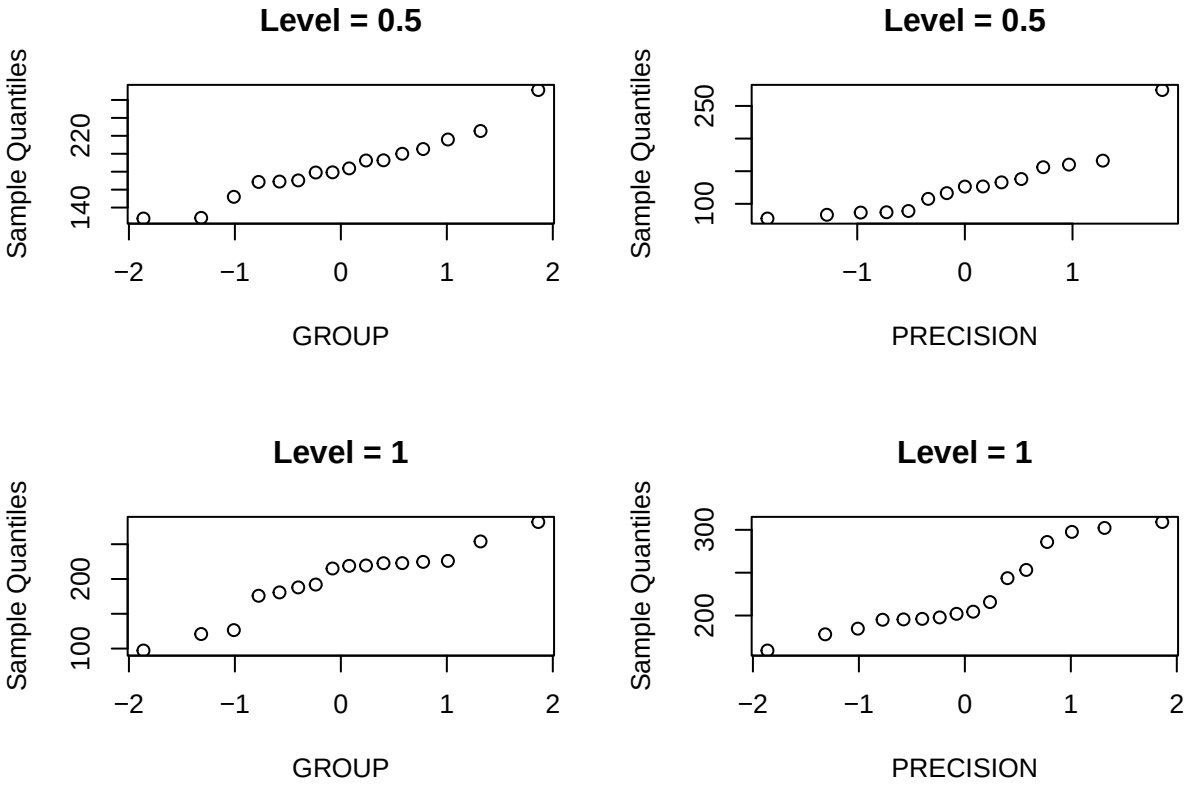
Let's look at the boxplots to get an idea of what the data looks like.



We can see there are a couple of outliers at the 0.5% level and the medians are considerably different. The GROUP side median is approximately 175 while the PRECISION is around 130. Comparing the two IQRs, they are not close and the PRECISION side is much larger showing larger variance. Based on these summaries, there does appear to be an association that the GROUP feeding of the heifers at the 0.5% level shows they will more likely weigh more after the 6 months.

In the 1% level of treatment, there is 1 low outlier on the GROUP side but the medians are close together. Since there are not a lot of heifers in each set, I do not want to eliminate the outliers.

We will look to see how normal the distributions of the data are.



The graphs appear to be fairly normal except for the outliers at the top or bottom end as we saw in the boxplots. The precision graph at the 1% level doesn't follow the normal line very closely.

```
## # A tibble: 4 x 3
##   Treatment level mean_tgain
##   <fct>     <dbl>      <dbl>
## 1 GROUP       0.5       185.
## 2 GROUP       1         198.
## 3 PRECISION   0.5       129.
## 4 PRECISION   1         226.
```

```
## # A tibble: 4 x 3
##   Treatment level sd_tgain
##   <fct>     <dbl>    <dbl>
## 1 GROUP       0.5     35.7
## 2 GROUP       1       49.1
```

```
## 3 PRECISION   0.5     49.9
## 4 PRECISION   1       48.7
```

Inferential Analysis

Since there was a balanced number of heifers in each set, we were going to use the equal variances method, but because the standard deviations for both sets are different, we will use the unequal variances method.

The 0.5% Level of Treatment. The GROUP feeding method, the mean of the total weight gain was 185.08 and the SD was 35.70 pounds. For the PRECISION feeding, the mean of the total weight gained was 128.54 and the SD was 49.89 pounds. The difference between the two means is 56.54, a good distance from zero, indicating a type of association between the total weight gain and the treatment method.

For the 0.5% level of treatment, we will test the hypothesis that the GROUP feeding of the treatment weight gain is different than the PRECISION feeding weight gain. H0: mu(group)-mu(precision)=0 HA: mu(group)-mu(precision)!=0 We will use a 95% confidence level.

```
##
##  Welch Two Sample t-test
##
## data:  level05$tgain by level05$Treatment
## t = 3.6074, df = 25.237, p-value = 0.001333
## alternative hypothesis: true difference in means between group GROUP and group PRECISION is not equal
## 95 percent confidence interval:
##  24.27326 88.79674
## sample estimates:
##     mean in group GROUP mean in group PRECISION
##                 185.075                 128.540
```

After running the 2-sample t-test, we found the p-value is 0.001333, which is much less than the alpha of 0.05. We will reject the null hypothesis in favor of the alternative. The confidence interval is (24.273, 88.797) We are 95% confident that the average weight gain of the GROUP feeding is between 24.27 and 88.79 more than the method of PRECISION feeding the heifers.

The 1% Level of Treatment For the GROUP feeding, the mean of the total weight gain was 197.91 and the SD was 49.14. For the PRECISION feeding, the mean of the total weight gained was 226.26 and the SD was 48.75. The variance between each of the standard deviations is less than 1 and the difference between the means is approximately 28 pounds.

For the 1% level of Treatment We will test the hypothesis that the GROUP feeding of the treatment is different to the PRECISION feeding. H0: mu(group)-mu(precision)=0 HA: mu(group)-mu(precision)!=0 Again, we will use a 95% confidence level.

```
##
##  Welch Two Sample t-test
##
## data:  level1$tgain by level1$Treatment
## t = -1.6387, df = 29.998, p-value = 0.1117
## alternative hypothesis: true difference in means between group GROUP and group PRECISION is not equal
## 95 percent confidence interval:
##  -63.696632   6.984132
## sample estimates:
##     mean in group GROUP mean in group PRECISION
##                197.9062                226.2625
```

The p-value is 0.1117, so we will not reject null hypothesis in favor of the alternative. The confidence interval is -63.6965, 6,9840. Because this includes zero, there is almost no association between the two types of feeding methods at the 1% level. The test statistic is 1.6387 which is almost 0, again showing the data closely aligns with the null hypothesis.

Conclusion

Feeding the heifers at the 0.5% level, the data provides evidence that the GROUP feeding method did a much better job on putting the weight on the heifers than the PRECISION feeding method. We thought it was very interesting that there was such a big difference in the amount of weight that was put on by the GROUP feeding method verses the PRECISION feeding method at the 0.5% level. We do not know of any other factors that may have caused the GROUP feeding heifers to gain additional weight. Were they kept in different pastures and one had better grass than the other? Did the PRECISION heifers have to walk farther to the barn to get the food?

Feeding the heifers at the 1% level, the data does not give us strong evidence that the GROUP feeding method was any better or worse than the PRECISION feeding method in helping the heifers to gain weight.

Note: *I found the process somewhere to get the mean & SD across multiple categories. I had saved the link at the bottom of the page and somewhere along the way I deleted it by accident.