STAT 285
Two Sample Inference

Richard Lockhart

Simon Fraser University

Fall 2014 — Surrey

Purposes of These Notes

- Describe two samples inference problems
- Tests for equality of two means.
- Confidence interval for difference of two means.
- Tests for equality of two proportions.
- Confidence interval for difference of two proportions.
- Scientific issues surrounding the techniques.

Comparison

- Common experimental design: get a sample of people.
- Split group into two groups at random.
- Put n in group 1, m in group 2.
- Give new treatment to group 1; control treatment to group 2.
- Observe responses $X_1, \ldots, X_n$ in group 1; $Y_1, \ldots, Y_n$ in group two.
- Question of interest: compare treatment mean to control mean.
- Introduce notation. X population mean, SD are μ_1 and σ_1.
- Subscript 2 for Y.
- Give confidence interval for $\mu_1 - \mu_2$.
- Test: H_0:$\mu_1 = \mu_2$ or H_0:$\mu_1 \leq \mu_2$.

Methodology

- All methods based on distribution of $\bar{X} - \bar{Y}$.
- Mean and SD:

$$\mathrm{E}(\bar{X} - \bar{Y}) = \mu_1 - \mu_2$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\mathrm{Var}(\bar{X} - \bar{Y})} = \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

- If n and m large, or population distributions normal,

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

  has approximately a $N(0,1)$ distribution.
- Standard error usually has to be estimated.
- Just replace σ_i by S_i.
- If n and m large can ignore impact of estimation.

Confidence intervals for $\mu_1 - \mu_2$

- Three procedures in use.
- Large sample normal approximation:

$$\bar{X} - \bar{Y} \pm z_{\alpha/2}\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}$$

- Normal populations, approximate t distribution

$$\bar{X} - \bar{Y} \pm t_{\alpha/2,\nu}\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}$$

  where df ν is estimated from the data; see page 357.
- Normal populations, equal variances $\sigma_1 = \sigma_2 = \sigma$:

$$\bar{X} - \bar{Y} \pm t_{\alpha/2,\nu}\sqrt{\frac{S^2}{n} + \frac{S^2}{m}}$$

  where df $\nu = n + m - 2$ and S^2 is pooled variance estimate:

$$S^2 = \frac{n-1}{n+m-2}S_1^2 + \frac{m-1}{n+m-2}S_2^2$$

  (a weighted average of two estimates of σ^2).

Hypothesis tests for $\mu_1 - \mu_2$

- Three procedures in use.
- All based on statistics of form

$$T = \frac{\bar{X} - \bar{Y}}{\hat{\sigma}_{\bar{X}-\bar{Y}}}$$

- Issues:
  - One tailed versus two
  - How to estimate standard error
  - Null distribution

Hypothesis tests for $\mu_1 - \mu_2$, II

- Normal populations, equal variances $\sigma_1 = \sigma_2 = \sigma$. Use t distribution and

$$\hat{\sigma}_{\bar{X}-\bar{Y}} = \sqrt{\frac{S^2}{n} + \frac{S^2}{m}}$$

  where degrees of freedom $\nu = n + m - 2$ and S^2 is pooled variance estimate.

- Normal populations, unequal variances. Use t distribution and

$$\hat{\sigma}_{\bar{X}-\bar{Y}} = \sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}$$

  where degrees of freedom ν estimated from data as for CI.

- Large Samples. Previous suggestion or replace t with normal.

Two sample inference for proportions

- Two independent sets of Bernoulli trials.
- X is Binomial(n, p_1) and Y is Binomial(m, p_2)
- Confidence intervals for $p_1 - p_2$.
- Hypothesis tests: H_0:$p_1 = p_2$ or H_0: $p_1 \le p_2$.
- One-sided or two sided alternatives.
- All based on $\hat{p}_1 - \hat{p}_2 = X/n - Y/m$ which has

$$\mathrm{E}(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$$

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\mathrm{Var}(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}}$$

- Replace standard error with appropriate estimate: either using $\hat{p}_1$ and $\hat{p}_2$ OR for testing $p_1 = p_2 = p$ (with p not specified) use

$$\hat{p}_1 = \hat{p}_2 = \hat{p} = \frac{X+Y}{n+m}.$$

- Make normal approximation even after estimation of SE.

Salk Polio Vaccine Example

- $n = m = 200,000$. X is number of polio cases in vaccine group. Y is number in control group.
- Model: assume X and Y independent.
- Mildly controversial. Polio is contagious.
- Assume each treated child has prob p_1 of getting polio. Control child prob is p_2.
- Does the vaccine work?
- Hypothesis testing problem. H_0:$p_1 = p_2$ vs H_{a}:$p_1 < p_2$.
- Clearly one-sided.
- Observe $X = 54$, $Y = 142$.
- Pooled estimate of p is $\hat{p} = 196/400,000 = 4.8 \times 10^{-4}$ so

$$T = \frac{54/2000000 - 142/200000}{\sqrt{\hat{p}(1-\hat{p})\frac{1}{200000} + \frac{1}{200000}}} = -6.29.$$

- P-value from normal curve is 1.6×10^{-10}.
- Overwhelming evidence against assertion vaccine doesn't work.

Page 286 Q 67

- Compression strength in pounds of boxes.
- Two methods to compare: fixed and floating.

Method	Sample Size	Sample Mean	Sample SD
Fixed	10	807	27
Floating	10	757	41

- Published article concludes "the difference between the compression strength using fixed and floating platen method was found to be small compared to normal variation strength between identical boxes"
- Agree or not?
- Main practical issue: what statistical analysis is relevant.
- We will measure the difference between the two methods by estimating the difference in average strength between the two methods.
- We will give a confidence interval, then interpret.

The interval

- Introduce notation: samples of $n = 10$ and $m = 10$ from two populations, means μ_1 (for fixed method) and μ_2 (for floating); SDs σ_1 and σ_2.
- Confidence interval for $\mu_1 - \mu_2$ is

$$\bar{X} - \bar{Y} + t_{0.025,\nu}\sqrt{\frac{27^2}{10} + \frac{41^2}{10}}$$

- Pooled or not? df ν?.
- Fact: estimated Standard Error is the same whether we pool or not.
- Formula for ν simplifies because $n = m$:

$$\nu = 9\frac{(s_1^2 + s_2^2)^2}{s_1^4 + s_2^4} = 15.6$$

- Advice from text, round down. For $\nu = 15$ get multiplier 2.131. For $\nu = 16$ get 2.12. For $\nu = 15.6$ get 1.125. No important difference.
- If you believed $\sigma_1 = \sigma_2$ you would use 18 df. No important difference.

Results and interpretation relative to question

- Interval is

$$50 \pm 2.12 \times 15.52 \text{ or } 17.1 \text{ to } 82.9.$$

- Is that small compared to normal variation?
- No.
- Normal variation is summarized by the SDs of 27 and 41.
- This interval is of similar sized numbers. (Numbers not unlike 27 and 41.)

Clear difference between methods?

- We could also ask if there is definitely a difference. I think this is not the way to answer the question asked. But book thinks it is.
- Different views of 'small compared to'.
- Test $H_0{:}\mu_1 = \mu_2$ against two sided alternative.
- Borrow almost all arithmetic from CI:

  $$T = \frac{807 - 757}{15.52} = 3.22$$

  with $\nu = 15$ df.
- Get P value from student's t curve. In tables: area to right of 3.22 is between 0.002 and 0.003.
- Closer to 0.003.
- Double to get P. A bit smaller than 0.006.
- From computer with $\nu = 15.6$ get 0.0055.
- Either way it is quite clear that there is a difference in mean compression strength between the two methods.
- So the claim in the paper is not credible, whatever it means.