1. Estimation

2. Testing

we've done 1 sample testing.

General setup

$$X_1, \dots, X_n \overset{iid}{\sim} N(\mu_x, \sigma_x^2)$$

SINGLE SAMPLE

$$Y_1, \dots, Y_m \overset{iid}{\sim} N(\mu_y, \sigma_y^2)$$

Assume $\sigma_y^2 = \sigma_x^2$

$$t = \frac{\bar{X}_n - \bar{Y}_m - (\mu_x - \mu_y)}{S\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

$$s^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{(n+m-2)}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Under assumptions

$$t \sim t_{n+m-2}$$

What is the t distribution?

$X \sim N(0,1)$

$W \sim \frac{\chi^2_n}{n}$ "Chi-squared"

$\Rightarrow \frac{X}{\sqrt{W}} \sim t_n$

$\bar{X}_n - \bar{Y}_m \pm S\sqrt{\frac{1}{n}+\frac{1}{m}}\; t_{n+m-2}\left(1-\frac{\alpha}{2}\right)$

is a $1-\alpha$ CI

$t_{n+m-2}\left(1-\frac{\alpha}{2}\right)$ is $1-\frac{\alpha}{2}$

quantile from t_{n+m-2} df.

(under equal variances assumption)

There is an adjustment for unequal variances

Welch's t-test by default in R

t.test(x, y)

Mann-Whitney

with t-test

1. Parametric Assumptions
2. Testing Difference in means

For t-test we test

$$H_0: \mu_X = \mu_Y \quad \Leftrightarrow \quad H_0: \mu_X - \mu_Y = 0$$

$$H_1: \mu_X \neq \mu_Y \quad \Leftrightarrow \quad H_1: \mu_X - \mu_Y \neq 0$$

For Mann-Whitney

$H_0: F = G$ where F and G are distributions

$H_1: F \neq G$

what should we know overall

1) Difference between what we're testing vs. t-test

2) General idea why this works.

General Idea

Suppose we have two samples

$(X_1, \ldots, X_n)$ $\qquad$ $(Y_1, \ldots, Y_m)$

$$Z = (X_1, \ldots, X_n, Y_1, \ldots, Y_m)$$

if X and Y have the same distribution then order of the X_i's within Z should be uniform

equivalent to

$$\sum R_i = \sum_{\text{within } Z} \text{Rank of } X_i$$

shouldn't be too big or too small.

Ch 11 Q41

$\hat{D} = \text{median}(X_i - Y_j)$

$X_1, \dots, X_n \overset{ind}{\sim} F$

$Y_1, \dots, Y_m \overset{ind}{\sim} G$

a. Suppose F and G are normal dist.

WTS $E(\hat{D}) = \mu_x - \mu_y$

Pf.

$X_i - Y_j \sim N(\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2)$

$\hat{\Delta} = \text{median}(X_i - Y_j)$

$\hat{\Delta} - \mu_x - \mu_y$

$= \text{median}(X_i - Y_j) - \mu_x - \mu_y$

$= \text{median}((X_i - \mu_x) - (Y_j - \mu_y))$

$(X_i - \mu_x) \stackrel{d}{=} -(X_i - \mu_x)$

similarly for Y

$= \text{median}(-(X_i - \mu_x) + (Y_j - \mu_y))$

$= \text{median}(-(X_i - Y_j - \mu_x + \mu_y))$

$$= -\left(\text{median}\,(X_i - X_j) - \mu_X - \mu_Y \right)$$

$$\Rightarrow E(\hat{\Delta}) = \mu_X - \mu_Y$$

b) It's the median!

Bootstrap 2 sample test

Define

$$\tilde{X}_i = X_i - \bar{X} + \bar{Z}$$

$$\tilde{Y}_i = Y_i - \bar{Y} + \bar{Z}$$

$\bar{z}$ mean of combined sample

$(x_1, \ldots, x_n, y_1, \ldots, y_m)$

$P(X > c)$

$= E(I(X > c))$

$\approx \frac{1}{B} \sum_i I(x_i > c)$