STAT 270- Chapter 7
Inference: two samples

July 15, 2012

Normal, known variances

$$X_1, \ldots, X_m \text{ iid } \mathit{Normal}(\mu_1, \sigma_1^2)$$

$$Y_1, \ldots, Y_n \text{ iid } \mathit{Normal}(\mu_2, \sigma_2^2)$$

Want to make inference about $\mu_1 - \mu_2$.
Assumptions:

- X_is and Y_is are independent.
- σ_1^2 and σ_2^2 known. (unrealistic)
- $n \neq m$ in general but they can be equal.

Distribution theory:

$$\bar{X} \sim \mathit{Normal}(\mu_1, \frac{\sigma_1^2}{m})$$

$$\bar{Y} \sim \mathit{Normal}(\mu_2, \frac{\sigma_2^2}{n})$$

Normal, known variances

$\bar{X} - \bar{Y}$ is a linear combination of two normal variables:

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

Standardizing: pivotal quantity:

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim \mathit{Normal}(0, 1)$$

Example 7.1

Example 7.3

Example 7.4

Example 7.5

Normal, approximately equal but unknown variances

$$X_1, \ldots, X_m \text{ iid } Normal(\mu_1, \sigma^2)$$
$$Y_1, \ldots, Y_n \text{ iid } Normal(\mu_2, \sigma^2)$$

Assumptions:

- X_is and Y_is are independent but similar measurments: they have approximately the same variance.
- σ^2 is unknown.

Estimate σ_2 by the pooled variance

$$s_p^2 = \frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}$$

then

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{(\frac{1}{m} + \frac{1}{n})s_p^2}} \sim t_{m+n-2}$$

Example

An instructor hypothesizes that getting enough sleep improves the students performance on the exam. To test this hypothesis she randomly selects 20 students and randomly devides them into two groups of equal sizes: group 1 sleep 4 hours per night for a week and group 2 sleep 8 hours per night for a week. The two groups write an exam at the end of the week. Two students in group 1 gave up the experiment. The results are given in the following table. Assuming normal distributions for the test scores test the instructors hypothesis.

group 1	5	4	-	14	18	-	12	10	3	17
group 2	6	7	14	15	3	11	18	5	16	10

The large sample case

$X_1, \ldots, X_m$ iid with $E(X_i) = \mu_1$ and $var(X_i) = \sigma_1^2$
$Y_1, \ldots, Y_n$ iid with $E(Y_i) = \mu_2$ and $var(Y_i) = \sigma_2^2$
m and n are latge (≥ 30).
Apply CLT to obtain:

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim Normal(0, 1)$$

When σ_1 and σ_2 are unknown replace them with the sample variances s_1 and s_2:

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} \sim Normal(0, 1)$$

Example

To compare two nutrition programs for cats, each program was randomly assigned to 400 cats and their weight was measured after a month. The results are given bellow. Do the two peograms have significantly different effects on cats' wieght? (test at $\alpha = .01$)

Data:

$n_1 = 270$, $\bar{x}_1 = 8.2$ lb, $s_1 = 1.4$ lb

$n_2 = 130$, $\bar{x}_1 = 7.8$ lb, $s_1 = 1.1$ lb

With the same sample sizes and sd's what should the minimum difference in means be to reject H_0 at $\alpha = .05$?

Example 7.8

Example 7.8 continued

Calculate the type two error rate of the $\alpha = .01$ test for $\mu_1 = \mu_2 + 7$.

The binomial case

- Two independent binomial experiments
- Want to make inference about the difference between the success rates

m trials, each resulting in either success or failure, p_1 = probability of success
X: number of successes

$$X \sim binomial(m, p_1)$$

if $mp_1 \geq 5$, $m(1 - p_1) \geq 5$

$$X \sim Normal(mp_1, mp_1(1 - p_1))$$

and

$$\hat{p}_1 \sim Normal(p_1, \frac{p_1(1 - p_1)}{m})$$

where $\hat{p}_1 = \frac{X}{m}$.

The binomial case

n trials, each resulting in either success or failure, $p_2 =$ probability of success

Y: number of successes

$$Y \sim (n, p_2)$$

if $np_2 \geq 5$ and $n(1 - p_2) \geq 5$

$$Y \sim Normal(np_2, np_2(1 - p_2))$$

and

$$\hat{p}_2 \sim Normal(p_2, \frac{p_2(1 - p_2)}{n})$$

where $\hat{p}_2 = \frac{Y}{n}$.

$\hat{p}_1$ and $\hat{p}_2$ are independent: normal theory:

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{m} + \frac{p_2(1-p_2)}{n}}} \sim Normal(0, 1)$$

Example 7.10

The paired case

$$(X_1, Y_1), \ldots, (X_n, Y_n)$$

measurments on the same objects
Calculate

$$D_i = X_i - Y_i$$

and proceed as you would in the single sample case.

Example 7.11

Example 7.13, the difference between paired and two sample problems

Placebo:

More on pairing

- Paired tests are more sensitive than non-paired tests:

- Pairing is a special case of **blocking**:

- Pairing and blocking is done at the price of losing degrees of freedom: