**Christopher Boone**

**STAT 430 – Project**

**Title:** Analysis of Concussions in Male and Female College Athletes from 1997 to 1999

**Data:** <http://users.stat.ufl.edu/~winner/data/concussion.dat>

**Description:** <http://users.stat.ufl.edu/~winner/data/concussion.txt>

> data=read.table("http://users.stat.ufl.edu/~winner/data/concussion.dat")

> names(data)=c("Gender","Sport","Academic\_Year","Concussion\_ID","Count")

S1 = Student-Athletes who are diagnosed with a concussion in all sports from 1997 to 1999 (Male and Female)

S2 = Student-Athletes who are not diagnosed with a concussion in all sports from 1997 to 1999 (Male and Female)

S3 = Female Student-Athletes data set

S4 = Female Student Athletes who are diagnosed with a concussion in all sports from 1997 to 1999

S5 = Male Student-Athletes data set

S6 = Male Student- Athletes who are diagnosed with a concussion in all sports from 1997 to 1999

S7 = Soccer data set

S8 = Student-Athletes, who played soccer, were diagnosed with a concussion from 1997 to 1999 (Male and Female)

S9 = Female Student-Athletes, who played soccer, were diagnosed with a concussion from 1997 to 1999

S10 = Male Student-Athletes, who played soccer, were diagnosed with a concussion from 1997 to 1999

S13 = Softball/Baseball data

S14 = Female Softball data

S15 = Male Baseball data

> s1=subset(data,Concussion\_ID==1)

> head(s1)

> boxplot(s1$Count~s1$Gender,main="Distribution in Concussion Counts based off Gender",col=c("pink","blue"))



> library(ggplot2)

> bp=ggplot(s1,aes(x="",y=Count,fill=Sport))+geom\_bar(width=1,stat="identity")

> bp



> ggplot(s1,aes(x=Count,y=Sport,fill=Sport))+geom\_density\_ridges()+theme\_ridges()+theme(legend.position="none")



> ggplot(s1,aes(x=Count,y=Gender,fill=Gender))+geom\_density\_ridges()+theme\_ridges()+theme(legend.position="none")



> boxplot(s1$Count~s1$Academic\_Year,main="Distribution in Concussion Counts based off Academic Year",col=c("purple","gold","red"))



> s3=subset(data,Gender=="Female")

> s4=subset(s3,Concussion\_ID==1)

> boxplot(s4$Count~s4$Academic\_Year,main="Female Distribution in Concussion Counts based off Academic Year",col=c("purple","gold","red"))



> bp=ggplot(s4,aes(x="",y=Count,fill=Sport))+geom\_bar(width=1,stat="identity")

> bp



> ggplot(s4,aes(x=Count,y=Sport,fill=Sport))+geom\_density\_ridges()+theme\_ridges()+theme(legend.position="none")



> s5=subset(data,Gender=="Male")

> s6=subset(s5,Concussion\_ID==1)

> boxplot(s6$Count~s6$Academic\_Year,main="Male Distribution in Concussion Counts based off Academic Year",col=c("purple","gold","red"))



> bp=ggplot(s6,aes(x="",y=Count,fill=Sport))+geom\_bar(width=1,stat="identity")

> bp



> ggplot(s6,aes(x=Count,y=Sport,fill=Sport))+geom\_density\_ridges()+theme\_ridges()+theme(legend.position="none")



> s7=subset(data,Sport=="Soccer")

> s8=subset(s7,Concussion\_ID=="1")

> bp=ggplot(s8,aes(x="",y=Count,fill=Gender))+geom\_bar(width=1,stat="identity")

> bp



> ggplot(s8,aes(x=Count,y=Gender,fill=Gender))+geom\_density\_ridges()+theme\_ridges()+theme(legend.position="none")



> boxplot(s8$Count~s8$Academic\_Year,main="Soccer Distribution of Concussion based off Acadermic Year",col=c("purple","gold","red"))



> s9=subset(s8,Gender=="Female")

> bp=ggplot(s9,aes(x="",y=Count,fill=Academic\_Year))+geom\_bar(width=1,stat="identity")

> bp



> s10=subset(s8,Gender=="Male")

> bp=ggplot(s10,aes(x="",y=Count,fill=Academic\_Year))+geom\_bar(width=1,stat="identity")

> bp



> bp=ggplot(s2,aes(x="",y=Count,fill=Sport))+geom\_bar(width=1,stat="identity")

> bp



> boxplot(s2$Count~s2$Gender,main="Sports Distribution of non-Concussion based off Gender",col=c("pink","blue"))



> boxplot(s2$Count~s2$Academic\_Year,main="Sports Distribution of non-Concussion based off Acadermic Year",col=c("purple","gold","red"))



> s13=subset(s2,Sport=="Softball/Baseball")

> bp=ggplot(s13,aes(x="",y=Count,fill=Gender))+geom\_bar(width=1,stat="identity")

> bp



> s14=subset(s13,Gender=="Female")

> bp=ggplot(s14,aes(x="",y=Count,fill=Academic\_Year))+geom\_bar(width=1,stat="identity")

> bp



> s15=subset(s13,Gender=="Male")

> bp=ggplot(s15,aes(x="",y=Count,fill=Academic\_Year))+geom\_bar(width=1,stat="identity")

> bp



> summary(aov(s1$Count~s1$Sport))

Df Sum Sq Mean Sq F value Pr(>F)

s1$Sport 4 5897 1474.1 23.26 4.07e-08 \*\*\*

Residuals 25 1585 63.4

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | DF | Sum of Squares | MS | F Value | Pr > F |
| Model | 4 | 5897 | 1474.25 | 23.253 | 4.07e-08 |
| Error | 25 | 1585 | 63.4 |  |  |
| Total | 29 | 7482 |  |  |  |

> TukeyHSD(aov(s1$Count~s1$Sport))

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = s1$Count ~ s1$Sport)

$`s1$Sport`

diff lwr upr p adj

Gymnastics-Basketball -20.000000 -33.4996725 -6.500328 0.0017175 \*

Lacrosse-Basketball -7.333333 -20.8330058 6.166339 0.5138506

Soccer-Basketball 23.000000 9.5003275 36.499672 0.0003288 \*

Softball/Baseball-Basketball -3.500000 -16.9996725 9.999672 0.9392251

Lacrosse-Gymnastics 12.666667 -0.8330058 26.166339 0.0734659

Soccer-Gymnastics 43.000000 29.5003275 56.499672 0.0000000 \*

Softball/Baseball-Gymnastics 16.500000 3.0003275 29.999672 0.0112117 \*

Soccer-Lacrosse 30.333333 16.8336609 43.833006 0.0000061 \*

Softball/Baseball-Lacrosse 3.833333 -9.6663391 17.333006 0.9174600 \*

Softball/Baseball-Soccer -26.500000 -39.9996725 -13.000328 0.0000479 \*

> plot(TukeyHSD(aov(s1$Count~s1$Sport)))



> TukeyHSD(aov(s1$Count~s1$Sport),ordered=T,conf.level=0.90)

Tukey multiple comparisons of means

90% family-wise confidence level

factor levels have been ordered

Fit: aov(formula = s1$Count ~ s1$Sport)

$`s1$Sport`

diff lwr upr p adj

Lacrosse-Gymnastics 12.666667 0.6958479 24.63749 0.0734659

Softball/Baseball-Gymnastics 16.500000 4.5291812 28.47082 0.0112117 \*

Basketball-Gymnastics 20.000000 8.0291812 31.97082 0.0017175 \*

Soccer-Gymnastics 43.000000 31.0291812 54.97082 0.0000000 \*

Softball/Baseball-Lacrosse 3.833333 -8.1374855 15.80415 0.9174600

Basketball-Lacrosse 7.333333 -4.6374855 19.30415 0.5138506

Soccer-Lacrosse 30.333333 18.3625145 42.30415 0.0000061 \*

Basketball-Softball/Baseball 3.500000 -8.4708188 15.47082 0.9392251

Soccer-Softball/Baseball 26.500000 14.5291812 38.47082 0.0000479 \*

Soccer-Basketball 23.000000 11.0291812 34.97082 0.0003288 \*

> plot(TukeyHSD(aov(s1$Count~s1$Sport),ordered=T,conf.level=0.90))



> summary(aov(s4$Count~s4$Sport))

Df Sum Sq Mean Sq F value Pr(>F)

s4$Sport 4 4850 1212.6 27.64 2.19e-05 \*\*\*

Residuals 10 439 43.9

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> TukeyHSD(aov(s4$Count~s4$Sport))

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = s4$Count ~ s4$Sport)

$`s4$Sport`

diff lwr upr p adj

Gymnastics-Basketball -23.666667 -41.464225 -5.869109 0.0094535 \*

Lacrosse-Basketball -15.333333 -33.130891 2.464225 0.1008796

Soccer-Basketball 28.666667 10.869109 46.464225 0.0024832 \*

Softball/Baseball-Basketball -8.333333 -26.130891 9.464225 0.5615867

Lacrosse-Gymnastics 8.333333 -9.464225 26.130891 0.5615867

Soccer-Gymnastics 52.333333 34.535775 70.130891 0.0000165 \*

Softball/Baseball-Gymnastics 15.333333 -2.464225 33.130891 0.1008796

Soccer-Lacrosse 44.000000 26.202442 61.797558 0.0000772 \*

Softball/Baseball-Lacrosse 7.000000 -10.797558 24.797558 0.7003914

Softball/Baseball-Soccer -37.000000 -54.797558 -19.202442 0.0003361 \*

> plot(TukeyHSD(aov(s4$Count~s4$Sport)))



> TukeyHSD(aov(s4$Count~s4$Sport),ordered=T,conf.level=0.90)

Tukey multiple comparisons of means

90% family-wise confidence level

factor levels have been ordered

Fit: aov(formula = s4$Count ~ s4$Sport)

$`s4$Sport`

diff lwr upr p adj

Lacrosse-Gymnastics 8.333333 -7.03114576 23.69781 0.5615867

Softball/Baseball-Gymnastics 15.333333 -0.03114576 30.69781 0.1008796

Basketball-Gymnastics 23.666667 8.30218757 39.03115 0.0094535 \*

Soccer-Gymnastics 52.333333 36.96885424 67.69781 0.0000165 \*

Softball/Baseball-Lacrosse 7.000000 -8.36447909 22.36448 0.7003914

Basketball-Lacrosse 15.333333 -0.03114576 30.69781 0.1008796

Soccer-Lacrosse 44.000000 28.63552091 59.36448 0.0000772 \*

Basketball-Softball/Baseball 8.333333 -7.03114576 23.69781 0.5615867

Soccer-Softball/Baseball 37.000000 21.63552091 52.36448 0.0003361 \*

Soccer-Basketball 28.666667 13.30218757 44.03115 0.0024832 \*

> plot(TukeyHSD(aov(s4$Count~s4$Sport),ordered=T,conf.level=0.90))



> summary(aov(s6$Count~s6$Sport))

Df Sum Sq Mean Sq F value Pr(>F)

s6$Sport 4 1703 425.7 10.49 0.00133 \*\*

Residuals 10 406 40.6

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> TukeyHSD(aov(s6$Count~s6$Sport))

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = s6$Count ~ s6$Sport)

$`s6$Sport`

diff lwr upr p adj

Gymnastics-Basketball -16.3333333 -33.4553973 0.7887307 0.0632442

Lacrosse-Basketball 0.6666667 -16.4553973 17.7887307 0.9999275

Soccer-Basketball 17.3333333 0.2112693 34.4553973 0.0469450 \*

Softball/Baseball-Basketball 1.3333333 -15.7887307 18.4553973 0.9988761

Lacrosse-Gymnastics 17.0000000 -0.1220640 34.1220640 0.0518542

Soccer-Gymnastics 33.6666667 16.5446027 50.7887307 0.0005298 \*

Softball/Baseball-Gymnastics 17.6666667 0.5446027 34.7887307 0.0424984 \*

Soccer-Lacrosse 16.6666667 -0.4553973 33.7887307 0.0572711

Softball/Baseball-Lacrosse 0.6666667 -16.4553973 17.7887307 0.9999275

Softball/Baseball-Soccer -16.0000000 -33.1220640 1.1220640 0.0698258

> plot(TukeyHSD(aov(s6$Count~s6$Sport)))



> TukeyHSD(aov(s6$Count~s6$Sport),ordered=T,conf.level=0.90)

Tukey multiple comparisons of means

90% family-wise confidence level

factor levels have been ordered

Fit: aov(formula = s6$Count ~ s6$Sport)

$`s6$Sport`

diff lwr upr p adj

Basketball-Gymnastics 16.3333333 1.552002 31.11466 0.0632442

Lacrosse-Gymnastics 17.0000000 2.218669 31.78133 0.0518542

Softball/Baseball-Gymnastics 17.6666667 2.885336 32.44800 0.0424984 \*

Soccer-Gymnastics 33.6666667 18.885336 48.44800 0.0005298 \*

Lacrosse-Basketball 0.6666667 -14.114664 15.44800 0.9999275

Softball/Baseball-Basketball 1.3333333 -13.447998 16.11466 0.9988761

Soccer-Basketball 17.3333333 2.552002 32.11466 0.0469450 \*

Softball/Baseball-Lacrosse 0.6666667 -14.114664 15.44800 0.9999275

Soccer-Lacrosse 16.6666667 1.885336 31.44800 0.0572711

Soccer-Softball/Baseball 16.0000000 1.218669 30.78133 0.0698258

> plot(TukeyHSD(aov(s6$Count~s6$Sport),ordered=T,conf.level=0.90))



> summary(aov(s2$Count~s2$Sport))

Df Sum Sq Mean Sq F value Pr(>F)

s2$Sport 4 1.006e+10 2.515e+09 27.46 7.94e-09 \*\*\*

Residuals 25 2.290e+09 9.159e+07

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> TukeyHSD(aov(s2$Count~s2$Sport))

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = s2$Count ~ s2$Sport)

$`s2$Sport`

diff lwr upr p adj

Gymnastics-Basketball -30488.833 -46715.916 -14261.750 0.0000892 \*

Lacrosse-Basketball -22808.667 -39035.750 -6581.584 0.0030019 \*

Soccer-Basketball -7635.000 -23862.083 8592.083 0.6443852

Softball/Baseball-Basketball 21795.833 5568.750 38022.916 0.0047285 \*

Lacrosse-Gymnastics 7680.167 -8546.916 23907.250 0.6394041

Soccer-Gymnastics 22853.833 6626.750 39080.916 0.0029413 \*

Softball/Baseball-Gymnastics 52284.667 36057.584 68511.750 0.0000000 \*

Soccer-Lacrosse 15173.667 -1053.416 31400.750 0.0749183

Softball/Baseball-Lacrosse 44604.500 28377.417 60831.583 0.0000002 \*

Softball/Baseball-Soccer 29430.833 13203.750 45657.916 0.0001449 \*

> TukeyHSD(aov(s2$Count~s2$Sport),ordered=T,conf.level=0.90)

Tukey multiple comparisons of means

90% family-wise confidence level

factor levels have been ordered

Fit: aov(formula = s2$Count ~ s2$Sport)

$`s2$Sport`

diff lwr upr p adj

Lacrosse-Gymnastics 7680.167 -6709.1802 22069.51 0.6394041

Soccer-Gymnastics 22853.833 8464.4865 37243.18 0.0029413 \*

Basketball-Gymnastics 30488.833 16099.4865 44878.18 0.0000892 \*

Softball/Baseball-Gymnastics 52284.667 37895.3198 66674.01 0.0000000 \*

Soccer-Lacrosse 15173.667 784.3198 29563.01 0.0749183

Basketball-Lacrosse 22808.667 8419.3198 37198.01 0.0030019 \*

Softball/Baseball-Lacrosse 44604.500 30215.1532 58993.85 0.0000002 \*

Basketball-Soccer 7635.000 -6754.3468 22024.35 0.6443852

Softball/Baseball-Soccer 29430.833 15041.4865 43820.18 0.0001449 \*

Softball/Baseball-Basketball 21795.833 7406.4865 36185.18 0.0047285 \*

> plot(TukeyHSD(aov(s2$Count~s2$Sport),ordered=T,conf.level=0.90))



> kruskal.test(s1$Count~s1$Sport)

Kruskal-Wallis rank sum test

data: s1$Count by s1$Sport

Kruskal-Wallis chi-squared = 22.903, df = 4, p-value = 0.0001324

> kruskal.test(s4$Count~s4$Sport)

Kruskal-Wallis rank sum test

data: s4$Count by s4$Sport

Kruskal-Wallis chi-squared = 12.478, df = 4, p-value = 0.01413

> kruskal.test(s6$Count~s6$Sport)

Kruskal-Wallis rank sum test

data: s6$Count by s6$Sport

Kruskal-Wallis chi-squared = 11.18, df = 4, p-value = 0.02462

> kruskal.test(s2$Count~s2$Sport)

Kruskal-Wallis rank sum test

data: s2$Count by s2$Sport

Kruskal-Wallis chi-squared = 25.643, df = 4, p-value = 3.735e-05