Stevens Institute of Technology

SYS-601 Homework #8

Due Apr. 9 2018

Submit the following using the online submission system: 1) Completed assignment cover sheet, 2) Written responses in PDF format, 3) All saved models (e.g. .xlsx or .py files).

8.1 Full Factorial Design [2 points]

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

8.2 Video Game Sales [7 points]

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.[1] 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) 1 pt Calculate the sample mean global sales (`GA_Sales`) $\bar{X}_j$ for each platform.

(b) 3 pts Calculate the following values for a one-way ANOVA:

(i) Sum of squares of columns (SSC)
(ii) Mean square of columns (MSC)
(iii) Sum of squares of error (SSE)
(iv) Mean square of error (MSE)

(c) 3 pt 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?).

[1] Source: GregorySmith at `https://www.kaggle.com/gregorut/videogamesales`

8.3 Baseball Salary Analysis [11 points]

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.[2] 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) 1 PT Calculate the mean salary $\bar{X}_i$ for each year.

(b) 1 PT Calculate the mean salary $\bar{X}_j$ for each league.

(c) 3 PTS 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?).

(d) 3 PTS 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?).

(e) 3 PTS 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?).

[2] Source: Sean Lahman at `http://www.seanlahman.com/baseball-archive/statistics/`