Statistics 305/605: Introduction to Biostatistical Methods for Health Sciences

Summary of Review Material

Jinko Graham

2018-10-03

Ch8-10: Inference of a Population Mean

- A statistic is computed from data on a sample; e.g. $\bar{X}$, the sample average.
- By contrast, a parameter is a population quantity; e.g. μ, the population average.
- Statistical inference: Learning about parameters from statistics that are subject to random variation.
  - e.g. Hypothesize about parameters such as μ.
  - Test $H_0 : \mu = 0$ vs. $H_a : \mu \neq 0$.

- Key point: Even though the population mean, μ, and the population SD, σ, are unknown, we know the (approximate) distribution of the pivotal quantity

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}},$$

  regardless of the shape of the population distribution for X.
- This result relies on the CLT, which tells us that (large) sample averages such as $\bar{X}$ are approximately normally distributed.
- Many of the statistics we will study are based on averages, so inference of a population mean is a useful template.
- Knowing the distribution of the pivotal quantity allows us to construct confidence intervals, calculate p-values, test statistical hypotheses, calculate power, etc.

Ch11: Inference for a Difference of Population Means

- Inference for the difference between two population means is based on either the pivotal quantity Z (SDs known) or T (SDs unknown).
- CIs are of the form estimate $\pm$ margin of error
  - the margin of error is a critical value (z^* for Z, t^* for T) times the SE for the estimate.
- To test $H_0 : \mu_1 - \mu_2 = 0$ against $H_a : \mu_1 - \mu_2 \neq 0$
  - We use our sample of data to compute the observed value t (or z if SDs known) of a test statistic.
  - We compare this observed value to a reference distribution for the test statistic obtained under H_0.
  - The p-value is the chance of seeing a value of the test statistic as or more extreme than the value that was observed, under H_0.
  - Compare the p-value to a significance level α to obtain a test of H_0 against H_a.
- Inference is considered reliable when the parent populations are normal, or when rules-of-thumb about sample sizes for the CLT are satisfied.

Ch14: Inference for Proportions

- Inference for the difference $p_1 - p_2$ between two population proportions is based on a pivotal quantity, also called Z.
- CIs are estimate $\pm$ margin of error, where
  - estimate is the difference between sample proportions, and
  - margin of error is a critical value (z^*) times the SE (estimated SD) of the difference in sample proportions.
- To test $H_0 : p_1 - p_2 = 0$ against $H_a : p_1 - p_2 \neq 0$
  - We use our sample of data to compute the observed value z of a test statistic.
  - We compare this observed value to a reference distribution for the test statistic obtained under H_0.
  - The p-value is the chance of seeing a value of the test statistic as or more extreme than the value that was observed, under H_0.
  - Compare the p-value to a significance level α to obtain a test of H_0 against H_a.
- Inference is considered reliable when there are sufficient numbers of successes and failures in each sample for the CLT to hold.

Ch6: Probability

- Discussed the basic definitions and rules of probability, including the definition of conditional probability.
- Use Bayes' Theorem to relate $P(A \mid B)$ to
  - $P(B \mid A)$, $P(A)$ and $P(B)$.
- Public-health and medical practitioners work with many conditional probabilities every day; e.g.,
  - diagnostic test sensitivity and specificity
  - relative risks and odds ratios
- Case-control data
  - Disease probabilities or risks in the exposure groups cannot be estimated, owing to oversampling of the cases in the study design.
  - However, exposure probabilities in the disease groups can be estimated, allowing us to estimate odds ratios.
  - For a rare disease, the odds ratio approximates the relative risk.