STAT 420 – Homework 9

1. Sports Programs (without R)

a. Consider the model $Y_{ij} = \mu_j + \varepsilon_{ij}$, where ε_{ij}'s are i.i.d. $N(0, \sigma^2)$.
Test $H_0: \mu_B = \mu_F = \mu_S$ with $\alpha = 0.05$.

$J = 3.$
$N = n_1 + n_2 + \ldots + n_J = 5 + 5 + 5 = 15.$
$$\overline{y} = \frac{n_1 \cdot \overline{y}_1 + n_2 \cdot \overline{y}_2 + \ldots + n_J \cdot \overline{y}_J}{N} = \frac{5 \cdot 3.0 + 5 \cdot 3.3 + 5 \cdot 2.4}{15} = 2.9.$$

$$\text{SSB} = n_1 \cdot \left(\overline{y}_1 - \overline{y}\right)^2 + n_2 \cdot \left(\overline{y}_2 - \overline{y}\right)^2 + \ldots + n_J \cdot \left(\overline{y}_J - \overline{y}\right)^2$$
$$= 5 \cdot (3.0 - 2.9)^2 + 5 \cdot (3.3 - 2.9)^2 + 5 \cdot (2.4 - 2.9)^2 = 2.1.$$

$$\text{SSW} = \left(n_1 - 1\right) \cdot s_1^2 + \left(n_2 - 1\right) \cdot s_2^2 + \ldots + \left(n_J - 1\right) \cdot s_J^2$$
$$= 4 \cdot 0.220 + 4 \cdot 0.145 + 4 \cdot 0.235 = 2.4.$$

SSTotal = SSB + SSW = 2.1 + 2.4 = 4.5.

Completing the ANOVA table,

Source	SS	df	MS	F
Between Groups	2.1	$J - 1 = 2$	1.05	**5.25**
Within Groups	2.4	$N - J = 12$	0.2	
Total	4.5	$N - 1 = 14$		

According to the F-distribution, the critical region is $F > F_{0.05}(2,12) =$ **3.89**. Since the test statistic lies in the critical region, we reject H_0 and conclude that the model does a significant job of predicting GPA.

b. The 95% confidence level using Tukey's pairwise comparison procedure is

$$\left(\overline{y}_i - \overline{y}_j\right) \pm \frac{q_\gamma(J, N-J)}{\sqrt{2}} \cdot s_{pooled} \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} = (3.3 - 2.4) \pm \frac{q_{0.05}(3,12)}{\sqrt{2}} \cdot \sqrt{0.2} \cdot \sqrt{\frac{1}{5} + \frac{1}{5}}$$
$$= 0.9 \pm \frac{3.77}{\sqrt{2}} \cdot \sqrt{0.2} \cdot \sqrt{\frac{1}{5} + \frac{1}{5}}$$
$$= 0.9 \pm 0.75$$
$$= (0.15, 1.65)$$

With this method, we are 95% confident that the mean difference between the average GPAs of Football and Soccer players is between 0.15 and 1.65.

c. The 95% confidence level using Scheffe's multiple comparison procedure is

$$\sum_{j=1}^{J} c_j \bar{y}_j \pm \sqrt{F_\alpha(J-1, N-J)} \cdot \sqrt{\text{MSW}} \cdot \sqrt{(J-1)\sum_{j=1}^{J} \frac{c_j^2}{n_j}} = (3.3-2.4) \pm \sqrt{F_\alpha(2,12)} \cdot \sqrt{0.2} \cdot \sqrt{2 \cdot \left(0 + \frac{1}{5} + \frac{1}{5}\right)}$$

$$= 0.9 \pm \sqrt{3.89} \cdot \sqrt{0.2} \cdot \sqrt{2 \cdot \left(0 + \frac{1}{5} + \frac{1}{5}\right)}$$

$$= 0.9 \pm 0.79$$

$$= (0.11, 1.69)$$

where $c_B = 0$, $c_F = 1$, and $c_S = -1$.

With this method, we are 95% confident that the mean difference between the average GPAs of Football and Soccer players is between 0.11 and 1.69.

d. The 95% confidence level using Scheffe's multiple comparison procedure is

$$\sum_{j=1}^{J} c_j \bar{y}_j \pm \sqrt{F_\alpha(J-1, N-J)} \cdot \sqrt{\text{MSW}} \cdot \sqrt{(J-1)\sum_{j=1}^{J} \frac{c_j^2}{n_j}}$$

$$= \left(\frac{3.0+3.3}{2} - 2.4\right) \pm \sqrt{F_\alpha(2,12)} \cdot \sqrt{0.2} \cdot \sqrt{2 \cdot \left(\frac{1}{20} + \frac{1}{20} + \frac{1}{5}\right)}$$

$$= 0.75 \pm \sqrt{3.89} \cdot \sqrt{0.2} \cdot \sqrt{2 \cdot \left(\frac{1}{20} + \frac{1}{20} + \frac{1}{5}\right)}$$

$$= 0.75 \pm 0.68$$

$$= (0.07, 1.43)$$

where $c_B = ½$, $c_F = ½$, and $c_S = -1$.

With this method, we are 95% confident that the difference between the average GPAs of Baseball and Football compared to the Soccer players is between 0.07 and 1.43.

e. First order the GPAs and rank them.

Sport	S	S	B	S	S	F	B	B	B	F	S	F	F	B	F
GPA	1.9	2.0	2.3	2.4	2.6	2.8	2.9	3.1	3.1	3.1	3.1	3.3	3.5	3.6	3.8
Rank	1	2	3	4	5	6	7	9.5	9.5	9.5	9.5	12	13	14	15

Then calculate the rank mean for each group.

$$\bar{r}_B = \frac{3+7+9.5+9.5+14}{5} = 8.6. \qquad \bar{r}_F = \frac{6+9.5+12+13+15}{5} = 11.1.$$

$$\bar{r}_S = \frac{1+2+4+5+9.5}{5} = 4.3. \qquad \bar{r} = \frac{N+1}{2} = 8.$$

Test Statistic:

$$\text{K} = \frac{12}{15 \cdot 16}\left[5 \cdot (8.6-8)^2 + 5 \cdot (11.1-8)^2 + 5 \cdot (4.3-8)^2\right] = \mathbf{5.915}.$$

Critical Value: $\chi^2_{\alpha}(J-1) = \chi^2_{0.10}(2) = \mathbf{4.605}$.

Since the test statistic does lie in the critical region (5.915 > 4.605), we reject H_0 and conclude that the model does a significant job of predicting GPA. This is the same result as the “parametric” test back in part a.

2. Product Data (without R)

For this exercise you are not to use R or any other software to solve the exercises. A calculator is allowed.

In order to rate three brands of a particular product, a consumer agency divided eighteen individuals at random into three groups and asked each one of them to rate one brand of the product on the scale from 0 to 100.

Brand							Mean	Variance
1	66	72	77	81	87	85	$\bar{y}_1 = 78$	$s_1^2 = 64$
2	83	73	69	77	67	87	$\bar{y}_2 = 76$	$s_2^2 = 62$
3	85	74	85	88	89	95	$\bar{y}_3 = 86$	$s_3^2 = 48$

a. Test H_0: $\mu_1 = \mu_2 = \mu_3$ with $\alpha = 0.10$.

$J = 3.$

$N = n_1 + n_2 + \ldots + n_J = 6 + 6 + 6 = 18.$

$$\bar{y} = \frac{n_1 \cdot \bar{y}_1 + n_2 \cdot \bar{y}_2 + \ldots + n_J \cdot \bar{y}_J}{N} = \frac{6 \cdot 78 + 6 \cdot 76 + 6 \cdot 86}{18} = 80.$$

$$\text{SSB} = n_1 \cdot \left(\bar{y}_1 - \bar{y}\right)^2 + n_2 \cdot \left(\bar{y}_2 - \bar{y}\right)^2 + \ldots + n_J \cdot \left(\bar{y}_J - \bar{y}\right)^2$$

$$= 6 \cdot (78 - 80)^2 + 6 \cdot (76 - 80)^2 + 6 \cdot (86 - 80)^2 = 336.$$

$$\text{SSW} = \left(n_1 - 1\right) \cdot s_1^2 + \left(n_2 - 1\right) \cdot s_2^2 + \ldots + \left(n_J - 1\right) \cdot s_J^2$$

$$= 5 \cdot 64 + 5 \cdot 62 + 5 \cdot 48 = 870.$$

$$\text{SSTotal} = \text{SSB} + \text{SSW} = 336 + 870 = 1206.$$

Completing the ANOVA table,

Source	SS	df	MS	F
Between Groups	336	$J - 1 = 2$	168	**2.90**
Within Groups	870	$N - J = 15$	58	
Total	1206	$N - 1 = 17$		

According to the F-distribution, the critical region is $F > F_{0.05}(2,15) = \mathbf{3.68}$. Since the test statistic does not lie in the critical region, we fail to reject H_0 and conclude that the average ratings of the three brands are not significantly different.

b. The 90% confidence level using Tukey's pairwise comparison procedure is

$$\left(\bar{y}_i - \bar{y}_j\right) \pm \frac{q_\gamma(J, N-J)}{\sqrt{2}} \cdot s_{pooled} \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} = (86 - 76) \pm \frac{q_{0.10}(3,15)}{\sqrt{2}} \cdot \sqrt{58} \cdot \sqrt{\frac{1}{6} + \frac{1}{6}}$$

$$= 10 \pm \frac{3.14}{\sqrt{2}} \cdot \sqrt{58} \cdot \sqrt{\frac{1}{6} + \frac{1}{6}}$$

$$= 10 \pm 9.76$$

$$= (0.24, 19.76)$$

With this method, we are 90% confident that the mean difference between the average rating for Brand 3 and the average rating for Brand 2 is between 0.24 and 19.76.

c. The 90% confidence level using Scheffe's multiple comparison procedure is

$$\sum_{j=1}^{J} c_j \bar{y}_j \pm \sqrt{F_\alpha(J-1, N-J)} \cdot \sqrt{\text{MSW}} \cdot \sqrt{(J-1)\sum_{j=1}^{J}\frac{c_j^2}{n_j}}$$

$$= \left(86 - \frac{78+76}{2}\right) \pm \sqrt{F_\alpha(2,15)} \cdot \sqrt{58} \cdot \sqrt{2 \cdot \left(\frac{1}{24} + \frac{1}{24} + \frac{1}{6}\right)}$$

$$= 9 \pm \sqrt{2.48} \cdot \sqrt{58} \cdot \sqrt{2 \cdot \left(\frac{1}{24} + \frac{1}{24} + \frac{1}{6}\right)}$$

$$= 9 \pm 8.48$$

$$= (0.52, 17.48)$$

where $c_1 = -½$, $c_2 = -½$, and $c_3 = 1$.

With this method, we are 95% confident that the difference between the average rating for Brand 3 as compared to the average rating for Brand 1 and Brand 2 is between 0.52 and 17.48.

d. First order the brand ratings and rank them.

Brand	1	2	2	1	2	3	1	2	1	1
Rating	66	67	69	72	73	74	77	77	81	66
Rank	1	2	3	4	5	6	7.5	7.5	9	1

Brand	2	1	3	3	1	2	3	3	3	2
Rating	83	85	85	85	87	87	88	89	95	83
Rank	10	12	12	12	14.5	14.5	16	17	18	10

Then calculate the rank mean for each group.

$\bar{r}_1 = \frac{48}{6} = 8.0$ $\quad\bar{r}_2 = \frac{42}{6} = 7.0$ $\quad\bar{r}_3 = \frac{81}{6} = 13.5$ $\quad\bar{r} = \frac{N+1}{2} = 9.5$

Test Statistic:

$$\text{K} = \frac{12}{18 \cdot 19} \cdot \left[6 \cdot (8-9.5)^2 + 6 \cdot (7-9.5)^2 + 6 \cdot (13.5-9.5)^2\right] = \mathbf{5.158}$$

Critical Value: $\quad\chi^2_\alpha(J-1) = \chi^2_{0.10}(2) = \mathbf{4.605}$.

Since the test statistic does lie in the critical region (5.158 > 4.605), we reject H_0 and conclude that the model does a significant job of product rating.