Shared Social Responsibility

Report by Group 4

Flat Rate Pricing We will begin by considering whether there is a difference between the proportion of purchases between the two flat-rate pricing conditions, `FR` and `FR Charity`.

a) Null and alternative hypotheses to compare population proportions

H_0 = the difference of proportion purchases between `FR` and `FR Charity` is 0
H_a = the difference of proportion purchases between `FR` and `FR Charity` is not 0

The data for `FR` and `FR Charity` has been summarized in the table below.

Condition	NumberSold	Riders	MerchandiseRevenues
All	All	All	All
FR	77.00	12,663.00	4,592.41
FR	63.00	15,561.00	6,688.57
FR Charity	79.00	14,796.00	6,476.78
FR Charity	101.00	15,796.00	5,845.94

Appropriate test statistic for the difference of population proportions is given by:

$$\frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}}$$

Using the test statistic above, we will first calculate $\bar{p}_1$ and $\bar{p}_2$.

$\bar{p}_1$: sum of **NumberSold** for `FR` divided by the total number of **Riders** for `FR`
$\bar{p}_2$: sum of **NumberSold** for `FR Charity` divided by the total number of **Riders** for `FR Charity`

n_1: sum of **Riders** for `FR`
n_2: sum of **Riders** for `FR Charity`

We will first load the `Sales.csv` file to RStudio and filter for the 2 fixed rate conditions for our calculations.

```
fr <- Sales %>%
  filter(Sales$Condition == 'FR' | Sales$Condition == 'FR Charity')
```

```
p1 <- (fr[[1,2]] + fr[[2,2]]) / (fr[[1,3]] + fr[[2,3]])
p2 <- (fr[[3,2]] + fr[[4,2]]) / (fr[[3,3]] + fr[[4,3]])

n1 <- fr[[1,3]] + fr[[2,3]]
n2 <- fr[[3,3]] + fr[[4,3]]

z1 <- (p1 - p2) / sqrt(((p1 * (1 - p1)) / n1) + ((p2 * (1 - p2)) / n2))
```

```
## [1] -1.526455
```

b) The test statistic resulted in a z-score of -1.526455.
And our concerned critical value when examining a 5% level is -1.96. Since the z-score is greater than the critical value at the 5% interval, we cannot reject the null hypothesis.

```
p1 <- 2 * pnorm(abs(z1), lower.tail = FALSE)
```

```
## [1] 0.1268965
```

c) The p-value associated with this test statistic was calculated to be 0.127.

d) The p-value 0.127 can be interpreted as follows:
there is a 12.7% chance that the proportion of the population will be observed in the bottom and top 2.5% of the curve.

Since this value is greater than 5%, we cannot reject the null hypothesis at this given level since there is no statistical significance in the differences between the two samples. And assuming the null hypothesis is true, there is a 12.7% chance we will obtain test results that are at least as extreme as the results actually observed in our sample.

NYOP Pricing We will continue by considering whether there is a difference between the proportion of purchases between the two NYOP pricing conditions, `NYOP` and `NYOP Charity`.

a) Null and alternative hypotheses to compare population proportions:

$H_0 =$ no difference of proportion purchases between `NYOP` and `NYOP Charity`
$H_a =$ the difference of proportion purchases between `NYOP` and `NYOP Charity` exists

The data for `NYOP` and `NYOP Charity` has been summarized in the table below.

```
Sales %>%
  filter(Condition %in% c('NYOP','NYOP Charity')) %>%
  select(Condition:MerchandiseRevenues) %>%
  dtab(dec = 2, nr = 100) %>% render()
```

Condition	NumberSold	Riders	MerchandiseRevenues
All	All	All	All
NYOP	1,137.00	14,077.00	4,845.27
NYOP	1,233.00	14,186.00	7,038.63
NYOP Charity	539.00	12,227.00	5,690.59
NYOP Charity	628.00	13,741.00	6,003.44
NYOP Charity	626.00	18,117.00	8,557.47

Using the same test statistic as above, we will first calculate $\bar{p}_1$ and $\bar{p}_2$.

$\bar{p}_1$: sum of **NumberSold** for `NYOP` divided by the total number of **Riders** for `NYOP`
$\bar{p}_2$: sum of **NumberSold** for `NYOP Charity` divided by the total number of **Riders** for `NYOP Charity`

n_1: sum of **Riders** for `NYOP`
n_2: sum of **Riders** for `NYOP Charity`

```
nyop_total <- Sales %>%
  filter(Sales$Condition == 'NYOP' | Sales$Condition == 'NYOP Charity')
```

```
p1 <- (nyop_total[[1,2]] + nyop_total[[2,2]]) / (nyop_total[[1,3]] + nyop_total[[2,3]])
p2 <- (nyop_total[[3,2]] + nyop_total[[4,2]] + nyop_total[[5,2]]) / (nyop_total[[3,3]] + nyop_total[[4,

n1 <- nyop_total[[1,3]] + nyop_total[[2,3]]
```

```
n2 <- nyop_total[[3,3]] + nyop_total[[4,3]] + nyop_total[[5,3]]

z2 <- (p1 - p2) / sqrt(((p1 * (1 - p1)) / n1) + ((p2 * (1 - p2)) / n2))
pval <- 2 * pnorm(abs(z2), lower.tail = FALSE)
```

```
## [1] 22.74971
```

b) Since the z-score of 22.750, is greater than the critical value at 5%, 1.96, we reject the null hypothesis and conclude that the difference in proportion of purchases between `NYOP` and `NYOP Charity` is significant.

```
## [1] 1.444616e-114
```

c) The p-value associated with the test statistic is $1.445e - 114$.

d) Assuming the null hypothesis is true, there is a minutely small chance we will obtain test results at least as extreme as the results actually observed in our sample.

Radiant Analysis Now, we will analyze the `NYOP` dataset using Radiant.

a) First, we will load the `NYOP.csv` file to Radiant.

```
## Load commands
NYOP <- readr::read_csv("NYOP.csv", n_max = Inf) %>%
  fix_names() %>%
  to_fct()
register("NYOP")
```

Then create a **UnitPrice** Variable.

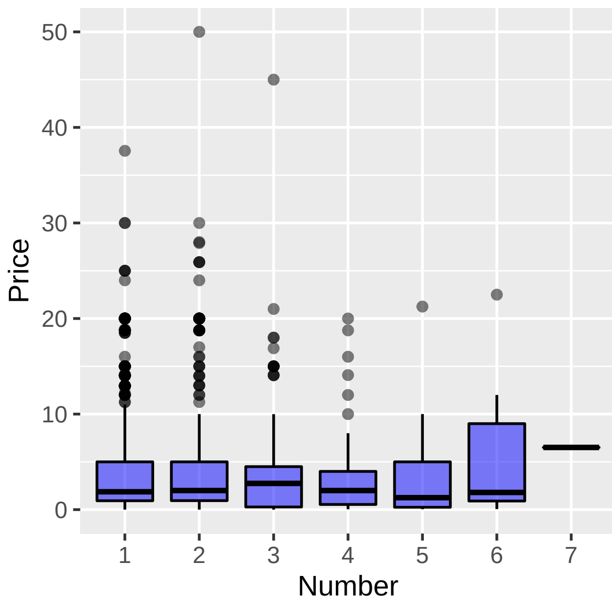
```
## create new variable(s)
NYOP <- mutate(NYOP, UnitPrice = Price / Number)
```

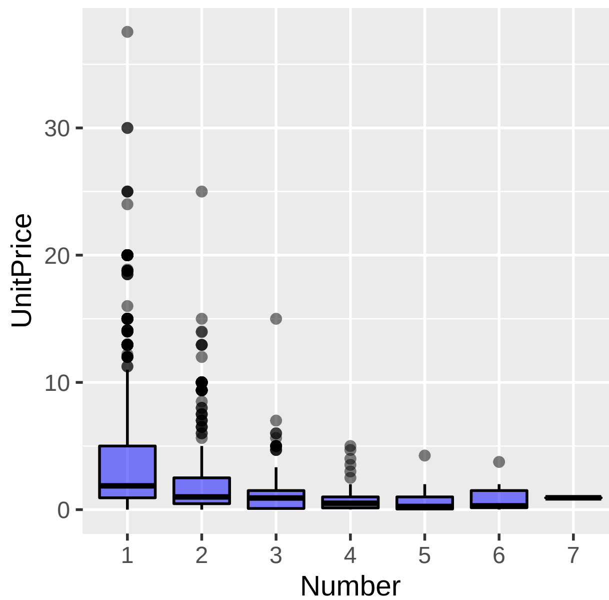
Finally change the type of **Number** to a factor.

```
## change variable type
NYOP <- mutate_at(NYOP, .vars = vars(Number), .funs = as_factor)
```

c) Both **Price** and **UnitPrice** can be visualized against the factor variable **Number** using a Box Plot.

```
visualize(
  NYOP,
  xvar = "Number",
  yvar = c("Price", "UnitPrice"),
  type = "box",
  custom = FALSE
)
```





These box plots show price against number and **UnitPrice** against **Number**.

The following observations can be made looking at the two plots:

Median: Is not as consistent for **Price** vs. **Number** as **UnitPrice** vs. **Number**. The median values observed for **UnitPrice** are fairly consistent, with a slight decrease after 1, but consistent after that. But the values observed for **Price** are more variable.

Quartiles: The interquartile ranges increase as the number of pictures increases when looking at the **Price**. And this range decreases, for a minimum at 4, before increasing again until 6 when looking at the **UnitPrice**. There is no range for 7, since there was only 1 recorded purchase.

Skew: As the **Number** increases, we observe a greater positive skew when considering **Price**. But the initially positive skew decreases as the number of pictures increases in the case of **UnitPrice**.

Outliers: The number of outliers decreases as the **Number** increases for both plots and the relativity in distance from the box decreases as the **Number** increases.

c) The average unit purchase price for `NYOP` and `NYOP Charity` Conditions

```r
result <- explore(
  NYOP,
  vars = "UnitPrice",
  byvar = "Condition",
  fun = c("n_obs", "mean", "min", "max", "sd"),
  nr = Inf
)
summary(result)
```

```
## Explore
## Data         : NYOP
## Grouped by   : Condition
## Functions    : n_obs, mean, min, max, sd
## Top          : Function
##
##     Condition  variable n_obs  mean min    max    sd
##          NYOP UnitPrice 1,641 1.040   0 10.000 1.305
##  NYOP Charity UnitPrice 1,457 5.680   0 37.550 4.670
```

The difference in average unit price for both conditions is substantial since the mean for `NYOP` is only 1.040, whereas the mean for `NYOP Charity` is 5.680. The variability in `NYOP Charity` prices is greater than `NYOP` since the standard deviation for this condition is much larger than that for `NYOP`.

For formulating a statistical test, we will begin by assuming that the null hypothesis is true, therefore there is no difference in averages between the two populations.

The appropriate test statistic for the mean of 2 distributions can be calculated as follows:

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Using the average unit purchase prices from above, we can calculate the necessary values for the test statistic above.

```r
x_1 <- result[[1]][["mean"]][1]
x_2 <- result[[1]][["mean"]][2]
sd_1 <- result[[1]][["sd"]][1]
sd_2 <- result[[1]][["sd"]][2]
n_1 <- result[[1]][["n_obs"]][1]
n_2 <- result[[1]][["n_obs"]][2]

z_nyop <- (x_1 - x_2) / (sqrt(((sd_1^2)/n_1) + ((sd_2^2)/n_2)))
```

```
## [1] -36.67573
```

d)Null and alternative hypotheses

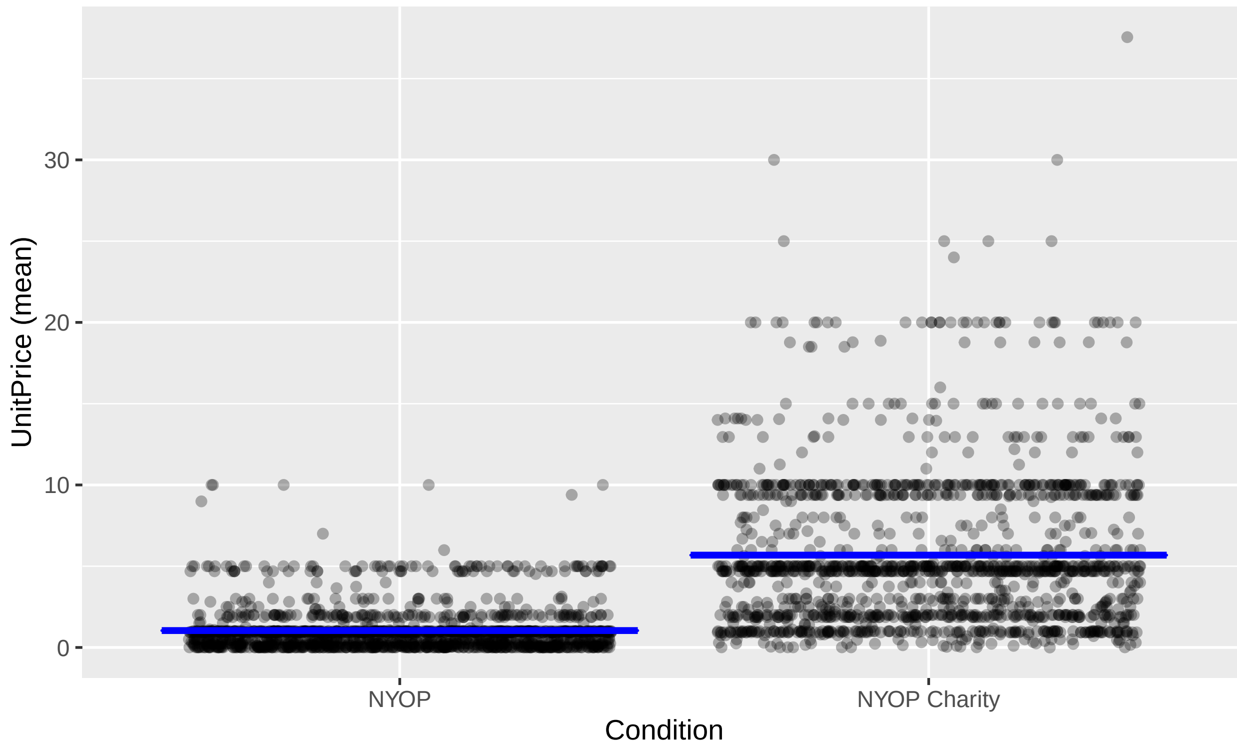
$H_0 =$ The difference in average unit purchase price of `NYOP` and `NYOP Charity` is 0 $H_a =$ The difference in average unit purchase price of `NYOP` and `NYOP Charity` is not 0

e) Using Radiant to perform test computations, we obtain the results below:

```r
result <- compare_means(
  NYOP,
  var1 = "Condition",
  var2 = "UnitPrice"
)
summary(result, show = TRUE)
```

```
## Pairwise mean comparisons (t-test)
## Data      : NYOP
## Variables : Condition, UnitPrice
## Samples   : independent
## Confidence: 0.95
## Adjustment: None
##
##     Condition  mean     n n_missing    sd    se    me
##          NYOP 1.040 1,641         0 1.305 0.032 0.063
##  NYOP Charity 5.680 1,457         0 4.670 0.122 0.240
##
##  Null hyp.             Alt. hyp.                        diff  p.value se
##  NYOP = NYOP Charity   NYOP not equal to NYOP Charity   -4.64 < .001  0.127
##  t.value df       2.5%   97.5%
##  -36.676 1657.943 -4.888 -4.392 ***
##
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plot(result, plots = "scatter", custom = FALSE)
```



Our calculated t-value matched the t-value output by Radiant.

p-value is less than 5% (0.05), so we reject the null hypothesis.
p-value is less than 0.01 (1%), meaning that the difference between the mean values is very significant. So the likelihood of a type 1 error misclassification is very low.

This scatter plot illustrates that `NYOP Charity` has a higher maximum value and variability in price, so the mean is much higher than that of `NYOP`, which has values grouped closely together resulting in a low mean. From the means of the scatter plot, it is clear to see that the proportions are not equal. The visibly large difference in means of the two conditions also supports our rejection of the null hypothesis, since the difference in means is clearly more than 0 and the values are more than 5% apart.

The distribution of values around the mean further supports the conclusion that the misclassification of a type 1 error is very low.

f) We will now use filtering to investigate differences between 1 and 6 pictures.

Results for 1 picture:

```r
result1 <- explore(
  NYOP,
  vars = "UnitPrice",
  byvar = "Condition",
  fun = c("n_obs", "mean", "min", "max", "sd"),
  data_filter = "Number == 1",
  nr = Inf
)
summary(result1)
```

```
## Explore
## Data        : NYOP
```

```
## Filter       : Number == 1
## Grouped by   : Condition
## Functions    : n_obs, mean, min, max, sd
## Top          : Function
##
##     Condition  variable n_obs  mean min    max    sd
##          NYOP UnitPrice 1,162 1.177   0 10.000 1.432
##  NYOP Charity UnitPrice 1,203 5.941   0 37.550 4.830
```

Result for 6 pictures:

```r
result6 <- explore(
  NYOP,
  vars = "UnitPrice",
  byvar = "Condition",
  fun = c("n_obs", "mean", "min", "max", "sd"),
  data_filter = "Number == 6",
  nr = Inf
)
summary(result6)
```

```
## Explore
## Data         : NYOP
## Filter       : Number == 6
## Grouped by   : Condition
## Functions    : n_obs, mean, min, max, sd
## Top          : Function
##
##     Condition  variable n_obs  mean   min   max    sd
##          NYOP UnitPrice     6 0.495 0.010 1.500 0.615
##  NYOP Charity UnitPrice     3 1.970 0.160 3.750 1.795
```

Looking at these result, we can conclude that there is more variability in people that chose 1 picture as opposed to those that bought 6 pictures since the volume of sales for 1 picture is much higher. The mean also differs substantially between 1 picture and 6 pictures, thus supporting our rejection of the null hypothesis.

The minimum and maximum values have a greater range at 1 picture since the sample size is much greater than that of 6 pictures, thus leading to more variability as we concluded initially.

```r
x_16 <- result6[[1]][["mean"]][1]
x_26 <- result6[[1]][["mean"]][2]
sd_16 <- result6[[1]][["sd"]][1]
sd_26 <- result6[[1]][["sd"]][2]
n_16 <- result6[[1]][["n_obs"]][1]
n_26 <- result6[[1]][["n_obs"]][2]

z_nyop6 <- (x_16 - x_26) / (sqrt(((sd_16^2)/n_16) + ((sd_26^2)/n_26)))
```

```
## [1] -1.383092
```

There are less people buying 6 pictures (only 9) compared to 1 picture (over 1000), so the sample sizes are drastically different, thus there is more variability.

Our calculated values are the same as that produced by Radiant, which is surprising given the low value of n. Since the sample size is so small, we would expect more variability in the t-score. But we calculated a value very close to the one on Radiant.

Economics **a)** We will first compute the average daily profit under each of the 4 pricing strategies.

```r
# FR only
fr_only <- Sales %>%
  filter(Sales$Condition == 'FR')

fr_char <- Sales %>%
  filter(Sales$Condition == 'FR Charity')

for (i in 1:nrow(fr_only)) {
  fr_total <- sum(fr_only$NumberSold) * (12.95 - 1.2)
}

avg_fr <- fr_total/sum(Sales$Condition == 'FR')
```

```r
# FR with charity
for (i in 1:nrow(fr_char)) {
  fr_char_total <- sum((fr_char$NumberSold * 0.5 * 12.95) - (fr_char$NumberSold * 1.2))
}

avg_fr_char <- fr_char_total/sum(Sales$Condition == 'FR Charity')
```

```r
# NYOP
nyop_only <- NYOP %>%
  filter(NYOP$Condition == 'NYOP')

nyop_char <- NYOP %>%
  filter(NYOP$Condition == 'NYOP Charity')

nyop_totalp <- sum(nyop_only$Price)
avg_nyop <- (nyop_totalp - sum(as.numeric(nyop_only$Number))*1.2)/2
```

```r
#NYOP Charity
nyop_char_total <- sum(nyop_char$Price)
avg_nyop_char <- (nyop_char_total*0.5 - sum(as.numeric(nyop_char$Number))*1.2)/3
```

The daily profits are summarized below.

```
## Daily profit for fixed price: $822.5
```

```
## Daily profit for fixed price with charity: $474.75
```

```
## Daily profit for name your own price: $-334.1
```

```
## Daily profit for name your own price with charity: $885.5183
```

b) The most profitable strategy, based on the average daily profit, is `NYOP Charity`.

Ranking of the strategies in terms of profits:
1) `NYOP Charity`
2) `FR`
3) `FR Charity`
4) `NYOP`

c) The societal profits can be defined as follows:

`Societal profits = park profit + revenue going to charity`

Therefore, the societal profit for `FR` and `NYOP` are the daily park profits calculated above.

For the 2 conditions with a charity model, the values can be calculated.

```
fr_char_total <- (sum(as.numeric(fr_char$NumberSold)) * 12.95 - sum(as.numeric(fr_char$NumberSold)) * 1
nyop_char_total <- (nyop_char_total - sum(as.numeric(nyop_char$Number)*1.2))/3
```

The average societal profits for each of the conditions can be summarized as seen below:

```
## Average societal profit for fixed price: $822.5
```

```
## Average societal profit for name your own price: $-334.1
```

```
## Average societal profit for fixed price with charity: $1057.5
```

```
## Average societal profit for name your own price with charity: $2488.237
```

The ranking of average daily societal profits in this case:
1) `NYOP Charity`
2) `FR Charity`
3) `FR`
4) `NYOP`

d) The leading strategy in terms of societal profit is `NYOP Charity`.
The additional profits generated, over the span of the entire year (assuming the year has 365 days), can be seen by the model below:

```
fr_year <- avg_fr * 365
fr_char_year <- fr_char_total * 365
nyop_year <- avg_nyop * 365
nyop_char_year <- nyop_char_total * 365
```

The additional profits are summarized below:

```
## FR: $607993.9
```

```
## FR Charity: $522218.9
```

```
## NYOP: $1030153
```

e) Merchandise sales increased when there was charity involved.

This was a trend after analyzing the merchandise sales data, when we observed that the strategies, when compared to their non-charity counterparts (`FR` vs `FR Charity` and `NYOP` vs `NYOP Charity`). Merchandise sales should not be a concern to societal profit if the strategy that is being implemented contains a charity component. This will increase merchandise revenue which will cause an increase in park profit, but it will also increase charity contributions, through the picture sales revenue, which will ultimately increase societal profits.

If we use a strategy that does not include charity, merchandise revenue will be less than its charity counterpart. This will result in a decreased park profit and smaller charity profit, which amounts to a smaller societal profit. Therefore, merchandise sales are a concern if an non-charity strategy is being implemented.