MDSC 643.01, Biostatistics I
Brant, Fall 2002

Relating P-values and Confidence Intervals

The Acccept/Reject Approach

- technical simplification to facilitate theoretical development
- choose desired cut-off for signficance ("alpha-level") α (.05?)
- if $P < \alpha$ "Reject" H_0 , if $P \geq \alpha$ then "Accept" H_0

If $P < \alpha$ then the $(1-\alpha) \times 100$ % confidence interval will not contain the Null value.

Comparing Means Based on Small Samples

When comparing means based on small samples

- cannot use CLT to justify normal sampling distributions for sample means
- cannot "ignore" the estimation of σ in SE's

The Two-sample t-test

Assumptions: Normality: Two populations: means μ_1, μ_2 , normally distributed
Equal variances: Same dispersion both populations $\sigma_1 = \sigma_2 = \sigma$
Independence: Independent random samples from each population

Always begin with inspection via boxplots ("normal"? , same spread ?)

Pooled variance estimate: $s^2 = \frac{(n_1 - 1) s_1{}^2 + (n_2 - 1) s_2{}^2}{n_1 + n_2 - 2}$

Standard error for the difference in means: $SE(\bar{x}_1 - \bar{x}_2) = s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

t-statistic for equal means: $t = \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)}$

Sampling distribution of t statistic

- Student's t-distribution - depends on degrees of freedom
- for two sample case, $d.f. = n_1 + n_2 - 2$

e.g.

Confidence interval for $\mu_1 - \mu_2$ - modification of normal theory "generic form":

$$\bar{x}_1 - \bar{x}_2 \pm (t-value) SE(\bar{x}_1 - \bar{x}_2)$$

where t-value chosen from t-tables with d.f. = $n_1 + n_2 - 2$ to cut off α in each tail.

N.B. NOT the *observed* t-statistic

The One-sample t-test

Assumptions:

Normality: One normal shaped population, mean= μ , s.d.= σ

Random sampling

t-statistic for comparing μ to a fixed value μ_0 ,: $t = \dfrac{\bar{x} - \mu_0}{SE(\bar{x})}$

where $SE(\bar{x})$ is just the usual, $\dfrac{s}{\sqrt{n}}$

- refer to t-distribution with n - 1 d.f.

Confidence Interval: $\bar{x} \pm t-value\, SE(\bar{x})$ - use (n-1) d.f.

Paired Samples

Two-sample matched study: one in which samples are chosen from two populations in such a way that each individual in sample 1 has a matching (i.e. similar) partner in sample 2. By necessity, this means that $n_1 = n_2$

Repeated measures experiment: An experiment in which multiple measurements of the same variable are taken (under different circumstances).
e.g.

The paired t-test: To compare means μ_1 and μ_2 for the two populations:

- assume (pretend) that the two populations can be paired.
- consider population of paired differences, $d = x_1 - x_2$
    - has mean $\mu = \mu_1 - \mu_2$ (mean of differences is difference of means).
- base inference for $\mu_1 = \mu_2$ on one sample methods using $\bar{d}$ and s_d

mean and s.d. of sample of differences