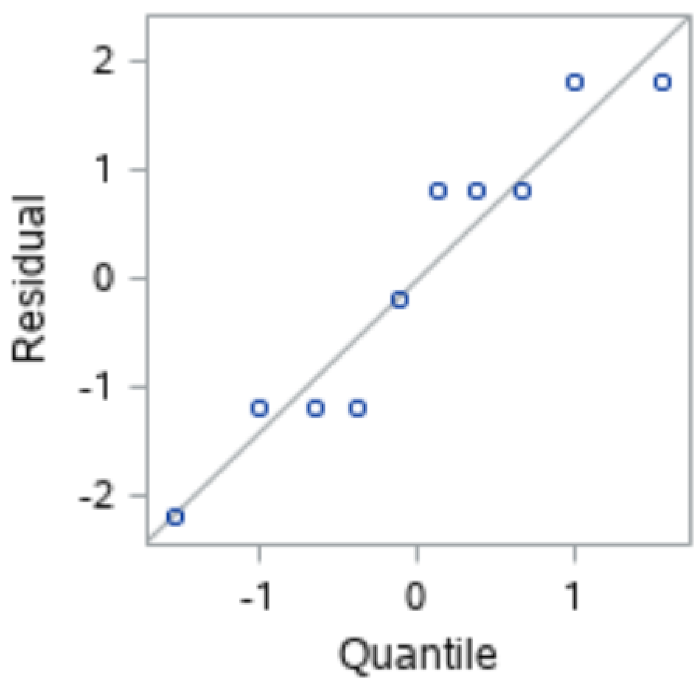
STT 4660
Homework #6

Timothy Stubblefield

October 14, 2020

3.5 (e)

1) First, the normal Probability Plot is below,



2) To calculate the correlation of coefficient consider the equation,

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

3) Using the following SAS Code, I obtained the Pearson Correlation Coeficcient (r),

```
/* Find Pearson's Correlation Coefficient */
proc corr data=airfreight;
    var y x;
run;
```

4) As a result, the associated printout in SAS is,

Pearson Correlation Coefficients, N = 10 Prob > |r| under H0: Rho=0		
	y	x
y	1.00000	0.94916 <.0001
x	0.94916 <.0001	1.00000

5) Therefore, we have the following,

$$\text{Correlation Coefficient} = r = 0.94916$$

6) We are setting up an F-test for normality where,

$$H_0 : F = normal$$

$$H_a : F \neq normal$$

7) Now, compare r to the value in Table B.6 when $\alpha = 0.01$ and n = 10

$$\text{Critical Value} = 0.879$$

8) So the critical region is $r < 0.879$.

9) Since $r = 0.9416 > 0.879$, our correlation coefficient is outside the critical range and thus, we fail to reject H_0, essentially concluding H_0.

10) Therefore, there is evidence to show that the data follows a normal distribution.

3.5 (g)

1) So we know that $\chi^2_{Bp} = 1.03$ from the problem.

2) I will now show how the B-P test statistic was computed.

3) From the notes, we have the equation,

$$\chi^2_{Bp} = \frac{SSR^*/2}{(SSE/n)^2}$$

4) So, SSR* is just the SSR for the model,

$$y_i' = e_i^2 = r_0 + r_1 x_i + \delta_i$$

5) When running the model in SAS, the residuals are provided by the print-out. Now, we can square each e_i to get e_i^2 or y_i' as it is referred to in the new model.

6) Below is the SAS code that shows the data of e_i^2 vs x_i and the generation of the linear model

```
/* Find the SSR* */
data res;
    input e x;
cards;
3.24 1
1.44 0
1.44 2
3.24 0
0.04 3
1.44 1
4.84 0
0.64 1
0.64 2
0.64 0
;
run;

proc reg data=res;
     model e=x; * obtain SSR* from the table and calculate the test s
run;
```

7) Now, the ANOVA Table for the e_i^2 linear regression model is,

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6.40000	6.40000	3.53	0.0970
Error	8	14.49600	1.81200		
Corrected Total	9	20.89600			

8) From the new ANOVA Table, we can see that SSR* = 6.4

9) From original ANOVA Table from previous problems, we know that SSE = 17.6

10) Now substitute our values in to the B-P test statistic formula as such,

$$\chi^2_{Bp} = \frac{6.4/2}{(17.6/10)^2}$$

$$\chi^2_{Bp} = \frac{3.2}{3.0976}$$

$$\chi^2_{Bp} = 1.033057851$$

Now, rounding this to two decimals we have,

$$\chi^2_{Bp} = 1.03$$

■

11) Now, to perform the hypothesis test of,

$$H_0 : r_1 = 0 \quad \text{(constant variance)}$$

$$H_a : r_1 \neq 0 \quad \text{(nonconstant variance)}$$

12) So we know $\chi^2_{Bp} \sim \chi^2(1) \quad \text{under} \quad H_0$

13) Now, using SAS, I computed the p-value,

$$p - value = 0.23487$$

14) So $p - value = 0.38487 > 0.1 = \alpha$.

15) Therefore, do not reject H_0, and thus conclude H_0.

16) Thus, there is evidence to say that the residuals have constant variance.

3.15

a) First, we want to fit a linear regression function.

   1) By using SAS, I found that the regression function is,

$$y = 2.57533 - 0.3240x$$

b) Now, perform and F-test for Lack of Fit using $\alpha = 0.025$.

   1) Let's perform an F-test to test the hypothesis,

$$H_0 : \text{Linear Model is Correct}$$

$$H_a : \text{Linear Model is Not Correct}$$

2) In SAS, you can use the /lackfit after the model type to generate the ANOVA Table for the Lack of Fit test.
3) The ANOVA Table for Lack of Fit test is below,

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	12.59712	12.59712	55.99	<.0001
Error	13	2.92465	0.22497		
Lack of Fit	3	2.76725	0.92242	58.60	<.0001
Pure Error	10	0.15740	0.01574		
Corrected Total	14	15.52177			

4) From the notes, we have that,

$$F^* = \frac{SSLF/df(SSLF)}{SSE(F)/df(SSE(F)} = \frac{MSLF}{MSPE}$$

5) By looking at the ANOVA Table, we can see that,

$$MSLF = 0.92242$$

$$MSPE = 0.01574$$

6) Thus, substitute in values to compute F*,

$$F^* = \frac{0.92242}{0.01574}$$

$$F^* = 58.60$$

(As you see, we could have also got F* from the ANOVA Table)

7) And, we know $F^* \sim F(df(SSLF), df(SSPE))$, so

$$F^* \sim F(3, 10)$$

8) From SAS we have that,

$$p - value = 0.000000096$$

9) Hence, $p - value = 0.000000096 < 0.025 = \alpha$.
10) Therefore, we reject H_0 and conclude H_a.
11) Thus, there is evidence to show that the linear model we built does not accurately suit the data.

c) Finally, do we know what regression model is appropriate for the data?

1) In Part (b), we determine that our simple linear regression model does not fit the data.

2) Even though we concluded that a linear regression function is not appropriate to model the data, we still do not know what model is appropriate for the data. Potentially, a quadratic regression model or exponential regression model would best fit the data. However, we do not know. In order to determine an appropriate model, we would need to build more models and then perform Lack of Fit tests until one of them is accepted by hypothesis testing.