Data Type

2019年1月31日 星期四　　下午10:55

Data Type

Nominal: Categorical, Qualitative, non-parametric (no order)

Ordinal: have meaningful order, Eg: satisfaction, fanciness, etc. Gap可能不一样，需要注意mean的定义

Interval/Ratio: Scale, Quantitative, Parametric. 可能有continuous或discrete

Main idea of ANOVA:

Figure out how much of the total variance comes from:
F = The variance between the groups / the variance within the groups
F(b, w)
b = #groups - 1 (d.f. for var between groups)
W = #observations - #groups (d.f. for var within groups)
若主要来自组里，accept H0; 反之，reject

Types of t-test:

1. **Independent samples (between samples, unpaired samples) t-test**: tests the man of two different groups
2. **Paired samples (within subjects, repeated measures, dependent samples) t-test:** tests the man of one group twice (比如同一组人服药前、服药后)
3. **One sample t-test:** tests the man of one group against a hypothetical value or a known population mean

问题：

1) Sample & population should be roughly normal in distribution. 如果skewed, p values may be inaccurate
2) Each group should have similar size of data points. 不能把大组和小组比。
3) Independent samples
4) Data should be approximately interval-level or higher. (不适用于order

之类)

Chi-squared test

Nominal/categorical data
对每一个cell, O=observed value, E=expected value, 对cell 求和
$\sum \frac{(O-E)^2}{E}$
再加上d.f.，就可以得到probability, 看是否reject H0
这里算E的方式：对每一个cell(i, j), i row total* j column total/总的total就是E(i, j)
d.f. = (#rows - 1) * (#columns - 1)

Choose test: 根据data, #samples, purpose来选

Data:

Nominal ->Test for proportion, Difference of two proportions, Chi-sq test for independence
* （可以表示为frequency）
Interval/Ratio -> Test for a mean, Difference of two means (independent samples), Difference of two means (paired), Regression analysis
Ordinal -> 可能取nominal，可能取interval/ratio, 取决于具体情况

Samples:

1 sample -> Test for proportion, Test for a mean
2 samples -> Difference of two proportions, Difference of two means (independent samples)
1 sample, 2 measures or variables -> Chi-sq test for independence, Regression analysis, Difference of two means (paired)

Purpose:

Test for value -> Test for proportion, Test for a mean, Difference of two means (paired)
Compare two statistics -> Difference of two proportions, Difference of two means (independent samples)

Find relationship -> Chi-sq test for independence, Regression analysis

注意 Chi-sq test for independence, Regression analysis的差别在于data, 前者用于nominal (用table表示), 后者用于interval/ratio（用散点图表示）