Vinay Maruri

HW5-STAT 135

1] let C be $X_{control}$. let O be X_{ozone}.

$H_0: \mu_C = \mu_O$ $\quad$ $H_A: \mu_C \neq \mu_O$

1st Assumption: Since I'm asked to use Mann-Whitney test, I assume the control group and ozone group are two independent samples.

2nd Assumption: Since α not specified, I assume $\alpha = 0.05$.

3rd Assumption: I assume equal variance for the 2 samples. (to leverage Theorem A + B of lecture 10.)

under Normal population assumption,

$$R = \left\{ \frac{|\bar{X}_c - \bar{Y}_o|}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \geq t_{n+m-2}(\alpha/2) \right\}$$

$\bar{X}_c$; Sample mean of control

$\bar{Y}_o$; Sample mean of ozone

n = # observations in control

m = # of observations in ozone

The following was computed in code.

$n = 23$, $m = 22$

$\bar{X}_c = 22.42$, $\bar{Y}_o = 11.01$

$S_p = 15.36$; plugging into rejection region, we get

$$R = \{2.491887 \geq 2.016692\}$$

and hence we reject the H_0, at a significance level of 0.05. There is enough evidence to suggest that $\mu_c \neq \mu_o$.

I go a step further and test

$H_0: \mu_c = \mu_o$ and $H_A: \mu_c > \mu_o$.

$$R = \left\{ \frac{\bar{X}_c - \bar{Y}_o}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \geq t_{n+m-2}(\alpha) \right\} = \{2.491887 \geq 1.681071\}$$

Hence, we reject the Ho against at significance level 0.05 = α. There is enough evidence to support the H_A, and conclude that ozone likely has a negative effect on weight in rats (compared to control)

under Mann-Whitney test,

$H_0: C = O$ $\quad$ $H_A: C \neq O$ $\quad$ where C: control
O: Ozone

~~R_C = sum of ranks in control~~

R_O = sum of ranks in ozone

~~n = # of obs in control~~

~~m = # of obs in ozone~~

$u = \min\{u_C, u_O\}$

where $u_C = mn + \frac{n(n+1)}{2} - R_C$

$u_O = mn + \frac{m(m+1)}{2} - R_O$

$R = \{u < c\}$ ← defined using Table of critical values

~~Computed in code~~: $U = 385$, p-value $= 0.002826$

Hence, we reject the Ho at a significance level of $\alpha = 0.05$. It's likely that there is enough evidence to support H_A.

7.1 (1) $H_0: \mu_c = \mu_o$ $H_A: \mu_c < \mu_o$

where c: cornflakes, o: oats

Assumption: we have paired samples. under normal theory, this implies

$$R = \left\{ \frac{\bar{D}_n}{\sqrt{\frac{1}{n} S_D^2}} \le -t_{n-1}(\alpha) \right\} \qquad \alpha = 0.05$$

where $\bar{D}_n = \frac{1}{n} \sum_{i=1}^{n} D_i$, $D_i = X_c - Y_o$

$$S_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - \bar{D}_n)^2 \qquad n-1 = 13$$

$\bar{D}_n = 0.3628571$

$t_{13}(0.05) = -1.770933$

$S_D^2 = 0.1648066$

p-value $= P\left(\frac{\sqrt{n}\,\bar{D}_n}{S_D} \le 3.344 \mid H_0\right)$

p-value = pt(3.344, 13, lower.tail = T)

$= 0.997$

$$\Rightarrow R = \left\{ \frac{0.3628571}{\sqrt{\frac{1}{14} \cdot S_D^2}} \le 1.770933 \right\}$$

$$\Rightarrow R = \{ 3.344 \le 1.770933 \}$$

we fail to reject the H_0 at a significance level of 0.05. There is not enough evidence of mean cholesterol on cornflakes being less than mean cholesterol on oatbran diet.

(7) Subject	Cornflk	Oatbran	Diff	|Diff|
1	4.61	3.84	-0.77	0.77
2	6.42	5.57	-0.85	0.85
3	5.40	5.85	-0.45	0.45
4	4.54	4.80	-0.26	0.26
5	3.95	3.68	-0.30	0.30
6	3.82	2.96	-0.86	0.86
7	5.01	4.41	-0.60	0.60
8	4.34	3.72	-0.62	0.62
9	3.80	3.49	-0.31	0.31
10	4.56	3.84	-0.72	0.72
11	5.35	5.26	-0.09	0.09
12	3.89	3.73	-0.16	0.16
13	2.25	1.84	-0.41	0.41
14	4.24	4.14	-0.10	0.10

Subject	rank	signed rank
1	12	-12
2	13	-13
3	8	8
4	4	4
5	5	-5
6	14	-14
7	9	-9
8	10	-10
9	6	-6
10	11	-11
11	1	-1
12	3	-3
13	7	-7
14	2	-2

$W_- = 93$

$W_+ = 12$

where C = cornflakes
O = oat bran

$\alpha = 0.05$ $H_0: C = O$ $H_A: C < O$

$\Rightarrow R = \{W_+ \geq C_1\} = \{W_- \leq C_2\}$

$n = 14$

$\Rightarrow C_1 = 26 = C_2$ ← from table of critical values.

$\Rightarrow R = \{12 \geq 26\} = \{93 \leq 26\}$

$\Rightarrow$ we fail ~~to~~ reject the H_0 at a significance level of 0.05. There is not enough evidence to support H_A, that cholesterol levels while on cornflake diet are less than on oat bran diet.

(3) p-value for normal theory hypothesis test: 0.997
from (1).

p-value for wilcoxon test: 0.9966
computed in code

The p-values are not very different.

3] H_0: $X \overset{\text{follows}}{\sim}$ poisson Distribution $\boxed{\text{Assume } \alpha = 0.05}$
H_A: H_0 is not true.

$\hat{\lambda}_{MLE} = \bar{X}_n$
≈ 5.57

$$2\sum_{i=1}^{9} O_i \log \frac{O_i}{E_i} \xrightarrow{d} \chi^2_{m-2} = \chi^2_7$$

$$E_i = np_i(\hat{\lambda}) = np_i(\bar{X}_n)$$

$$\Rightarrow R = \{-2\log \lambda(\bar{X}_n) \geq \chi^2_7(\alpha)\}$$

category	X	obs'd	prob	Prob$_{MLE}$
1	0,1,2	5	$e^{-\lambda}\left(\frac{\lambda^0}{0!}+\frac{\lambda^1}{1!}+\frac{\lambda^2}{2!}\right)$	$e^{-\hat{\lambda}}\left(\frac{\hat{\lambda}^0}{0!}+\frac{\hat{\lambda}^1}{1!}+\frac{\hat{\lambda}^2}{2!}\right)$
2	3	13	$e^{-\lambda}\left(\frac{\lambda^3}{3!}\right)$	$e^{-\hat{\lambda}}\left(\frac{\hat{\lambda}^3}{3!}\right)$
3	4	19		
4	5	16		
5	6	15		
6	7	9		
7	8	12		
8	9	7	$e^{-\lambda}\left(\frac{\lambda^9}{9!}\right)$	
9	10,11,12	4	$e^{-\lambda}\left(\frac{\lambda^{10}}{10!}+\frac{\lambda^{11}}{11!}+\frac{\lambda^{12}}{12!}\right)$	$e^{-\hat{\lambda}}\left(\frac{\hat{\lambda}^{10}}{10!}+\frac{\hat{\lambda}^{11}}{11!}+\frac{\hat{\lambda}^{12}}{12!}\right)$

expected	8.41	10.97	15.28	17.02	15.80	12.57	8.76	5.42	5.26
category	1	2	3	4	5	6	7	8	9

$$-2\log(\bar{X}_n) = 2\sum_{i=1}^{m} O_i \log\frac{O_i}{E_i} = 6.888$$

qchisq(0.05, df=7, lower.tail=False)

$\chi^2_7(0.05) = 14.06714$

$R = \{6.888 \geq 14.06714\}$

~~We fail to reject Ho~~ at significance level 0.05. There is not enough evidence to suggest X does not follow a Poisson distribution.

4] H_0: Multinomial variables have equal probabilities of having headaches of certain severity.
H_A: H_0 is false. Assume $\alpha = 0.05$
$df = (\#\text{groups}-1)(\#\text{levels}-1) = (2-1)(4-1) = 3$

expected:	Pfizer	Placebo
No headache	1303.4468	1306.5532
Mild	428.9888	430.0112
Moderate	374.6132	375.3868
Severe	40.9512	41.0488

see code for computation.

$$\chi^2 = \frac{\sum_{g,l} (O_{g,l} - E_{g,l})^2}{E_{g,l}} = 366.3075$$

qchisq(0.05, df=3, lower.tail=F) = 7.814728

pchisq(366.3075, df=3, lower.tail=F) = 4.389×10^{-79}

We reject the H_0 at a significance level of 0.05. There is enough evidence to suggest there are not equal probabilities of having headaches of certain severity for the 2 columns.

$$R = \{\lambda(\bar{x}_n) \le c\} = \left\{ \frac{\sum_{g,l} (O_{g,l} - E_{g,l})^2}{E_{g,l}} \ge \chi^2_{(I-1)(J-1)}(\alpha) \right\}$$

5] H_0: Age is independent of the desire to ride a bicycle.

H_A: H_0 is false [Age is not independent of desire to ride a bicycle.]

Assume $\alpha = 0.05$

expected		Age		
	18-24	25-34	35-49	50-64
Yes	50.886	49.868	50.377	49.868
No	49.114	48.132	47.623	48.132

See code for computation

$$\chi^2 = \sum_{q,l} \frac{(O_{q,l} - E_{q,l})^2}{E_{q,l}} = 8.006066$$

qchisq(0.05, df=3, lower.tail=F) = 7.814728

pchisq(8.006066, df=3, lower.tail=F) = 0.04585651

$$R = \{\lambda(\bar{x}_n) \le c\} = \left\{ \sum_{q,l} \frac{(O_{q,l} - E_{q,l})^2}{E_{q,l}} \ge \chi^2_{(I-1)(J-1)}(\alpha) \right\}$$

We reject the H_0 at a significance level of 0.05. There is enough evidence to suggest that Age is not independent of desire to ride a bicycle.