Assignment 5

Kyle Ligon

2018-11-27

Problem A- Chi Square Test for Differences in Probabilities

```
men <- c(32, 68)
women <- c(26, 74)

taste <- data.frame(men, women, row.names = c("no likey",
    "likey"))
taste_test <- chisq.test(x = taste, correct = FALSE) %>%
    tidy()
```

Hypotheses:

H_0: $p_{men} = p_{women}$
H_1: $p_{men} \neq p_{women}$

Test Statistic

The test statistic of this Chi-Square test is 1

Critical Region

We are looking for a $\chi^2_{0.975,1}$ and $\chi^2_{0.025,1}$, which are equal to 5.0238862 and 0.001.

Conclusion

With the test statistic not greater than or less than the critical region, we cannot reject the Null hypothesis that the probabilities are the same. There is not enough evidence to suggest that the two probabilities are different.

Problem B- Fisher's Exact Test

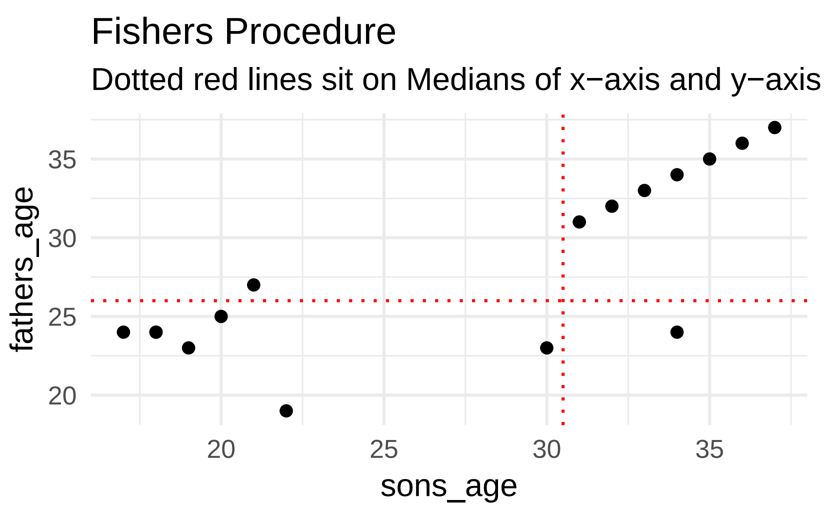
Suppose that 16 observations pairs of X = age of marriage of a husband, and Y=age of marriage of his father, resulted on 7 pairs where both ages were above the median. Are the two variables positively correlated?

```
sons_age <- c(19, 18, 18, 20, 21, 34, 17, 30,
    31, 32, 33, 34, 35, 36, 37, 22)
fathers_age <- c(23, 24, 24, 25, 27, 24, 24, 23,
```

```
    31, 32, 33, 34, 35, 36, 37, 19)

# make the data frame
age <- data.frame(sons_age = sons_age, fathers_age = fathers_age)

# make the scatterplot
ggplot(data = age, aes(x = sons_age, y = fathers_age)) +
    geom_point() + labs(title = "Fishers Procedure",
    subtitle = "Dotted red lines sit on Medians of x-axis and y-axis") +
    geom_vline(xintercept = median(sons_age),
        linetype = "dotted", color = "red") +
    geom_hline(yintercept = median(fathers_age),
        linetype = "dotted", color = "red") +
    theme_minimal()
```




```
counter <- function(vector1, vector2) {
    ### Arguments: vector1- {Vector of sons_ages}
    ### vector2- {Vector of fathers_ages}
    q1_cnt = 0
    q2_cnt = 0
    q3_cnt = 0
    q4_cnt = 0
    for (i in 1:length(vector1)) {
        if ((vector1[i] >= median(vector1)) &
            (vector2[i] > median(vector2))) {
            q1_cnt = q1_cnt + 1
        } else if ((vector1[i] <= median(vector1)) &
            (vector2[i] > median(vector2))) {
            q2_cnt = q2_cnt + 1
        } else if ((vector1[i] <= median(vector1)) &
```

```
            (vector2[i] <= median(vector2))) {
            q3_cnt = q3_cnt + 1
        } else q4_cnt = q4_cnt + 1
    }
    cnt_table = c(q2_cnt, q1_cnt, q3_cnt, q4_cnt)
    cnt_table <- as.table(matrix(cnt_table, nrow = 2))

    fisher.test(cnt_table, alternative = "less") %>%
        tidy()
}

test <- counter(sons_age, fathers_age)
```

To accomplish this, we will run a Lower Tailed Fisher Exact Test. The reason for lower tailed is due to the fact the p_1 would be lower than or equal to p_2 if they were positively correlated.

Hypotheses:

$H_0 : p_1 \geq p_2$
$H_1 : p_1 < p_2$

Test Statistic:

The test statistic is 0.031639.

P-Value:

The p-value is 0.0050505.

Conclusion:

With a p-value less than 0.05, there is enough evidence to suggest that we reject the null hypothesis that p_1 is less than or equal to p_2. There does seem evidence to support the claim that p_2 is of higher likelihood.[1]

[1] Where p_2 is the likelihood that an Observation from Column 1 lies in Row 2

Problem C- Chi Square Test for Differences in Probabilities

```
ase <- c(11, 11, 1)
nyse <- c(24, 11, 0)

stocks <- data.frame(ase, nyse, row.names = c("A",
    "B", "C")) %>% t()

stock_test <- chisq.test(x = stocks) %>% tidy()
```

Hypotheses:

H_0: $p_{ASE} = p_{NYSE}$
H_1: At least two of the populations have different populations.

Test Statistic

The test statistic of this Chi-Square test is 3.4954392.

Critical Region

We are looking for a $\chi^2_{0.95,2}$, which is equal to 5.9941.

Conclusion

With the test statistic less than the critical region, we cannot reject the Null hypothesis that the ratings percentages are the same. There is not enough evidence to suggest that the two stock groups have different ratings groups.

Problem D- Chi-Square Test

```
products <- data.frame(fishing_rod = c(6, 14,
    21), kitchen_tool = c(73, 65, 58), music_cd = c(55,
    82, 48), exercise_machine = c(7, 8, 8), row.names = c("daytime",
    "nighttime", "weekend"))

products_test <- chisq.test(x = products) %>%
    tidy()
```

1) What does your analysis look like?

I'm going to test to see whether the time of day has any affect on the proportion on the product being sold.

Hypotheses:

H_0: Products' sales are independent of what time their ad airs.
H_1: Products' sales depend upon the time their ad airs.

Test Statistic

The test statistic of this Chi-Square Test for Independence is 16.5425.

Critical Region

We are looking for a $\chi^2_{0.95,6}$, which is equal to 12.5916.

Conclusion

With the test statistic greater than the critical region, we can reject the Null hypothesis that the product sales are independent of what time their ad airs. There is evidence to suggest that the products' sales depend upon their time their ad airs.

2) Calculate and comment on Cramer's Contingency Coefficient

```
products_cramers_coef <- cramersV(products)
```

With Cramer's Contingency Coefficient showing us the amount of association between the variables, our coefficient of 0.1363 tells us there is little association between the prouducts and the times they air.

Problem E- Median Test

```
sampl_1 <- c(35, 42, 42, 30, 15, 31, 29, 29, 17)
sampl_2 <- c(34, 38, 26, 17, 42, 28, 35, 33, 16)
sampl_3 <- c(17, 29, 30, 36, 41, 30, 31, 23, 38)

sample_col <- c(rep("samp_1", length(sampl_1)),
    rep("samp_2", length(sampl_2)), rep("samp_3",
        length(sampl_3)))

medians <- c(sampl_1, sampl_2, sampl_3)

med_frame <- data.frame(sample_col = as.factor(sample_col),
    medians = medians) %>% as.tibble()

med_test <- Median.test(trt = med_frame$sample_col,
    y = med_frame$medians)
```

Hypotheses:

H_0: All *3* populations have the same median.
H_1: At least two of the populations have different medians.

Test Statistic

The test statistic of this Median test is 0.8269.

Critical Region

We are looking for a $\chi^2_{0.95,2}$, which is equal to 5.9941.

Conclusion

With the test statistic greater than the critical region, we can reject the Null hypothesis that the medians are the same. There is evidence to suggest that at least two medians are different.