Homework 8

8.1 Full Factorial Design

An expanded version of the dice quality assurance example introduced in class considers the effect of two controlled factors:

- Dice color: red, blue, green, or purple
- Rolling hand: right or left

on the outcome variable measuring the sum of two dice.

a) List required experimental trials for a full factorial design.

Solution:

Full Factorial Design for Dice & Rolling Hand

Trial	Run	Dice Color	Hand
1	2	Red	Right
2	5	Red	Left
3	7	Blue	Right
4	1	Blue	Left
5	8	Green	Right
6	3	Green	Left
7	6	Purple	Right
8	4	Purple	Left

8.2 Video Game Sales

The attached file vgsales.csv contains a list of video game titles released between 2008 and 2010 having greater than 100,000 sales on Sony PlayStation 3 (PS3), Nintendo Wii, or Microsoft Xbox 360 (X360) platforms. Organize the data in a format suitable for a one-way ANOVA to investigate whether the **platform** is a significant factor for **global sales.**

a) Calculate the sample mean global sales (GA Sales) $\overline{X}_j$ for each platform.

Solution:

Platform	Sum of Global_Sales	Count of Global_Sales	Average
PS3	396.45	482	0.8225
Wii	516.4	861	0.5998
X360	427.66	501	0.8536

Data is gathered using Pivot table.

b) Calculate the following values for a one-way ANOVA:

i. Sum of squares of columns (SSC)

Solution:

Sum square of columns: 26.37

ii. Mean square of columns (MSC)

Solution:

Mean square of columns: 13.18

iii. Sum of squares of error (SSE)

Solution:

Sum square of error: 6796.01

iv. Mean square of error (MSE)

Solution:

Mean square of error: 3.69

c) Calculate the F statistic and p-value for the following hypothesis test:

H_0: all platforms have equal mean global sales

H_a: at least one platform differs from the others.

Discuss the results of the hypothesis test (what can be concluded?).

Solution:

P-value: 0.0283 or 2.83%

P value is very small so we can conclude that all platforms have UNEQUAL global sale. We can reject the H_0 hypothesis.

8.3 Baseball Salary Analysis

The attached file Salaries.csv contains a list of the top 200 salaries for baseball players in the American League (AL) and National League (NL) during the 2013-2016 seasons. Organize the data in a format suitable for a two-way ANOVA to investigate whether the year and/or league are significant factors for salaries controlling for interaction effects.

a) Calculate the mean salary $\bar{X}i$ for each year.

Solution:

Year	Sum of salary	Count of salary	Average of salary
2013	2,782,949,864	400	6,957,374.66
2014	2,926,378,857	400	7,315,947.14
2015	3,202,569,748	400	8,006,424.37
2016	3,429,054,931	400	8,572,637.33

b) Calculate the mean salary $\bar{X}j$ for each league.

Solution:

League Name	Sum of salary	Count of salary	Average of salary
AL	6,331,513,237.00	800.00	7,914,391.55
NL	6,009,440,163.00	800.00	7,511,800.20

c) Calculate the F statistic and p-value for the following hypothesis test:

H_0: the mean salary for each year are all equal

H_a: at least one year's mean salary differs from the others

Discuss the results of the hypothesis test (what can be concluded?).

Solution:

The p-value this hypothesis is 0.0004 or 0.04% which is quite low. We can reject the H_0 hypothesis.

d) Calculate the F statistic and p-value for the following hypothesis test:

H_0: the mean salary for both leagues are all equal

H_a: at least one league's mean salary differs from the others

Discuss the results of the hypothesis test (what can be concluded?).

Solution:

The p-value this hypothesis is 0.168 or 16.8% which is acceptable. We can accept the H_0 hypothesis.

e) Calculate the F statistic and p-value for the following hypothesis test:

H_0: all interactions between year and league are zero

H_a: there is at least one nonzero interaction between year and league

Discuss the results of the hypothesis test (what can be concluded?).

Solution:

The p-value this hypothesis is 0.213 or 21.3% which is acceptable. We can accept the H_0 hypothesis.