Inferential statistics

Shehu, Nathan Y.

Research Methodology Training by The African Centre for Phytomedicine Research and Development, University of Jos.

24th-28th June, 2019

Inferential statistics

- Using data collected from samples of a population, and making **generalizations** which go **beyond** the sample to the population from which they came.
- To:
  - Draw an inference
  - Make a statement
  - Reach a conclusion

Inferential statistics: Types

- Confidence interval estimation

- Hypothesis testing

Some concepts:

- Population>> Complete set of individuals or objects that the investigator is interested in studying
- Sample – A subset of the population that is actually being studied
- Variable- Xcteristic of an individual or object that can have different values
- Independent variable – The variable that is systematically manipulated or measured by the
- Dependent variable – The outcome variable of interest
- Data – The measurements that are collected by the investigator
- Statistic – Summary measure of a sample
- Parameter – Summary measure of a population

Statistical Inference: Hypothesis Testing

- Is there evidence that the population parameter, e.g., RR, OR, is different from the null value?
- Interval Estimation –How do we determine the precision of the point estimate by accounting for sampling variability?

The null and alternative hypotheses

- H0: No effect of smoking on mortality. (RR=1) ,
- HA: There is effect of smoking on mortality. (RR≠1)
- H0: No relationship between eating raw rice and getting diarrhea. (OR=1)
- HA: Relationship between eating raw hamburger and getting diarrhea. (OR ≠ 1)

Hypothesis Testing

		True state of nature	
		Ho true	Ho false
Data-based decision	Fail to reject	GOOD	Type II error
	Reject H0	Type I error	Power

Our Judicial System

		Truth	
		Ho true	Ho false
Jury's verdict	Innocent	GOOD	**Bad**
	Guilty	Bad	Good

Steps of Hypothesis Testing

- 1. State the null and alternative hypothesis.
- 2 . Specify the significance level (α).
- 3. Determine the appropriate statistical test and compute the value of the test statistic.
- 4. Determine the p-value associated with the test statistic.
- 5. State a conclusion: Reject or Fail to reject the null hypothesis.

p-value

- If the p-value is less than the specified significance level (α) then reject the null hypothesis and conclude that there is significant evidence from the data to support the alternative hypothesis.
- If the p-value is greater than the specified significance level (α), then fail to reject the null hypothesis and conclude that there is not significant evidence from the data to support the alternative hypothesis.

Common Statistical Tests of Significance

- Large Sample Chi Square tests
  - Chi-Square tests Pearson's uncorrected $\chi 2$
  - Mantel-Haenszel $\chi 2$
  - Yates corrected $\chi 2$
- Small Sample Test:
- Fisher's Exact Test
- • Each of these tests provides us with a p-value.

Cohort Study : Relative risk

	Smokers	Non smokers
Death	A	B
Survival	C	D

- RR = Risk in the exposed (A/A+C)/ (B/B+D)

Cohort Study : Relative risk

- 1. Ho: RR=1, Ha: RR≠1
- 2. $\alpha = 0.05$
- 3. Large–sample chi-square test (MantelHaenszel)
- 4. P value < 0.05?

Inferential statistics

- Odds ratio>> case control

- Fishers exact expected cell value < 5

Confidence Intervals

- Goal: Use sample information to compute two numbers
- L(lower limit) and U(upper limit) about which we can claim with a certain amount of confidence (95%) that they surround the true value of the parameter.

Properties of Confidence Intervals

- The wider the CI, the less precise the estimate.

- The more narrow the CI, the more precise the estimate.

Confidence Interval vs. P-value

- Similarities:
  - Multiple formulas, (approximate and exact)
  - Neither account for bias –
  - Statistically equivalent (Theoretically!)
- Differences:
- CI provides same information as a statistical test, plus more
- CI reminds reader of variability
- CI provides range of compatible values (interval estimation)
- CI more clearly shows influence of sample size

Cohort Study : Relative risk

R = 1 : Group 1 (exposed)has the same risk with group 2 (unexposed)

R<1 : Group 1 has more Less risk than group 2

R> : Group 1 has more risk than group 2

Understanding P value and RR(CI)

- .

Study #	Risk Ratio	Risk Ratio 95% CI	P value
1	2.0	(1.2, 3.3)	0.008
2	7.0	(1.2, 40.8)	0.03
3	7.0	(0.8, 61.3)	0.07
4	0.97	(0.9, 1.07)	0.65
5	0.98	(0.97, 0.99)	0.0001

Comparing:	Dependent variable	Independent variable	Parametric test (Dependent variable is normally distributed)	Non-parametric test
The means of two INDEPENDENT groups	Continuous/ scale	Categorical/ nominal	Independent t-test	Mann-Whitney test
The means of 2 paired (matched) samples e.g. weight before and after a diet for one group of subjects	Continuous/ scale	Time variable (time 1 = before, time 2 = after)	Paired t-test	Wilcoxon signed rank test
The means of 3+ independent groups	Continuous/ scale	Categorical/ nominal	One-way ANOVA	Kruskal-Wallis test
The 3+ measurements on the same subject	Continuous/ scale	Time variable	Repeated measures ANOVA	Friedman test
Relationship between 2 continuous variables	Continuous/ scale	Continuous/ scale	Pearson's Correlation Co-efficient	Spearman's Correlation Co-efficient (also use for ordinal data)
Predicting the value of one variable from the value of a predictor variable	Continuous/ scale	Any	Simple Linear Regression	
Assessing the relationship between two categorical variables	Categorical/ nominal	Categorical/ nominal		Chi-squared test

Questions and comments

?