STA409

Answer to Assignment 4

12112627
李乐平

1. Solution.

(1).

$$\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}\mathrm{E}^2(\overline{Y}_{ij}-\overline{Y}_{i\cdot}-\overline{Y}_{\cdot j}+\overline{Y})$$
$$=\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}(\mu+\alpha_i+\beta_j+\gamma_{ij}-(\mu+\alpha_i)-(\mu+\beta_i))$$
$$=\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}\gamma_{ij}.$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}\mathrm{Var}(\overline{Y}_{ij}-\overline{Y}_{i\cdot}-\overline{Y}_{\cdot j}+\overline{Y})$$
$$=\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}[$$
$$\mathrm{Var}(\overline{Y}_{ij})+\mathrm{Var}(\overline{Y}_{i\cdot})+\mathrm{Var}(\overline{Y}_{\cdot j})+\mathrm{Var}(\overline{Y})$$
$$-2(\mathrm{Cov}(\overline{Y}_{ij},\overline{Y}_{i\cdot})+\mathrm{Cov}(\overline{Y},\overline{Y}_{i\cdot})+\mathrm{Cov}(\overline{Y}_{ij},\overline{Y}_{\cdot j})+\mathrm{Cov}(\overline{Y},\overline{Y}_{\cdot j}))$$
$$+2(\mathrm{Cov}(\overline{Y}_{ij},\overline{Y})+\mathrm{Cov}(\overline{Y}_{i\cdot},\overline{Y}_{\cdot j}))$$
$$]$$
$$=\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}[\frac{\sigma^2}{n_{ij}}+\frac{\sigma^2}{n_{i\cdot}}+\frac{\sigma^2}{n_{\cdot j}}+\frac{\sigma^2}{n}-2(\frac{\sigma^2}{n_{i\cdot}}+\frac{\sigma^2}{n}+\frac{\sigma^2}{n_{\cdot j}}+\frac{\sigma^2}{n}-\frac{\sigma^2}{n}-\frac{n_{ij}\sigma^2}{n_{i\cdot}n_{\cdot j}})]$$
$$=\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}(\frac{\sigma^2}{n_{ij}}-\frac{\sigma^2}{n_{i\cdot}}-\frac{\sigma^2}{n_{\cdot j}}-\frac{\sigma^2}{n}+2\frac{n_{ij}\sigma^2}{n_{i\cdot}n_{\cdot j}})$$
$$=\sigma^2(ab-a-b+1+2\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{n_{ij}^2}{n_{i\cdot}n_{\cdot j}}).$$

Because n_{ij}'s are all the equal,

$$\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}\mathrm{Var}(\overline{Y}_{ij}-\overline{Y}_{i\cdot}-\overline{Y}_{\cdot j}+\overline{Y})=(a-1)(b-1)\sigma^2.$$

$$\mathrm{SSAB}=\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}(\overline{Y}_{ij}-\overline{Y}_{i\cdot}-\overline{Y}_{\cdot j}+\overline{Y})^2.$$

$$\mathrm{E(SSAB)}=\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}(\mathrm{E}^2(\overline{Y}_{ij}-\overline{Y}_{i\cdot}-\overline{Y}_{\cdot j}+\overline{Y})+\mathrm{Var}(\overline{Y}_{ij}-\overline{Y}_{i\cdot}-\overline{Y}_{\cdot j}+\overline{Y}))$$
$$=(a-1)(b-1)\sigma^2+\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}\gamma_{ij}^2.$$

(2).

When n_{ij}'s are all equal (say $n_{ij} = n_0, \forall i, j$), we have

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(\bar{y}_{i\cdot}-\bar{y})(\bar{y}_{\cdot j}-\bar{y}) = n_0\sum_{i=1}^{a}(\bar{y}_{i\cdot}-\bar{y})\sum_{j=1}^{b}(\bar{y}_{\cdot j}-\bar{y}) = 0$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(\bar{y}_{i\cdot}-\bar{y})(\bar{y}_{ij}-\bar{y}_{i\cdot}-\bar{y}_{\cdot j}+\bar{y}) = n_0\sum_{i=1}^{a}(\bar{y}_{i\cdot}-\bar{y})\sum_{j=1}^{b}(\bar{y}_{ij}-\bar{y}_{i\cdot}-\bar{y}_{\cdot j}+\bar{y}) = 0$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(\bar{y}_{\cdot j}-\bar{y})(\bar{y}_{ij}-\bar{y}_{i\cdot}-\bar{y}_{\cdot j}+\bar{y}) = n_0\sum_{i=1}^{a}(\bar{y}_{ij}-\bar{y}_{i\cdot}-\bar{y}_{\cdot j}+\bar{y})\sum_{j=1}^{b}(\bar{y}_{\cdot j}-\bar{y}) = 0$$

$$\mathrm{SSM} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(\bar{y}_{ij}-\bar{y})^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}[(\bar{y}_{i\cdot}-\bar{y})+(\bar{y}_{\cdot j}-\bar{y})+(\bar{y}_{ij}-\bar{y}_{i\cdot}-\bar{y}_{\cdot j}+\bar{y})]^2$$

$$= \mathrm{SSA}+\mathrm{SSB}+\mathrm{SSAB}$$

$$+2\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(\bar{y}_{i\cdot}-\bar{y})(\bar{y}_{\cdot j}-\bar{y})$$

$$+2\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(\bar{y}_{i\cdot}-\bar{y})(\bar{y}_{ij}-\bar{y}_{i\cdot}-\bar{y}_{\cdot j}+\bar{y})$$

$$+2\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(\bar{y}_{\cdot j}-\bar{y})(\bar{y}_{ij}-\bar{y}_{i\cdot}-\bar{y}_{\cdot j}+\bar{y})$$

$$= \mathrm{SSA}+\mathrm{SSB}+\mathrm{SSAB}$$

2. Solution.

(1). By simple calculation and we get

$$\bar{y}_{\mathrm{A}} = 43, \bar{y}_{\mathrm{B}} = 31, \bar{y}_{\mathrm{C}} = 43, \bar{y} = 39$$

$$S_{\mathrm{A}}^2 = 12.8, S_{\mathrm{B}}^2 = 8.4, S_{\mathrm{C}}^2 = 51.2$$

$$\mathrm{SSB} = \sum_{\text{Group } g} n_g(\bar{Y}_g - \bar{Y})^2 = 576$$

$$\mathrm{SSW} = \sum_{\text{Group } g} (n_g - 1)S_g^2 = 362$$

ANOVA Table					
Source	DF.	SS.	MS.	F-Value	p-Value
Between	2	576	288	11.934	7.9×10^{-4}
Within	15	362	24.133		
Total	17	938			

Since p-value = $7.9\times10^{-4} < 0.05$, we have to reject H_0: $\mu_A=\mu_B=\mu_C$ under the significance level of 0.05.

(2). We can easily compute that

Z_{ij}	1	2	3	4	5	6
A	9	16	25	1	4	9
B	16	16	1	4	1	4
C	16	9	25	49	64	121

$$\bar{Z}_{A} = 10.67, \bar{Z}_{B} = 7, \bar{Z}_{C} = 42.67, \bar{Z} = 20.11$$

$$S_{ZA}^{2} = 75.47, S_{ZB}^{2} = 50.4, S_{ZC}^{2} = 1675.47$$

$$\mathrm{SSB}_{Z} = \sum_{\text{Group } g} n_{g}(\bar{Z}_{g} - \bar{Z})^{2} = 4619.11$$

$$\mathrm{SSW}_{z} = \sum_{\text{Group } g} (n_{g} - 1)S_{Zg}^{2} = 9006.67$$

ANOVA Table of Z_{ij}					
Source	DF.	SS.	MS.	F-Value	p-Value
Between	2	4619.11	2309.56	3.85	4.47×10^{-2}
Within	15	9006.67	600.44		
Total	17	13625.78			

Since p-value = 0.0447 < 0.05, we have to reject H_0: $\sigma_A^2 = \sigma_B^2 = \sigma_C^2$.

(3).

y	27	29	30	32	32	33	35	38	39	40	40	42	45	46	47	48	49	50
R	1	2	3	4.5	4.5	6	7	8	9	10.5	10.5	12	13	14	15	16	17	18
G	B	B	B	B	C	B	B	A	C	A	C	A	A	A	A	C	C	C

$$\bar{r}_{A} = 12.083, \bar{r}_{B} = 3.917, \bar{r}_{C} = 12.5, \bar{r} = 9.5$$

$$\mathrm{KW} = \frac{(n-1)\sum_{i=1}^{k} n_{i}(\bar{R}_{i} - \bar{R})^{2}}{\sum_{i=1}^{k}\sum_{j=1}^{n_{i}} (\bar{R}_{i} - \bar{R})^{2}} = 9.883$$

The p-value is given by $\chi^2(9.883, 2) = 7.14 \times 10^{-3} < 0.05$, hence H_0: $m_A = m_B = m_C$ must be rejected.

3. Solution.

(1). The null hypothesis H_0: there is no difference among the mean number of glassware produced by the four workers is accepted since p-value = 0.4841 > 0.1.

源	自由度	平方和	均方	F 值	Pr > F
模型	3	20.9206349	6.9735450	0.84	0.4841
误差	26	215.7460317	8.2979243		
校正合计	29	236.6666667			

(2). The null hypothesis H_0: there is no difference among the mean number of glassware produced by the four workers is rejected since p-value = 0.0056 < 0.1.

源	自由度	平方和	均方	F 值	Pr > F
模型	2	75.4303030	37.7151515	6.32	0.0056
误差	27	161.2363636	5.9717172		
校正合计	29	236.6666667			

(3). By the result below, we can obtain

$$\mu_A - \mu_B \in [1.461, 6.273]$$
$$\mu_A - \mu_C \in [-1.270, 3.306]$$
$$\mu_B - \mu_C \in [-5.202, -0.495]$$

That implies differences do exist between tool pairs (A, B) and (B, C).

Comparisons significant at the 0.1 level are indicated by ***.				
tool 比较	均值间差值	Simultaneous 90% 置信限		
ToolA - ToolC	1.018	-1.270	3.306	
ToolA - ToolB	3.867	1.461	6.273	***
ToolC - ToolA	-1.018	-3.306	1.270	
ToolC - ToolB	2.848	0.495	5.202	***
ToolB - ToolA	-3.867	-6.273	-1.461	***
ToolB - ToolC	-2.848	-5.202	-0.495	***

(4). The null hypothesis H_0: there is no interaction effect between worker and tool is rejected by p-value = 0.0874 < 0.1.

源	自由度	I 型 SS	均方	F 值	Pr > F
worker	3	20.92063492	6.97354497	1.73	0.1964
tool	2	89.32525656	44.66262828	11.09	0.0007
worker*tool	6	53.92077518	8.98679586	2.23	0.0874

(5). The null hypothesis H_0: the residuals follow a normal distribution is accepted by a Shapiro-Wilk test with p-value = 0.6901 > 0.1.

正态性检验				
检验	统计量		p 值	
Shapiro-Wilk	W	0.97525	Pr < W	0.6901
Kolmogorov-Smirnov	D	0.1	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.036512	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.235878	Pr > A-Sq	>0.2500

4. Solution.

(1). The Type I Error Rate is 0.0498.

表: WORK.TASK4_ST1 视图: 列名

列 总行数: 1 总列数: 1

全选

type_I_err_rate

	type_I_err_rate
1	0.0498

(2). The Type I Error Rate is 0.0802 > 0.0498.

表: WORK.TASK4_ST2 视图: 列名

列 总行数: 1 总列数: 1

全选

type_I_err_rate

	type_I_err_rate
1	0.0802

From the result, we can see that heteroscedasticity may increase the Type I Error Rate, so homoscedasticity is necessary to be guaranteed if we want to handle an F-test in one-way ANOVA to test the equality of group means.

5. Solution.

(1).

Since the r is known, so the parameter is p. Let

$$h(y)=\frac{(y-r-1)!}{y!(r-1)!},\eta(p)=\ln p,T(y)=y,A(q)=r\ln\frac{1}{1-p}.$$

Then

$$\begin{aligned}f(y;r,p)&=\mathrm{C}_{y+r-1}^{y}(1-p)^r p^y\\&=\frac{(y+r-1)!(1-p)^r p^y}{y!(r-1)!}\\&=\frac{(y+r-1)!}{y!(r-1)!}\exp(y\ln p-r\ln\frac{1}{1-p})\\&=h(y)\exp(\eta(p)\cdot T(y)-A(p))\end{aligned}$$

Hence NB(r, p) belongs to the exponential family when r is known.

(2). Rewrite the distribution of NB(r, p) as a canonical form.

Let

$$q=\ln p,h(y)=\frac{(y-r-1)!}{y!(r-1)!},\eta_c(q)=q,T(y)=y,A_c(q)=r\ln\frac{1}{1-e^q}$$

then

$$\begin{aligned}f(y;r,p)&=\mathrm{C}_{y+r-1}^{y}(1-p)^r p^y\\&=\frac{(y+r-1)!(1-p)^r p^y}{y!(r-1)!}\\&=\frac{(y+r-1)!}{y!(r-1)!}\exp(yq-r\ln\frac{1}{1-e^q})\\&=h(y)\exp(\eta_c(q)\cdot T(y)-A_c(q))\end{aligned}$$

Hence

$$\mathrm{E}(y)=A_c{}'(q)=\frac{re^q}{1-e^q}=\frac{rp}{1-p}$$

$$\mathrm{Var}(y)=A_c{}''(q)=\frac{re^q}{(1-e^q)^2}=\frac{rp}{(1-p)^2}$$