APS 2014 and 2020

Overall results of 2020

1. <https://www.apsc.gov.au/initiatives-and-programs/workforce-information/aps-employee-census-2020/2020-aps-employee-census-overall-results>

When to use an unpaired t-test?

* An unpaired t-test is used to compare the mean between two independent groups. You use an unpaired t-test when you are comparing two separate groups with equal variance.

Examples of appropriate instances during which to use an unpaired t-test:

* Research, such as a pharmaceutical study or other treatment plan, where ½ of the subjects are assigned to the treatment group and ½ of the subjects are randomly assigned to the control group.
* Research during which there are two independent groups, such as women and men, that examines whether the average bone density is significantly different between the two groups.
* Comparing the average commuting distance traveled by New York City and San Francisco residents using 1,000 randomly selected participants from each city.

Mann-Whitney test

1. <https://www.sheffield.ac.uk/polopoly_fs/1.885207!/file/99_Mann_Whitney_U_Test.pdf>

Spearman correlation

1. The Spearman correlation method computes the correlation between the rank of x and the rank of y variables.
   * cor(x, y, method = c("pearson", "kendall", "spearman"))
   * where x,y: numeric vectors with the same length

Chi-square statistics )

1. Used for nominal data (Male / Female)
   * Hypotheses
     1. Null: Assumes that there is no association between the two variables.
     2. Alt: Assumes that there is an association between the two variables.
   * <http://www.sthda.com/english/wiki/chi-square-test-of-independence-in-r>

Welch t-test

1. 4 conditions required to meet
   1. Two data need identical interval and continuity
   2. Two groups are independent (independence)
   3. Two data need to follow normality
   4. Two groups have unequal variance
2. Welch's t-test performs better than Student's t-test whenever sample sizes and variances are unequal between groups.

Unpaired (2014 and 2020 APS data) two-samples t-tests for numerical values

1. Unpaired two-samples t-test can be used only under certain conditions:

* When the two groups of samples (A & B), being compared, are **normally distributed**. This can be checked using **Shapiro-Wilk test**.
* And when the **variances** of the two groups are **equal**. This can be checked using **F-test**

1. Shapiro-Wilk test in R
   * <http://www.sthda.com/english/wiki/normality-test-in-r>

install.packages("dplyr") # for data manipulation

**if**(!**require**(devtools)) install.packages("devtools") devtools::install\_github("kassambara/ggpubr")# ggpubr for an easy ggplot2-based data visualization

or

install.packages("ggpubr")

# Load required packages

**library**("dplyr")

**library**("ggpubr")

# Import data

my\_data <- read.csv(file.choose())

1. Check for **normality**

# Random sample of 10 rows

set.seed(1234)

dplyr::sample\_n(my\_data, 10)

# Normality test if sample size n < 10, skip if n > 30 because of the central limit theorem

shapiro.test(my\_data$responseV)

* From the output, the p-value > 0.05 implying that the distribution of the data are not significantly different from normal distribution. In other words, we can assume the normality.

2. Check for **homoscedasticity**

* F-test is sensitive to the outliers and normal assumption. Therefore, Shapiro-Wilk test should be preceded.
* If data is not normally distributed, try Levene’s test or Fligner-Killeen test instead of F-test because those two test are less sensitive to the normal assumption

# F-test hypotheses

* Whether the variance of group A is equal to the variance of group B? (two-tailed)
* Whether the variance of group A is less than the variance of group B? (one-tailed)
* Whether the variance of group A is greater than the variance of group B? (one-tailed)

Ex) var.test(values ~ groups, data, alternative = “two.sided”)

3. Back to t-test for unpaired two-samples t-test

# T-test hypotheses

* Whether the variance of group A is equal to the variance of group B? (two-tailed)
* Whether the variance of group A is less than the variance of group B? (one-tailed)
* Whether the variance of group A is greater than the variance of group B? (one-tailed)

# 4. Preliminary test to check independent t-test assumptions

* Assumption 1: Are the two samples independents?
* Assumption 2: Are the data from each of the 2 groups follow a normal distribution?
* Assumption 3: Do the two populations have the same variances?

# 4a. Check for homoscedasticity or heteroscedasticity

1. If data is homoscedasticity
   * Classical t-test
2. If data is heteroscedasticity
   * Welch t-statistic

# 4b. Compute in R

t.test(x, y, alternative=”two.sided”, var.equal=TRUE)

# var.equal=TRUE for classical test

# var.equal=FALSE for Welch test

# 4c. Check data and summary statistics using library(dplyr)

# 4d. Visualize the data using library(“ggpubr”)

# 5. Compute t-test

Res = t.test(data1\_variable, data2\_variable, var.equal = TURE)

**How to correlate data in likert scale:**

Non parametric correlation: Generally when we want to find correlation between variables, we use pearson correlation. However, pearson correlation is used when the data is continuous like sales data or advertisement expenditure.

When the data is ORDINAL like likert scale having strongly agree to strongly disagree, we must use spearman correlation. Spearman correlation is also the non parametric test to determine correlation between ordinal variables like likert scale. It uses rankings of data rather than absolute values of the variables.

Spearman test can also be done on 7 point likert scale

Basically the 2 variables data to correlate can be in the form of 5 point likert scales each or 7 point likert scales each. The type should be ordinal

Graphical user interface

Description automatically generated with medium confidence

What is likert scale:

One of the most popular methods of measuring respondent’s attitude on a question is likert scale. It is the measure of attitude which is designed to allow the respondents to rate how strongly they agree or disagree to each constructed statement. It is also called SUMMATED SCALE because the overall scale score can be calculated by summating the attribute values of each items as selected by the respondents.

How is score from likert scale calculated:

Example: are you satisfied with your job ?

1. Strongly disagree
2. disagree
3. Neutral
4. agree
5. Strongly agree

There will be approximately 30 statements with respect to job. It a respondent strongly agree with all 30 statements, then the total score will be 30\*5 = 150. If the respondent strongly disagree with all 30 stataments, then the total score will be 30\*1 = 30. If the attitude is neutral, that is if respondent answers all 30 in a neutral way, total score = 30\*3 = 90.

Therefore for a likert scale on the basis of this, the total score will range between 30 to 250. A respondent if score above 90, then he is considered to have a favourable attitude for the question. If he scores less than 90, then unfavourable attitude and if the score is 90, then neutral attitude.

Table

Description automatically generated

Why do we use only non parametric tests and analysis for likert type data?

Because all parametric tests like pearson correlation and t tests etc require interval type data. Likert scale is ordinal data and therefore for ordinal data parametric tests cannot be used.

However,

**When multiple likert question responses are summed together,** it may form interval data and therefore in this case, parametric tests can be used like t tests, factor analysis and ANOVA

Also,

**Sometimes, the data from likert scale reproduced to nominal level** by combining agree and disagree responses into 2 categories: example 5 point likert scale is grouped to (positive response or non positive response), in this case, we can use chi square test, Cochran q test and Mcnemar test.

Chi square test is used to compare proportion between 2 or more independent groups or investigate if there is any association between 2 nominal scale variables.

For likert scale regression analysis, ordinal logistic regression and multinomial logistic regression is used, or for binary variables, can use binary logistic regression.