

Homework3_SDS375

Caroline Zhuang

2025-10-01

```
data <- read.csv("homework3_data.csv")
colors <- get_colors("~/Desktop/redbull.png")
cols <- make_palette(colors[1:100,])
```



Summarizing sales, nps, items by design

```
means <- data %>%
  group_by(design) %>%
  summarize(mean(sales), median(sales), mean(nps), median(nps), mean(items), median(items))
means
```

```
## # A tibble: 2 x 7
##   design `mean(sales)` `median(sales)` `mean(nps)` `median(nps)` `mean(items)`
##   <int>      <dbl>      <dbl>      <dbl>      <int>      <dbl>
## 1     0        31.8        32.0        4.60         5         2.90
## 2     1        35.5        35.3        5.62         6         4.19
## # i 1 more variable: `median(items)` <int>
```

From the naked eye, it seems that there is an increase in sales, nps and items when comparing between the old design and the new design. We will perform some t-tests to confirm these results.

T-tests

```
t_test_sales <- t.test(sales ~ design, data = data)
t_test_sales
```

```
##
## Welch Two Sample t-test
##
## data: sales by design
## t = -8.1554, df = 186.01, p-value = 5.042e-14
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -4.551445 -2.778364
## sample estimates:
## mean in group 0 mean in group 1
## 31.84819 35.51309
```

Based off of the t-tests, I would recommend switching from the old design to the new design. For the two sample t-test between sales by design, the 95 percent confidence interval is (-4.55, -2.778), meaning we are 95% confident that the true mean difference of the sales in the old website and sales in the new website lies in that interval. This entire interval is below 0 and the p-value is also far smaller than 0.05, meaning these results are statistically significant.

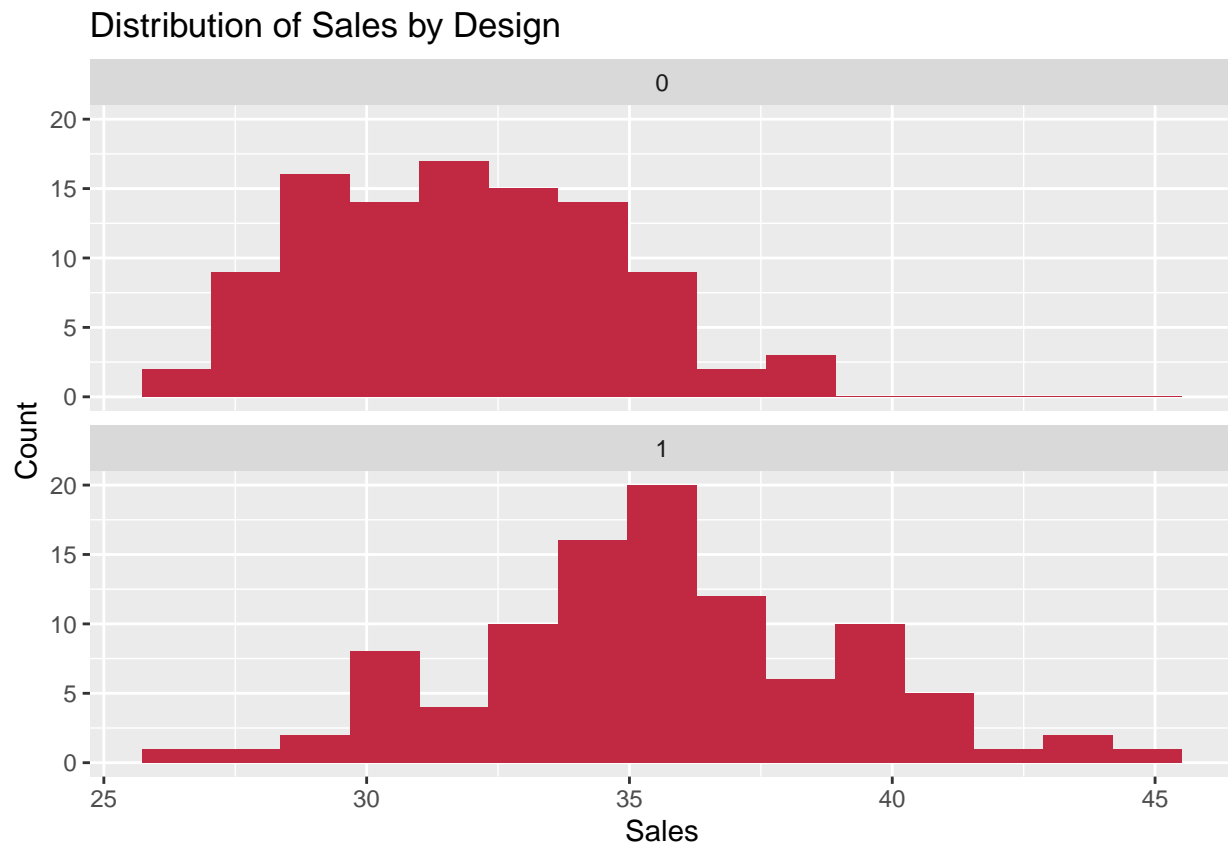
This means that the new design would increase sales per customer by about \$2.78 to \$4.55 on average, which is at least 1.80 per customer. The point estimate is around 3.66, so if the company were to continue having 200 customers per time period, the estimated sales increase would be \$732.

```
t_test_nps <- t.test(nps ~ design, data = data)
t_test_nps
```

```
##
## Welch Two Sample t-test
##
## data: nps by design
## t = -7.7396, df = 197.36, p-value = 5.054e-13
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -1.2701097 -0.7542928
## sample estimates:
## mean in group 0 mean in group 1
## 4.603960 5.616162
```

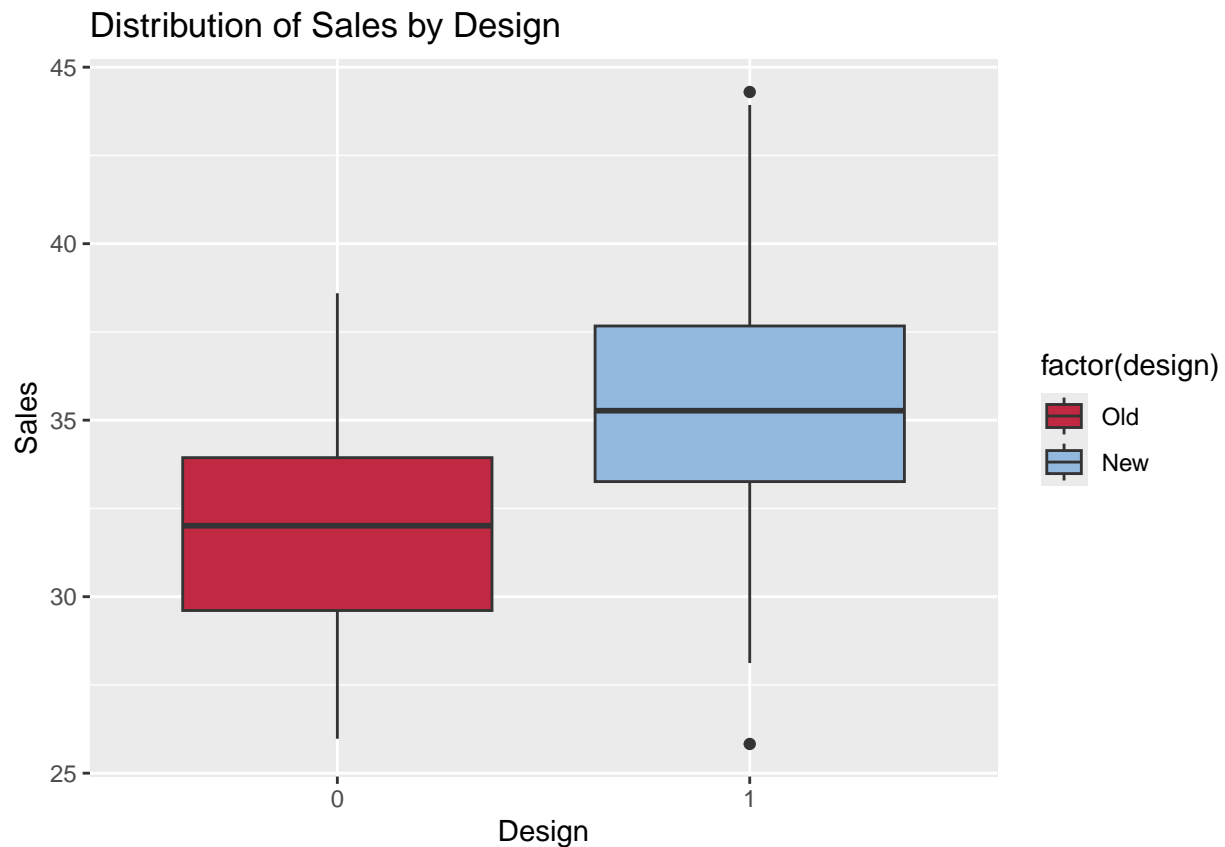
The same goes for the two-sample t-test of net promoter score by design. The 95% confidence interval is (-1.26, -0.75), meaning we are 95% confident that the true mean difference of the nps in the old website and sales in the new website falls between that interval. The entire interval is also below 0 and the p-value implies that these results are statistically significant.

```
sales_plot <- data %>%
  ggplot(aes(sales)) +
  geom_histogram(bins = 15, fill = "#C12841") +
  facet_wrap(~design, ncol = 1) +
  labs(title = "Distribution of Sales by Design", x = "Sales", y = "Count")
sales_plot
```



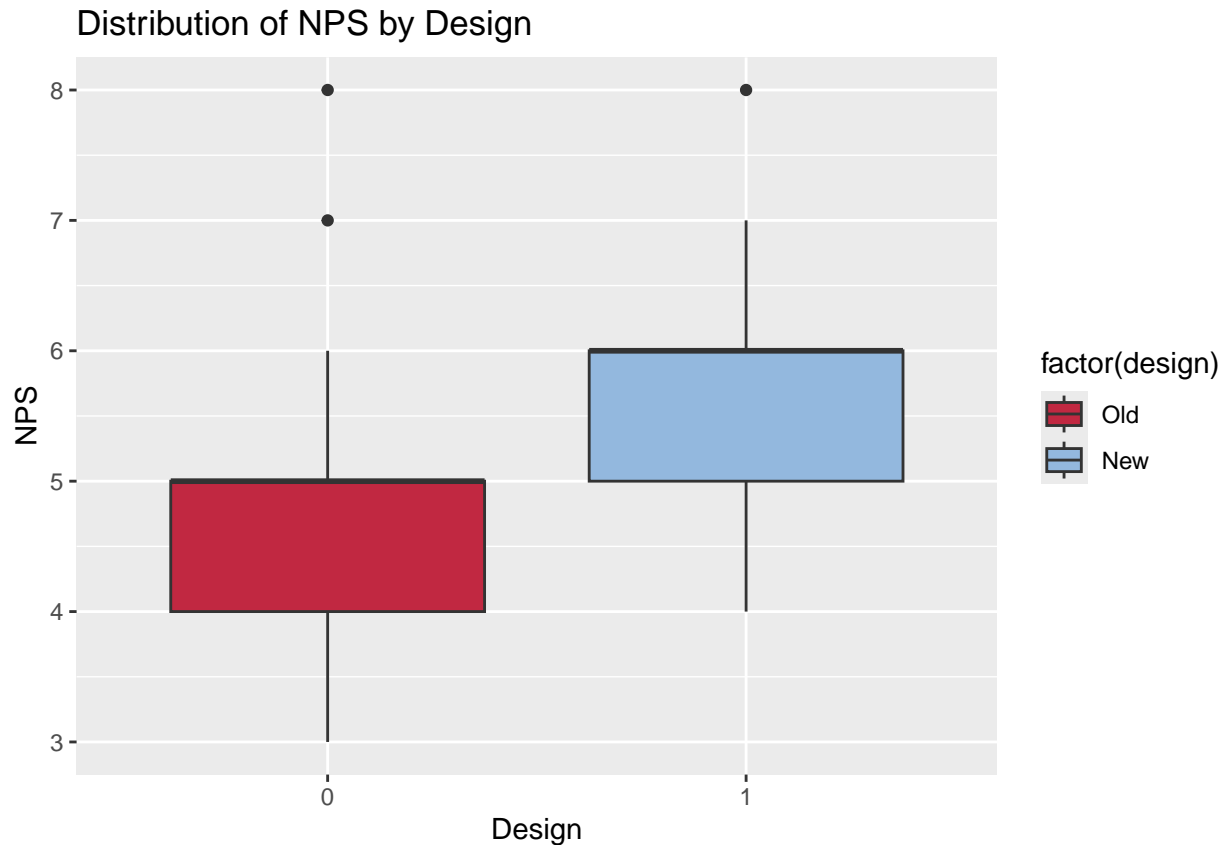
From the plot above, you can see the shift of the center to the right in the new design choice. This agrees with our t-test. However, I would like to note that the new website's range of sales is much wider than the previous website's, meaning there is more variance.

```
redbull_colors <- c("#C12841", "#93B8DE")
sales_box_plot <- data %>%
  ggplot(aes(factor(design), sales, fill = factor(design))) +
  geom_boxplot() +
  scale_fill_manual(values = redbull_colors, labels = c("Old", "New")) +
  labs(title = "Distribution of Sales by Design", x = "Design", y = "Sales")
sales_box_plot
```



From the boxplot above, it seems that both designs have pretty well balanced distributions, with the new design's median being higher than the previous design. The new design's distribution isn't completely above the previous design.

```
redbull_colors <- c("#C12841", "#93B8DE")
nps_box_plot <- data %>%
  ggplot(aes(factor(design), nps, fill = factor(design))) +
  geom_boxplot() +
  scale_fill_manual(values = redbull_colors, labels = c("Old", "New")) +
  labs(title = "Distribution of NPS by Design", x = "Design", y = "NPS")
nps_box_plot
```



From the boxplot above, an immediate thing of notice is that both designs have a very minimal Q3 (the top 25%). It seems very squished down to the median, suggesting that the distributions are skewed. However, the new design does have a higher median, suggesting that the newer design does better in terms of NPS.

Primary Statement

My final recommendation: Switching from the old website to the new website version would lead to an increased amount of revenue and NPS.

Alternative statement: The redesign will actually cost the company to lose money and have a declined NPS, even though the data suggests that it will do the opposite.

One thing that I thought of when analyzing this is thinking about the cost per item. Maybe looking at the average of sales and nps suggests an increase when redesigning, but customers are actually buying less profitable items. For example, the customers on the old website were more willing to buy more expensive items, which were marked up more by the company, and therefore led to more revenue.

```
# look at cost per item too
mean_cost_per_item <- data %>%
  mutate(cost_per_item = sales / items) %>%
  group_by(design) %>%
  summarize(mean(sales), mean(items), mean(cost_per_item))
mean_cost_per_item
```

```
## # A tibble: 2 x 4
##   design `mean(sales)` `mean(items)` `mean(cost_per_item)`
##   <int>     <dbl>       <dbl>         <dbl>
## 1     0     31.8       2.90         14.1
## 2     1     35.5       4.19         10.3
```