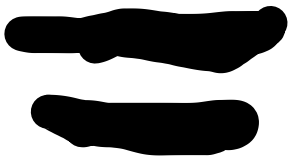
Final Project



Dataset: sexlierel.dat

Source: R.R. Clayton (1971). "Religiosity and Premarital Sexual Permissiveness: Elaboration of the Relationship and Debate," Sociological Analysis, Vol. 32, #2, pp81-96

Description: Tabulation among 4 scales of premarital sexual permissiveness, religiosity, propensity to lie and gender.

Variables/Columns Gender 8 /* 1=Female, 2=Male */ *Scale: 16* / 1=Ritualistic, 2=Experiential, 3=Ideological, 4=Composite/ *Premarital sexual permissiveness 24* / 1=Low, 2=High / *Propensity to Lie 32* / 1=Lower, 2=Higher / *Religiosity 40* / 1=Low, 2=High */ Count 46-48

Introduction

Given the data above, I would like to determine the variation between the variables above. More specifically, is there a correlation between Religiosity and Premarital sexual permissiveness. Quantities are given to us through the value of count.

```
data_set <- read.csv("project_data.csv", header = TRUE)

names(data_set)
```

```
## [1] "Gender"                          "Scale"
## [3] "Premarital.Sexual.Permissiveness" "Propensity.to.Lie"
## [5] "Religiosity"                     "Count"
```

```
summary(data_set)
```

```
##      Gender        Scale      Premarital.Sexual.Permissiveness
##  Min.   :1.0   Min.   :1.00   Min.   :1.0
##  1st Qu.:1.0   1st Qu.:1.75   1st Qu.:1.0
##  Median :1.5   Median :2.50   Median :1.5
##  Mean   :1.5   Mean   :2.50   Mean   :1.5
##  3rd Qu.:2.0   3rd Qu.:3.25   3rd Qu.:2.0
##  Max.   :2.0   Max.   :4.00   Max.   :2.0
##  Propensity.to.Lie  Religiosity      Count
##  Min.   :1.0       Min.   :1.0   Min.   :  6.00
```

```
##  1st Qu.:1.0       1st Qu.:1.0   1st Qu.: 19.50
##  Median :1.5       Median :1.5   Median : 30.50
##  Mean   :1.5       Mean   :1.5   Mean   : 35.59
##  3rd Qu.:2.0       3rd Qu.:2.0   3rd Qu.: 50.25
##  Max.   :2.0       Max.   :2.0   Max.   :112.00
```

eda

```
kable(data_set[1:65,])
```

	Gender	Scale	Premarital.Sexual.Permissiveness	Propensity.to.Lie	Religiosity	Count
1	1	1	1	1	1	52
2	1	1	1	1	2	74
3	1	1	1	2	1	50
4	1	1	1	2	2	51
5	1	1	2	1	1	34
6	1	1	2	1	2	13
7	1	1	2	2	1	41
8	1	1	2	2	2	20
9	1	2	1	1	1	57
10	1	2	1	1	2	69
11	1	2	1	2	1	50
12	1	2	1	2	2	51
13	1	2	2	1	1	38
14	1	2	2	1	2	9
15	1	2	2	2	1	43
16	1	2	2	2	2	18
17	1	3	1	1	1	17
18	1	3	1	1	2	112
19	1	3	1	2	1	26
20	1	3	1	2	2	75
21	1	3	2	1	1	16
22	1	3	2	1	2	31
23	1	3	2	2	1	30
24	1	3	2	2	2	31
25	1	4	1	1	1	44
26	1	4	1	1	2	83
27	1	4	1	2	1	43
28	1	4	1	2	2	58
29	1	4	2	1	1	30
30	1	4	2	1	2	17
31	1	4	2	2	1	38
32	1	4	2	2	2	23
33	2	1	1	1	1	25
34	2	1	1	1	2	17
35	2	1	1	2	1	30
36	2	1	1	2	2	28
37	2	1	2	1	1	59
38	2	1	2	1	2	6
39	2	1	2	2	1	58
40	2	1	2	2	2	10

	Gender	Scale	Premarital.Sexual.Permissiveness	Propensity.to.Lie	Religiosity	Count
41	2	2	1	1	1	21
42	2	2	1	1	2	20
43	2	2	1	2	1	32
44	2	2	1	2	2	26
45	2	2	2	1	1	51
46	2	2	2	1	2	14
47	2	2	2	2	1	56
48	2	2	2	2	2	12
49	2	3	1	1	1	15
50	2	3	1	1	2	28
51	2	3	1	2	1	16
52	2	3	1	2	2	43
53	2	3	2	1	1	38
54	2	3	2	1	2	27
55	2	3	2	2	1	48
56	2	3	2	2	2	22
57	2	4	1	1	1	24
58	2	4	1	1	2	18
59	2	4	1	2	1	23
60	2	4	1	2	2	35
61	2	4	2	1	1	52
62	2	4	2	1	2	12
63	2	4	2	2	1	57
64	2	4	2	2	2	11
NA	NA	NA	NA	NA	NA	NA

```
lmOut <- lm(Religiosity ~ Count, data = data_set)
summary(lmOut)
```

```
##
## Call:
## lm(formula = Religiosity ~ Count, data = data_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5573 -0.4933 -0.0086  0.4671  0.7125
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.598984   0.126272  12.663   <2e-16 ***
## Count       -0.002781   0.003073  -0.905    0.369
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5047 on 62 degrees of freedom
## Multiple R-squared:  0.01304,    Adjusted R-squared:  -0.002883
## F-statistic: 0.8189 on 1 and 62 DF,  p-value: 0.369
```

Interperting the data

About 1344 females and 934 males in this data.

335 females say they are ritualistic. 227 had low sexual permissiveness. 108 had high sexual permissiveness.

335 females say they are experimental. 227 had low sexual permissiveness. 108 had high sexual permissiveness.

338 females say they are ideological. 230 had low sexual permissiveness. 108 had high sexual permissiveness.

336 females say they are composite. 228 had low sexual permissiveness. 108 had high sexual permissiveness.

---

233 males say they are ritualistic. 100 had low sexual permissiveness. 133 had high sexual permissiveness.

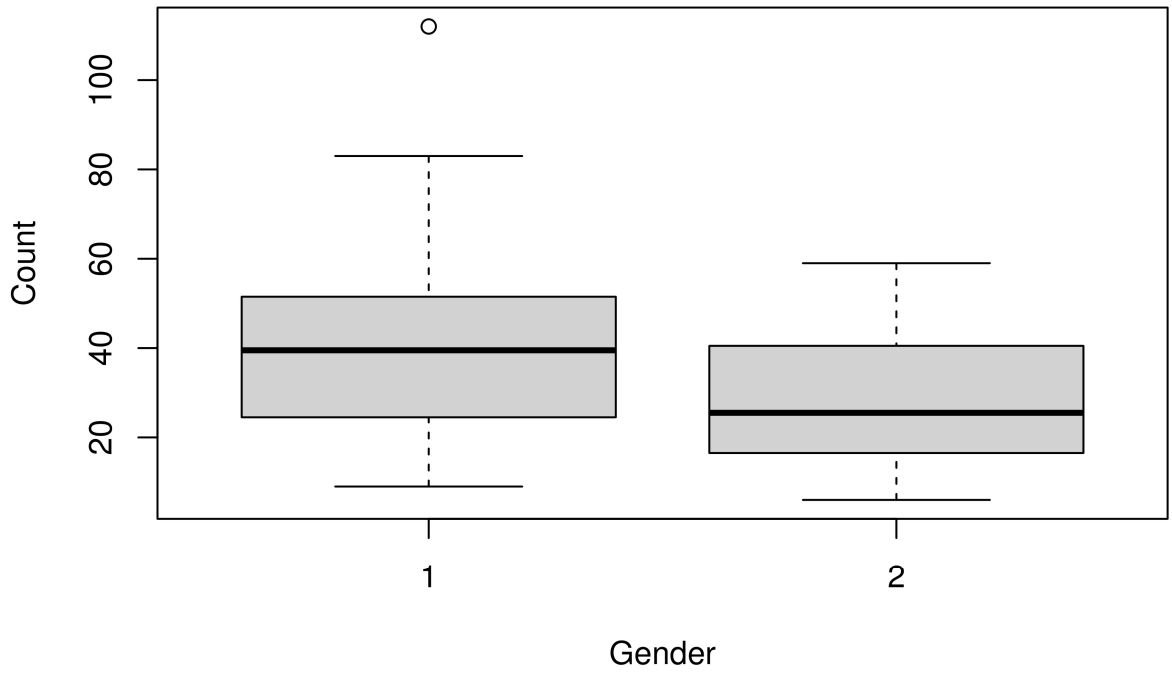
232 males say they are experimental. 99 had low sexual permissiveness. 133 had high sexual permissiveness.

237 males say they are ideological. 102 had low sexual permissiveness. 135 had high sexual permissiveness.

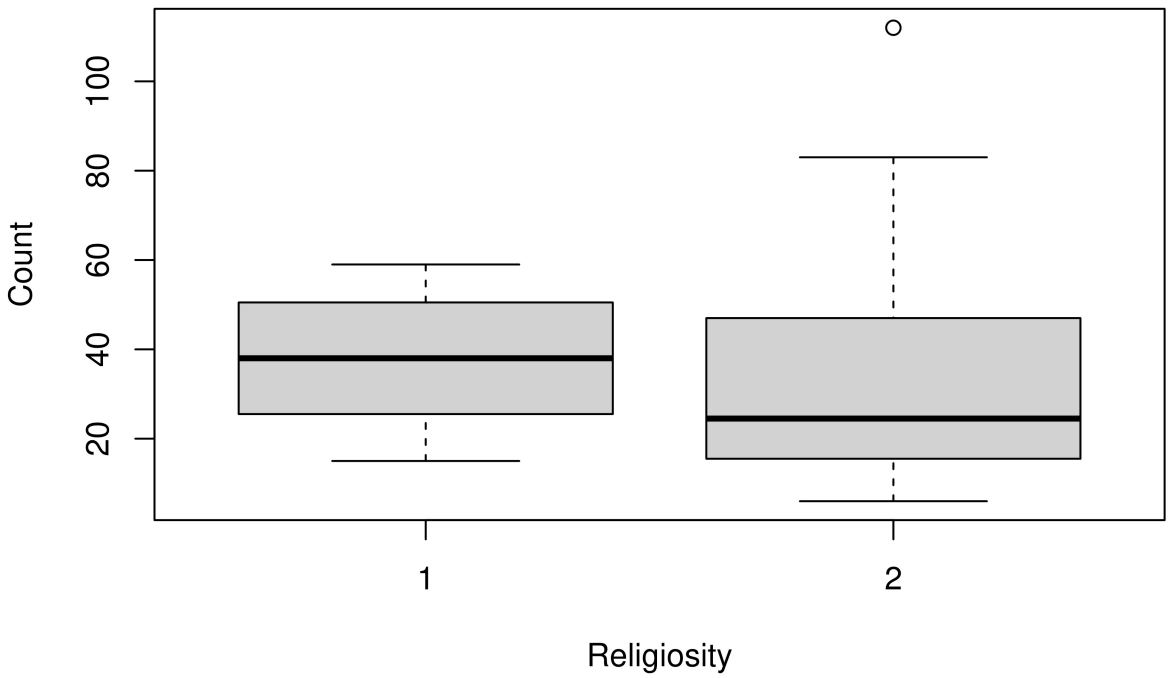
232 males say they are composite. 100 had low sexual permissiveness. 132 had high sexual permissiveness.

Boxplots

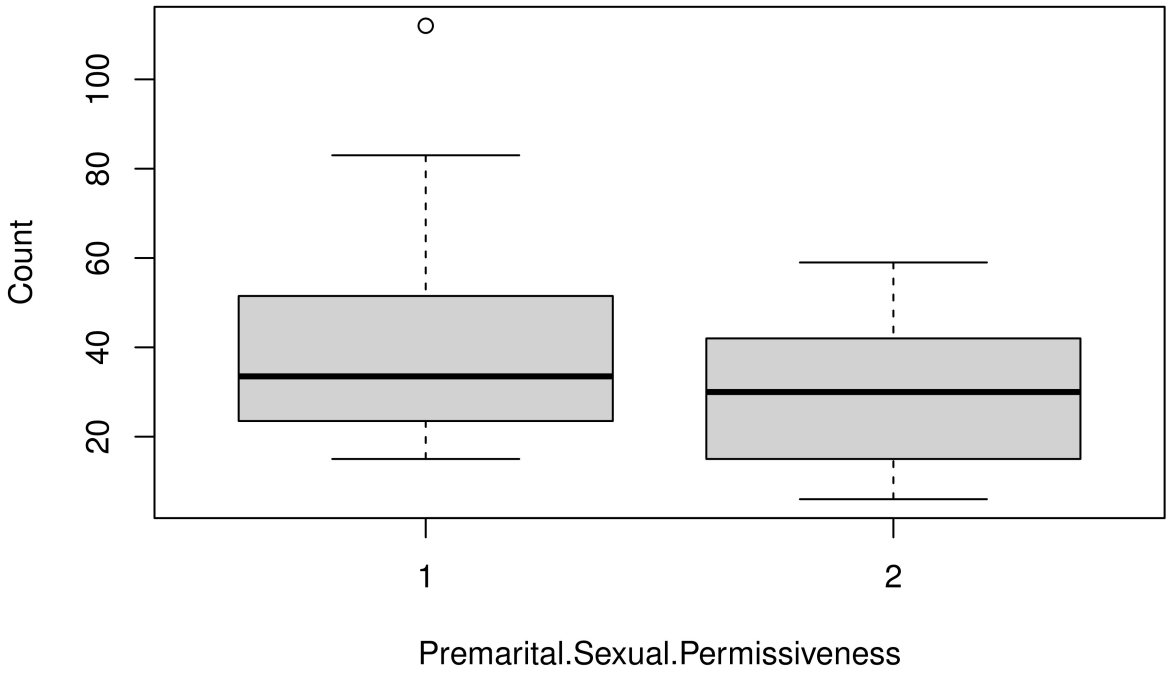
```
boxplot(Count~Gender, data_set)
```



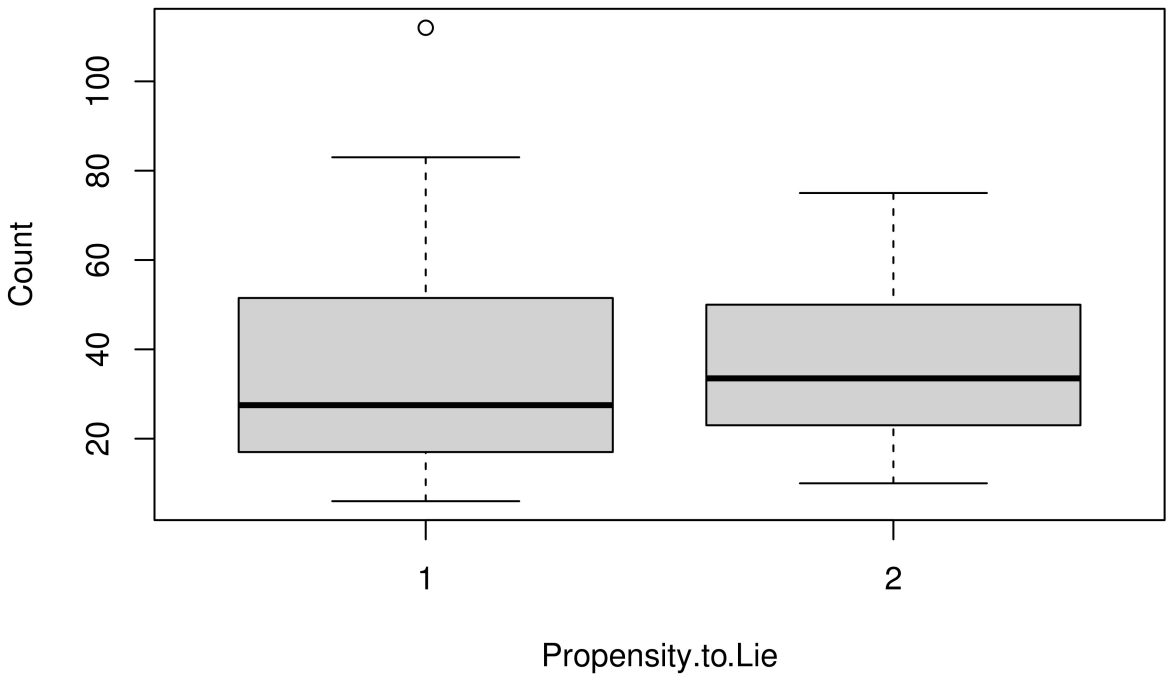

```
boxplot(Count~Religiosity, data_set)
```




```
boxplot(Count~Premarital.Sexual.Permissiveness, data_set)
```




```
boxplot(Count~Propensity.to.Lie, data_set)
```



Analysis From the data given, we are look at four distinct groups. 1. High religiosity individuals with low sexual permissiveness. 2. Low religiosity individuals with high sexual permissiveness. 3. High sexual permissiveness individuals with lower propensity to lie. 4. Low sexual permissiveness individuals with higher propensity to lie.

Simple Linear Regression

```
data_1<-data_set
Linear_fit<-lm(Count~Religiosity,data = data_1)
Linear_fit$coefficients
```

```
## (Intercept) Religiosity
##     42.6250     -4.6875
```

CI

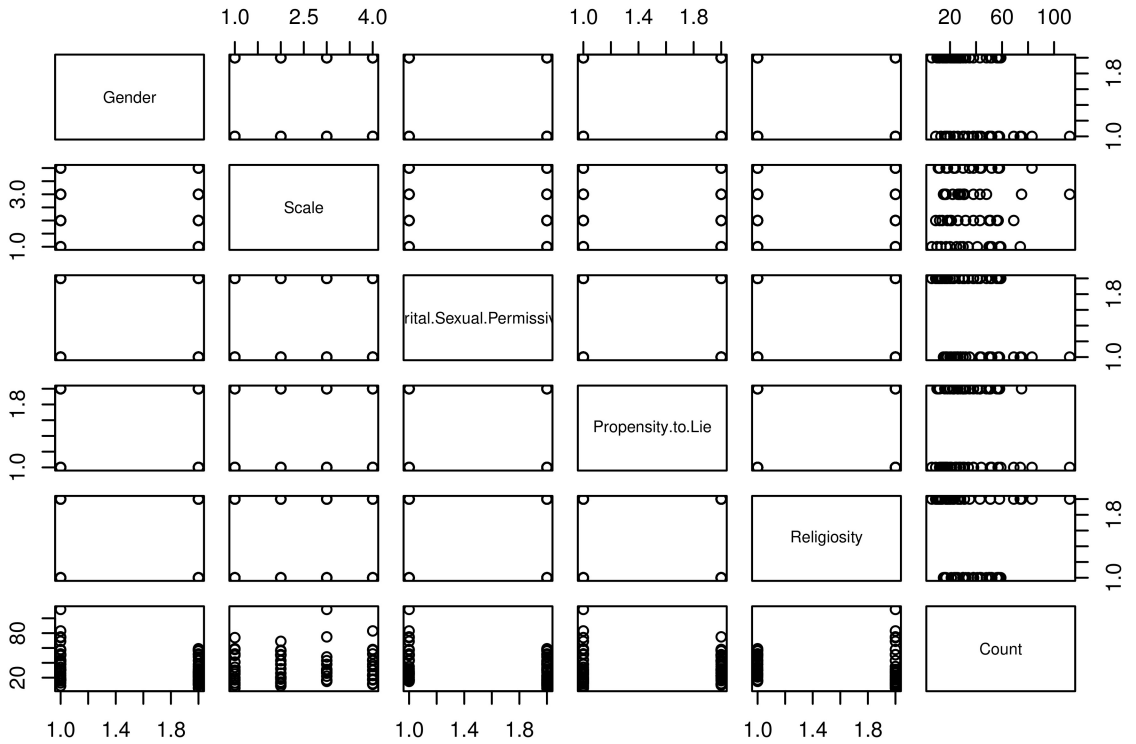
```
confint(Linear_fit, level=0.95)
```

```
##                 2.5 %    97.5 %
## (Intercept)  26.25278 58.997216
## Religiosity -15.04220  5.667199
```

Above is the confidence Interval

Matrix

```
pairs(data_set)
```



The count column is going to be the most hepful here. Indicates the number of indivuals in each group.

```
#linear regression
fit1<-lm(Count ~., data=data_set)
fit1
```

```
##
## Call:
## lm(formula = Count ~ ., data = data_set)
##
## Coefficients:
##                       (Intercept)                             Gender
##                            76.531                            -12.812
##                             Scale  Premarital.Sexual.Permissiveness
##                             0.050                            -10.875
##                  Propensity.to.Lie                         Religiosity
##                             1.000                             -4.688
```

```
fit2<-lm(Propensity.to.Lie ~., data=data_set)
fit2
```

```
##
## Call:
## lm(formula = Propensity.to.Lie ~ ., data = data_set)
##
## Coefficients:
##                       (Intercept)                             Gender
##                         1.4435039                          0.0092765
##                             Scale  Premarital.Sexual.Permissiveness
##                        -0.0000362                          0.0078737
##                       Religiosity                              Count
##                         0.0033938                          0.0007240
```

```
fit3<-lm(Religiosity ~., data=data_set)
fit3
```

```
##
## Call:
## lm(formula = Religiosity ~ ., data = data_set)
##
## Coefficients:
##                       (Intercept)                             Gender
##                         1.7323438                         -0.0428332
##                             Scale  Premarital.Sexual.Permissiveness
##                         0.0001672                         -0.0363560
##                 Propensity.to.Lie                              Count
##                         0.0033431                         -0.0033431
```

```
# Obtaining CI for model coefficients
confint(fit1, level=0.95)
```

```
##                                     2.5 %     97.5 %
## (Intercept)                      44.890182 108.172318
## Gender                          -22.577145  -3.047855
## Scale                            -4.316882   4.416882
## Premarital.Sexual.Permissiveness -20.639645  -1.110355
## Propensity.to.Lie                -8.764645  10.764645
## Religiosity                      -14.452145   5.077145
```

```
confint(fit2, level=0.95)
```

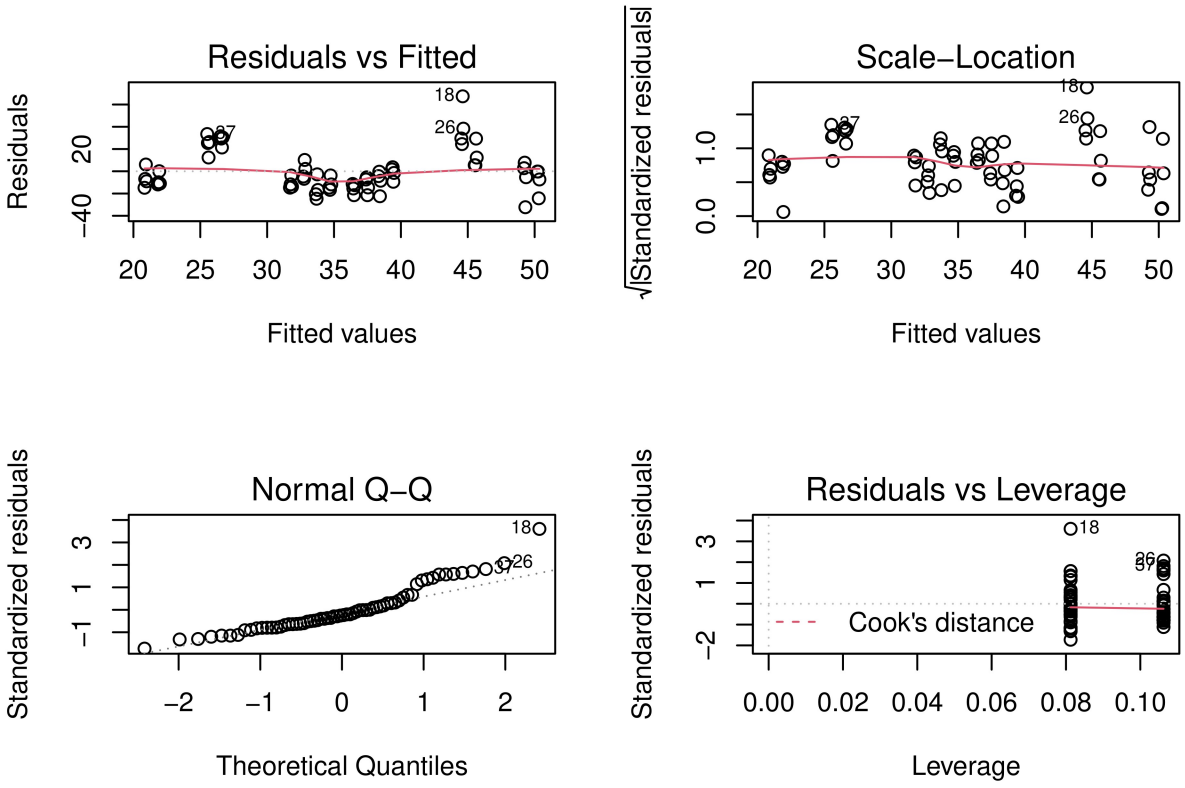
```
##                                       2.5 %     97.5 %
## (Intercept)                      0.508696791 2.37831097
## Gender                          -0.268642629 0.28719562
## Scale                           -0.117539071 0.11746667
## Premarital.Sexual.Permissiveness -0.265887396 0.28163481
## Religiosity                     -0.261431070 0.26821875
## Count                           -0.006345771 0.00779381
```

```
confint(fit3, level=0.95)
```

```
##                                       2.5 %      97.5 %
```

```
## (Intercept)                          0.84056539 2.624122252
## Gender                              -0.31844675 0.232780415
## Scale                               -0.11645362 0.116787931
## Premarital.Sexual.Permissiveness -0.30790164 0.235189726
## Propensity.to.Lie                   -0.25752072 0.264206873
## Count                               -0.01030712 0.003620968
```

```
# Diagnostic plots
layout(matrix(c(1,2,3,4),2,2))
plot(fit1)
```



Analysis

The points on the Q-Q plot roughly fall in a straight line. There are not enough observations that deviate from the line. We can declare that the residuals are normally distributed. We can also tell from the Scale-Location graph that equal variance is not violated since the values on the plot don't deviate to far from the horizontal line. The residual vs. fitted plot tells us that the residuals follow a linear pattern since the red line is roughly horizontal. However, the Residuals vs Leverage plot has multiple values that fall with cook's distance. This means there could be overly influential point in our data set.

Anova Table

```
# compare models
anova(fit1, fit2)
```

```
## Warning in anova.lmlist(object, ...): models with response '"Propensity.to.Lie"'
## removed because response differs from model 1
```

```
## Analysis of Variance Table
##
## Response: Count
##                                 Df  Sum Sq Mean Sq F value  Pr(>F)
## Gender                           1  2626.6 2626.56  6.8986 0.01102 *
## Scale                            1     0.2    0.20  0.0005 0.98179
## Premarital.Sexual.Permissiveness  1  1892.3 1892.25  4.9699 0.02968 *
## Propensity.to.Lie                1    16.0   16.00  0.0420 0.83829
## Religiosity                      1   351.6  351.56  0.9234 0.34058
## Residuals                       58 22082.9  380.74
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Analysis

since a lot of the F values are larger than alpha, we would reject the null hypothesis for the different variables.

Conclusion

Through multiple linear regression, we are able to depict a model that estimates the relationship between one independent variable and one dependent variable using lines. We also used ANOVA tables to compare the population means.