A6

Stat 301

Summer 2019

Instructions: Follow submission directions in BbLearn

(1) The article "Characterization of Highway Runoff in Austin, Texas, Area"[1] gave a scatterplot, along with the regression equation that modeled runoff by rainfall (with units of measurements volumes of water in m^3) for a particular location. All output needed for this is included below.

(a) Write the population model; identify all components of the model

(b) Does the relationship between rainfall volume and runoff volume appear to be linear? Describe the scatterplot

(c) State the regression equation (with the values from the regression output)

(d) Interpret the slope and intercept in context. If one does not make logical sense, state why not.

(e) Using the regression equation, estimate the runoff volume for rainfall amounts of 5, 47, and 127 m^3

(f) Calculate the residuals for each estimated value. The observed values of y for $x = 5,\ 47,\ 127$ are (respectively): $y = 4,\ 46,\ 100$

(g) Conduct a test of significance for the slope (results from `R` output). List hypotheses, test statistic, pvalue, result, and conclusion.

(h) List the values of r and R^2 (r will have to be computed). Interpret both values in context. What do they tell us about the model?

(i) List assumptions of regression. Check the assumptions (briefly describe how each assumption is met or not).

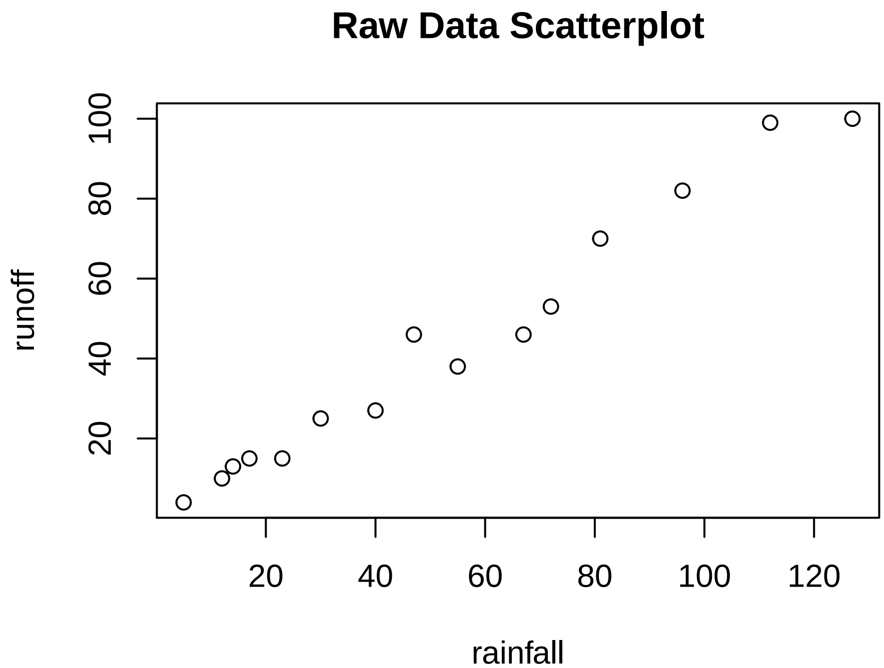
(j) Overall, what is your assessment of the model. Good, bad? Use results from significance test, r, R^2, and assumptions to make this assessment.

```r
head(old.man.is.snoring)
```

```
  rainfall runoff
1        5      4
2       12     10
3       14     13
4       17     15
5       23     15
6       30     25
```

```r
with(old.man.is.snoring,plot(runoff~rainfall,main='Raw Data Scatterplot'))
```

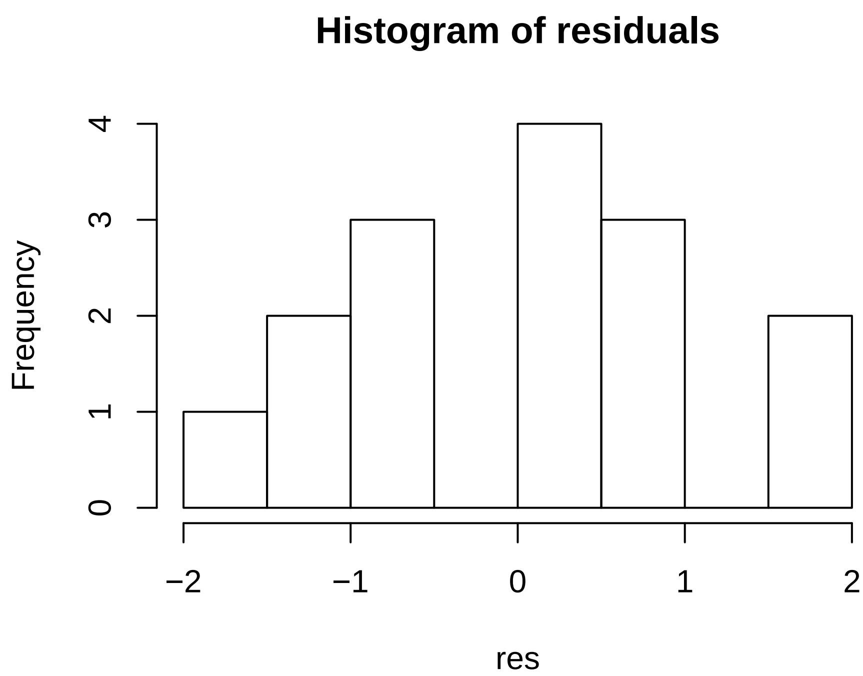
[1] "Characterization of Highway Runoff in Austin, Texas Area", *J. of Envir. Engr.*, 1998: 131-137



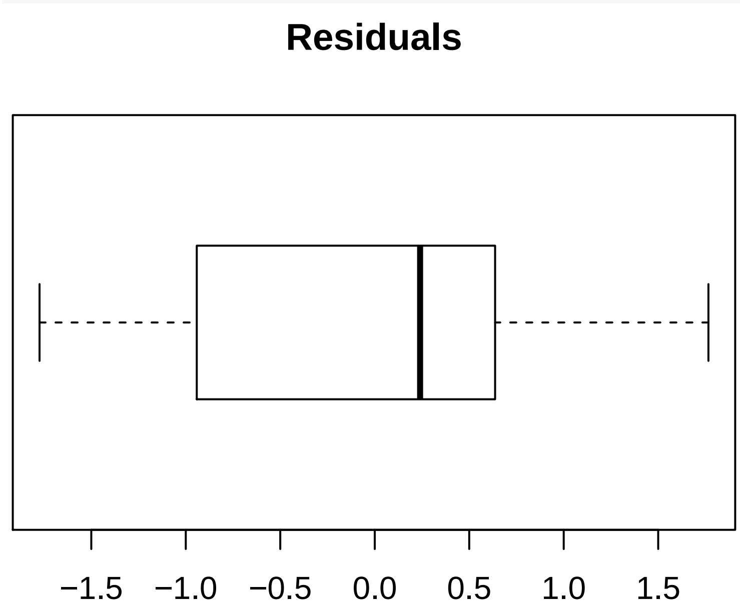

```
texas=lm(runoff~rainfall,data=old.man.is.snoring)
summary(texas)
```

```

Call:
lm(formula = runoff ~ rainfall, data = old.man.is.snoring)

Residuals:
   Min     1Q Median     3Q    Max 
-8.279 -4.424  1.205  3.145  8.261 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.12830    2.36778  -0.477    0.642    
rainfall     0.82697    0.03652  22.642  7.9e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.24 on 13 degrees of freedom
Multiple R-squared:  0.9753,	Adjusted R-squared:  0.9734 
F-statistic: 512.7 on 1 and 13 DF,  p-value: 7.896e-12
```

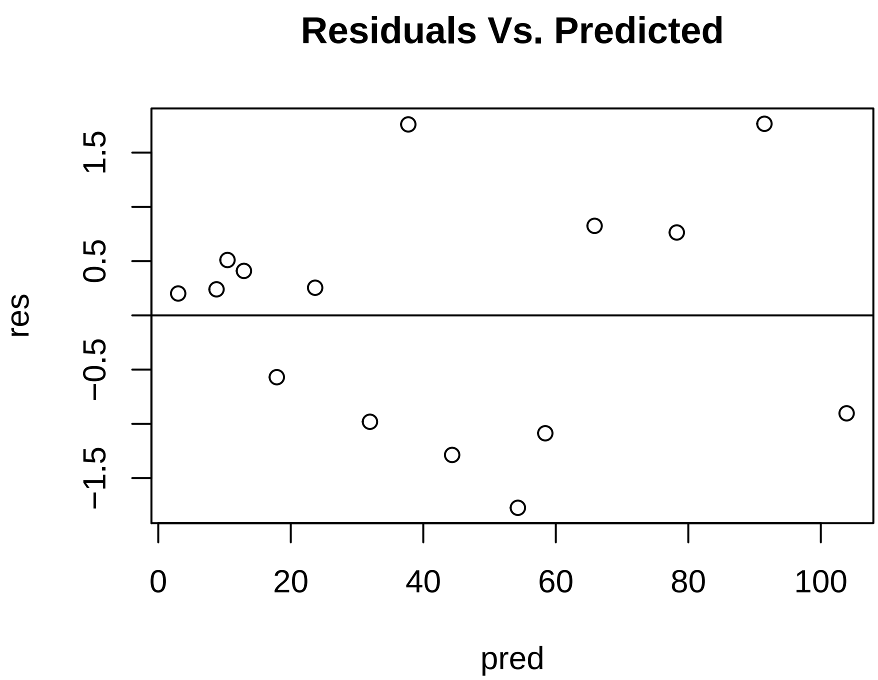
```
res=rstudent(texas); pred=fitted(texas)
hist(res,breaks=6,main='Histogram of residuals')
```



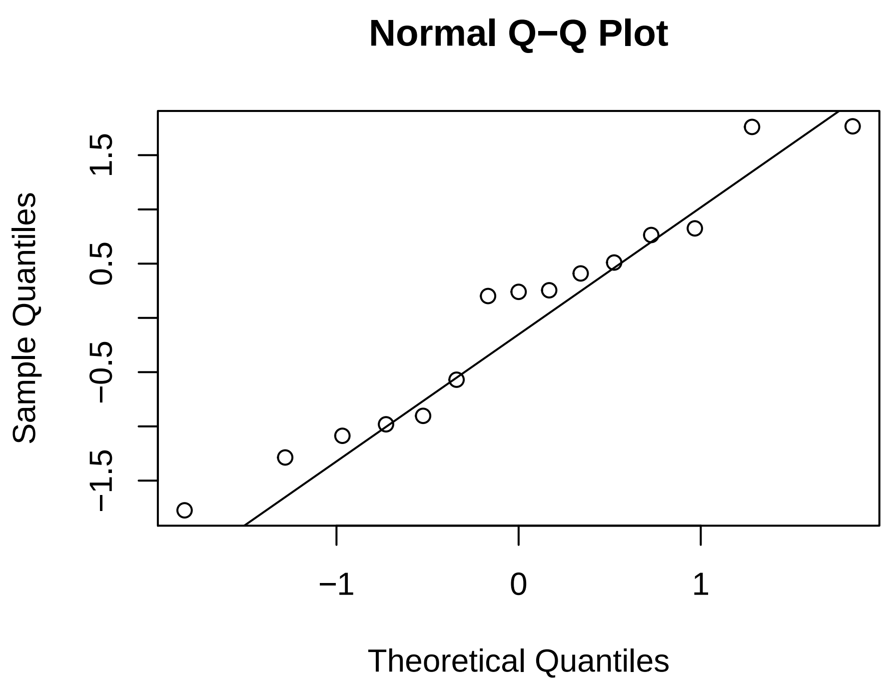

```
boxplot(res,horizontal=T,main='Residuals')
```




```
plot(pred,res,main='Residuals Vs. Predicted'); abline(0,0)
```




```r
qqnorm(res); qqline(res)
```



(2) A six-sided die is rolled 120 times. Fill in the expected frequency column. Then conduct a hypothesis test to determine if the die is fair; include the kind of error that could have been made.

Face value	Frequency	Expected
1	15	
2	29	
3	16	
4	15	
5	30	
6	15	

(3) A major food manufacturer is concerned that the sales for its skinny french fries have been decreasing. As a part of a feasibility study, the company conducts research into the types of fries sold across the country to determine if the type of fries sold is independent of the area of the country. The results of the study are shown in the table below. Conduct a test of independence.

Type of Fries	*Northeast*	*South*	*Central*	*West*	**Total**
skinny	70	50	20	25	165
curly	100	60	15	30	205
steak	20	40	10	10	80
Total	190	150	45	65	450