STAT 230 REVISED PROJECT

STAT 290

How To Make a Free Throw Shooter Miss

Author:
Kevin Toney, Chase Weight, Nick Mena

May 2, 2017

Contents

1 Abstract

In this study, we performed an experiment to determine what factors significantly influence the amount of free throws made by a basketball player. We decided on this experiment because we are basketball fans that regularly attend games. We were wondering how much our crowd noise and ball inflation levels mattered in a player's shooting performance.

The factors of interest were whether there were large amounts of noise in the room or not, and whether the basketball was fully inflated or not. Therefore, this experiment was a 2^2 factorial design. We hypothesized that none of these factors would make a significant impact.

We invited a group of 20 college students to participate in our study. These students varied in gender and skill, but we used a poisson distribution to account for each shooter independently. During the experiment, we assigned the students to one of four groups. One group had deflated balls with no noise, another had deflated balls with lots of noise. Then, the last two groups used inflated balls with silence, or inflated balls with lots of noise. Each student shot 20 free throws and recorded the proportion of made shots.

After gathering data, we performed poisson regression. At the significance level of alpha=0.05, we found shooting free throws with a different level of inflation made the only statistically significant difference. As a matter of fact, shooting with deflated balls improved the free throw percentage of a player.

The experiment is practically significant because a team can be assured that crowd noise, alone, won't significantly impact their players at the free throw line. Other factors, which can be studied more, such as the player's emotional state, visual distractions, fatigue, or confidence level may affect their performance. Therefore, we would recommend that basketball players practice regularly on their free throw shooting and not worry about the crowd's noise.

Now we can answer our question "What factors significantly influence a players free throw percentage?" We conclude the inflation level of the ball, not the noise level of the crowd, makes a significant impact on free throw shooting.

2   Introduction

In this study, we performed an experiment to determine what factors significantly influence the proportion of free throws made by a basketball player.We decided on performing this experiment because we are huge basketball fans, and wanted to know if our presence made a difference in the outcome of a game.

We were wondering how much a fan's noise mattered in the outcome of the game. Additionally, we wondered if the amount of air in a ball would make a difference. The reason why we were interested in the inflation level is because a scandal with a famous football player, named Tom Brady, accused the player of deflating footballs so he would have more control over the ball he threw. Our original hypothesis was that none of these factors made a significant impact.

3   Experimental Design

We randomly split up 20 students into four groups and they had an objective to shoot 20 free throws. Each group reflected a combination of the factors of interest. For your information, these students varied in gender and skill, but we used a poisson distribution to account for each shooter independently.

The factors of interest (main effects) were whether there were large amounts of noise or not, and whether the basketball is fully inflated or not. For a table of these factors, see the figure below. We chose to test only these two factors out of the many that would influence a player's shooting. For example, a player might struggle shooting free throws because they are tired, sick, or angry. We ignored those factors for this experiment.

Inflated/No Noise	Inflated/Noise	Deflated/No Noise	Deflated/Noise
Shooter	""	""	""
""	""	""	""
""	""	""	""
""	""	""	""
""	""	""	""

The reponse variable has a unit of makes in 20 attempts. The variable is measured in counts, thus supporting our chosen method of analysis.

We decided to run a poisson regression for analysis. The data has a sample size of 20 for each of the 20 shooters. That sample size is enough for a poisson distribution and

a poisson regression model. Each of the five people in a group will have a p-value. The central limit theorem says that for each group, 5*20 shots will help estimate p-values.

4 Results

Here is the data we gathered for each of the 20 shooters.

```
freethrows <- read.table("Free Throws.csv", header=T, sep=",")
freethrows$Response <- freethrows$Response*20
```

```
summary(freethrows$Response)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.00    5.75    9.50    8.80   11.00   17.00
```

| Group	Response	Silence	Inflated
1	3.00	1	1
1	6.00	1	1
1	5.00	1	1
1	7.00	1	1
1	6.00	1	1
2	16.00	1	0
2	5.00	1	0
2	12.00	1	0
2	17.00	1	0
2	10.00	1	0
3	8.00	0	1
3	9.00	0	1
3	11.00	0	1
3	10.00	0	1
3	2.00	0	1
4	12.00	0	0
4	10.00	0	0
4	11.00	0	0
4	11.00	0	0
4	5.00	0	0

Table 1: The response variable is the amount of free throw makes in 20 attempts. In the 'Silence' and 'Inflated' columns, a one means 'Yes' and a zero means 'No'.

According to our results, group two had the highest free throw percentage. This group shot in silence and used a underinflated ball.

5   Analysis

Here are the boxplots showing how each group's data relates to each other.
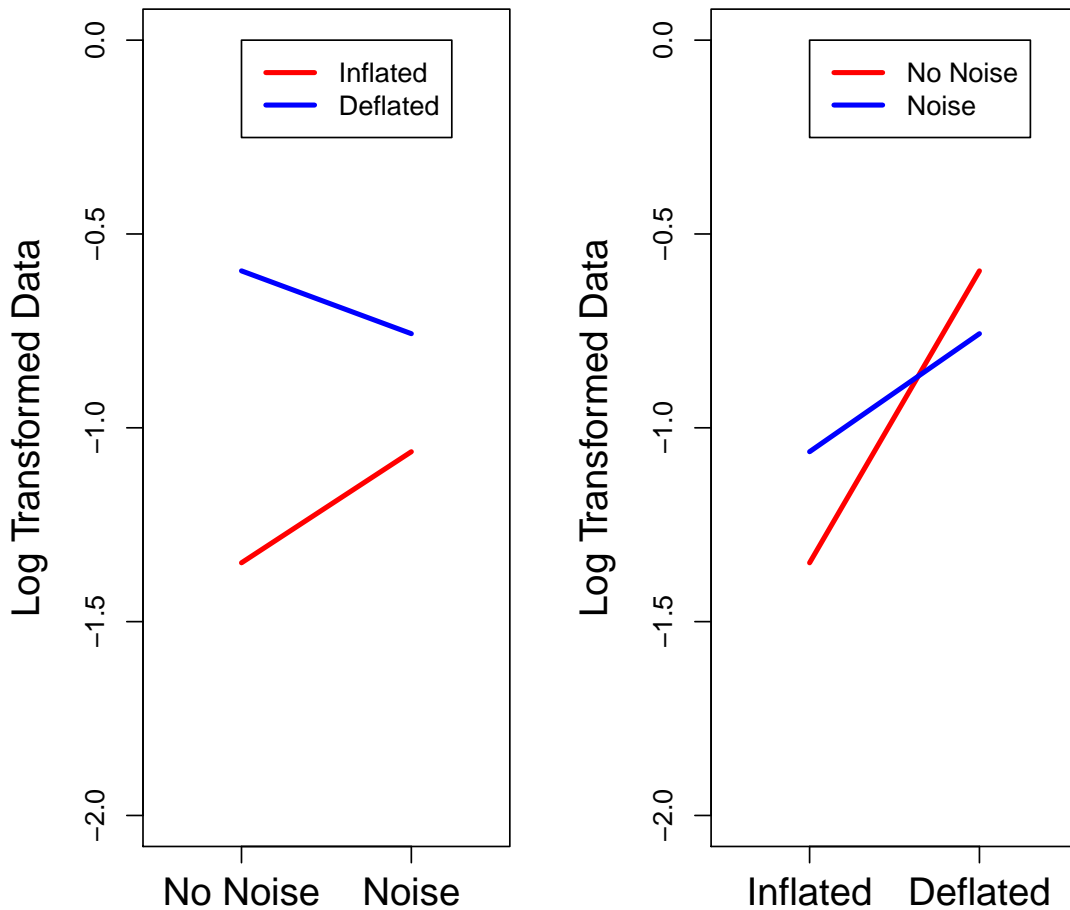
Group Distributions
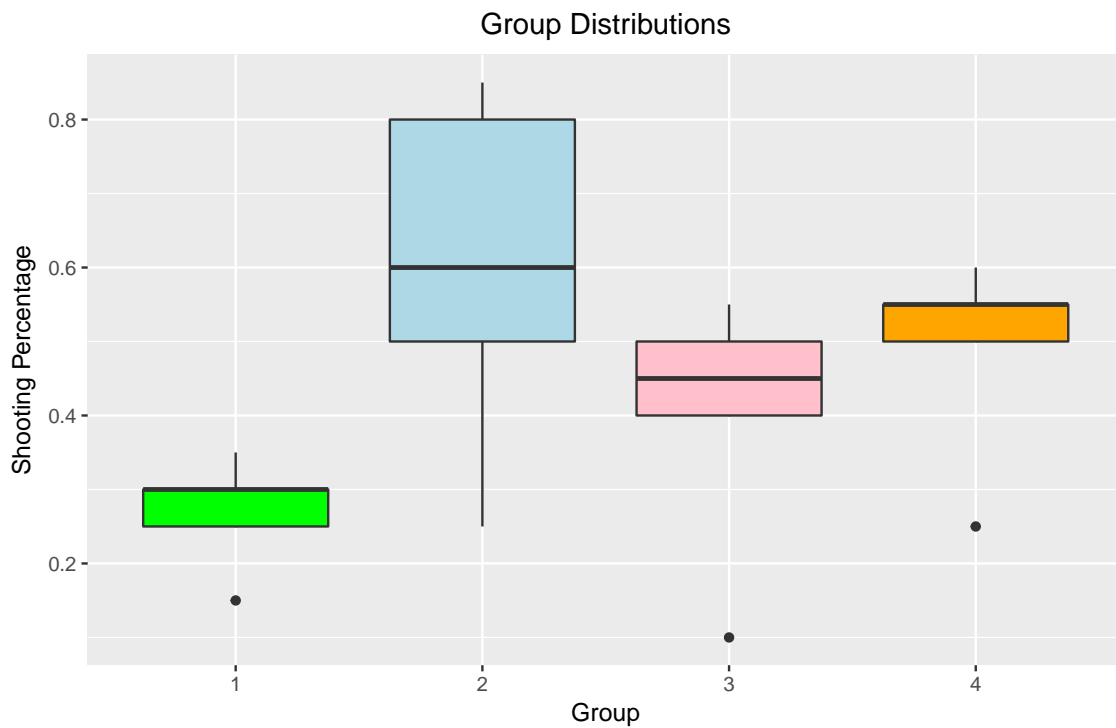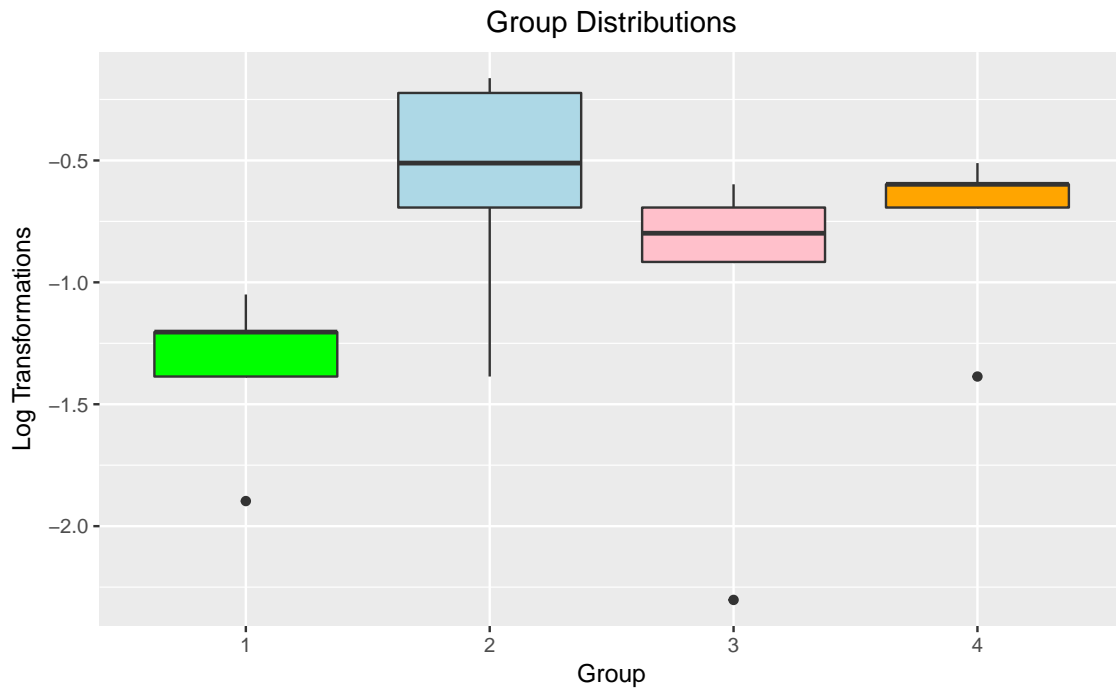
You can see that the variances of each group are not similar. We made a log transformation to the data so the data had variances that were more alike.

Furthermore, we found an apparent interaction between noise and inflation. The interaction graphs are below.

Interaction Plots for the Two Factors

I performed a poisson regression model to compare the average amount of shots made per player. If we do regression for all three factors, the degrees of freedom is 16. Using this model, we looked at a chi square of 1.2.

```
#anova test on a log transformation
anova(lm(logsurv~Inflated+Silence+Inflated*Silence, data=freethrows))

## Analysis of Variance Table
##
## Response: logsurv
##                  Df Sum Sq Mean Sq F value  Pr(>F)
```

```
## Inflated          1 1.3980 1.39798  5.7439 0.02911 *
## Silence           1 0.0194 0.01939  0.0797 0.78134
## Inflated:Silence  1 0.2515 0.25151  1.0334 0.32449
## Residuals        16 3.8941 0.24338
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#anova(lm(freethrows£Response~Inflation+Noise+Inflation*Noise, data=freethrows))

#TukeyHSD(aov(lm(logsurv ~ Group, data=freethrows)))

#try running a poisson regression
#I am running GLM on the regular data, not log transformations

#null deviance is the base model.
#16 degrees of freedom come after taking the three factors.
#3 degree of freedom and a chi square of 1.2.

reg_pois1 <- glm(freethrows$Response*20~Inflated+Silence+Inflated*Silence,
              family=poisson, data=freethrows)
summary(reg_pois1)

##
## Call:
## glm(formula = freethrows$Response * 20 ~ Inflated + Silence +
##     Inflated * Silence, family = poisson, data = freethrows)
##
## Deviance Residuals:
##    Min       1Q   Median       3Q      Max
## -2.5406  -0.2794   0.2536   0.6633   1.3574
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)         2.2824     0.1429  15.977   <2e-16 ***
## Inflated1          -0.2029     0.2131  -0.952   0.3409
## Silence1            0.2025     0.1925   1.052   0.2929
## Inflated1:Silence1 -0.5956     0.3148  -1.892   0.0585 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
```

```
##
##     Null deviance: 35.945  on 19  degrees of freedom
## Residual deviance: 22.174  on 16  degrees of freedom
## AIC: 108.53
##
## Number of Fisher Scoring iterations: 4

pois_diff <- 35.945 - 22.174
1 - pchisq(pois_diff, df=3)

## [1] 0.003234027

predict.glm(reg_pois1, type = "link")

##        1        2        3        4        5        6        7        8
## 1.686399 1.686399 1.686399 1.686399 1.686399 2.484907 2.484907 2.484907
##        9       10       11       12       13       14       15       16
## 2.484907 2.484907 2.079442 2.079442 2.079442 2.079442 2.079442 2.282382
##       17       18       19       20
## 2.282382 2.282382 2.282382 2.282382

#The counts are not the same = conclusion.
#use the counts because they have more variability n*p*q.


reg_pois2 <- glm(freethrows$Response*20~Inflated+Silence,
             family=poisson, data=freethrows)
summary(reg_pois2)

##
## Call:
## glm(formula = freethrows$Response * 20 ~ Inflated + Silence,
##     family = poisson, data = freethrows)
##
## Deviance Residuals:
##     Min       1Q    Median       3Q      Max
## -2.16130  -0.39995  -0.00719   0.54598   1.74705
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.40006    0.12136  19.776  < 2e-16 ***
## Inflated1   -0.48666    0.15524  -3.135  0.00172 **
## Silence1    -0.02273    0.15077  -0.151  0.88017
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 35.945  on 19  degrees of freedom
## Residual deviance: 25.802  on 17  degrees of freedom
## AIC: 110.16
##
## Number of Fisher Scoring iterations: 4

pois_diff_2 <- 25.802 - 22.174
1 - pchisq(pois_diff_2, df=2)

## [1] 0.1630008

#The interaction and the main effects are significant.


reg_pois3 <- glm(freethrows$Response*20~Inflated,
                 family=poisson, data=freethrows)
summary(reg_pois3)

##
## Call:
## glm(formula = freethrows$Response * 20 ~ Inflated, family = poisson,
##     data = freethrows)
##
## Deviance Residuals:
##      Min        1Q     Median        3Q        Max
## -2.13639  -0.37936    0.03024    0.57635    1.70626
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.38876    0.09578  24.939  < 2e-16 ***
## Inflated1   -0.48666    0.15524  -3.135  0.00172 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 35.945  on 19  degrees of freedom
```

```
## Residual deviance: 25.825  on 18  degrees of freedom
## AIC: 108.19
##
## Number of Fisher Scoring iterations: 4

pois_diff_3 <- 25.825 - 22.174
1 - pchisq(pois_diff_3, df=1)

## [1] 0.05603556

reg_pois4 <- glm(freethrows$Response*20~Silence,
                 family=poisson, data=freethrows)
summary(reg_pois4)

##
## Call:
## glm(formula = freethrows$Response * 20 ~ Silence, family = poisson,
##     data = freethrows)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7979  -1.0687   0.1975   0.6786   2.4852
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.18605    0.10600  20.623   <2e-16 ***
## Silence1    -0.02273    0.15076  -0.151     0.88
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 35.945  on 19  degrees of freedom
## Residual deviance: 35.922  on 18  degrees of freedom
## AIC: 118.28
##
## Number of Fisher Scoring iterations: 4

pois_diff_4 <- 35.922 - 22.174
1 - pchisq(pois_diff_4, df=1)

## [1] 0.0002090433
```

11

```
#I am trying the predict function for a glm object
#What are the results telling me?


#I'll also try the linear model for log transformed data.

#the difference from when we threw out the interaction (25.802 - 22.174) is not
#significantly different
#The difference from when we threw out the inflaction factor is 35.922.
#There is a significant difference in that factor.
#the difference from the base model of 22.174 and when we threw out noise, is
#not statistically significant. 25.825-22.174 isn't a difference.
```

6   Conclusion

```
pois_diff <- 35.945 - 22.174
1 - pchisq(pois_diff, df=3)

## [1] 0.003234027

pois_diff_2 <- 25.802 - 22.174
1 - pchisq(pois_diff_2, df=2)

## [1] 0.1630008

pois_diff_3 <- 25.825 - 22.174
1 - pchisq(pois_diff_3, df=1)

## [1] 0.05603556

pois_diff_4 <- 35.922 - 22.174
1 - pchisq(pois_diff_4, df=1)

## [1] 0.0002090433
```

The difference from when we threw out the inflaction factor (35.922), is a significant difference from the model resulting from all three factors (22.174). The difference from the model of 25.802, when we threw out noise, is not significant enough if it's compared to the model with all factors (22.174). The interaction is also not statistically significant.

Therefore, only the inflation levels in the ball make a statistically significant difference on the amount of made free throws by a free throw shooter after 20 attempts.

```
predict.glm(reg_pois1, type = "link")
##        1        2        3        4        5        6        7        8
## 1.686399 1.686399 1.686399 1.686399 1.686399 2.484907 2.484907 2.484907
##        9       10       11       12       13       14       15       16
## 2.484907 2.484907 2.079442 2.079442 2.079442 2.079442 2.079442 2.282382
##       17       18       19       20
## 2.282382 2.282382 2.282382 2.282382
```

Now, does the statistically significant factor have a practical significance? The predict.glm function helped me determine practical significance. The predict function gave four different amounts of made free throws. I looked at these four different results (1.6864, 2.4849, 2.0794, 2.2824) and compared them to each other.

In my opinion, the difference between the null model (1.6864) and the inflation factor (2.4849) is big enough (0.7985) to be practically significant. I believe this is true because that difference would be a four percent difference in a shooter's free throw percentage. However, the difference between the null model (1.6864) and the interaction factor (2.2824), which is 0.5960 is not big enough to be practically significant.

Finally, my research concludes that basketball players will not be affected enough by crowd noise alone or by the interaction of noise and inflation. When players practice free throw shooting, they will need to worry about other factors, such as the inflation levels of the ball, their health, their confidence level, or their emotional state.

Bibliography

[1] Donald Arseneau. Url package version 3.4.

[2] Javier Bezos and Johannes L. Braams. Babel package version 3.9r.

[3] David Carlisle. Graphicx package version 1.0g.

[4] David B. Dahl. Xtable package version 1.8-2.

[5] Patrick W. Daly and Arthur Ogawa. Natbib package version 8.31b.

[6] Robin Faribairns and Hubert Partl. Parskip package version 2.0.

[7] Wintston Chang Hadley Wickham and RStudio. ggplot2 package version 2.2.1.

[8] Alan Jeffrey and Frank Mittelbach. Inputenc package version 1.2c.

[9] Voker Kuhlmann. Vmargin package version 2.5.

[10] Unknown. Amsmath package version 2.16a.

[11] Piet van Oostrum. Fancyhdr package version 3.8.

[12] Yihui Xie. Knitr package version 1.15.1.