Overview on selection of test statistics

Dr Thida
MBBS, MMedSc (Pubic Health), PhD (Epidemiology)
Deputy Director/Head
Epidemiology Research Division
Department of Medical Research (POLB)

Choosing the Right Statistic

- Matching research design to appropriate analysis
- Useful to follow a step-by-step process with 3 basic questions to be answered
  1. What type of research questions are you asking?
  2. What type and number of variables do you want to analyze?
  3. What type of data do you have and what characteristic they have?

Step 1: What type of RQ are you asking

- What is the typical BP for people in a giving population
  - Descriptive
- Is there a relationship between BP and SES?
- Can neighborhood you live in predict your BP?
  - Correlational/Predictive
- Is the typical blood BP higher in one area compared to another?
- Do male have a greater incidence of depression vs females?
  - Group difference/cause and effect

Appropriate analysis for your research question

- Descriptive analysis
  - Frequency, percentiles, central tendency, Standard scored
- Correlational analysis
  - Correlation
  - Regression
- Analyzing differences between groups
  - t-tests
  - One-Way ANOVA/ Two-way ANOVA (Factorial ANOVA)
  - Chi-square test/ logistic regression/ multinomial regression

Step 2: What type and number of variables do you want to analyze?

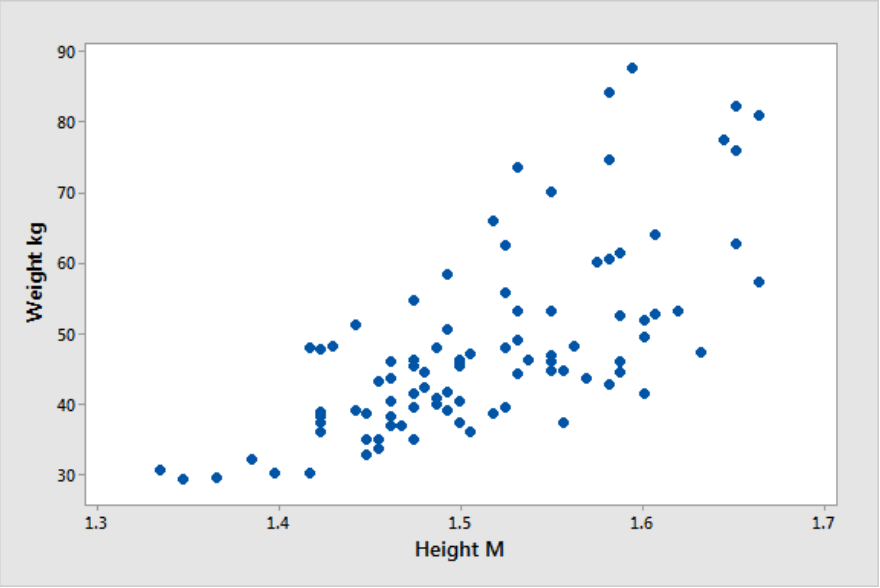
- Dependent vs independent
- Number of variables
- Operational definition of each variable
- Normal/expected range of scores/levels of each variable
  - Dependent variables
  - Independent variables

Step 3: What type of data do you have and what characteristics they have

- Measurement for each variable
  - Nominal/categorical
  - Ordinal
  - Interval/ratio
- Additional information
  - Interval/ratio
    - Normally distributed?
    - Appropriate range of scores?/ Any outliers?
  - Nominal/categorical
    - Are the groups equal/balanced?
    - Are some of the categories empty?

Step 4: Draw a diagram of your design

- Example 1: is there a relationship between height and body weight?
  - Body weight-continuous: 20-90 kgs (Continuous variable)
  - Height: 1.3-1.7 m (Continuous variable)
- Both are Interval
- Pearson's correlation test



Step 4: Draw a diagram of your design

- Example 2: Do people with BMI values below 25 have lower SBP than people with BMI above 25?
- SBP: Dependent Variable; continuous: mmHg from 100-220
- BMI value: Independent variable, categorical (two groups/ levels): BMI <= 25/BMI >=25.1

	BMI<25	BMI>=25
Mean SBP (mmHg)		

Step 4: Draw a diagram of your design

- Example 3: Is the effect of sex on SBP different for people with BMI values below 25 than people with BMI above 25?
  - SBP: Dependent Variable; continuous: mmHg from 100-220
  - Sex: Independent Variable, categorical: male/female
  - BMI value: Independent Variable, categorical (two groups' levels): BMI <= 25/BMI >25.1

	BMI<=25	BMI>25
Mean SBP (mmHg) for males		
Mean SBP (mmHg) for females		

Step 5: Determine need for parametric or non-parametric test

- Does your data set meet the assumptions of parametric testing?
- What if it doesn't ?
  - Use parametric testing anyway (large sample size)
  - Transform the data (log transformation)
  - Use a non-parametric technique
    - Pearson rho/spearman's rho
    - Independent t-test/ Mann-Whitney U test/ Median test
    - Paired t-test/ Wilcoxon Signed Ranks or Sign test
    - Repeated measure ANOVA/ Friedman test
    - One-Way ANOVA/Kruskal-Wallis test/Median test

Step 6: Make the final decision

- Make determinations about your variables
- Make sure you meet all the assumptions
- Are there other approaches that could be taken?
- What approach have other studies with similar designs used?

Example 1

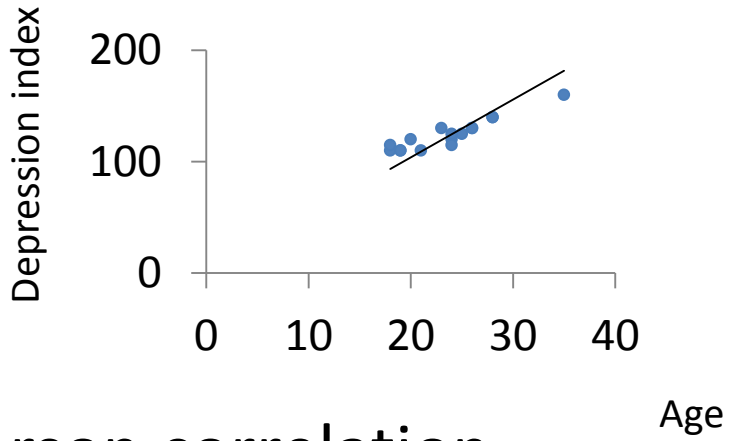
- RQ: what is the relationship between gender and having a diagnosis of clinical depression?
- What you have:
  - One categorical dependent variable: diagnosis of depression: yes/no (i.e. # of people in each category)
  - One categorical independent variable: sex: male/female

	Males	Females
Has depression		
Does not have depression		

- Technique: Chi square test of independence

Example 2

- RQ: Is there a relationship between age and depression index? Does depression increase with age?
- What you have:
  - Two continuous variables (age, score on a depression index)



- Technique: Pearson correlation
- Non parametric alternative: Spearman's correlation

Example 3

- RQ: Are male more depressed than females?
- What you have:
  - One continuous Dependent Variable (depression index score)
  - One categorical Independent Variable with two levels/groups (male/female)

	Male	females
Mean BDI score		

- Technique: independent t test
- Non parametric alternative: Mann-Whitney U-test/ Median test

Example 4

- RQ: Will 10 weeks of exercise training reduce the depression score?
- What we have
  - One continuous Dependent Variable (Depression score)
  - One categorical Independent Variable (pre-test/post-test)

	Pre-test	Post-test
Mean Depression score		

- Technique: Dependent/paired-sample t-test
- Non parametric alternative: Wilcoxon Signed Rank test

Example 5

- RQ: Is there are difference in depression score for people under 30, 31-49 and 50 years and over?
- What we have
  - One continuous Dependent Variable (Depression score
  - One categorical IV with two or more groups/levels (age: under 30, 31-49, 50+)

	<30 years	31-49 years	50+ years
Mean BDI score			

- Technique: One-Way ANOVA
- Non parametric alternative: Kruskal-Wallis test/ Median test

Example 6

- RQ: What is the effect of age on Depression scores for males and females?
- What we have
  - Two categorical Independent Variables (Sex: male/female; age: under 30, 31-49, 50+)
  - One continuous Dependent Variable (Depression score)

	<30 years	31-49 years	50+ years
Mean male Depression score			
Mean male Depression score			

- Technique: Two-Way (Factorial) ANOVA
- Non parametric alternative: None

Example 7

- RQ: Which of two therapy interventions is more effective in reducing Depression score across three time periods (pre-treatment, post-treatment, 3 months post-treatment)??
- What we have
  - One continuous Dependent Variable (Depression score)
  - One between groups Independent Variable: type of intervention (exercise, meditation)
  - One within-groups Independent Variable : (3 measurement points)

	Time 1	Time 2	Time 3
Mean BDI with exercise			
Mean BDI with meditation			

- Technique: Factorial Repeated Measures(Mixed between-within or Split-plot) ANOVA
- Non parametric alternative: None

Example 8

- RQ: Does the SBP change with the change of sex, income and BMI?
- What we have
  - One categorical Dependent variable: (Normal/High SBP)
  - Two or more categorical/Continuous variables

	Normal SBP	High SBP	Adj OR (95% CI)
Sex Male Female			
Income Low High			
BMI (continuous)			

Technique: Multivariable logistic regression
Non parametric alternative: None

Type of outcome variable determines choice of multivariable analysis.

Type of outcome	Example of outcome variable	Type of bivariate analysis	Type of multivariable analysis
Interval	Blood pressure, weight, temperature	Correlation coefficient, linear regression, *t* test, ANOVA	Multiple linear regression, analysis of variance (and related procedures)
Dichotomous	Death, cancer, intensive care unit admission	Chi-squared, Fisher's exact, *t* test, chi-squared for trend, Mann–Whitney test	Multiple logistic regression
Ordinal	Stage of disease, severity of symptoms	Chi-squared for trend, Mann–Whitney test, Spearman's rank correlation coefficient	Proportional odds regression
Nominal	Cause of death, site of cancer	Chi-squared, ANOVA, Kruskal–Wallis	Multinomial logistic regression
Time to outcome	Time to death, time to cancer	Log-rank	Proportional hazards analysis

References

- https://www.google.com/search?rlz=1C1SQJL_enTH810TH810&ei=5GMtXf_cKNem9QOo1ryADQ&q=choosing+the+appropriate+statistical+analysis+technique+step+by+step&oq=Choosing+the+Appropriate+Statistical+Analysis+Technique+step&gs_l=psy-ab.1.1.33i21j33i160l3.36100.44289..48452...0.0..0.329.1587.2-5j1.....0....2j1..gws-wiz.....0i71j35i39j33i22i29i30.Cu8TECNp2VU#kpvalbx=1
- Mitchell H. Katz (2009).Study Design and Statistical Analysis: A Practical Guide for Clinicians (2nd Edition), the United States of America by Cambridge University Press, New York
- Bland JM, Altman DG. The logrank test. BMJ. 2004 May 1;328(7447):1073.
- Plans for Data Analysis ppt, 2019 Research Methodology Workshop

Thank you!