|  |  |
| --- | --- |
| ***Depth at which drilling***  ***begins, Feet*** | ***Time to drill 5 feet,***  ***Minutes*** |
| **0** | **4.90** |
| **25** | **7.41** |
| **50** | **6.19** |
| **75** | **5.57** |
| **100** | **5.17** |
| **125** | **6.89** |
| **150** | **7.05** |
| **175** | **7.11** |
| **200** | **6.19** |
| **225** | **8.28** |
| **250** | **4.84** |
| **275** | **8.29** |
| **300** | **8.91** |
| **325** | **8.54** |
| **350** | **11.79** |
| **375** | **12.12** |
| **395** | **11.02** |

1. Visualization can always help us understand the behavior of the data. For this question, use Excel to plot the scattergram (scatterplot) of the data, and copy the resulting plot in your response. In the scatterplot, you can see a linear relationship between the depth at which drilling begins and the time to drill 5 feet. Is that a positive or negative relationship?

Independent Variable- Depth Dependent Variable- Time

***We often see patterns or relationships in scatterplots. When the y variable (Depth at which drilling begins) tends to increase as the x variable (Time to drill 5 feet) increases, we say there is a positive correlation between the Depth of drilling and the time to drill.***

**2.** Write a general regression equation (with 𝛽̂s) for the above linear relationship in this problem.

***E (Time to Drill) = 𝛽̂0 + 𝛽̂1(Depth of Drilling)***

***OR***

***Time to Drill = 𝛽̂0 + 𝛽̂depth \* Depth of Drilling +*** ε

**3.** Write the Null Hypothesis that corresponds to the above linear relationship (write both the Null Hypothesis equation and its English explanation, e.g., “There is no…”).

Null Hypothesis: There is no linear correlation between the depth of drilling and the time to drill in the proposed model.

H0: ***𝛽̂Depth = 0***

4. What is the t-value that SAS has computed for 𝛽̂𝐷𝑒𝑝𝑡ℎ? Interpret the t-value that you got from SAS in layman’s terms.

Data Drill;

Infile '\\files\users\sonikaprakashs\Documents\drillrock.dat';

Input Depth Time;

Proc Reg;

Model Time=Depth /CLB

| **Parameter Estimates** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **95% Confidence Limits** | |
| **Intercept** | **1** | 4.78960 | 0.66631 | 7.19 | <.0001 | 3.36939 | 6.20981 |
| **Depth** | **1** | 0.01439 | 0.00285 | 5.05 | 0.0001 | 0.00832 | 0.02046 |

*=*

***Where,***

*= 0.01439*

*Hypothesized value of = 0*

*= 0.00285*

*= = 5.05*

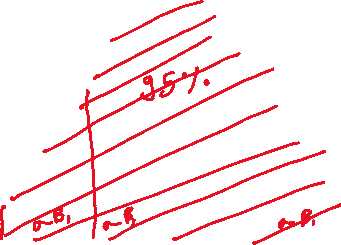
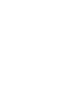
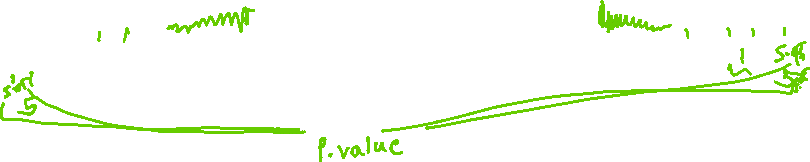
***In terms of Standard Deviation of Sampling Distribution of 𝛽̂1, the observed 𝛽̂*𝐷𝑒𝑝𝑡ℎ *(From the sample) is approximately 5.05 Standard deviation away from zero (0) [which is Null Hypothesis of 𝛽̂*𝐷𝑒𝑝𝑡ℎ*].***

**5. Based on the computed t-value, do you expect the related p-value to be big or small? Why? Do the SAS results (e.g., p-value) corroborate your guess? Can you reject the Null Hypothesis using this p-value?**

***If H0: 𝛽̂*𝐷𝑒𝑝𝑡ℎ *= 0(If this is true)***

***P-value= p(|t|>5.05) <0.0001Chart, line chart, histogram

Description automatically generated***



***The P-value is super-small, the probability of having another sample that would give us a value as contradictory or more contradictory than the observed, so we reject the Null Hypothesis.***

H0: ***𝛽̂Depth = 0 – Reject***

**At any level of α, there is enough evidence to indicate that the depth of the drilling and time to drill are linearly correlated. (As the p-value is less than 0.0001).**

**6. Write the final (specific) least squares prediction equation for this problem (you need to replace 𝛽̂s with their SAS estimates).**

***E (Time to Drill) = 𝛽̂0 + 𝛽̂1(Depth of Drilling)***

***OR***

***Time to Drill = 𝛽̂0 + 𝛽̂depth (Depth of Drilling) +*** ε

***Where,***

***𝛽̂0/intercept = 4.80***

***𝛽̂1/ 𝛽̂Depth = 0.01439***

***E (Time to Drill) = 4.80 + 0.01439 \* (Depth)***

***OR***

***Time to Drill = 4.80 + 0.01439(Depth)+* ε**

**7. Can you interpret the intercept in the least squares prediction equation? Why? If yes, write your interpretation of 𝛽̂0 in layman's terms.**

***As our data set encloses 0, so the least squares prediction equations say that –***

***We estimate that the minimum average time to drill is 4.80 minutes as per the proposed model.***

**8. Interpret 𝛽̂𝐷𝑒𝑝𝑡ℎ (that you got from SAS) in the least squares prediction equation in layman's terms.**

***Holding other variables fixed, we estimate the average time to drill to* increase by 0.01439 minutes for every additional 5 feet to the drilling.**

**9. Interpret the 𝑟2 (R-square) value (from SAS) in layman’s terms.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 1.43219 | **R-Square** | 0.6300 |
| **Dependent Mean** | 7.66294 | **Adj R-Sq** | 0.6053 |
| **Coeff Var** | 18.68987 |  |  |

**63% of the variability of the time can be explained by the proposed linear relationship between the depth of drilling and the time to drill.**

**10. Interpret the Root Mean Square Error (RMSE) value (from SAS) in layman’s terms.**

***We are 95% confident that the interval of [ 0.00832, 0.02046] encloses the true increase (𝛽̂Depth) in the mean of time to drill for every additional 5 feet of the depth of drilling, holding other variables fixed.***

***We expect most (95%) of the actual time to drill to fall within the (2\*1.43) 2.86 minutes of their respective least squares predicted values.***