University of Illinois at Chicago

Department of Statistics

Bowling Analysis

Author:
Yvonne Ho

A project report submitted for the course of

STAT-481 Statistical Methods II

May 1, 2019

Contents

1 Introduction **2**

2 Data Exploration **2**
2.1 Response Variable . . . 2
2.2 Explanatory Variables . . . 2

3 Methods **3**
3.1 Fixed Effects Model . . . 3
3.1.1 Assumptions Check . . . 3
3.1.2 Testing for Main Effects . . . 4
3.2 Multiple Comparisons of Pairwise Differences . . . 4
3.2.1 Bonferroni's Test . . . 4
3.2.2 Tukey's Test . . . 5
3.3 Random Effects Model . . . 5
3.3.1 Assumptions Check . . . 6
3.3.2 Estimates of Variances . . . 7

4 Conclusion **7**

5 SAS Code **8**

1 Introduction

The Professional Bowlers Association (PBA) collects bowling scores from female professional bowlers with the PBA memberships.This statistical approach is used to investigate different averages and variabilities of their bowling scores among the three bowlers.

2 Data Exploration

2.1 Response Variable

The first variable we want to investigate is *Bowling Score*:

Table 2.1: Summary Statistics for Bowling Score

	N	Mean	Std Deviation	Median	Min	Max	Skewness	Kurtosis
Bowling Score	168	204.2380952	28.0718151	201.5	126	276	0.23242279	-0.2058985

The bowling score had a maximum of 276 points and a minimum of 126 points. The average of the score is 204.2380952 points and the median is 201.5 points. The score is slightly right skewed. According to Figure **??** and Table 2.2 below, the histogram, Q-Q plot, and Shapiro-Wilk test show that bowling score is normally distributed.

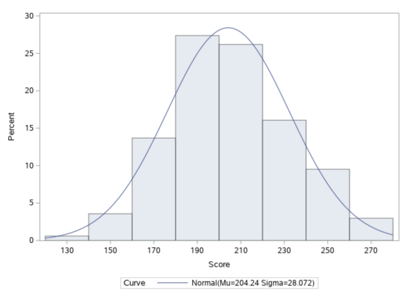

Figure 2.1: Histogram

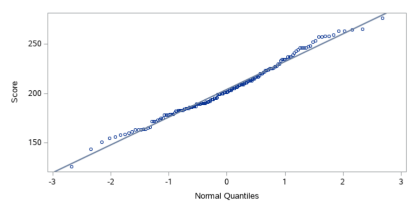

Figure 2.2: Q-Q Plot

Table 2.2: Normality Tests

Test	Statistic		p Value	
Shapiro-Wilk	W	0.988233	Pr < W	0.1739
Kolmogorov-Smirnov	D	0.057527	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.113621	Pr > W-Sq	0.0774
Anderson-Darling	A-Sq	0.698606	Pr > A-Sq	0.0708

2.2 Explanatory Variables

Here are two explanatory variable is collected:

1. **Game Number**: The number of games is counted by real positive integers from 1 to 56.
2. **Bowler**: Three of fifteen bowlers are selected, who are labeled in numbers.
   - Wendy MacPherson=7
   - Shannon Pluhowski=8
   - Aumi Guerra=12

Table 2.3: Summary Statistics for Bowlers in Bowling Scores

	N	Mean	Std Deviation	Median	Min	Max	Skewness	Kurtosis	Sum
Bowler 7	56	209.535714	29.6941771	206.5	156	276	0.24020276	-0.720802	11734
Bowler 8	56	208.4285714	25.8652854	208	144	264	0.0763187	0.20628042	11672
Bowler 12	56	194.75	26.5304285	190.5	126	263	0.35220397	0.4333532	10906

Bowler is an indicator variable. The selected bowlers are assigned and labeled as a specific number, which are seven, eight, and twelve.Each bowlers fairly played 56 times.

Table 2.3 simply summarizes each bowler's performance. Wendy got the highest score in a total of 11734 points; Shannon had total 11672 points, and Aumi got 10906 points in total. Wendy had the highest average score, but Aumi had the lowest. The median and mean of Shannon's score are close, and her score is less skewed to the right, so Shannon is the best player who has the most stable performance in bowling game.

3 Methods

3.1 Fixed Effects Model

In a bowling game, bowling scores are affected by different bowlers, who can be regarded as a treatment effect. Thus, the model statement is constructed in fixed effects model to investigate different bowling scores on average among the three bowlers, which is shown below:

$$\begin{aligned} \text{Bowling Score}_{ij} &= \mu_i + \varepsilon_{ij} \\ &= \mu + \tau_i + \varepsilon_{ij}, \begin{cases} i = 1, 2, 3 \\ j = 1, 2, \ldots, 56 \\ \varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2) \end{cases} \end{aligned}$$

- Bowling score is the value of the response variable in the j^{th} trial / replicate for the i^{th} treatment
- Bowlers are the only one factor and divided into i levels of treatment, denoted by τ_i
- μ_i is the i^{th} treatment mean
- ε_{ij} is the error of the j^{th} trial / replicate at the i^{th} treatment
- The number of treatments (k) is 3
- The number of observations per treatment (n_i) is 56 $\Rightarrow$ This study is a balanced design.
- $N = \sum_{i=1}^{k} n_i = 168$ is the total number of observations
- $\sum_{i=1}^{k} n_i \tau_i = 0$ is a constraint in this model $\Rightarrow \sum_{i=1}^{k} \tau_i = 0$ in a balanced design

3.1.1 Assumptions Check

3.1.1.1 Normality of Errors

The Q-Q plot and Shapiro-Wilk test are one of the methods to determine whether residual are normally distributed.

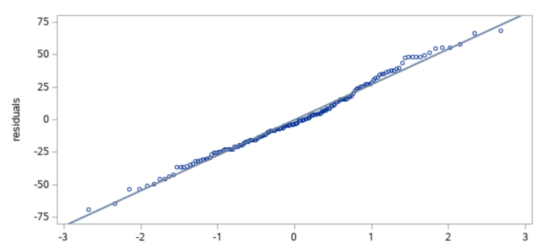

Figure 3.1: Q-Q Plot

Table 3.1: Tests for Normality

Test	Statistic		p Value	
Shapiro-Wilk	W	0.989136	Pr < W	0.2247
Kolmogorov-Smirnov	D	0.064338	Pr > D	0.0878
Cramer-von Mises	W-Sq	0.115235	Pr > W-Sq	0.0737
Anderson-Darling	A-Sq	0.661633	Pr > A-Sq	0.0861

The Shapiro-Wilk test assumes that residual is normal distributed, where is denoted by a null hypothesis (H_0) below.

$$\text{H}_0 : e_i \sim \text{Normal} \quad \textit{versus} \quad \text{H}_1 : e_i \nsim \text{Normal}$$

Table 3.1 shows that its test statistic is equal to 0.989136 with the p-value of 0.2247, where is greater than the significance level (α)at 0.05. The Q-Q plot also shows that most of residual are on the straight line in Figure 3.1. Therefore, H_0 is fail to reject, and the assumption of errors' normality is met.

3.1.1.2 Equal Variance of Errors

Levene's test and the plot of Residual vs $\widehat{\text{Bowling Scores}}_{ij}$ can examine whether residual has equal variance. The Levene's test assumes that there is equal variance among residual. Here are the hypotheses:

$$\text{H}_0 : \sigma_i^2 = 0, i = 1, 2, \ldots, k \quad \textit{versus} \quad \text{H}_1\text{: at least one } \sigma_i^2 \text{ is different}$$

Table 3.2 shows that test statistics is equal to 0.71 with p-value of 0.4935, which is greater than α at 0.05. It is difficult to recognize patterns in Figure 3.1. Hence, H_0 is fail to reject, and the assumption of equal variance in residual is met.

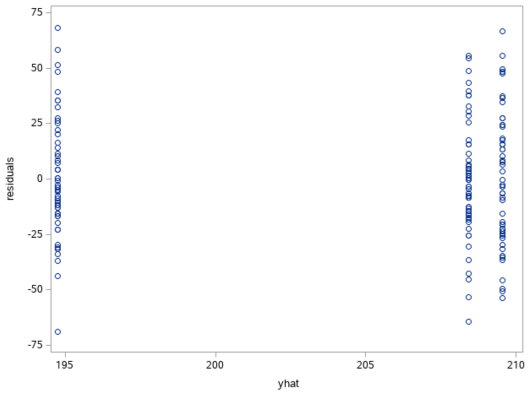

Figure 3.2: Residual Vs $\widehat{\text{Bowling Scores}}_{ij}$

Table 3.2: Levene's Test

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Bowler	2	1406456	703228	0.71	0.4935
Error	165	1.6359E8	991483		

3.1.2 Testing for Main Effects

The ANOVA table is used to conclude whether all treatment means and effects are the same. The assumptions are made, based on the hypotheses below.

$$\text{Mean Model: } \text{H}_0 : \mu_1 = \cdots = \mu_k \quad \textit{versus} \quad \text{H}_1\text{: At least one mean } \mu_i \text{ is different}$$
$$\text{Effects Model: } \text{H}_0 : \tau_1 = \cdots = \tau_k = 0 \quad \textit{versus} \quad \text{H}_1\text{: At least one treatment effect } \tau_i \neq 0$$

Table 3.3: ANOVA Table

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	7596.3333	3798.1667	5.05	0.0074
Error	165	124004.1429	751.5403		
Corrected Total	167	131600.4762			

Table 3.3 displays a test statistic in a value of 5.05 with p-value of 0.0074, which is smaller than α at 0.05. Consequently, H_0 is rejected, and at least one treatment means and effects are different. In other words, the averages of bowling scores, corresponding to the relevant bowlers, is different.

3.2 Multiple Comparisons of Pairwise Differences

Since each bowler has different performance than the others, we want to have a further investigation in which bowler is different from each other via Bonferroni's and Tukey's tests. We prefer to use Bonferroni's test first and then Tukey's test because we know there are differences among the bowlers in the data and Tukey's test is used to double check the results from Bonferroni's test.

3.2.1 Bonferroni's Test

Bonferroni's test inspects all pairwise comparisons by developing confidence intervals in advance of the study. It also controls the familywise error (α). Since there are less than 50 comparisons, α^* will not be too low to reject and find a significant difference between the averages of bowling scores from two of the three bowlers.

$$\mathrm{H}_0 : \mu_i = \mu_j \text{for any pair} \, (i, j) \quad \textit{versus} \quad \mathrm{H}_1 : \text{at least one pair is not the same}$$

Figure 3.4b lists the differences between bowlers in a pair. As a result, only Wendy and Shannon do not have differences in their bowling scores on average. Therefore, H_0 is rejected, and at least one pair is not the same.

Table 3.4: Bonferroni Tests

Alpha	0.05
Error Degrees of Freedom	165
Error Mean Square	751.5403
Critical Value of t	2.41863
Minimum Significant Difference	12.53

(a)

Comparisons significant at the 0.05 level are indicated by ***.				
Bowler Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
7 - 8	1.107	-11.423	13.638	
7 - 12	14.786	2.255	27.316	***
8 - 7	-1.107	-13.638	11.423	
8 - 12	13.679	1.148	26.209	***
12 - 7	-14.786	-27.316	-2.255	***
12 - 8	-13.679	-26.209	-1.148	***

(b) Confidence Intervals

3.2.2 Tukey's Test

Tukey's test can investigate differences in treatment means when the study has a balanced design.This procedure helps to control the familywise error, α. There are $\binom{k}{2} = \frac{k(k-1)}{2} = \frac{3(3-1)}{2} = 3$ pairwise comparisons.

$$\mathrm{H}_0 : \mu_i = \mu_j \text{for any pair} \, (i, j) \quad \textit{versus} \quad \mathrm{H}_1 : \text{at least one pair is not the same}$$

Table 3.5: Tukey's Tests

Alpha	0.05
Error Degrees of Freedom	165
Error Mean Square	751.5403
Critical Value of Studentized Range	3.34469
Minimum Significant Difference	12.253

(a)

Comparisons significant at the 0.05 level are indicated by ***.				
Bowler Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
7 - 8	1.107	-11.146	13.360	
7 - 12	14.786	2.533	27.039	***
8 - 7	-1.107	-13.360	11.146	
8 - 12	13.679	1.426	25.931	***
12 - 7	-14.786	-27.039	-2.533	***
12 - 8	-13.679	-25.931	-1.426	***

(b) Confidence Intervals

Figure 3.5a shows the same result as Figure 3.4b in Bonferroni test. Consequently, there are no differences between Wendy's and Shannon's bowling scores on average, which indicates H_0 is rejected, and at least one pair is different.

3.3 Random Effects Model

Random effects model is similar to the fixed effects model. However, this model is focus on comparing with the variability of bowling scores in each level of treatment. The random effects model is shown below:

$$\text{Bowling Score}_{ij} = \mu + \tau_i + \varepsilon_{ij}, \begin{cases} i = 1, 2, 3 \\ j = 1, 2, \ldots, 56 \\ \left.\begin{array}{l} \varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2) \\ \tau_i \overset{\text{iid}}{\sim} N(0, \sigma_\tau^2) \end{array}\right\}\text{independent} \end{cases}$$

3.3.1 Assumptions Check

3.3.1.1 Normality of Errors

The Q-Q plot and Shapiro-Wilk test check if residual is normal distributed.

$$\mathrm{H}_0 : e_i \sim \text{Normal} \quad \textit{versus} \quad \mathrm{H}_1 : e_i \nsim \text{Normal}$$

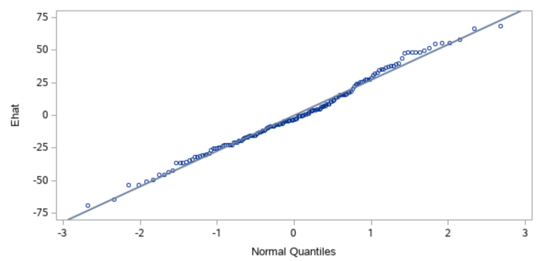

Figure 3.3: Q-Q Plot

Test	Statistic		p Value	
Shapiro-Wilk	W	0.989136	Pr < W	0.2247
Kolmogorov-Smirnov	D	0.064338	Pr > D	0.0878
Cramer-von Mises	W-Sq	0.115235	Pr > W-Sq	0.0737
Anderson-Darling	A-Sq	0.661633	Pr > A-Sq	0.0861

Table 3.6: Tests for Normality

The Shapiro-Wilk test in random effects model is the same as in the fixed effects model, according to Figure 3.6 and Figure 3.1. The Q-Q plot in Figure 3.3 also reflects that residual forms a straight line. These show that the residual is normally distributed.

3.3.1.2 Equal Variance of Errors

The plot of Residual Vs $\widehat{\text{Bowling Scores}}_{ij}$ and Levene's test determine whether residual has homogenous variance.

$$\mathrm{H}_0 : \sigma_i^2 = 0, i = 1, 2, \ldots, k \quad \textit{versus} \quad \mathrm{H}_1\text{: at least one } \sigma_i^2 \text{ is different}$$

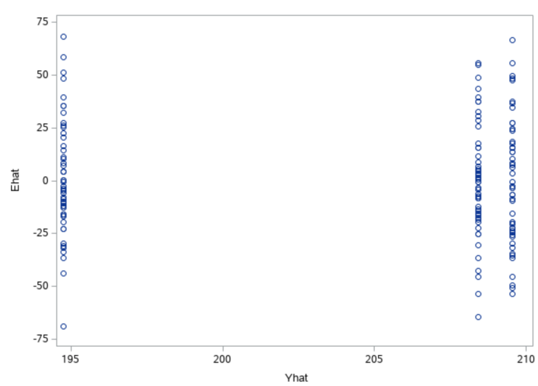

Figure 3.4: Residual Vs $\widehat{\text{Bowling Scores}}_{ij}$

Table 3.7: Levene's Test

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Bowler	2	1406456	703228	0.71	0.4935
Error	165	1.6359E8	991483		

Table 3.7 and Table 3.2 show that random and fixed effects models got the same result in Levene's test. Additionally, the plot in Figure 3.4 does not show any patterns. Thus, the assumption of homogenous variance in residual is met.

3.3.1.3 Testing for Random Effects

The ANOVA table in random effects model analyzes whether the variability of bowling scores exists.

$$\mathrm{H}_0 : \sigma_\tau^2 = 0, i = 1, 2, \ldots, k \quad \textit{versus} \quad \mathrm{H}_1 : \sigma_\tau^2 > 0$$

Table 3.8: ANOVA Table

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	7596.3333	3798.1667	5.05	0.0074
Error	165	124004.1429	751.5403		
Corrected Total	167	131600.4762			

Table 3.8 and Table 3.3 are the same but in different models. In Table 3.8, test statistic is 5.05 with p-value of 0.0074, which is smaller than α at 0.05. Hence, H_0 is rejected, and the variability of bowling scores among the bowlers is greater than zero. Additionally, the magnitude of the variability reflects how different the scores are from game to game or how bowlers' scores differ from each other.

3.3.2 Estimates of Variances

The variability of bowling scores is significant, which is calculated below:

$$\begin{aligned}\widehat{\sigma}_\tau^2 &= \frac{1}{\tilde{n}}\left(MSTR - MSE\right) \\ &= \frac{1}{\frac{1}{k-1}\left(N - \frac{\sum_{i=1}^{k}\left(n_i^2\right)}{\sum_{i=1}^{k} n_i}\right)}\left(MSTR - MSE\right) \\ &= \frac{(k-1)\left(MSTR - MSE\right)}{\left(N - \frac{\sum_{i=1}^{k}\left(n_i^2\right)}{\sum_{i=1}^{k} n_i}\right)} \\ &= \frac{(3-1)\left(3798.1667 - 751.5403\right)}{\left(168 - \frac{56^2\times 3}{168}\right)} \\ &= \frac{223383}{4106} \\ &\approx 54.40404286\end{aligned} \tag{1}$$

$$\widehat{\sigma}^2 = MSE = 751.5403 \tag{2}$$

Equation (1) describes how different the bowlers' performances are. The larger $\widehat{\sigma}_\tau^2$ is, the larger differences among the bowling scores from the bowlers.

Equation (2) states how different all the bowling scores are among the bowlers.

4 Conclusion

The data shows that different bowlers have different performances in bowling games, based on main effects tests. The average bowling scores among the bowlers evaluate their performances. Equal variance is assumed, so it suggests that all bowlers have the same "stableness". However, it is only found that Wendy and Shannon do not show differences in average score when they are compared in a group. The ANOVA table also provides evidence that the bowler is significant in both the fixed and random models. The estimated variances reflect how large different performances among the bowlers are. These is evidence bowlers are not consistent in their scores from game to game.

5 SAS Code

```
FILENAME REFFILE '/folders/myfolders/Project_2/P2_Dataset2.xlsx';

PROC IMPORT DATAFILE=REFFILE
        DBMS=XLSX
        OUT=Bowling;
        GETNAMES=YES;
RUN;

PROC CONTENTS DATA=Bowling; RUN;

PROC MEANS DATA=Bowling NMISS N; RUN;

PROC MEANS DATA=Bowling;
VAR Score Bowler GameNum;
RUN;

PROC MEANS DATA=Bowling;
CLASS Bowler;
VAR Score GameNum;
RUN;

PROC MEANS DATA=Bowling sum;
CLASS Bowler;
VAR Score GameNum;
RUN;

PROC UNIVARIATE DATA=Bowling Normal Plot;
        Var Score;
        qqplot Score;
        histogram Score / normal;
RUN;

PROC UNIVARIATE DATA=Bowling Normal Plot;
        CLASS Bowler;
        Var Score GameNum;
        qqplot Score;
        histogram Score / normal;
RUN;

/* Fixed Effects*/
PROC GLM DATA=Bowling;
CLASS Bowler;
MODEL Score = Bowler;
MEANS Bowler/ LSD BON TUKEY CLDIFF HOVTEST=LEVENE WELCH;
OUTPUT OUT=result1 R=residuals P=y_hat;
RUN;

PROC UNIVARIATE DATA=result1 NORMAL PLOT;
VAR residuals;
RUN;

PROC SGPLOT DATA=result1;
 scatter x=y_hat y=residuals;
RUN;
```

```
/* Random Effects*/
PROC GLM DATA = Bowling;
        CLASS Bowler;
        MODEL Score = Bowler;
        RANDOM Bowler ;
        MEANS Bowler/hovtest = levene;
        OUTPUT OUT=ResPred P=Yhat R=Ehat;
        TITLE 'Results from Proc GLM';
RUN;

PROC UNIVARIATE DATA=ResPred NORMAL PLOT;
VAR Ehat;
RUN;

PROC SGPLOT DATA=ResPred;
 scatter x=Yhat y=Ehat;
RUN;

PROC MIXED DATA = Bowling;
        CLASS Bowler;
        MODEL Score=;
        RANDOM Bowler;
        TITLE 'Results from Proc Mixed';
        RUN;
```