Chapter 3 Exercise 4

Allen Church

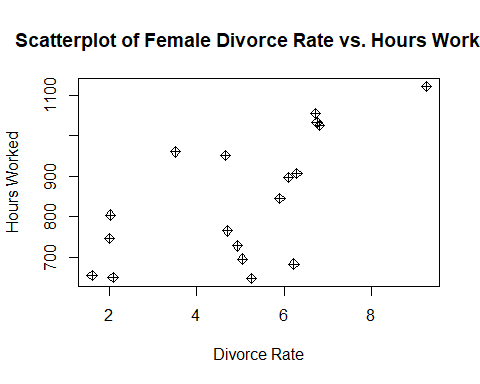
September 24, 2019

4a. For each data set (women and men), create scatterplot of hours worked on y axis and divorce rates on x axis

#Load datasets and view  
load("C:/Users/WB537822/Desktop/stats1/Ch3\_Exercise4\_Divorce\_rates\_Women.RData")  
load("C:/Users/WB537822/Desktop/stats1/Ch3\_Exercise4\_Divorce\_rates\_Men.RData")  
  
View(Mdata)  
View(Wdata)

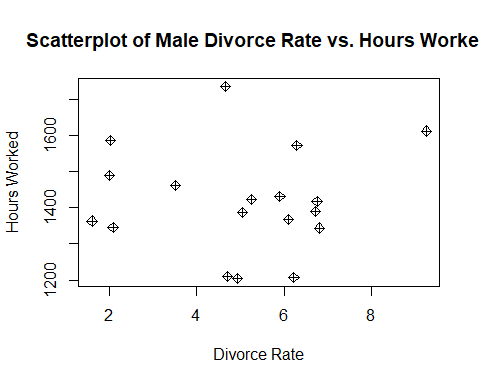
Scatterplot for women dataset

plot(Wdata$divorcerate, Wdata$hours, main="Scatterplot of Female Divorce Rate vs. Hours Worked",  
 xlab="Divorce Rate", ylab = "Hours Worked", pch=9)



Scatterplot for men dataset

plot(Mdata$divorcerate, Mdata$hours, main="Scatterplot of Male Divorce Rate vs. Hours Worked",  
 xlab="Divorce Rate", ylab = "Hours Worked", pch=9)



4b. For each data set estimate OLS regression in which hours worked is regressed on divorce rates. Report estimated regression equation and interpret coefficients. Explain differences in coefficients.

Women dataset OLS regression: The formula for this regression is the dependent variable = hours worked and the independent variable = divorce rate.

#Create OLS regression for women dataset  
ols\_women <- lm(Wdata$hours ~ Wdata$divorcerate)  
summary(ols\_women)

##   
## Call:  
## lm(formula = Wdata$hours ~ Wdata$divorcerate)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -218.645 -60.793 0.662 101.665 188.968   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 601.86 77.25 7.791 7.8e-07 \*\*\*  
## Wdata$divorcerate 48.28 14.36 3.362 0.00396 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 122.7 on 16 degrees of freedom  
## Multiple R-squared: 0.414, Adjusted R-squared: 0.3774   
## F-statistic: 11.3 on 1 and 16 DF, p-value: 0.003965

Men dataset OLS regression: The formula for this regression is the dependent variable = hours worked and the independent variable = divorce rate.

#Create OLS regression for men dataset  
ols\_men <- lm(Mdata$hours ~ Mdata$divorcerate)  
summary(ols\_men)

##   
## Call:  
## lm(formula = Mdata$hours ~ Mdata$divorcerate)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -214.89 -64.49 -18.40 67.69 317.80   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1410.642 93.087 15.154 6.55e-11 \*\*\*  
## Mdata$divorcerate 1.798 17.302 0.104 0.919   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 147.8 on 16 degrees of freedom  
## Multiple R-squared: 0.0006743, Adjusted R-squared: -0.06178   
## F-statistic: 0.0108 on 1 and 16 DF, p-value: 0.9185

The results between the two datasets above suggest that, since the beta hat 1 for women is 48.28 vs. beta hat 1 for men is 1.798, hours worked and divorce rate have a much higher positive correlation for women.

4c. What are fitted value and residual for men in Germany?

#Create a new dataframe only containing men data from Germany  
german\_men <- subset(Mdata, country == "Germany")

Fitted value for men in Germany

ols\_german\_men <- lm(german\_men$hours ~ german\_men$divorcerate)  
ols\_german\_men$fitted.values

## 1   
## 1204.6

Residual for men in Germany

summary(ols\_german\_men)

##   
## Call:  
## lm(formula = german\_men$hours ~ german\_men$divorcerate)  
##   
## Residuals:  
## ALL 1 residuals are 0: no residual degrees of freedom!  
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 1205 NA NA NA  
## german\_men$divorcerate NA NA NA NA  
##   
## Residual standard error: NaN on 0 degrees of freedom

The results above show there are 0 residuals for men in Germany.

4d. Fitted value and residual for women in Spain

#Create a new dataframe only containing women data from Spain  
spanish\_women <- subset(Wdata, country == "Spain")  
ols\_spanish\_women <- lm(spanish\_women$hours ~ spanish\_women$divorcerate)

Fitted value for women in Spain

ols\_spanish\_women$fitted.values

## 1   
## 650.4

Residual for women in Spain

summary(ols\_spanish\_women)

##   
## Call:  
## lm(formula = spanish\_women$hours ~ spanish\_women$divorcerate)  
##   
## Residuals:  
## ALL 1 residuals are 0: no residual degrees of freedom!  
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 650.4 NA NA NA  
## spanish\_women$divorcerate NA NA NA NA  
##   
## Residual standard error: NaN on 0 degrees of freedom

The results above show there are 0 residuals for women in Spain.