Assignment #4

Due March 10

The following table shows the age (in years) and systolic blood pressure for a sample of 16 men. Assume that the sample was selected randomly from the population of U.S. males between the ages of 30 and 70.

|  |  |
| --- | --- |
| Age | Systolic BP |
| 39 | 134 |
| 47 | 135 |
| 45 | 137 |
| 47 | 145 |
| 65 | 162 |
| 46 | 135 |
| 67 | 170 |
| 42 | 133 |
| 67 | 158 |
| 56 | 154 |
| 64 | 162 |
| 56 | 150 |
| 36 | 129 |
| 59 | 150 |
| 34 | 120 |
| 42 | 128 |

For this assignment, address all of the following:

1. Use proc reg in SAS to conduct a regression analysis to predict team systolic BP from age. The proc reg code also needs to include a statement that produces a dataset containing the predicted systolic BP values and residuals.5
2. Specify the null and alternative hypothesis of primary interest for the linear regression analysis. Referring to the SAS results, describe whether the analyst would reject or fail to reject the null hypothesis? Explain.
3. Specify the linear prediction model indicated by the SAS results.
4. Assume that you are 55 years of age and your actual systolic blood pressure is 145. Comparing the predicted value to the actual value, would you rather have a negative residual or positive residual if you are concerned about having high blood pressure? Explain.
5. Do the scatter (fit plot), residual and normal-probability plots suggest that the conditions of linear regression are satisfied in these data? Explain.
6. Referring to R2, explain why 100% of the variability in systolic blood pressure cannot be explained by age. Describe two additional predictor or explanatory variables that you believe would account for a portion of the unexplained variability.

Students shall turn in their SAS code, the pdf file containing the SAS results, and their written responses to the above items in class on the due date. The written responses must be in the form of grammatical and complete sentences. The results of statistical tests must be reported using APA format. For guidance on using APA format, students can check out <https://depts.washington.edu/psych/files/writing_center/stats.pdf> or other similar pages,

1. H0: There is no relationship between age and systolic blood pressure. (β1 = 0)

Ha: There is a relationship between age and systolic blood pressure. (β1 ≠ 0)

With the p< .05 for both model and age, I would reject the null hypothesis and conclude that there is a relationship between age and systolic blood pressure.

1. The linear model shows the relationship with the independent variable age and the dependent variable systolic blood pressure. There looks like a strong linear positive trend between the relationship of systolic blood pressure and age.
2. I would have a negative residual. Based on the linear model I would expect to have a blood pressure of 149. Since my actual blood pressure was 145 I would have a negative residual.
3. Based on the scatter fit plot, residual plots, and normal-probability plots, the conditions of the linear regression are satisfied. There is linearity between age and systolic blood pressure. In the residual plot, the residuals are in a horizontal direction and do not have a pattern. Based on the residual plot the errors do look independent of each other and do not have clumps or patterns. The variances are equal and the spread is constant among the predicted values. The normal-probability does show a normal distribution among the residuals.
4. The reason why R2 does not equal in 100% because there might be other factors that can cause the variation of Y to occur. Weight could be another predictor variable that can affect the systolic blood pressure or amount of sodium intake can affect the systolic blood pressure.

Code:

title "Age and Systolic Blood Pressure";

data systolic\_blood\_pressure;

input age systolic\_bp;

datalines;

39 134

47 135

45 137

47 145

65 162

46 135

67 170

42 133

67 158

56 154

64 162

56 150

36 129

59 150

34 120

42 128

;

run;

title "Systolic Blood Pressure on Age";

proc reg;

model systolic\_bp = age;

output out=predres p=pred\_iq r=resid;

run;

proc print; run;

proc means; run;