Last Lecture

Hypothesis testing
  Philosophical underpinning
  Fisherian hypothesis testing
  Refinements

Example of hypothesis testing- one sample t test

Remarks

• $P$ value represents probability that a result would occur, over the long run, if $H_0$ is true, not the probability that a specific result would occur by chance.
• Failing to reject $H_0$ does not mean you accept it, just that you must suspend judgement for the moment.
• There may be bias in the literature against ‘negative’ results.

Decision errors (FigS 3.2, 3.3)

• Type I error-
  – Occurs when $H_0$ mistakenly rejected
  – Probability can be measured as area under curve
• Type II error-
  – Occurs when $H_0$ should be rejected, but isn’t
  – Probability more difficult to determine because we often don’t know what value to associate with $H_a$.

Assumptions

• Sampling assumptions must be valid
• Continuous data must conform to underlying normal distribution
• Variance in the response variable is not related to the predictor variables

How to detect violations of assumptions

• Run exploratory analysis
  – Plot treatment means and standard deviations or if both variables continuous, plot Y by X
  – Plot predicted values versus residuals
  – Examine frequency distributions
• Run preliminary statistical tests
  – Compare variance among treatment groups
  – Check for normality of distributions

Worrying about assumptions

• Parametric techniques fairly robust to violations of normality
• Somewhat less robust to violations of variance assumption
Coping with violation of assumptions

- Make sure to record confounding factors and include them in analysis (deals with multimodal distributions)
- Transform data
  - Often improves distribution greatly and reduces variance/mean relationships
    - Right skew often common - dealt with by log-transformation or square root transform
    - Left skew can be dealt with by power transformations or by 'reflecting' data (adding constant and subtracting all observations from it)
  - Check transformed data for assumptions too and compare untransformed and transformed data!
- Use non-parametric test (page 47)

Intro to correlation analysis

- Evaluating relationship between two random variables (X, Y).
- Both variables considered to be continuous.
- Need to determine whether variables covary (vary together)
- Quantify with covariance, but this measure incorporates magnitude of variables.
- Covariance formula: Table 5.1
- Special case: covariance of a variable with itself is equal to its variance