Comparing independent samples

STAT 135 Lab 9

0722/2021

(a) The sample means for the two sites are:

$$\bar{X}_b = 140.2105, \quad \bar{X}_c = 220.4 .$$

The pooled variance estimate is:

$$S_p^2 = 2581.458$$

To test $H_0: \mu_b = \mu_c$ versus $H_1: \mu_b \neq \mu_c$, the rejection region is

$$R = \left\{ \frac{|\bar{X}_b - \bar{X}_c|}{S_p\sqrt{\frac{1}{19}+\frac{1}{20}}} \geq t_{37}(\alpha/2) \right\}$$

Since $t_{37}(0.05/2) = 2.0262$, and $\frac{|\bar{X}_b - \bar{X}_c|}{S_p\sqrt{\frac{1}{19}+\frac{1}{20}}} = 4.927$, we reject H_0 at $\alpha = 0.05$ and conclude that there is enough evidence of difference in concentrations of sulfate.

(b) Let's first sort the observations from both sites, and mark the background samples as green and contaminated samples as red:

61	66	68	73	99	101	121	129	131	140	143	147	
1	2	3	4	5	6	7	8	9	10	11	12	
154	155	157	159	169	174	177	179	183	190	190	192	204
13	14	15	16	17	18	19	20	21	22.5	22.5	24	25
214	218	219	224	227	249	259	260	260	266	273	274	
26	27	28	29	30	31	32	33.5	33.5	35	36	37	
282	285											
38	39											

under which we also wrote down the corresponding ranks.

Next we calculate the rank sums:

$$R_1 = 1+2+3+\cdots+37+39 = 242,$$

$$R_2 = 9+13+17+\cdots+25+38 = 538 .$$

The U statistic can be calculated as follows:

$$U_1 = 19\times20 + \frac{19\times20}{2} - 242 = 328,$$

$$U_2 = 19\times20 + \frac{20\times21}{2} - 538 = 52$$

Thus $U = \min\{U_1, U_2\} = 52$.

Looking up the table of critical values, we know the reject region at $\alpha = 0.05$ is: $R = \{ u < 119 \}$.

Thus, the observed U statistic is in the rejection region, and we reject Ho.

(c) To calculate the p-value for the t test in (a), we write

$$p\text{-value} = P(R \mid H_0) = P\left(\frac{|\bar{X}_b - \bar{X}_c|}{S_p \sqrt{\frac{1}{19} + \frac{1}{20}}} \geq 4.927 \;\middle|\; H_0 \right)$$

$$= 2 * pt(-4.927, df = 37, lower.tail = TRUE)$$

$$= 1.767 \times 10^{-5}.$$

To calculate the p-value for the Mann-whitney test, we use

wilcox.test(background, contaminated)

in R, which in return gives p-value $= 1.116 \times 10^{-4}$.

The two p-values are both very significant, while the p-value for the t-test is much more significant.

Comparing paired samples

(a) The mean of the differences is:

$$\bar{D}_n = \bar{X}_n - \bar{Y}_n = 0.36286.$$

The sample variance of the differences is:

$$S_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - \bar{D}_n)^2 = 0.1648$$

The rejection region of the test $H_0: \mu_c = \mu_0$ vs $H_1: \mu_c \neq \mu_0$

is $R = \left\{ \frac{\sqrt{n}\,|\bar{D}_n|}{S_D} \geq \underbrace{t_{n-1}(\alpha/2)}_{= t_{13}(0.05/2) = 2.1604} \right\}$

Since $\frac{\sqrt{n}\,|\bar{D}_n|}{S_D} = \sqrt{\frac{14}{0.1648}} \times 0.36286 = 3.344$,

we reject Ho and conclude that there is enough evidence of difference in cholesterol levels.

(b) To perform the Wilcoxon ranked sum test, we first write out the ranks of absolutes differences:

Diff	abs(Diff)	rank	sign
-0.77	0.77	12	−
-0.85	0.85	13	−
0.45	0.45	8	+
0.26	0.26	4	+
-0.30	0.30	5	−
-0.86	0.86	14	−
-0.60	0.60	9	−
-0.62	0.62	10	−
-0.31	0.31	6	−
-0.72	0.72	11	−
-0.09	0.09	1	−
-0.16	0.16	3	−
-0.41	0.41	7	−
-0.10	0.10	2	−

From the table, we can calculate

$$W_+ = 8+4 = 12$$

$$W_- = 12+13+5+14+9+10+6+11+1+3+7+2 = 93$$

From the table of critical value, we know the rejection region at $\alpha=0.05$ is $R = \{ W_+ \le 21 \text{ or } W_- \le 21 \}$.

Therefore, we again reject Ho.

(c) The p-value for the paired t test is:

$$\text{p-value} = P\left(\frac{\sqrt{n}\,|\bar{D}_n|}{S_D} \geq 3.344 \,\middle|\, H_0 \right) = 0.005278.$$

The p-value for the Wilcoxon ranked sum test can be obtained in R using: wilcox.test(cornflk, oatbran, paired=TRUE).

$$\Rightarrow \text{p-value} = 0.008545.$$

The two p-values in this case are close to each other.

(d) The pooled variance estimate is:

$$S_p^2 = \frac{\Sigma(X_i-\bar{X})^2 + \Sigma(Y_i-\bar{Y})^2}{14+14-2} = 1.0279$$

$$\Rightarrow S_p = 1.01387 > S_D = \sqrt{0.1648} = 0.4059.$$