Date: 12/2/2022

1. Statistical tests and when they are used

a. **T test:** used when comparing the means of precisely two groups. The t-test assumes the data:

1. are independent
2. are normally distributed.
3. have a similar amount of variance within each group being compared

b. **Z test:** used to determine whether two population means are different when the variances are known and the sample size is large. The test statistic is assumed to have a normal distribution, and nuisance parameters such as standard deviation should be known in order for an accurate z-test to be performed. Z-tests are closely related to t-tests, but t-tests are best performed when an experiment has a small sample size. Z-tests assume the standard deviation is known, while t-tests assume it is unknown.

$$Z = \frac{(\bar{X} - \mu_0)}{s}$$

c. **Chi squared test:** used to compare observed results with expected results. The purpose of this test is to determine if a difference between observed data and expected data is due to chance, or if it is due to a relationship between the variables being studied. Therefore, a chi-square test is an excellent choice to help us better understand and interpret the relationship between our two categorical variables.

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

d. **Rank test:** is any test involving ranks. Examples include:

i. Wilcoxon signed-rank test

ii. Kruskal–Wallis one-way analysis of variance

iii. Mann–Whitney U (special case)

iv. Page's trend test

v. Friedman test

vi. Rank products

e. ANOVA: ANOVA stands for Analysis of Variance. It's a statistical test that was developed by Ronald Fisher in 1918 and has been in use ever since. Put simply, ANOVA tells you if there are any statistical differences between the means of three or more independent groups. One-way ANOVA is the most basic form. There are other variations that can be used in different situations, including:

i. Two-way ANOVA

ii. Factorial ANOVA

iii. Welch's F-test ANOVA

iv. Ranked ANOVA

v. Games-Howell pairwise test

ANOVA helps in finding out whether the differences between groups of data are statistically significant. It works by analyzing the levels of variance within the groups through samples taken from each of them. If there is a lot of variance within the data groups, then there is more chance that the mean of a sample selected from the data will be different due to chance.

As well as looking at variance within the data groups, ANOVA takes into account sample size (the larger the sample, the less chance there will be of picking outliers for the sample by chance) and the differences between sample means. All these elements are combined into a F value, which can then be analyzed to give a probability (p-value) of whether or not differences between the groups are statistically significant.

A one-way ANOVA compares the effects of an independent variable on multiple dependent variables. Two-way ANOVA does the same thing, but with more than one independent variable, while a factorial ANOVA extends the number of independent variables even further.

f. ARIMA model (autoregressive integrated moving average) It's a model used in statistics and econometrics to measure events that happen over a period of time. The model is used to understand past data or predict future data in a series. It's used when a metric is recorded in regular intervals, from fractions of a second to daily, weekly or monthly periods.

Date: 12/2/2022

2. How to download multiple packages at once using anaconda

To install scipy and curl: `conda install scipy curl`

To install scipy and curl with specified versions: `conda install scipy=0.15.0 curl=7.26.0`