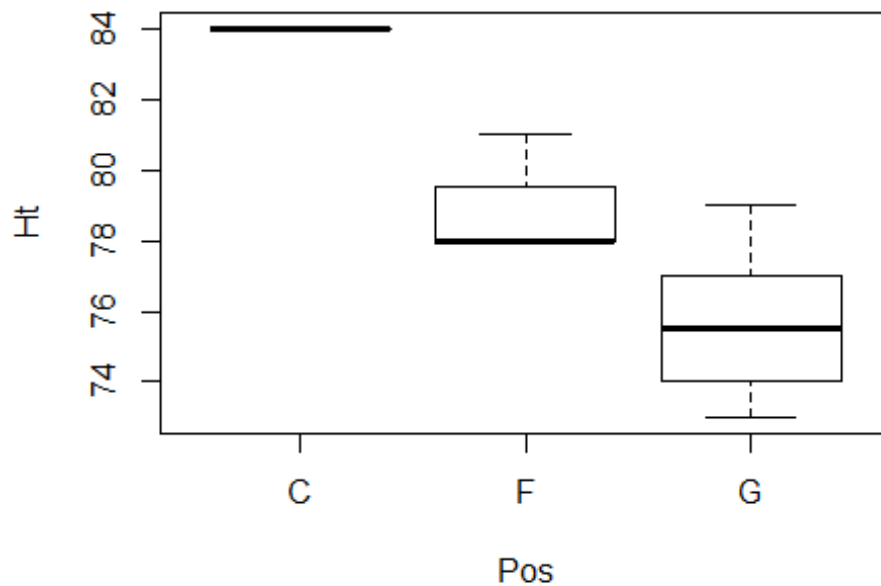# HW6

Aman Arora

4/26/2020

Answer 1

```
df = read.csv("C://Users//amana//Downloads//illinimensbb.csv",header=TRUE)
plot(Ht~Pos, data = df)
```


There does seems to a correlation between height and position based on the box plots. With tallest players at center, shorter players at guard and most middle height player in the front.

Answer 2 (a)

JAGS − Model

Model

{

for (i in 1:length(FGM))

{

FGM[i] ~ dbin(prob[i], FGA[i])

logit(prob[i]) <- betapos[Pos[i]] + betaht*htscaled[i]

FGMrep[i] ~ dbin(prob[i], FGA[i]) }

 for (j in 1:max(Pos)) {

 betapos[j] ~ dt(0, 0.01, 1)

}

betaht ~ dt(0, 0.16, 1)

}

```r
d1 <- list(FGM = df$FGM,FGA = df$FGA, Pos = unclass(df$Pos),htscaled =
as.vector(scale(df$Ht, scale=2*sd(df$Ht))))

inits1 <-
list(list(betapos=c(10,10,10),betaht=10,.RNG.name="base::Wichmann-
Hill",.RNG.seed=7),
list(betapos=c(10,10,-10), betaht=-10,.RNG.name="base::Wichmann-
Hill",.RNG.seed=8),
list(betapos=c(10,-10,10), betaht=-10,.RNG.name="base::Wichmann-
Hill",.RNG.seed=9),
list(betapos=c(10,-10,-10), betaht=10,.RNG.name="base::Wichmann-
Hill",.RNG.seed=10))

library(rjags)

## Warning: package 'rjags' was built under R version 3.6.3

## Loading required package: coda

## Warning: package 'coda' was built under R version 3.6.3

## Linked to JAGS 4.3.0
```

```
## Loaded modules: basemod,bugs

m1 <- jags.model("C://Users//amana//Downloads//FGM.bug", d1, inits1,
n.chains=4, n.adapt=2000)

## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 15
##    Unobserved stochastic nodes: 19
##    Total graph size: 110
##
## Initializing model

update(m1, 1000) # burn-in
x1 <- coda.samples(m1, c("betapos","betaht"),n.iter=10000)
gelman.diag(x1, autoburnin=FALSE)

## Potential scale reduction factors:
##
##             Point est. Upper C.I.
## betaht             1          1
## betapos[1]         1          1
## betapos[2]         1          1
## betapos[3]         1          1
##
## Multivariate psrf
##
## 1
```

As we can see the gelam-Rubin statistic is converging to 1 indicating no convergence problemeffectiveSize(x1[,1:4])

```
effectiveSize(x1[,1:4])

##     betaht betapos[1] betapos[2] betapos[3]
##   5531.385   8571.410   8668.900   7320.004
```
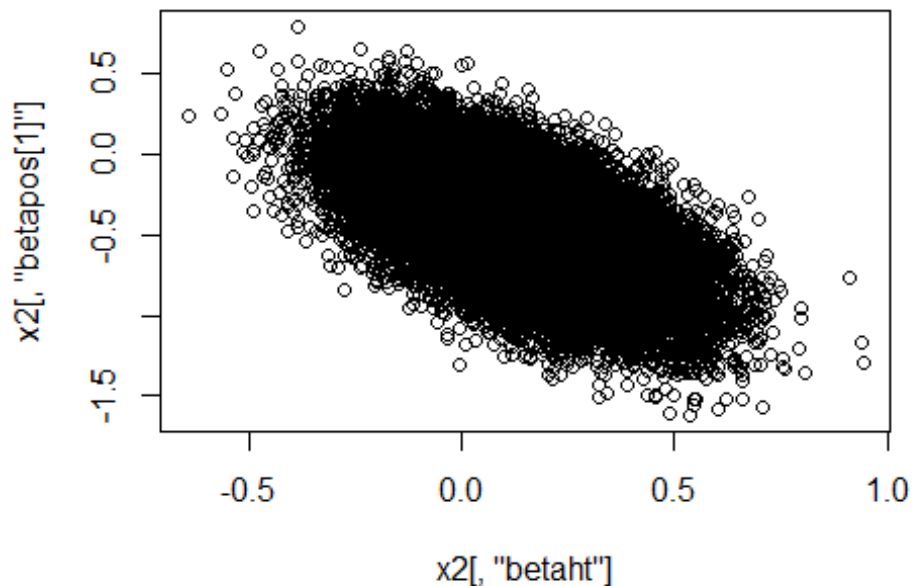
Answer 2 (b)

```
summary(x1[,1:4])

##
## Iterations = 3001:13000
## Thinning interval = 1
## Number of chains = 4
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
```

```
##               Mean      SD  Naive SE Time-series SE
## betaht       0.13788 0.17973 0.0008986      0.0024189
## betapos[1]  -0.45597 0.28828 0.0014414      0.0031166
## betapos[2]  -0.06212 0.11189 0.0005595      0.0012033
## betapos[3]  -0.33519 0.07137 0.0003569      0.0008346
##
## 2. Quantiles for each variable:
##
##                 2.5%      25%      50%      75%    97.5%
## betaht       -0.2155  0.01872  0.13698   0.2577   0.4931
## betapos[1]   -1.0221 -0.64916 -0.45681  -0.2619   0.1086
## betapos[2]   -0.2821 -0.13738 -0.06209   0.0132   0.1575
## betapos[3]   -0.4765 -0.38252 -0.33539  -0.2872  -0.1959
```
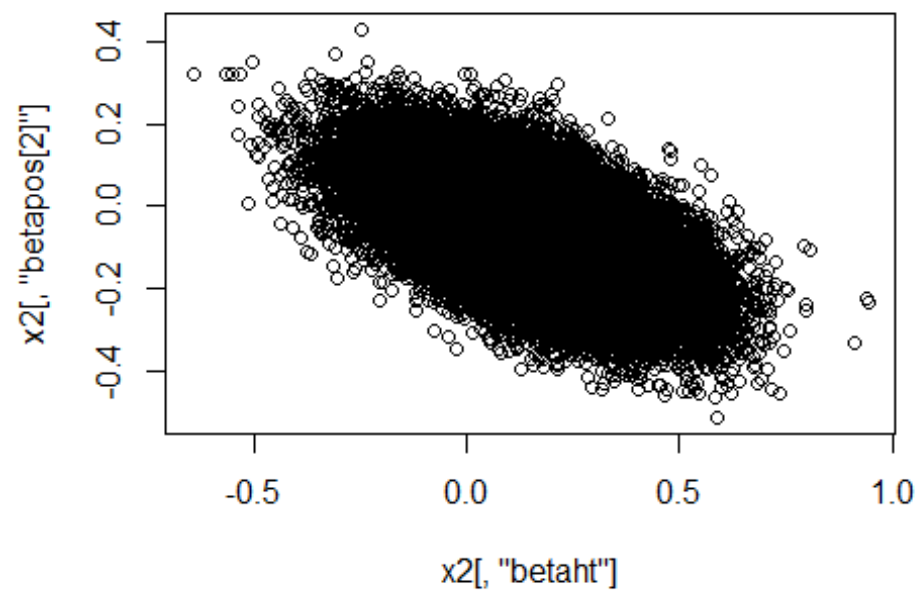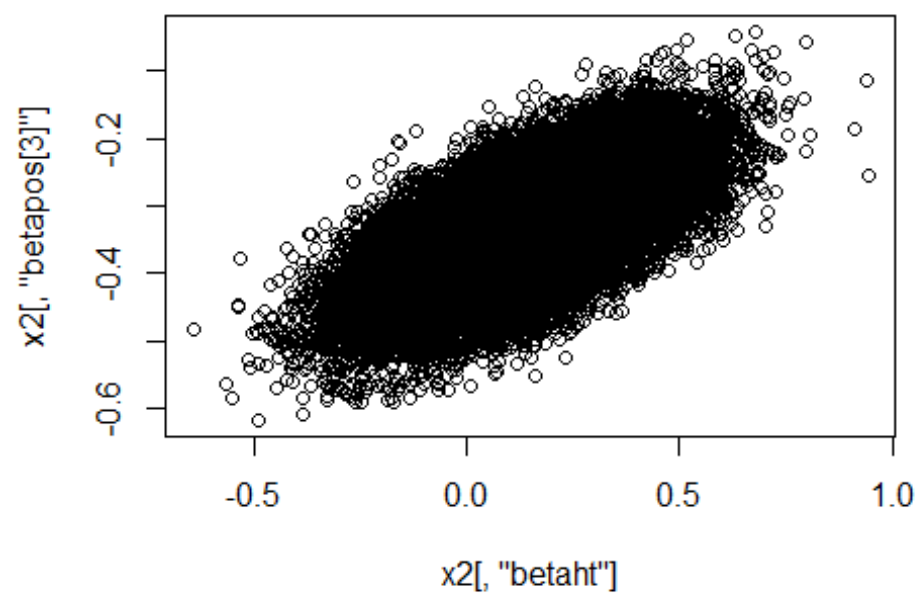
Answer 2 (c)

```
x2 = as.matrix(x1)

plot(x2[,'betaht'],x2[,'betapos[1]'])
```



```
plot(x2[,'betaht'],x2[,'betapos[2]'])
```

```
plot(x2[,'betaht'],x2[,'betapos[3]'])
```
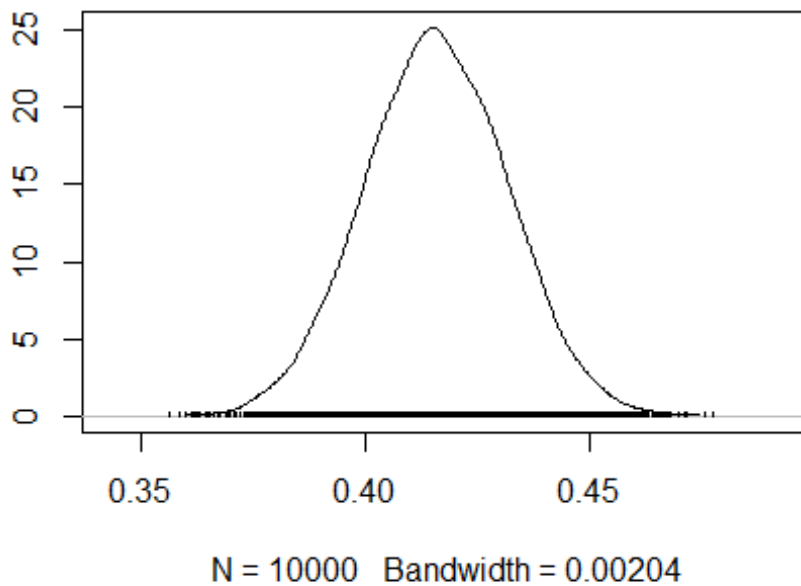
There seems to be negative correlation between beta[1] (center)/beta[2] (forward) and beta height. Also there is positive correlation between beta[3] (guard) and beta height.

Answer 2 (d)

Posterior density of FGM probability of Ayo Dosunmu (at index 4)

```
x3 <- coda.samples(m1, c("betapos","betaht","prob","FGMrep"),n.iter=10000)

x4 <- as.matrix(x3)
densplot(x3[,c("prob[4]")])
```


N = 10000   Bandwidth = 0.00204

Answer 2 (e)

Posterior probability of forward having a higher probability of making a goal vs guards.

```
cat("Posterior probability:", mean(x4[,'betapos[2]']>x4[,'betapos[3]']),"\n")

## Posterior probability: 0.962975

bayes_factor <-
mean(x4[,'betapos[2]']>x4[,'betapos[3]'])/mean(x4[,'betapos[2]']<
x4[,'betapos[3]'])
cat("Bayes factor:", bayes_factor)

## Bayes factor: 26.00878
```

Bayes factor indicates strong evidence in favor FGM for forward vs guards.

Answer 2 (f)

```r
probs <- as.matrix(x3)[, paste("prob[",1:nrow(df),"]", sep="")]
FGMrep <- as.matrix(x3)[, paste("FGMrep[",1:nrow(df),"]", sep="")]
Tchi <- numeric(nrow(df))
Tchirep <- numeric(nrow(df))
for(s in 1:nrow(FGMrep)){

 Tchi[s] <- sum((df$FGM - df$FGA*probs[s,])^2 /
 (df$FGA*probs[s,]*(1-probs[s,])))

 Tchirep[s] <- sum((FGMrep[s,] - df$FGA*probs[s,])^2 /
 (df$FGA*probs[s,]*(1-probs[s,])))
 }
mean(Tchirep >= Tchi)

## [1] 0.0483
```

THe p value is close to 0, the discrepency measure of replicated data has very low probability to be above discrepency measure of data indicating evidence overdispersion.

Answer 2 (g)(i)

JAGS − Model
‑

model {

for (i in 1:length(FGM)) {

FGM[i] ~ dbin(prob[i], FGA[i])

logit(prob[i]) <- betapos[Pos[i]] + betaht*htscaled[i] + epsilon[i]

epsilon[i] ~ dnorm(0,1/sigmaepsilon^2)

FGMrep[i] ~ dbin(prob[i], FGA[i])

}

for (j in 1:max(Pos)) {

betapos[j] ~ dt(0, 0.01, 1)

}

betaht ~ dt(0, 0.16, 1)

sigmaepsilon ~ dunif(0,10)

}

```r
library(rjags)

inits2 <-
list(list(betapos=c(10,10,10),betaht=10,sigmaepsilon=0.01,.RNG.name="base::Wi
chmann-Hill",.RNG.seed=7),
list(betapos=c(10,10,-10), betaht=-
10,sigmaepsilon=9,.RNG.name="base::Wichmann-Hill",.RNG.seed=8),
list(betapos=c(10,-10,10), betaht=-
10,sigmaepsilon=.01,.RNG.name="base::Wichmann-Hill",.RNG.seed=9),
list(betapos=c(10,-10,-10),
betaht=10,sigmaepsilon=9,.RNG.name="base::Wichmann-Hill",.RNG.seed=10))

m1 <- jags.model("C://Users//amana//Downloads//FGM_dis.bug", d1, inits2,
n.chains=4, n.adapt=2000)

## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 15
##    Unobserved stochastic nodes: 35
##    Total graph size: 142
##
## Initializing model

update(m1, 1000) # burn-in
x5 <- coda.samples(m1, c("betapos","betaht","sigmaepsilon"),n.iter=10000)
gelman.diag(x5, autoburnin=FALSE)

## Potential scale reduction factors:
##
##             Point est. Upper C.I.
## betaht            1.01       1.02
## betapos[1]        1.00       1.01
## betapos[2]        1.00       1.01
## betapos[3]        1.03       1.03
## sigmaepsilon      1.08       1.08
##
## Multivariate psrf
##
## 1.01
```
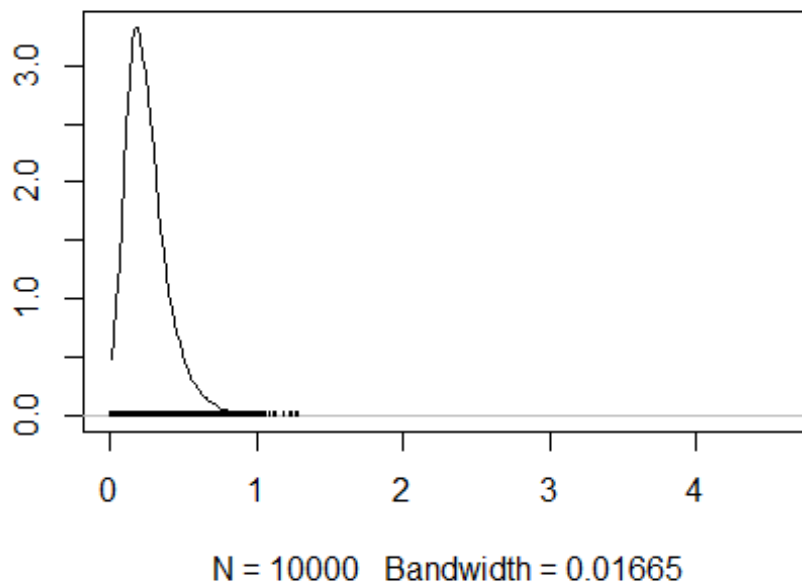
Answer 2 (g)(ii)

```r
densplot(x5[,c("sigmaepsilon")])
```

N = 10000   Bandwidth = 0.01665

Answer 2 (g)(iii)

```
x6 = as.matrix(x5)

cat("Posterior probability:", mean(x6[,'betapos[2]']>x6[,'betapos[3]']),"\n")

## Posterior probability: 0.7849

bayes_factor <-
mean(x6[,'betapos[2]']>x6[,'betapos[3]'])/mean(x6[,'betapos[2]']<
x6[,'betapos[3]'])
cat("Bayes factor:", bayes_factor)

## Bayes factor: 3.649
```

The conclusions for new model based on p value and bayes value is different. The evidence is now less strong vs previous model that forward players probability of making a field goal is higher than guard position player. The bayes factor also indicates a positive evidence not strong like previous model.

Answer 3(a)

model {

for (i in 1:length(blocks))

{

  blocks[i] ~ dpois(lambda[i])

  log(lambda[i]) <- logmin[i] + beta.pos[pos[i]] + beta.height*htscaled[i]

  blocksrep[i] ~ dpois(lambda[i])

}

beta.height ~ dnorm(0, 0.0001)

for (i in 1:max(pos)) {

beta.pos[i] ~ dnorm(0, 0.0001)

}

}}

```r
d2 <- list(blocks = df$BLK, logmin = log(df$MIN), htscaled =
as.vector(scale(df$Ht)), pos = unclass(df$Pos))

inits2 <-
list(list(beta.pos=c(100,100,100),beta.height=10,.RNG.name="base::Wichmann-
Hill",.RNG.seed=1981),
list(beta.pos=c(100,100,-100), beta.height=-10,.RNG.name="base::Wichmann-
Hill",.RNG.seed=1982),
list(beta.pos=c(100,-100,100), beta.height=-10,.RNG.name="base::Wichmann-
Hill",.RNG.seed=1983),
list(beta.pos=c(100,-100,-100), beta.height=10,.RNG.name="base::Wichmann-
Hill",.RNG.seed=1984))
library(rjags)

m2 <- jags.model("C://Users//amana//Downloads//Shots_blocked.bug", d2,
inits2, n.chains=4, n.adapt=1000)

## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 15
##    Unobserved stochastic nodes: 19
##    Total graph size: 120
##
## Initializing model
```

```
update(m2, 1000) # burn-in
x6 <- coda.samples(m2,
c("beta.pos","beta.height","blocksrep","lambda"),n.iter=50000)
gelman.diag(x6[,1:19], autoburnin=FALSE)

## Potential scale reduction factors:
##
##                Point est. Upper C.I.
## beta.height            1          1
## beta.pos[1]            1          1
## beta.pos[2]            1          1
## beta.pos[3]            1          1
## blocksrep[1]           1          1
## blocksrep[2]           1          1
## blocksrep[3]           1          1
## blocksrep[4]           1          1
## blocksrep[5]           1          1
## blocksrep[6]           1          1
## blocksrep[7]           1          1
## blocksrep[8]           1          1
## blocksrep[9]           1          1
## blocksrep[10]          1          1
## blocksrep[11]          1          1
## blocksrep[12]          1          1
## blocksrep[13]          1          1
## blocksrep[14]          1          1
## blocksrep[15]          1          1
##
## Multivariate psrf
##
## 1

effectiveSize(x6[,1:4])

## beta.height beta.pos[1] beta.pos[2] beta.pos[3]
##    11279.08    12486.00    14084.74    41034.70
```

Answer 3(b)

```
summary(x6[,1:4])

##
## Iterations = 2001:52000
## Thinning interval = 1
## Number of chains = 4
## Sample size per chain = 50000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##                Mean     SD  Naive SE Time-series SE
```

```
## beta.height  1.002 0.2766 0.0006185       0.0026047
## beta.pos[1] -5.286 0.6084 0.0013605       0.0054464
## beta.pos[2] -4.507 0.2886 0.0006454       0.0024324
## beta.pos[3] -4.451 0.1788 0.0003998       0.0008847
##
## 2. Quantiles for each variable:
##
##                  2.5%     25%     50%    75%   97.5%
## beta.height   0.4735  0.8131  0.9966  1.184   1.561
## beta.pos[1]  -6.5144 -5.6891 -5.2732 -4.871  -4.125
## beta.pos[2]  -5.0939 -4.6969 -4.4990 -4.307  -3.965
## beta.pos[3]  -4.8143 -4.5693 -4.4472 -4.328  -4.115
```

Answer 3(c)

```
x7 <- as.matrix(x6)

post.samp <- x7[,"beta.height"]
post.samp.change = exp(post.samp)
quantile(post.samp.change,c(0.025,0.975))

##     2.5%    97.5%
## 1.605604 4.762949
```

Since the lower end of 95 central inerval is greater than 1 (~1.6) this indicates greater height is associated with higher blocking rate.

Answer 3(d)

```
lambdas <- as.matrix(x6)[, paste("lambda[",1:nrow(df),"]", sep="")]
blocksrep <- as.matrix(x6)[, paste("blocksrep[",1:nrow(df),"]", sep="")]
Tchi <- numeric(nrow(blocksrep))
Tchirep <- numeric(nrow(blocksrep))

for(s in 1:nrow(blocksrep)){
 Tchi[s] <- sum((df$BLK - lambdas[s,])^2 / lambdas[s,])
 Tchirep[s] <- sum((blocksrep[s,] - lambdas[s,])^2 / lambdas[s,])
}
 mean(Tchirep >= Tchi)

## [1] 0.00707
```

The low p values indicates evidence of a problem.

Answer 3 (e)(i)

```
for (i in 1:nrow(df))
{
  df$p_value[i] <-  mean(blocksrep[,i] > df$BLK[i])
}

df[,c("Player","p_value")]
```

```
##                   Player  p_value
## 1  Bezhanishvili, Giorgi 0.533540
## 2           Cayce, Drew 0.058630
## 3    De La Rosa, Adonis 0.985950
## 4          Dosunmu, Ayo 0.705930
## 5         Feliz, Andres 0.829950
## 6        Frazier, Trent 0.821880
## 7         Griffin, Alan 0.007880
## 8        Griffith, Zach 0.185145
## 9         Jones, Tevian 0.899035
## 10        Jordan, Aaron 0.132660
## 11          Kane, Samba 0.001975
## 12       Nichols, Kipper 0.238850
## 13     Oladimeji, Samson 0.184070
## 14      Underwood, Tyler 0.269500
## 15     Williams, Da'Monte 0.046210
```

Answer 3 (e)(ii)

One of the player with p values less than .05 is Kane, Samba

Answer 3 (e)(iii)

Some of the players have probability close to 1 - "De La Rosa, Adonis" and "Jones, Tevian" since they blocked less shots than the play time and height would suggests. Hence the replicated values are greater than actual values in majority of replicates samples.