Lec 4/7

Friday, April 7, 2017 15:00

§16.4 Rank-sum test

The U test (two-sample)

Setting: two-independent sample independent problem; want a nonparametric alternative to t test

Goal: test H_0: two continuous populations are the same

vs H_1: the populations have different means $\neq$, $>$, $<$.

Data: sample 1: $X_1, \ldots, X_{n_1}$

sample 2: $Y_1, \ldots, Y_{n_2}$

Procedure:

1) rank All observations $\{X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}\}$ from 1 (smallest) to n_1+n_2 (largest)

2) Compute W_1 = sum of ranks for sample 1, W_2 = " " sample 2.

3) Compute test statistics:

stats used

$$U_1 = W_1 - \frac{n(n_1+1)}{2} \qquad U_2 = W_2 - \frac{n_2(n_2+1)}{2}$$

note $W_1 + W_2 = \frac{(n_1+n_2)(n_1+n_2+1)}{2}$. Similarly, $U_1 + U_2 = n_1 n_2$

$$U = \min(U_1, U_2)$$

4)

Alternative	statistic	CR
$\mu_1 \neq \mu_2$	U	$U \leq U_\alpha$
$\mu_1 > \mu_2$	U_2	$U_2 \leq U_{2\alpha}$
$\mu_1 < \mu_2$	U_1	$U_1 \leq U_{2\alpha}$

Table XI

Example 16.6:

Compare two kinds of emergency flares, A and B.

Record the burning times (rounded to nearest 10^{th} of a minute)

Brand A: $n_1 = 9$ Brand B: $n_2 = 10$

H_0: brands same, H_1: $\mu_A < \mu_B$.

$W_1 = 69$, $U_1 = 24$, $U_{2\alpha} = 24$, so reject H_0

under H_0: chance to see this data is 0.05.

Observations:

* Under H_0, the ranks are uniformly distributed.
* table only goes up to 15
* for "large" samples, $n_1, n_2 > 8$, U_1 approx normally distributed.

So we could use normal approx once we know μ, σ^2 of U_1 under H_0.

Theorem 16.2:

under assumptions required by U test and under H_0,

$$\mathbb{E}U_1 = \mathbb{E}U_2 = \frac{n_1 n_2}{2}.$$

$$Var(U_1) = Var(U_2) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

Proof:

under H_0: W_1 is sum of n_1 numbers drawn uniformly from $\{1, \dots, n_1+n_2\}$

$\rightarrow$ see Ex 8.15

$$\mathbb{E}W_1 = n_1 \left(\frac{n_1 + n_2 + 1}{2} \right) \quad \mathbb{E}U_1 = \frac{n_1(n_1 + n_2 + 1)}{2} - \frac{n_1(n_1 + 1)}{2} = \frac{n_1 n_2}{2}$$