MATH 503: Mathematical Statistics
Dr. Kimberly F. Sellers, Instructor

Homework 9 Solutions

1. **Let $X_{1j}, X_{2j}, \ldots, X_{a_j j}$ represent independent random samples of sizes a_j from a normal distribution with means μ_j and variances σ^2, $j = 1, 2, \ldots, b$. Show that**

$$\sum_{j=1}^{b}\sum_{i=1}^{a_j}(X_{ij} - \bar{X}_{..})^2 = \sum_{j=1}^{b}\sum_{i=1}^{a_j}(X_{ij} - \bar{X}_{.j})^2 + \sum_{j=1}^{b} a_j(\bar{X}_{.j} - \bar{X}_{..})^2,$$

or $Q' = Q'_3 + Q'_4$. Here,

$$\bar{X}_{..} = \frac{\sum_{j=1}^{b}\sum_{i=1}^{a_j} X_{ij}}{\sum_{j=1}^{b} a_j} \text{ and } \bar{X}_{.j} = \frac{\sum_{i=1}^{a} X_{ij}}{a_j}.$$

If $\mu_1 = \mu_2 = \ldots = \mu_b$, show that Q'/σ^2 and Q'_3/σ^2 have chi-square distributions. Note that Q'_3 and Q'_4 are independent, and hence Q'_4/σ^2 also has a chi-square distribution.

Solution:

$$\begin{aligned}
\sum_{j=1}^{b}\sum_{i=1}^{a_j}(X_{ij} - \bar{X}_{..})^2 &= \sum_{j=1}^{b}\sum_{i=1}^{a_j}[(X_{ij} - \bar{X}_{.j}) + (\bar{X}_{.j} - \bar{X}_{..})]^2 \\
&= \sum_{j=1}^{b}\sum_{i=1}^{a_j}(X_{ij} - \bar{X}_{.j})^2 + \sum_{j=1}^{b}\sum_{i=1}^{a_j}(\bar{X}_{.j} - \bar{X}_{..})^2 + 2\sum_{j=1}^{b}\sum_{i=1}^{a_j}(X_{ij} - \bar{X}_{.j})(\bar{X}_{.j} - \bar{X}_{..}),
\end{aligned}$$

where

$$\sum_{j=1}^{b}\sum_{i=1}^{a_j}(X_{ij} - \bar{X}_{.j})(\bar{X}_{.j} - \bar{X}_{..}) = \sum_{j=1}^{b}(\bar{X}_{.j} - \bar{X}_{..})\sum_{i=1}^{a_j}(X_{ij} - \bar{X}_{.j}) = \sum_{j=1}^{b}(\bar{X}_{.j} - \bar{X}_{..})\underbrace{\left(\sum_{i=1}^{a_j} X_{ij} - a_j\bar{X}_{.j}\right)}_{=0}$$

Thus,

$$\begin{aligned}
Q' = \sum_{j=1}^{b}\sum_{i=1}^{a_j}(X_{ij} - \bar{X}_{..})^2 &= \sum_{j=1}^{b}\sum_{i=1}^{a_j}(X_{ij} - \bar{X}_{.j})^2 + \sum_{j=1}^{b}\sum_{i=1}^{a_j}(\bar{X}_{.j} - \bar{X}_{..})^2 \\
&= \sum_{j=1}^{b}\sum_{i=1}^{a_j}(X_{ij} - \bar{X}_{.j})^2 + \sum_{j=1}^{b}(\bar{X}_{.j} - \bar{X}_{..})^2\sum_{i=1}^{a_j}1 \\
&= \sum_{j=1}^{b}\sum_{i=1}^{a_j}(X_{ij} - \bar{X}_{.j})^2 + \sum_{j=1}^{b} a_j(\bar{X}_{.j} - \bar{X}_{..})^2 = Q'_3 + Q'_4.
\end{aligned}$$

Note that

$$\frac{Q'}{\sigma^2} = \frac{\sum_{j=1}^{b}\sum_{i=1}^{a_j}(X_{ij} - \bar{X}..)^2}{\sigma^2} = \sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{X_{ij} - \bar{X}..}{\sigma}\right)^2,$$

where

$$\begin{aligned}
\chi^2_{\sum_{j=1}^{b} a_j} &= \sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{X_{ij} - \mu}{\sigma}\right)^2 = \sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{(X_{ij} - \bar{X}..) + (\bar{X}.. - \mu)}{\sigma}\right)^2 \\
&= \sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{X_{ij} - \bar{X}..}{\sigma}\right)^2 + \sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{\bar{X}.. - \mu}{\sigma}\right)^2 + 2\sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{X_{ij} - \bar{X}..}{\sigma}\right)\left(\frac{\bar{X}.. - \mu}{\sigma}\right),
\end{aligned}$$

where

$$\begin{aligned}
\sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{X_{ij} - \bar{X}..}{\sigma}\right)\left(\frac{\bar{X}.. - \mu}{\sigma}\right) &= \left(\frac{\bar{X}.. - \mu}{\sigma}\right)\sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{X_{ij} - \bar{X}..}{\sigma}\right) \\
&= \left(\frac{\bar{X}.. - \mu}{\sigma}\right)\left[\frac{1}{\sigma}\left(\underbrace{\sum_{j=1}^{b}\sum_{i=1}^{a_j} X_{ij} - \bar{X}..\sum_{j=1}^{b} a_j}_{=0}\right)\right],
\end{aligned}$$

so

$$\chi^2_{\sum_{j=1}^{b} a_j} = \sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{X_{ij} - \bar{X}..}{\sigma}\right)^2 + \sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{\bar{X}.. - \mu}{\sigma}\right)^2,$$

where

$$\sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{\bar{X}.. - \mu}{\sigma}\right)^2 = \left(\frac{\bar{X}.. - \mu}{\sigma}\right)^2\sum_{j=1}^{b} a_j = \left(\frac{\bar{X}.. - \mu}{\sigma/\sqrt{\sum_{j=1}^{b} a_j}}\right)^2 \sim \chi^2_1,$$

thus we see that $\frac{Q'}{\sigma^2} = \sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{X_{ij} - \bar{X}..}{\sigma}\right)^2 \sim \chi^2_{\left(\sum_{j=1}^{b} a_j\right)-1}$. Similarly,

$$\begin{aligned}
\sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{X_{ij} - \mu}{\sigma}\right)^2 &= \sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{X_{ij} - \bar{X}.j + \bar{X}.j - \mu}{\sigma}\right)^2 \\
&= \sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{X_{ij} - \bar{X}.j}{\sigma}\right)^2 + \sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{\bar{X}.j - \mu}{\sigma}\right)^2 \\
&= \sum_{j=1}^{b}\sum_{i=1}^{a_j}\left(\frac{X_{ij} - \bar{X}.j}{\sigma}\right)^2 + \sum_{j=1}^{b} a_j\left(\frac{\bar{X}.j - \mu}{\sigma}\right)^2 = \frac{Q_3'}{\sigma} + \chi^2_b,
\end{aligned}$$

thus $\frac{Q_3'}{\sigma} \sim \chi^2_{\left(\sum_{j=1}^{b} a_j\right)-b}$.

Finally, since $\frac{Q'}{\sigma} = \frac{Q_3'}{\sigma} + \frac{Q_4'}{\sigma}$ where $\frac{Q'}{\sigma} \sim \chi^2_{\left(\sum_{j=1}^{b} a_j\right)-1}$ and $\frac{Q_3'}{\sigma} \sim \chi^2_{\left(\sum_{j=1}^{b} a_j\right)-b}$, this implies that $\frac{Q_4'}{\sigma} \sim \chi^2_{\left(\sum_{j=1}^{b} a_j\right)-1-\left[\left(\sum_{j=1}^{b} a_j\right)-b\right]} = \chi^2_{b-1}$.

2. **Solve the following using R or SAS: The following are observations associated with independent random samples from three normal distributions having equal variances and respective means μ_1, μ_2, μ_3.**

I	II	III
0.5	2.1	3.0
1.3	3.3	5.1
-1.0	0.0	1.9
1.8	2.3	2.4
	2.5	4.2
		4.1

Compute the F-statistic that is used to test $H_0 : \mu_1 = \mu_2 = \mu_3$.

As shown in the below output, F=6.388, whether determined via R or SAS.

R code and output:

```
> data <- read.table("C:/courses/MATH 503/data/Exercise9.2.4data.txt",header=TRUE)
> summary(aov(obs ~ factor(group),data=data))
               Df   Sum Sq   Mean Sq  F value Pr(>F)
factor(group)   2    19.10     9.548    6.388 0.0129 *
Residuals      12    17.94     1.495
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

SAS code shown below; corresponding output provided in Figure 1.

```
filename exer924 'C:\courses\MATH 503\data\Exercise9.2.4data.txt';
data exercise924;
infile exer924 delimiter='09'x firstobs=2;
input obs group;
run;

proc anova data=exercise924;
class group;
model obs = group;
;

run;
```

The SAS System

The ANOVA Procedure

Class Level Information

Class	Levels	Values
group	3	1 2 3

Number of Observations Read	15
Number of Observations Used	15

The SAS System

The ANOVA Procedure

Dependent Variable: obs

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	19.09633333	9.54816667	6.39	0.0129
Error	12	17.93700000	1.49475000		
Corrected Total	14	37.03333333			

R-Square	Coeff Var	Root MSE	obs Mean
0.515653	54.74327	1.222600	2.233333

Source	DF	Anova SS	Mean Square	F Value	Pr > F
group	2	19.09633333	9.54816667	6.39	0.0129

Figure 1: Problem 2 SAS output

3. **Solve the following by hand, and by using either R or SAS: Let μ_1, μ_2, μ_3 be, respectively, the means of three normal distributions with a common but unknown variance σ^2. In order to test, at the $\alpha = 0.05$ significance level, the hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$ against all possible alternative hypotheses, we take an independent random sample of size 4 from each of these distributions. Determine whether we reject or fail to reject H_0 if the observed values from these three distributions are, respectively:**

X1	**5**	**9**	**6**	**8**
X2	**11**	**13**	**10**	**12**
X3	**10**	**6**	**9**	**9**

Solved in any of the three manners, we reject H_0 at the 5% significance level (F = 7.875; p-val = 0.0105).

R code and output:

```
> data <- read.table("C:/courses/MATH 503/data/Exercise9.2.6data.txt",header=TRUE)
> summary(aov(obs ~ factor(group),data=data))
                 Df Sum Sq    Mean Sq F value Pr(>F)
factor(group)    2      42     21.000   7.875 0.0105 *
Residuals        9      24      2.667
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

SAS code shown below; corresponding output provided in Figure 2.

```
filename exer926 'C:\courses\MATH 503\data\Exercise9.2.6data.txt';
data exercise926;
infile exer926 delimiter='09'x firstobs=2;
input obs group;
run;

proc anova data=exercise926;
class group;
model obs = group;
;

run;
```

The SAS System

The ANOVA Procedure

Class Level Information

Class	Levels	Values
group	3	1 2 3

Number of Observations Read	12
Number of Observations Used	12

The SAS System

The ANOVA Procedure

Dependent Variable: obs

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	42.00000000	21.00000000	7.88	0.0105
Error	9	24.00000000	2.66666667		
Corrected Total	11	66.00000000			

R-Square	Coeff Var	Root MSE	obs Mean
0.636364	18.14437	1.632993	9.000000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
group	2	42.00000000	21.00000000	7.88	0.0105

Figure 2: Problem 3 SAS output

4. **Let $X_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$ denote independently normally distributed random variables, where $\epsilon_{ij} \sim \mathbf{N}(0, \sigma^2)$; $i = 1, 2, \ldots, a$ and $j = 1, 2, \ldots, b$. Show that the maximum likelihood estimator of α_i, β_j, and μ are $\hat{\alpha}_i = \bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot}$, $\hat{\beta}_j = \bar{X}_{\cdot j} - \bar{X}_{\cdot\cdot}$, and $\hat{\mu} = \bar{X}_{\cdot\cdot}$, respectively.**

We have that $X_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$ where $\epsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$, i.e. $X_{ij} \sim \mathrm{N}(\mu + \alpha_i + \beta_j, \sigma^2)$, therefore $f(x_{ij}; \mu, \alpha_i, \beta_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2}(x_{ij} - \mu - \alpha_i - \beta_j)^2}$.

$$
\begin{aligned}
L(\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{x}) &= (2\pi\sigma^2)^{-ab/2} \exp\left[\frac{-1}{2\sigma^2} \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \mu - \alpha_i - \beta_j)^2\right] \\
\ln L(\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{x}) &= \frac{-ab}{2} \ln(2\pi) - \frac{ab}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \mu - \alpha_i - \beta_j)^2.
\end{aligned}
$$

Minimizing $\ln L(\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{x})$ with respect to $\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}$ respectively reduces to minimizing

$\sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \mu - \alpha_i - \beta_j)^2$

$$
\begin{aligned}
&= \sum_{j=1}^{b} \sum_{i=1}^{a} [(x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}_{\cdot\cdot}) + (\bar{x}_{\cdot\cdot} - \mu) + (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot} - \alpha_i) + (\bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot} - \beta_j)]^2 \\
&= \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}_{\cdot\cdot})^2 + \sum_{j=1}^{b} \sum_{i=1}^{a} (\bar{x}_{\cdot\cdot} - \mu)^2 \\
&\quad + \sum_{j=1}^{b} \sum_{i=1}^{a} (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot} - \alpha_i)^2 + \sum_{j=1}^{b} \sum_{i=1}^{a} (\bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot} - \beta_j)^2, \qquad (1)
\end{aligned}
$$

which, because $\sum_{i=1}^{a} \alpha_i = 0$ and $\sum_{j=1}^{b} \beta_j = 0$, Equation (1) is minimized iff $\hat{\mu} = \bar{x}_{\cdot\cdot}$, $\hat{\alpha}_i = \bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot}$, and $\hat{\beta}_j = \bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot}$.
Further, note that

$$
\begin{aligned}
E(\hat{\mu}) &= E(\bar{X}_{\cdot\cdot}) = \frac{1}{ab} \sum_{j=1}^{b} \sum_{i=1}^{a} E(X_{ij}) = \frac{1}{ab} \sum_{j=1}^{b} \sum_{i=1}^{a} (\mu + \alpha_i + \beta_j) = \frac{ab\mu}{ab} = \mu \\
E(\hat{\alpha}_i) &= E\left(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot}\right) = E\left(\bar{X}_{i\cdot}\right) - \mu = \frac{\sum_{j=1}^{b} E(X_{ij})}{b} - \mu = \frac{\sum_{j=1}^{b} (\mu + \alpha_i + \beta_j)}{b} - \mu \\
&= \frac{b\mu}{b} + \frac{b\alpha_i}{b} - \mu = \alpha_i \\
E(\hat{\beta}_j) &= E\left(\bar{X}_{\cdot j} - \bar{X}_{\cdot\cdot}\right) = \frac{\sum_{i=1}^{a} E(X_{ij})}{a} - \mu = \frac{\sum_{i=1}^{a} (\mu + \alpha_i + \beta_j)}{a} - \mu = \frac{a\mu}{a} + \frac{a\beta_j}{a} - \mu = \beta_j
\end{aligned}
$$

5. **Solve the following using either R or SAS: Given the following observations in a two-way classification with $a = 3$, $b = 4$, and $c = 2$, compute the F-statistics used to test that all interactions are equal to zero ($\gamma_{ij} = 0$), all column means are equal ($\beta_j = 0$), and all row means are equal ($\alpha_i = 0$), respectively.**

		Column			
		1	2	3	4
Row	1	3.1	4.2	2.7	4.9
		2.9	4.9	3.2	4.5
	2	2.7	2.9	1.8	3.0
		2.9	2.3	2.4	3.7
	3	4.0	4.6	3.0	3.9
		4.4	5.0	2.5	4.2

Whether solved via R or SAS, we produce the same ANOVA table; see below for code and associated output.

R code and output:

```
> data <- read.table("C:/courses/MATH 503/data/Exercise9.5.7data.txt",header=TRUE)
> summary(aov(obs~factor(row)+factor(col) + factor(row)*factor(col), data=data))
                        Df Sum Sq Mean Sq F value   Pr(>F)
factor(row)              2  7.298   3.649  30.726 1.90e-05 ***
factor(col)              3  8.131   2.710  22.825 3.01e-05 ***
factor(row):factor(col)  6  3.412   0.569   4.789   0.0102 *
Residuals               12  1.425   0.119
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

SAS code shown below; corresponding output provided in Figure 3.

```
filename exer957 'C:\courses\MATH 503\data\Exercise9.5.7data.txt';
data exercise957;
infile exer957 delimiter='09'x firstobs=2;
input obs row col;
run;

proc anova data=exercise957;
class row col;
model obs = row col row*col;
;

run;
```

The SAS System

The ANOVA Procedure

Class Level Information

Class	Levels	Values
row	3	1 2 3
col	4	1 2 3 4

Number of Observations Read	24
Number of Observations Used	24

The SAS System

The ANOVA Procedure

Dependent Variable: obs

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	18.84125000	1.71284091	14.42	<.0001
Error	12	1.42500000	0.11875000		
Corrected Total	23	20.26625000			

R-Square	Coeff Var	Root MSE	obs Mean
0.929686	9.881039	0.344601	3.487500

Source	DF	Anova SS	Mean Square	F Value	Pr > F
row	2	7.29750000	3.64875000	30.73	<.0001
col	3	8.13125000	2.71041667	22.82	<.0001
row*col	6	3.41250000	0.56875000	4.79	0.0102

Figure 3: Problem 5 SAS output