Hi there,

In this project, I am doing some statistical analysis to investigate the attitude of people towards using their cellphone for day-to-day shopping or online banking. I get into how this attitude varies in various age categories and how this belabour is different between men and women. I did all the analysis using the magic \*\*SPSS\*\* tool! It rocks!

There are 4 variables in the datafile:

Sex: 0 for man, 1 for women

Age:

* 1: 20 or younger
* 2: between 21 and 30
* 3: between 31 and 40
* 4: between 41 and 50
* 5: 50 and older

Shopping: respondent attitude towards using cell phone for online shopping

Banking: respondent attitude towards using cellphone for online banking

Both shopping and banking variables were each constructed by averaging three items from the questionnaire. Items were measured on a 5-point scale (1 = strongly disagree, 2 = disagree, 3 neither agree nor disagree, 4 = agree, 5 = strongly agree).

**Statistical description of sex and age variables:**

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

**Comparison men and women:**

**A picture containing table, parked, stop, clock

Description automatically generated**

The plot box above as well as below table shows that the average of shopping score is different between men and women. In order to prove that this saying is scientifically true we do the Levene test as men and women are two independent samples.

**A screenshot of a cell phone

Description automatically generated**

The result of the Levene test shows that the variance of two groups are not equal (p-value = 0.273 > 5%).

The P-value of the non-equal variances is 0.21 which is smaller than 5%. So, by a confidence level of 95%, we estimate that the behaviours of the women towards using cellphone for online shopping is significantly different than that of men.

**A screenshot of a cell phone

Description automatically generated**

**A picture containing parked, clock

Description automatically generated**

Based on the plot box and case summary table for banking, the average of banking score for men and women is different. Just like shopping, we do Levene test to find out the truth.

**A screenshot of a cell phone

Description automatically generated**

With a p-value of 0.073, we conclude that by 95% level of confidence, the variance of banking score is different between men and women. On the other hand, the p-value of banking score mean comparison between men and women (0.142 < 5%) shows that by 95% confidence we estimate that there is no significant difference between men and women in terms of their behaviour towards using cellphone for online banking.

**A screenshot of a cell phone

Description automatically generated**

**Comparing shopping score and banking score:**

**A screenshot of a cell phone

Description automatically generated**

The boxplot if difference of shopping scores and banking scores (shop\_bank) shows that this difference has extreme values.

Here shopping score and banking score are two dependent samples as each respondent has one shopping and one banking score. I subtract these two scores for each respondent, and I created a paired sample. The p-value of paired sample t-test is zero. This proves that by 95% chance, the average score of banking and shopping for all respondent, regardless of their sex, is significantly different.

**A screenshot of a cell phone

Description automatically generated**

Since this difference has some extreme values, I test the results with non-parametric sign test to ensure the liability of the aforementioned estimation. The p-value of sign test is 0. So, here again by a confidence level of 95% I estimate that the average banking and shopping scores are significantly different.

A screenshot of a cell phone

Description automatically generated

**Comparison of the attitude if different age group towards using cellphone for online shopping:**

Here I use ANOVA test to know if there is any difference between the 5 age groups in their attitude towards using cellphone for shopping.

I start the analysis by plotting the boxplot to have a general idea about the average and variance of shopping score of each age group.

A screenshot of a cell phone

Description automatically generated A picture containing table, parked

Description automatically generated

= the mean score of attitudes towards using cellphone for online shopping of each age category *i.*

I use Levene test to check the variance homogeneity:

at least two variances are different

P-value of the homogeneity test 0.828 which is greater than 5%. So, I do no reject H0 and I estimate that by 95% chance, the variances of the five age groups are similar in the population.

A screenshot of a cell phone

Description automatically generated

Since the size of the population is greater than 30, variances amongst 5 age groups are equal and age group samples are independent, the 3 conditions of Fisher test are satisfied, and I can use Fisher T-test for comparing the average of shopping score of each age group.

Fisher test hypothesis:

At least two averages are different

A screenshot of a cell phone

Description automatically generated

The p-value pf ANOVA test is 0.00 < 0.05. So, is not rejeted and i can estimate by 95% of confidence level that the average attitude towards using cellphone for shopping id different for each age group.

Since the boxplot shoe a few extreme cases, I verify this conclusion through Kruskal-Wallis non-parametric test:

A screenshot of a cell phone

Description automatically generated

The p-value of Kruskal-Wallis test is 0.00 < 0.05. I do not reject and the D\Fisher test result is confirmed.

I continue the analysis by comparing each age group to each of the others (multiple comparisons) to find out where the differences lie. Since there are several tests to do, I use Tukey method that controls the overall level.

A close up of a piece of paper

Description automatically generated

I interpret the attitude of using cellphone for shopping for the following groups ages as an example:

* : By 95% confidence, I estimate that the average score of using cellphone for shopping for users of 20 years and younger is between 0.1245 to 0.9938 greater than that of users who are between 41 and 50 years old
* : By 95% confidence, I estimate that the average score of using cellphone for shopping for users of 20 years and younger is between 0.1089 and 0.9884 greater than that of users who are 51 years and older.

**Comparison of the attitude if different age group towards using cellphone for online banking:**

I repute the analysis above but this time to investigate the attitude of these 5 different age groups toward using cellphone for online banking. Just like the previous section, I start with homogeneity test to see if the variance of banking score of different age groups are equal or not.

A screenshot of a cell phone

Description automatically generated

A screenshot of a video game

Description automatically generated

According to the case summary table, the average of the banking score of these 5 different age groups are different. Same is evident in the boxplot as well. Moreover, boxplot shows a few extreme values for age groups 4 and 5.

I go ahead and do the homogeneity Levene test and investigate the Fisher test conditions. The p-value of the Levene test is 0.081 > 0.05. So, the H0 is not rejected and by a confidence level of 95%, I can say that the banking score variances of 5 age groups are equal.

A screenshot of a cell phone

Description automatically generated

On the other hand, the size of the population (154) is greater than 30 and all age group samples are independent. So, the Fisher test conditions are satisfied, and I can use Fisher test to decide if the banking score average of these group ages are equal or not.

Fisher test hypothesis:

At least two averages are different

A screenshot of a cell phone

Description automatically generated

The p-value of Fisher test is 0.25 > 0.05. Therefore, I cannot reject H0 and I can estimate that by a 95% chance, the banking average does not vary amongst different age groups.

Since there are some extreme value sin the dataset, I validate the result of Fisher test by Kruskal-Wallis test.

A screenshot of a cell phone

Description automatically generated

As seen, the p-value of this test is greater than 5% (0.8087 > 0.05) and the results confirm the conclusion of Fisher test.

**Working with binary variables:**

What if my dependent variable is binary? ;)

I convert the banking and shopping variables to binary variables according to the following classification rules:

shopping\_binary:

* 1 if shopping >=3
* 0 if shopping < 3

banking\_binary:

* 1 if shopping >=3
* 0 if shopping < 3.

Here, I am willing to investigate if the proportion of men with a rather favorable attitude towards using cellphones for shopping is different from that of women.

= proportion of men having a more favourable attitude toward using cellphone for shopping

= proportion of men having a more favourable attitude toward using cellphone for shopping

Voila the test hypothesis:

Since two samples are independent (men and women), I use Chi-Square test to investigate if these 2 proportions are equal or not.

A screenshot of a cell phone

Description automatically generated

The p-value of Pearson Chi Square test is 0.018 < 0.05. I reject and I can say that by a confidence level of 95%, the proportion of men having a more positive attitude toward using cellphone for shopping is different than that of women.

A screenshot of a cell phone

Description automatically generated

Same analysis for the attitude of men and women toward banking though shows that by a confidence level of 95%, the proportion of men having a more positive attitude toward using cellphone for shopping is equal to that of women (P-value = 0.55).

A screenshot of a cell phone

Description automatically generated

At this step, I am curious to know if the proportion of people with a rather favorable attitude to cellphone use for shopping is different from the proportion of people with a rather favorable attitude to use cell phone for banking?

Here, two shopping and baking samples are independent, so I use McMemar test to answer my curiosity.

= proportion of people having a favourable attitude toward using cellphone for shopping

= proportion of people having a favourable attitude toward using cellphone for banking

My hypothesis tests:

The p-value of McNemar test is 0 < 0.05, so H0 is not rejected and by a 95% confidence I can estimate that the proportion of people having favourable attitude towards shopping and banking are significantly different.

A screenshot of a cell phone

Description automatically generated

That’s it! Hope you enjoyed this statistical journey. Do not hesitate to write me for comments or questions ☺

Cheers,

Niloo