**Applied Statistics assignment-8**

**Shiva Kumar Peddapuram 811235874** [**speddapu@kent.edu**](mailto:speddapu@kent.edu)

**9.3 Alaska pipeline faults (Data file: pipeline) This example compares in-field ultrasonic measurements of the depths of defects, Field, in the Alaska oil pipeline with measurements of the same defects in a laboratory, Lab. The lab measurements were done in six different batches, in the variable Batch. The goal is to decide if the field measurement can be used to predict the more accurate lab measurement. The lab measurement is the response variable, and the field measurement is the predictor variable. The data are from the National Institute of Science and Technology (2012, section 6).**

**9.3.1 Draw the scatterplot of Lab versus Field, and comment on the applicability of the simple linear regression model**

**Answer:**

***Code part:***

**Text

Description automatically generated with medium confidence**

**Execution part:**

**Graphical user interface, text, application, email

Description automatically generated**

***Plot:***

**Chart, scatter chart

Description automatically generated**

***Explanation:***

**From the above code we can conclude that the estimated standard error is -1.96750,**

**And we can also interpret that the 45º line is represented as a solid line and it should match the data, if on average, Lab and life measured the same quantity. Majority of the points lie above this line.**

**The ols line is dashed, it suggests that the deeper faults depth may be understand by the field of measurement.**

**9.3.2: Fit the simple regression model, get the residual plot, and summarize. Explain why the plot suggests nonconstant variance and provide a test for nonconstant variance.**

***Code:***

**Graphical user interface, text, application

Description automatically generated**

***Execution part:***

**Graphical user interface, text, application, email

Description automatically generated**

**Plot:**

Chart, scatter chart

Description automatically generated

**Chart, scatter chart

Description automatically generated**

**Explanation:**

**Here the chi squares us 29.58568with 1 as degree of freedom it is the very low p value**

**The slope estimate of intercept is 1.22 which is greater than 1 which represents that the field data especially for large field measurements underestimate the lab measurements.**

**Also, by observing the residual plot we can’t find that the variance is not constant.**

**As a final result deeper flaws are harder to measure, and the variance gets rises with field.**

**9.11 (Data file: fuel2001) In the fuel consumption data, consider fitting the mean function**

Table

Description automatically generated

**Solution:**

**Code part:**

**Graphical user interface, text, application

Description automatically generated**

****

**Execution part:**

**Graphical user interface, text, application

Description automatically generated**

**To be continued in next page**

Table

Description automatically generated

A picture containing text

Description automatically generated

**Explanation:**

**From the output we can conclude that the regression model got most effected by Alaska**

**9.19 (Data file: drugcost) Health plans use many tools to try to control the cost of prescription medicines. For older drugs, generic substitutes that are equivalent to name-brand drugs are sometimes available at a lower cost. Another tool that may lower costs is restricting the drugs that physicians may prescribe. For example, if several similar drugs are available for treating the same symptoms, a health plan may require physicians to prescribe only a few of them. Since the usage of the chosen drug will be higher, the health plan may be able to negotiate a lower price for that drug. The data described in Table 9.6, provided by Mark Siracuse, can be used to explore the effectiveness of these two strategies in controlling drug costs. The response variable is COST, the average cost of drugs per prescription per day. The data are from the mid-1990s, and are for 29**

**plans throughout the United States with pharmacies administered by a national insurance company. Provide a complete analysis of these data, paying regard to possible outliers and influential cases. Summarize your results about the importance of GS and RI. Can we infer that more use of GS and RI will reduce drug costs?**

***Code part:***

Text

Description automatically generated

***Code execution:***

Table

Description automatically generated

***Plots:***

Table

Description automatically generated with medium confidence

**Explanation:**

**This is the unweighted analysis. Depending on some conditions of the sites, if we see the increase in GS of 10% it will ultimately result in a reduction in prescription costs as per day of approximately $0.09 to $0.11. complexity is being increased with the restricted formulary RI.**

**Cost is reduced by around $0.04 if RI is increased by 10% without the three Minnesota clinics.**

**The interpretation of these data is more difficult, but they also lend itself to a weighted analysis. When utilizing MM as weights, the unit of analysis is the patient month rather than the clinic. Var(y|x) = means that the reported value of y is the average when all x values are the same.**