Discussion #9

Review :

1. Mann-Whitney test for two sample comparison

   (a) Assumption : $X_1, X_2, ..., X_{n_1}$ is a random sample from a distribution with μ_1 and $Y_1, Y_2, ..., Y_{n_2}$ is a random sample from a distribution with μ_2. Those two samples are independent.

   (b) Procedure

      i. Let n_1 be the sample size of the sample with the smaller sample size and n_2 be the sample size of the sample with the larger sample size. If two sample sizes are equal, then $n_1 = n_2$.

      ii. T^*=rank-sum for the sample size n_1.

      iii. $T^{**} = n_1(n_1 + n_2 + 1) - T^*$

      iv. $T = \min(T^*, T^{**})$

      v. Enter tables with T. If T less than the cutoff, then reject. Otherwise, accept H_0.

2. Binomial test for two sample comparison

   (a) Assumptions : $X_1, X_2, ..., X_{n_1}$ is a random sample from $B(n_1, p_1)$ and $Y_1, Y_2, ..., Y_{n_2}$ is a random sample from $B(n_2, p_2)$. Those two samples are independent.

   (b) Test statistic for $H_o : p_1 = p_2$

$$Z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}, \quad \text{where} \quad \hat{p} = \frac{X+Y}{n_1+n_2} = \frac{n_1\hat{p_1} + n_2\hat{p_2}}{n_1+n_2}.$$

   (c) (1-α) C.I. for $p_1 - p_2$

$$\hat{p_1} - \hat{p_2} - Z_{\alpha/2}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}} \le p_1 - p_2 \le \hat{p_1} - \hat{p_2} + Z_{\alpha/2}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}$$

   (d) $Var(\hat{p_1} - \hat{p_2}) = Var(\hat{p_1}) + Var(\hat{p_2}) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}, \quad E(\hat{p_1} - \hat{p_2}) = p_1 - p_2$

Practice Problem

1. A psychologist gives an aptitude test to 12 men and 12 women, the scores being as follows;

Men	80	79	92	65	83	84	95	78	81	85	73	52
Women	82	87	89	91	93	76	74	70	88	99	61	94

   Use the Mann-Whitney test to test the null hypothesis that the distribution of scores on this test is the same for men and as for women.

2. An antibiotic for pneumonia was injected into 100 patients with kidney malfunctions (called uremic patients) and into 100 patients with no kidney malfunctions (called normal patients). Some allergic reaction developed in 38 of the uremic patients and in 21 of the normal patients.

   (a) State the assumptions.

   (b) Perform a test of the claim that the rate of incidence of allergic reaction to the antibiotic is the same in uremic patients and in normal patients.

   (c) Construct a 98% C.I. for the difference between the 2 population proportions.

3. An experiment is planned to compare the rates at which two species of pine are infested with bark beetles. Let p_1 be the rate for the first species and p_2 be the rate of the second species. The investigator will test $H_o : p_1 = p_2$ against $H_A : p_1 > p_2$. To test this he will experiment 90 pines of the first species and 120 pines of the second species. He will reject H_o if $\hat{p_1} - \hat{p_2} > 0.12$.

   (a) If in fact, $p_1 = 0.48$ and $p_2 = 0.42$, find the power of his test.

   (b) When he does the actual experiment, he finds $\hat{p_1} = 0.45$ and $\hat{p_2} = 0.34$. Should he reject the null hypothesis at $\alpha = 0.10$?

4. Suppose that the sample variances (s^2) computed from random samples from $k = 3$ (independent) populations are as follows:

population	1	2	3
sample variance	7.9	3.3	2.1

   If we assume that each population has the same (true) population variance σ^2, then each separate sample variance s^2 is an estimate of σ^2. To combine these estimates into a single "pooled" estimate of σ^2, we use a weighted mean of the sample variances:

$$s_p^2 = \frac{\sum_{i=1}^{3}(n_i - 1)s_i^2}{(N - 3)}$$

   where $N = n_1 + n_2 + n_3$.

(a) Suppose $n_1 = n_2 = n_3 = 8$. Calculate the pooled estimate s_p^2 and find 95% confidence limits for σ^2.

(b) Repeat part (a) for the following three sets of sample sizes: $(n_1, n_2, n_3) = (4,4,16)$; (14,5,5); (7,10,7). Note that all 3 sets (and the set in (a)) have 24 total observations each.

Practice Problem Solutions

1. 1) $n_1 = 12$, $n_2 = 12$.

   2)

$$\begin{aligned} T^* &= \text{rank sum of sample of men} \\ &= 1+3+5+8+9+10+11+13+14+15+20+23 = 132 \end{aligned}$$

   3) $T^{**} = n_1(n_1+n_2+1) - T^* = 12(12+12+1) - 132 = 168$

   4) $T = min(T^*, T^{**}) = 132$

   5) From the table, at $\alpha = 5\%$, reject Ho if $T \leq 115$,
   We have no evidence to reject H_0.

2. (a). Assumptions:

   1. Both samples constitute independent binomial experiements;
   2. We assume independence of the trials from one sample to the other;
   3. The sample size must be large enough so that we can use the normal approximations.

   (b). p_1 is the proportion of reactions in uremic patients,
   p_2 is the proportion of reactions in normal patients.
   $H_0 : p_1 = p_2$, $\quad H_A : p_1 \neq p_2$.
   Use Z statistic

$$Z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}.$$

   From the data, we have

$$\hat{p_1} = 0.38, \hat{p_2} = 0.21, \hat{p} = \frac{38+21}{100+100} = 0.295.$$

   The one we've observed is: $z = 2.64$, then,
   p-value $= 2\ P(Z \geq 2.64) = 2(0.0041) = 0.0082.$,
   We have strong evidence against H_0.

(c). 98 % C. I. for $p_1 - p_2$ is given by

$$(\hat{p_1} - \hat{p_2}) \pm Z_{0.01}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}$$

$$0.023 \leq p_1 - p_2 \leq 0.317.$$

3. a.

$$\begin{aligned} & \text{power}(p_1 = 0.48, p_2 = 0.42) \\ = \ & P(\text{reject Ho}|p_1 = 0.48, p_2 = 0.42) \\ = \ & P(\hat{p}_1 - \hat{p}_2 > 0.12|p_1 = 0.48, p_2 = 0.42) \\ = \ & P(Z > \frac{0.12 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}|p_1 = 0.48, p_2 = 0.42) \\ = \ & P(Z > 0.87) = 0.1922. \end{aligned}$$

b. Test statistic for $H_o : p_1 = p_2, \quad H_A : p_1 > p_2$

$$Z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}, \quad \text{where} \quad \hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{n_1\hat{p_1} + n_2\hat{p_2}}{n_1 + n_2}.$$

From the data, we calculate $\hat{p} = 0.387$, and

$$z = \frac{0.45 - 0.34}{\sqrt{0.387(1 - 0.387)(\frac{1}{90} + \frac{1}{120})}} = 1.62.$$

p-value $= P(Z \geq 1.62) = 0.0526 < 0.1$, so she will reject Ho at $\alpha = 0.1$.

4. a. $S_p^2 = \frac{S_1^2 + S_2^2 + S_3^2}{3} = 4.433.$
$V^2 = \frac{(N-3)S_p^2}{\sigma^2} \sim \chi^2_{N-3}, N = 3(8) = 24.$
95 % CI for σ^2 is

$$\frac{(24-3)(4.433)}{\chi^2_{21,0.025}} \leq \sigma^2 \leq \frac{(24-3)(4.433)}{\chi^2_{21,0.975}}$$

$$2.624 \leq \sigma^2 \leq 9.056.$$

b.

$$s_p^2 = 3.1, 1.83 \leq \sigma^2 \leq 6.33,$$

$$s_p^2 = 5.919, 3.50 \leq \sigma^2 \leq 12.09,$$

$$s_p^2 = 4.27, 2.53 \leq \sigma^2 \leq 8.72.$$