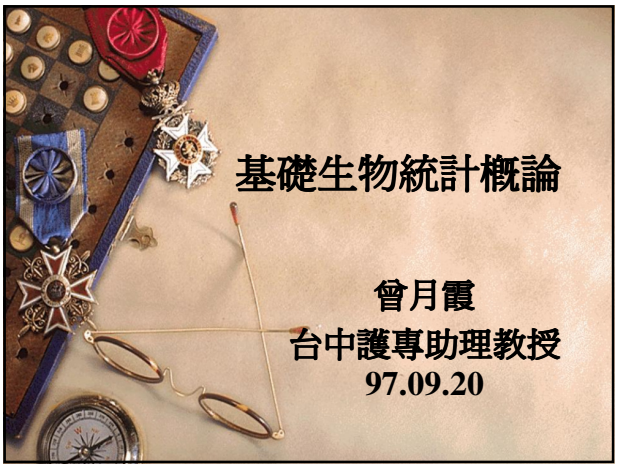

基礎生物統計概論

曾月霞
台中護專助理教授
97.09.20

Data level

- 類別資料(Nominal data)
  - Name categories with no implied order among the categories
- 序位資料(Ordinal data)
  - Ordered categories, where the differences between categories cannot be considered to be equal
- 等距資料(Interval data)
  - Has equal distances between values, but the zero point is arbitrary (參考用，不具意義)
- 等比資料(Ratio data)
  - Has equal intervals between values and a meaningful zero point

Data level

- ICD-9疾病分類
- N1, N2, N3
- 婚姻狀態
- 疼痛 yes/no
  1 to 10
- 身高、體重
- BUN, Hb, Ht
- 健康狀態 poor/fair/good/excellent
  SF-36

Level of measurement

- Have variables measured on the highest level of measurement possible because of the advantages
  - Greater analytic flexibility
  - Availability of more powerful statistical techniques
  - Greater amount of information than at lower levels
- The disadvantages of lower level of measurement
  - Loss of much information; can use "Recode" to change higher level of measurement into lower level of measurement

Variable

- Independent variable
- Dependent variable

母數分析 vs. 非母數分析

- 樣本數太少
- Assumptions are not met

Central tendency

- The general location of a "typical" data value, i.e., the data value around which other scores tend to cluster.
  - Mode
  - Median: divided the distribution into 2 equal halves
  - Mean: the most commonly used index of central tendency

Central tendency

- In a normal distribution, the mode= the median = the mean
- If mean > median, indicating the distribution is positive skew
- If mean < median, indicating the distribution is negative skew
- If distributions are close to being normal, the researcher only reports the mean. If not, researchers report 2 or more indicators of central tendency to reveal more information about the data.
- When the data are highly skewed or when there extreme, but valid, outlier, or data were measured on an ordinal scale, the median is a better index of central tendency.

Variability

- How similar or different the sample members are from one another
  - The range
  - SD：以mean爲參考點，分數離mean的遠近程度

Example

- 2 2 2 2 2 2 2 2 Mean = 2
- 0 1 1 1 1 1 1 8 Mean = 2
- Variability?
- Homogenous?

描述性統計

- 多用在人口學之描述
- 次數(Frequency)、百分比(percentage)
  - 研究樣本中男女各有幾位?百分比爲何?
- 平均數(Mean)及SD
  - 研究樣本的平均年齡是多少? SD= ?
- 眾數(Mode)
  - 出現最多次的年齡是多少?
- 中位數(Median)

卡方檢定

- Data level
  - Nominal vs. nominal data
- 男女與罹患疾病(CVA, DM, H/T)有無差異?

McNemar test

- For 2x2 tables, nominal and nominal data
- Coding 的方式要前後一致
- 同一組人之前測、後測
  - 在衛教後執行breast self-exam人數比例有無差異

相關(Correlation)

- Pearson's correlation
  - Interval/ration vs. interval/ration data
  - 年齡與血壓是否有關?
- Spearman's correlation
  - Ordinal vs. ordinal data
  - 學歷與職位是否有關?

Independent t- test vs. Mann-Whitney test t-test

- Assumption
  - Random sampling
  - The variable is normally distributed
  - The variance of the 2 populations are equal
- 2 groups vs. interval/ratio for dependent variables
- 男女的血壓有無差異?

Dependent (paired) t- test vs. Wilcoxon Signed ranks test

- Assumption
- 某班同學上課前後其知識有無差異?
- 丈夫與妻子對小孩教養態度有無差異?
- 病人與其主護護士對疼痛處理的態度有無差異?

1-way ANOVA vs. Kruskal-Wallis test

- Assumption
  - Random sampling
  - The dependent variable is normal distributed
  - Homogeneity of variance
    - Keep the sample size as equal as possible
- Interval/ration level for dependent variables
  - 組別 ≧ 3組
- 五個醫學中心N3之薪資有無差異?

Repeated ANOVA vs. Friedman test

- ≧ 3次重複性測量
- 同一組病人三次治療之疼痛分數有無差異性?

Regression

- Assumption
  - 線性關係假設
  - 依變項常態性假設
- 多個獨立變項(independent variable)預測一個依變項(dependent variable)
  - 年齡、手術後之天數、傷口大小與疼痛的關係

Regression

- 當自變項爲類別層次時，要先以虛擬變項處理之（dummy coding）
- 依變項爲類別層次時，要以Logistic regression進行分析