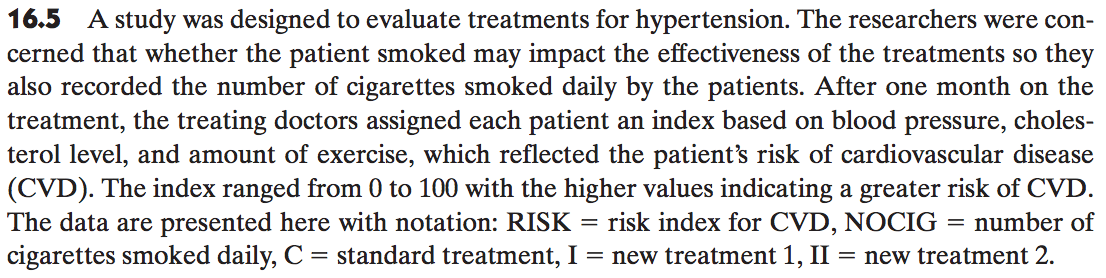
**CSUEB – STAT 6305 – Winter 2017 - Prof Yan Zhou**

**Homework 5 - Henry Lankin, Gui Larangeira**

February 14, 2017

**HW 5: 16.5, 16.6**

16.5



1. Write a model for the above experiment. Make sure to identify all variables and parameters in your model.

ANCOVA design model:

Using an ANCOVA treatment structure with:

– the risk index given to each patient representing the observed response variable: 30 observations from 10 replications of 3 possible treatments

– regression intercept of on

– the effect due to the treatment: 3 treatment levels

– regression (slope) coefficient of on

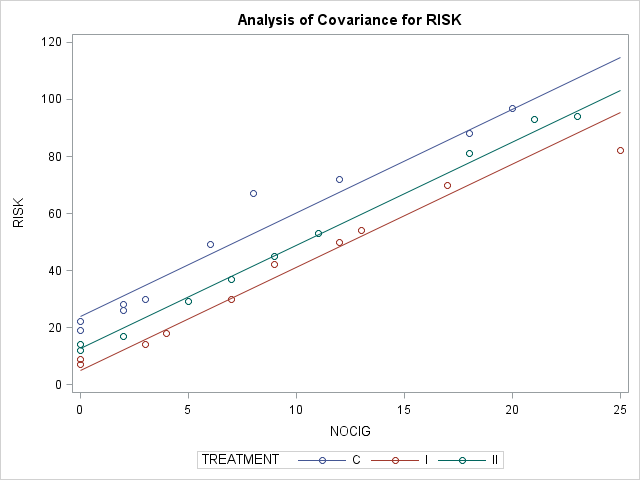
– the covariate effect due to the number of cigarettes smoked daily: 30 total values with experimental unit receiving level of treatment

– random error associated with each treatment-cigarette value combination: 30 residual errors

Reduced Model I (ANOVA):

Reduced Model II (ANOVA):

1. Provide a scatterplot of the data with regression lines which would allow a visual assessment of whether there is a significant relationship between risk index of CVD and the number of cigarettes smoked.



1. From your scatterplot in (b), do the three lines appear to have similar slopes?

As we can see in the scatterplot above, the three lines have very similar slopes, implying the relationship between covariate and the observed response is linear.

16.6



1. Test the hypothesis that the relationships between risk index and number of cigarettes have equal slopes for the three treatments at the level.

Consider the GLM model:

Treatment I:

Treatment II:

Control:

Thus, our hypotheses are:

Test statistic:

No interaction/equal slopes model:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 3 | 23230.96256 | 7743.65419 | 327.90 | <.0001 |
| **Error** | 26 | 614.00411 | 23.61554 |  |  |
| **Corrected Total** | 29 | 23844.96667 |  |  |  |

Interaction model/unequal slopes model:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 5 | 23358.04637 | 4671.60927 | 230.26 | <.0001 |
| **Error** | 24 | 486.92029 | 20.28835 |  |  |
| **Corrected Total** | 29 | 23844.96667 |  |  |  |

Rejection region:

At the significance level, the rejection region is given by .

Conclusion:

Since , we fail to reject the null hypothesis and conclude that the relationships between the risk index and number of cigarettes has equal slopes for each level of treatment.

This can also be seen with the -value of for the interaction term in the following table, confirming that we fail to reject the null hypothesis and the slopes are equal for each level of treatment.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **NOCIG** | 1 | 22164.04588 | 22164.04588 | 1092.45 | <.0001 |
| **TREATMENT** | 2 | 469.53955 | 234.76978 | 11.57 | 0.0003 |
| **NOCIG\*TREATMENT** | 2 | 127.08382 | 63.54191 | 3.13 | 0.0619 |

1. Does there appear to be a difference in the mean risk index for the three treatments?

Hypothesis:

Test statistic:

Reduced model (regression):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 21410.85065 | 21410.85065 | 246.29 | <.0001 |
| **Error** | 28 | 2434.11602 | 86.93271 |  |  |
| **Corrected Total** | 29 | 23844.96667 |  |  |  |

Full model AOV table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 3 | 23230.96256 | 7743.65419 | 327.90 | <.0001 |
| **Error** | 26 | 614.00411 | 23.61554 |  |  |
| **Corrected Total** | 29 | 23844.96667 |  |  |  |

Rejection region:

At the significance level, the rejection region is given by .

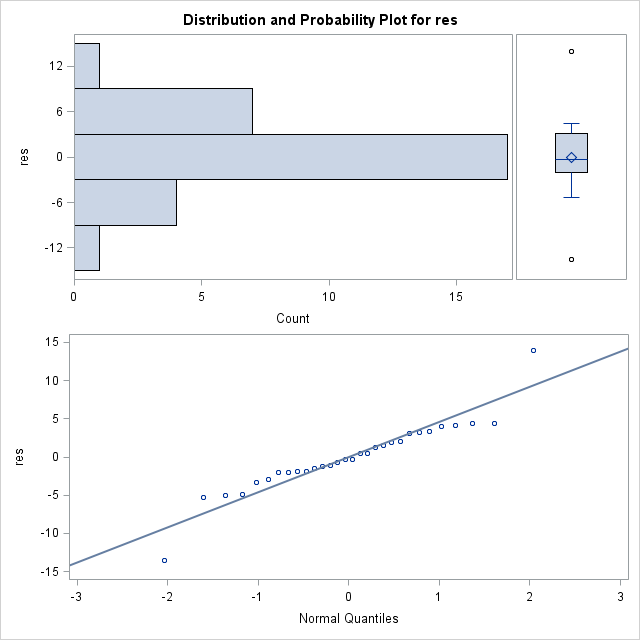
Conclusion:

Since , we reject the null hypothesis and conclude that there is not a significant difference in the mean risk index between the three treatments.

This can also be seen with the -value for the treatment term in the following table, confirming that we reject the null hypothesis and that there is not a significant difference in the mean risk index between the three treatments.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| **NOCIG** | 1 | 21410.85065 | 21410.85065 | 906.64 | <.0001 |
| **TREATMENT** | 2 | 1820.11191 | 910.05595 | 38.54 | <.0001 |

1. Are the necessary conditions for conducting the tests of hypotheses in parts (a) and (b) satisfied with this data set?
2. Normality of residuals:



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tests for Normality** | | | | |
| **Test** | **Statistic** | | **p Value** | |
| **Shapiro-Wilk** | **W** | 0.916843 | **Pr < W** | 0.0222 |

The QQ-plot above shows the residuals do not deviate significantly from the line, but the -value of implies that we should actually reject the null hypothesis of normality and conclude that the residuals may not follow a normal distribution.(???)

1. Equal variances:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Levene's Test for Homogeneity of risk\_adj Variance ANOVA of Squared Deviations from Group Means** | | | | | |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **TREATMENT** | 2 | 3997.0 | 1998.5 | 0.91 | 0.4138 |
| **Error** | 27 | 59183.8 | 2192.0 |  |  |

The Levene test for homogeneity of variances shows the -value of , implying that we fail reject the null hypothesis of equal variances. Thus, we conclude that the residuals have a common equal variance.

1. Independence (test is skipped)
2. Each slope: The covariate and the response, risk index, have an equal slope for every level of treatment as shown in question 16.6(b).
3. Linearity of covariate: The scatterplots in question 16.6(a) show a straight line is sufficient to describe the relationship between the owe for each level of treatment.

SAS code:

**data** hypertension;

input Patient TREATMENT$ RISK NOCIG;

cards;

1 C 22 0

2 C 26 2

3 C 49 6

4 C 67 8

5 C 72 12

6 C 19 0

7 C 28 2

8 C 97 20

9 C 88 18

10 C 30 3

11 I 7 0

12 I 9 0

13 I 14 3

14 I 18 4

15 I 30 7

16 I 42 9

17 I 50 12

18 I 54 13

19 I 70 17

20 I 82 25

21 II 12 0

22 II 14 0

23 II 17 2

24 II 29 5

25 II 37 7

26 II 45 9

27 II 53 11

28 II 81 18

29 II 93 21

30 II 94 23

;

**run**;

**proc** **print** data=hypertension;

**run**;

\* full ANCOVA;

**proc** **glm** data=hypertension;

class treatment;

model risk = nocig treatment / solution;

output out=residuals r=res;

**run**;

**quit**;

**proc** **print** data = residuals;

**run**;

\* test normality;

**proc** **univariate** normal plot data=residuals;

var res;

**run**;

**quit**;

\* reduced model 1 ANOVA;

**proc** **glm** data=hypertension;

class treatment;

model risk = treatment;

output out=residuals\_treat r=res;

**run**;

**quit**;

\* test normality;

**proc** **univariate** normal plot data=residuals\_treat;

var res;

**run**;

**quit**;

\* reduced model 2 regression;

**proc** **glm** data=hypertension;

model risk = nocig;

output out=residuals\_reg r=res;

**run**;

**quit**;

\* test normality;

**proc** **univariate** normal plot data=residuals\_reg;

var res;

**run**;

**quit**;

\* interaction model to provide SSEinter for check for equal slopes;

**proc** **glm** data=hypertension;

class treatment;

model risk = nocig treatment nocig\*treatment;

output out=residuals\_interact r=res;

**run**;

**quit**;

\*\* results from full ANCOVA gives SSEnointeract and interaction model gives SSEinteract

used to find F=MSbeta4beta5/MSEinteraction ~ (MSEnointeraction-MSEinteraction)/MSEinteraction

\* test normality;

**proc** **univariate** normal plot data=residuals\_interact;

var res;

**run**;

**quit**;

\* plot scatterplot to see regression relationship -- already plotted in full model output;

**proc** **plot** data=hypertension;

plot risk\*nocig=treatment;

**run**;

**quit**;

**proc** **means** data = hypertension;

var nocig;

**run**;

\* add new column with risk-mean;

**data** hypertension; set hypertension;

risk\_adj = risk - **3.62352665**\*(nocig - **8.5666667**);

**run**;

\* levene test;

**proc** **glm** data=hypertension;

class treatment;

model risk\_adj = treatment;

means treatment / hovtest = levene;

**run**;

**quit**;

**proc** **glm** data=hypertension;

class treatment;

model risk = treatment nocig nocig\*nocig;

**run**;

**quit**;