T Test Project: Customer Satisfaction

Alec Brooks

11/02/2024

Q1) Test whether the means of customer satisfaction are different for premium and standard members.

1) Which T-test to run.

```r
var.test(subset(info, Membership == "Premium")$Satisfaction,
         subset(info, Membership == "Standard")$Satisfaction)
```

```
##
##  F test to compare two variances
##
## data:  subset(info, Membership == "Premium")$Satisfaction and subset(info, Membership == "Standard")$
## F = 0.51946, num df = 46, denom df = 52, p-value = 0.02533
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.2959236 0.9209727
## sample estimates:
## ratio of variances
##          0.5194566
```

Since these variables are independent and the variance ratio is below 4.0, I will use an Independent Pooled Two-sample T-test.

2) The Hypotheses.

Null Hypothesis: The mean of the satisfaction score is equal between Standard and Premium members.

Alternative Hypothesis: The mean of the satisfaction score is **not** equal between Standard and Premium members.

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_a : \mu_1 - \mu_2 \neq 0$$

3) t-Value, df & p-value.

```r
t.test(subset(info, Membership == "Premium")$Satisfaction,
       subset(info, Membership == "Standard")$Satisfaction,var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  subset(info, Membership == "Premium")$Satisfaction and subset(info, Membership == "Standard")$
## t = 13.202, df = 98, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.846498 3.853622
## sample estimates:
## mean of x mean of y
##  8.425532  5.075472
```

t = 13.202

df = 98

p-value ~ 0

4) Conclusion, Interpretation, and Contextual Implication

With a p-value close to zero, we reject the null hypothesis, concluding there is a statistically significant difference in mean satisfaction scores between Standard and Premium members. The confidence interval (2.846, 3.854) suggests that, with 95% confidence, we know the true difference in mean satisfaction scores between groups is between 2.8 and 3.8 points.

Q2) Test whether the proportion of Premium and Standard members is equal.

1) Which test to run.

Since there are only two membership options (Premium and Standard), I will use a 1-sample proportion test on the subset of Premium members, with the proportion set at 0.5.

2) The Hypotheses.

Null Hypothesis: The proportion of Premium members is 50% of all members.

Alternative Hypothesis: The proportion of Premium members is not 50% of all members.

$$H_0 : p = 0.5$$
$$H_a : p \neq 0.5$$

3) t-Value, df & p-value.

```r
Premium_pop <- sum(info$Membership == "Premium")
Full_pop <- nrow(info)
prop.test(x = Premium_pop, n = Full_pop, p = 0.5)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  Premium_pop out of Full_pop, null probability 0.5
## X-squared = 0.25, df = 1, p-value = 0.6171
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.3703535 0.5719775
## sample estimates:
##    p
## 0.47
```

X-squared = 0.25 p-value = 0.6171

4) Conclusion, Interpretation, and Contextual Implication

With a p-value of approximately 0.62, we fail to reject the null hypothesis. This suggests that there is no statistically significant evidence to conclude that the proportion of Premium members is different from 50%. Therefore, the proportion of Premium and Standard members appears to be consistent with a 50-50 split.

Q3) Test if there is a significant difference in customer satisfaction before and after upgrading to Premium.

1) Which T-test to run.

Since we are comparing customer satisfaction scores before and after the upgrade for the same customers, these are paired variables. Therefore, I will use a Paired T-test.

2) The Hypotheses.

Null Hypothesis: The mean of satisfaction scores for customers before and after upgrade are equal.

Alternative Hypothesis: The mean of satisfaction scores for customers before and after upgrade are not equal.

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_a : \mu_1 - \mu_2 \neq 0$$

3) t-Value, df & p-value.

```r
t.test(score$Satisfaction_Before,score$Satisfaction_After,paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  score$Satisfaction_Before and score$Satisfaction_After
## t = -5.6525, df = 74, p-value = 2.795e-07
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
```

```
##  -0.9197028 -0.4402972
## sample estimates:
## mean difference
##           -0.68
```

t = -5.6525

df = 74

p-value = 0.0000002795

4) Conclusion, Interpretation, and Contextual Implication

With a p-value close to 0, we reject the null hypothesis and conclude there is a significant difference in customer satisfaction before and after upgrading to Premium. Additionally, since the mean difference is negative (-0.68), we can conclude that satisfaction after the upgrade is higher than before.

Q4) Test whether the changes in customer satisfaction before and after upgrading to Premium are different between genders.

1) Which T-test to run.

```r
score$Satisfaction_Change <- score$Satisfaction_After - score$Satisfaction_Before
var.test(subset(score, Gender == "Male")$Satisfaction_Change,
         subset(score, Gender == "Female")$Satisfaction_Change)
```

```
##
##  F test to compare two variances
##
## data:  subset(score, Gender == "Male")$Satisfaction_Change and subset(score, Gender == "Female")$Sati
## F = 1.0869, num df = 29, denom df = 44, p-value = 0.7883
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.5670038 2.1838973
## sample estimates:
## ratio of variances
##           1.086898
```

Since the ratio of variances is below 4.0, we will use an Independent Pooled Two-sample T-test.

2) The Hypotheses.

Null Hypothesis: The mean change in satisfaction scores for male customers is equal to that for female customers.

Alternative Hypothesis: The mean change in satisfaction scores for male customers is not equal to that for female customers.

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 \neq 0$$

3) t-Value, df & p-value.

```r
score$Satisfaction_Change <- score$Satisfaction_After - score$Satisfaction_Before
male_changes <- subset(score, Gender == "Male")$Satisfaction_Change
female_changes <- subset(score, Gender == "Female")$Satisfaction_Change

t.test(male_changes, female_changes, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  male_changes and female_changes
## t = 0.5856, df = 73, p-value = 0.5599
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.3471457  0.6360346
## sample estimates:
## mean of x mean of y
## 0.7666667 0.6222222
```

t = 0.5856

df = 73

p-value = 0.5599

4) Conclusion, Interpretation, and Contextual Implication

With a p-value of 0.559, which is greater than 0.05, we fail to reject the null hypothesis. This suggests that there is no statistically significant difference in the mean satisfaction change between male and female customers after upgrading to Premium.

Q5) Based on what you found through #1-4, how do you conclude your project regarding customer behavior?

we learned, approximately half of the customers upgraded to Premium. There is a significant difference in satisfaction between Standard and Premium members, with Premium members rating 2.8 to 3.8 points higher on average. Most of these higher ratings come after upgrading, and there is no significant difference in this increase between genders.

```r
library(readr)
info <- read_csv("CustomerInformation.csv")
score <- read_csv("CustomerScoreChange.csv")

#1
var.test(subset(info, Membership == "Premium")$Satisfaction,
         subset(info, Membership == "Standard")$Satisfaction)
t.test(subset(info, Membership == "Premium")$Satisfaction,
       subset(info, Membership == "Standard")$Satisfaction,var.equal = TRUE)
#2
Premium_pop <- sum(info$Membership == "Premium")
```

```r
Full_pop <- nrow(info)
prop.test(x = Premium_pop, n = Full_pop, p = 0.5)
#3
t.test(score$Satisfaction_Before,score$Satisfaction_After,paired = TRUE)
#4
score$Satisfaction_Change <- score$Satisfaction_After - score$Satisfaction_Before
var.test(subset(score, Gender == "Male")$Satisfaction_Change,
         subset(score, Gender == "Female")$Satisfaction_Change)
score$Satisfaction_Change <- score$Satisfaction_After - score$Satisfaction_Before
male_changes <- subset(score, Gender == "Male")$Satisfaction_Change
female_changes <- subset(score, Gender == "Female")$Satisfaction_Change
t.test(male_changes, female_changes, var.equal = TRUE)
```