440 Reproducibility and Statistics Assignment Homework

Find the minimum, mean, and variance for posttest optimism scores.

The variance for the optimism scores is found by using command:

var(politics$optimismscore)

The result is: [1] 344.6696

The result shows that the variance in the optimism scores is 344.6696.

The minimum optimism score is 17, which is found by using the following command:

min(politics$optimismscore)

The result is: [1] 17

The mean optimism score is 57.40152. The mean score was found by using the command:

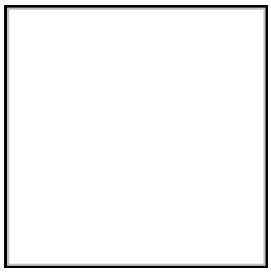
mean(politics$optimismscore)

The result is: [1] 57.40152

Create a histogram of posttest optimism scores. Feel free to use the default ugly settings since we didn't cover how to do this in ggplot. If you're a real go-getter, see if you can figure out how to make a prettier histogram with better labels, etc. using ggplot2.

The basic command to be used to create a histogram is:

hist(politics$optimismscore)



this link

However, if you would like to customize the graph there are several options. Some options used in the following graph are the color and the labeling of the axis. The following command is an example of what the same histogram using chartreuse color and foregoing the main title and adding the x-axis.

hist(politics$optimismscore, col="orange", xlab="Optimism Score",main="Histogram Showing Optimism Scores")

[This link]

Run a test to determine whether political affiliation is independent of one's gender.Present the data supporting your analyses in graphical or tabular form.

> tab<-table(polstesttime=="pre"],polstesttime=="pre"])

female male

democrat 14 12

independent 7 10

republican 12 11

chisq.test(polssex)

Pearson's Chi-squared test

data: polssex

X-squared = 1.4535, df = 2, p-value = 0.4835

The P-Value of 0.48 from the Chi Square test indicates there is no statistical significance- meaning sex and political affiliation are independent of each other.

Run a test to determine whether or not males and females have different incomes. Create a graph to display your data that includes the relevant means and standard errors.

> chisq.test(polstesttime=="pre"],polstesttime=="pre"])

Pearson's Chi-squared test

data:pols$sex[pols$testtime == "pre"] and polstesttime == "pre"]

X-squared = 66, df = 65, p-value = 0.4421

Sex is not independent of income.

See if male and females with different political affiliations have different posttest optimism scores. Create a graph to display your data that includes the relevant means and standard errors.

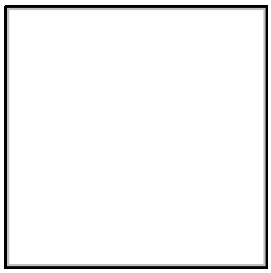
>
**summary(aov(optimismscore~party*sex,data=pols[pols$testtime==*
*="post",])) ### Df Sum Sq Mean Sq F value Pr(>F)**
###party 2 10147 5074 27.063 4.2e-09 ###sex 1 7 7 0.040 0.843
###party:sex 2 455 227 1.213 0.304
###Residuals 60 11248 187
--- ###Signif. codes: 0 ‘*’ 0.001 ‘’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

***The result of the ANOVA shows that there is a relation between optimism scores and the political party affiliation, F(2,60)=**

27.06, p<.001. ##On the other hand, the ANOVA shows there is no significant difference between post-test optimism scores and gender ,F(1,60)=0,40, p=.843.

The following command can be used to put this data into graphical form.

> temp<-pols[pols$testtime=="post",]%>%group_by(party,sex)%>%
summarize(means=mean(optimismscore),sems=sd(optimismscore)/sqrt(length(optimismscore))) > library("gplots") > col1=col2hex("deeppink") > col2=col2hex("deepskyblue2") >
f<-ggplot(temp, aes(x=party, y=means, fill=sex))+
geom_bar(stat="identity",position=position_dodge())+
scale_fill_manual(values=c(col1,col2),name="Sex",breaks=c("female","male"),labels=c("Female", "Male"))+
theme(legend.key=element_rect(color="black"))+
geom_errorbar(aes(ymax=means+sems, ymin=means-sems),width=.2,position=position_dodge(.9))+
ggtitle("Posttest Optimism Scores by Sex and Political Affiliation")+
labs(x="Political Party Affiliation",y="Posttest Optimism Scores)")+
scale_x_discrete(breaks=c("democrat","independent","republican"),labels=c("Democrat","Independent","Republican"))+
theme(plot.title=element_text(size=15,face="bold",vjust=.5))+
theme(axis.title.x=element_text(size=12,face="bold",vjust=-.25))+
theme(axis.title.y=element_text(size=12,face="bold",vjust=1))+
theme(axis.text.x=element_text(size=10,face="bold",color="black"))+
theme(axis.text.y=element_text(size=10,face="bold",color="black"))+
coord_cartesian(ylim=c(min(tempsems),max(tempsems)))+
theme(panel.border=element_blank(),axis.line=element_line())+
theme(panel.grid.major.x=element_blank())+
theme(panel.grid.major.y=element_line(color="darkgrey"))+
theme(panel.grid.minor.y=element_blank())+
theme(legend.position=c(.2,.76))+
theme(legend.background=element_blank())+
theme(legend.background=element_rect(color="black"))+
theme(legend.title=element_blank())+
theme(legend.title=element_text(size=12))+
theme(legend.title.align=.5)+
theme(legend.text=element_text(size=10,face="bold")) > f

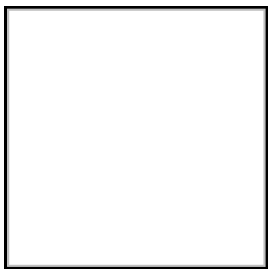


[This link]

Create a line graph for the same data. With error bars of course!

f<-ggplot(temp, aes(x=party, y=means, group=sex, color=sex))+
geom_line(size=1)+
geom_point(size=2)+
scale_color_manual(values=c(col1,col2),name="Sex",breaks=c("female","male"),labels=c("Female", "Male"))+
geom_errorbar(aes(ymax=means+sems, ymin=means-sems),width=.2)+
ggtitle("Optimism Scores by Sex and Political Affiliation")+
labs(x="Political Party Affiliation",y="Optimism Scores")+
scale_x_discrete(breaks=c("democrat","independent","republican"),labels=c("Democrat","Independent","Republican"))+
theme(plot.title=element_text(size=15,face="bold",vjust=.5))+

```
theme(axis.title.x=element_text(size=12,face="bold",vjust=-.25))+
theme(axis.title.y=element_text(size=12,face="bold",vjust=1))+
theme(axis.text.x=element_text(size=10,face="bold",color="black"))+
theme(axis.text.y=element_text(size=10,face="bold",color="black"))+
coord_cartesian(ylim=c(min(tempsems),max(tempsems)))+
theme(panel.border=element_blank(),axis.line=element_line())+
theme(panel.grid.major.x=element_blank())+
theme(panel.grid.major.y=element_line(color="darkgrey"))+
theme(panel.grid.minor.y=element_blank())+
theme(legend.position=c(.2,.76))+
theme(legend.background=element_blank())+
theme(legend.background=element_rect(color="black"))+
theme(legend.title=element_blank())+
theme(legend.title=element_text(size=12))+
theme(legend.title.align=.5)+
theme(legend.text=element_text(size=10,face="bold")) > f
```



[This Link]

Run a test to see if optimism scores pre- and post watching videos depends on sex.

summary(aov(optimismscore~testtime*sex+Error(subject/testtime),data=pols))

Error: subject

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

sex 1 80 80.4 0.119 0.731

Residuals 64 43105 673.5

Error: subject:testtime

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

testtime 1 770.9 770.9 41.299 1.87e-08 *### testtime:sex 1 0.9 0.9 0.049 0.825*
Residuals 64 1194.7 18.7
--- ### Signif. codes: 0 ‘’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

When using the ANOVA to determine if pre- and post-test optimism scores varies depending on gender it is shown that there is no relation between sex and optimism scores, F(1,64)= 0.12, p=0.73. However, the subjects were affected by watching the videos, F(1,64)= 41.3, p < .001. In addition, there is no relation to the pre-and post-test and gender, F(1,64)=0.049, p=0.825.

Run a Multiple Regression Analysis to see whether you can predict someones posttest optimism scores from their pretest optimism scores and party affiliation. Create a figure that shows the distributions of pre- and posttest optimism scores and sex

and create lines that best fit the data for both sexes regardless of whether the lines are warranted based on the statistical analyses

>
summary(lm(optimismscore[pols$testtime=="post"]~optimismscore[pols$testtime=="pre"]+party[pols$testtime=="pre"],data=pols))

Call:

lm(formula = optimismscore[pols$testtime == "post"] ~ optimismsco###re[pols$testtime == "pre"] + party[pols$testtime == "pre"], data = pols)

Residuals:

Min 1Q Median 3Q Max

-22.866 -2.562 1.267 3.901 8.948

Coefficients:

Estimate Std. Error t value

(Intercept) 8.1219 4.2616 1.906

optimismscore[pols$testtime == "pre"] 0.9419 0.0583 16.154

party[pols$testtime == "pre"]independent -1.3310 2.1284 -0.625

party[pols$testtime == "pre"]republican 0.7210 2.5000 0.288

Pr(>|t|)

(###Intercept) 0.0613 .

optimismscore[pols$testtime == "pre"] <2e-16 *###party[pols$testtime == "pre"]independent 0.5340*
###party[pols$testtime == "pre"]republican 0.7740
###--- ###Signif. codes: 0 ‘’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.022 on 62 degrees of freedom

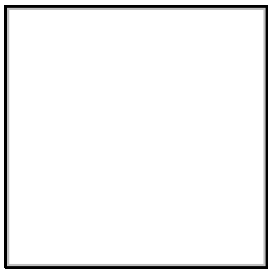
Multiple R-squared: 0.8971, Adjusted R-squared: 0.8922

F-statistic: 180.3 on 3 and 62 DF, p-value: < 2.2e-16

With the result of R*2 = 0.90, F(3,62) = 180.3, p<.001, it can been seen that post-test optimism scores can be predicted by pre-test optimism scores and political party affiliation.

f<-
ggplot(pols,aes(x=optimismscore[pols$testtime=="post"],y###=optimismscore[pols$testtime=="pre"],color=sex[pols$testtime=="pre"]))+geom_point(size=2)+
geom_abline(intercept=60.2+2.51/2, slope=-.092,color=col1)+
geom_abline(intercept=60.2-2.51/2, slope=-.092,color=col2)+
scale_color_manual(values=c(col1,col2),breaks=c("female","male"),labels=c("Female","Male"))+
ggtitle("Optimism Predicted by Sex and Pre-Test Optimism Score")+
labs(x="Pre-Test Optimism Score",y="Optimism Score(Higher=More)")+
theme(plot.title=element_text(size=15,face="bold", vjust=.5))+
theme(axis.title.x=element_text(size=12,face="bold", vjust=-.25))+
theme(axis.title.y=element_text(size=12,face="bold", vjust=1))+
theme(axis.text.x=element_text(size=10,face="bold",color="black"))+
theme(axis.text.y=element_text(size=10,face="bold",color="black"))+
theme(panel.border=element_blank(), axis.line=element_line())+
theme(panel.grid.major.x=element_blank())+

```
theme(panel.grid.minor.x=element_blank())+
theme(panel.grid.major.y=element_line(color="darkgrey"))+
theme(panel.grid.minor.y=element_blank())+
theme(legend.position=c(.83,.86))+
theme(legend.background=element_blank())+
theme(legend.title=element_blank())+
theme(legend.text=element_text(size=10,face="bold")) > f
```



this link