# Correlation and Regression



| | |
|---|---|
| **Started on** | Tuesday, June 2, 2020, 11:43 PM |
| **State** | Finished |
| **Completed on** | Tuesday, June 2, 2020, 11:57 PM |
| **Time taken** | 14 mins 7 secs |
| **Grade** | **80** out of 100 |

## Question 1

Correct
10 points out of 10

You are trying to understand the potential causes of a serious manufacturing defect. You compile historical data and use logistic regression to model the probability of the outcome, **Defective?**, as a function of many variables and their interactions. The whole model test results for this analysis are displayed below. What are the potential conclusions from this output?

**Whole Model Test**

| Model | -LogLikelihood | DF | ChiSquare | Prob>ChiSq |
|---|---|---|---|---|
| Difference | 30.13607 | 8 | 60.27215 | <.0001* |
| Full | 212.37582 | | | |
| Reduced | 242.51190 | | | |

| | |
|---|---|
| RSquare (U) | 0.1243 |
| AICc | 443.237 |
| BIC | 478.237 |
| Observations (or Sum Wgts) | 381 |

**Fit Details**

| Measure | Training | Definition |
|---|---|---|
| Entropy RSquare | 0.1243 | 1-Loglike(model)/Loglike(0) |
| Generalized RSquare | 0.2032 | (1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n)) |
| Mean -Log p | 0.5574 | $\sum$ -Log(ρ[j])/n |
| RASE | 0.4358 | $\sqrt{\sum(y[j]-\rho[j])^2/n}$ |
| Mean Abs Dev | 0.3775 | $\sum$ \|y[j]-ρ[j]\|/n |
| Misclassification Rate | 0.3071 | $\sum$ (ρ[j]≠ρMax)/n |
| N | 381 | n |

Select one:

a. The whole logistic model is not significant.

b. More than 30% of the observations are misclassified by the model.

c. None of the variables in the model, or their interactions, are significant.

d. 43.75% of the variation in **Defective?** is explained by the model.

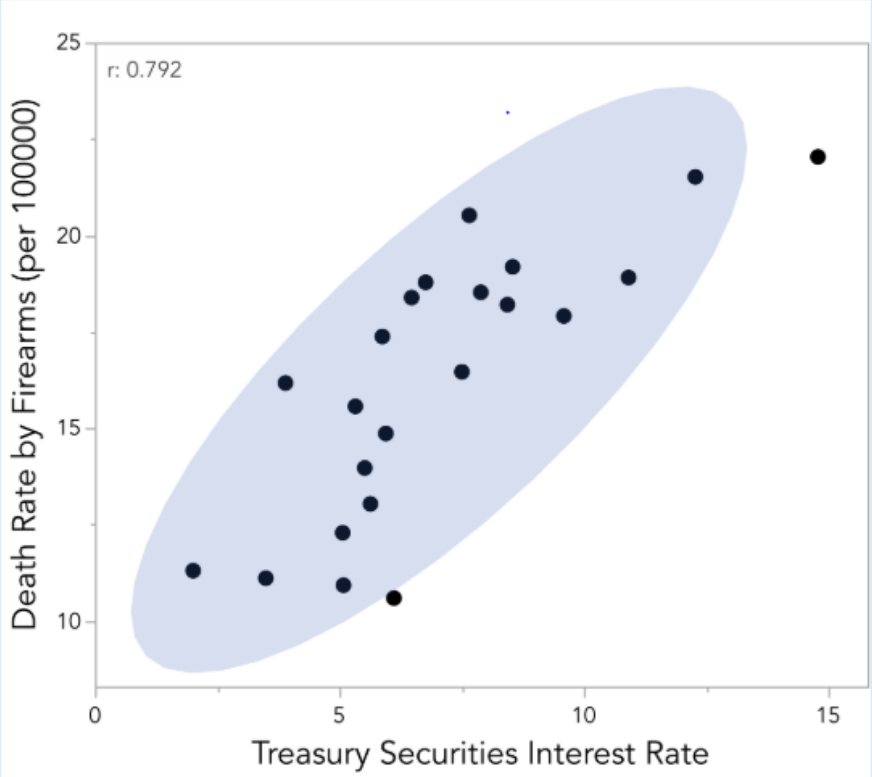
## Question **2**

Correct

10 points out of 10

### Question text

The graph below shows the historical data of the annual death rate by firearms (per 100000) versus the annual average treasury securities interest rate (for one-year constant maturity). What can you conclude from this analysis?




Select one:

a. Higher interest rates correspond to higher death rates by firearm.

b. Raising interest rates causes more firearm deaths.

c. An increase in firearm deaths causes higher taxes.

d. The correlation between the two variables is negative.

## Question **3**

Correct

10 points out of 10

### Question text

Which of the following are assumptions for the errors in multiple linear regression?

Select one:

a. equality and constant variance

b. independence, normality, and constant variance

c. independence, normality, and equality

d. independence and heteroscedasticity

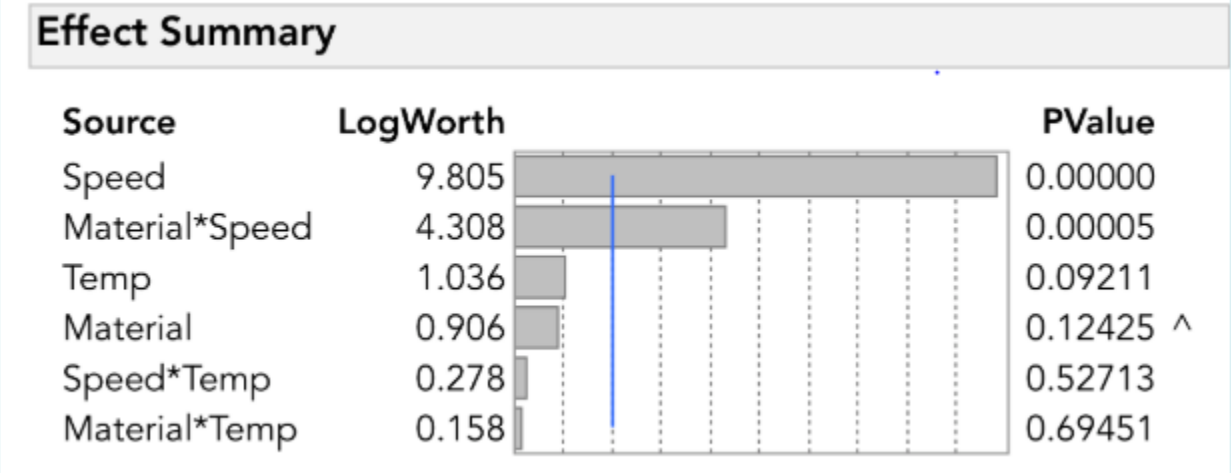
e. independence and normality

## Question **4**

Correct
10 points out of 10

### Question text

You fit a multiple linear regression model for **Strength** as a function of **Speed**, **Material**, **Temp**, and all possible two-way interactions. You use variable selection to reduce the model to only the significant terms. Using a *p*-value cutoff (or significance level) of 0.05, which terms are candidates for removal?




Select one:

a. **Material**, **Temp**, and **Material*Temp**

b. **Speed** and **Material**

c. **Temp** and **Material**

d. **Material*Temp**, **Speed*Temp**, and **Temp**

e. **Speed**, **Material**, and **Material*Speed**

## Question **5**

Correct
10 points out of 10

### Question text

You fit a simple linear regression model for a response, **Thickness**, as a function of one predictor, **Speed**. Using the Parameter Estimates table, what is the estimated regression equation?

| Parameter Estimates | | | | |
|---|---|---|---|---|
| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
| Intercept | 7.1196113 | 3.493887 | 2.04 | 0.0532 |
| Speed | 2.8072394 | 0.365323 | 7.68 | <.0001* |

Select one:

a. $\hat{Speed} = 2.807 + 7.12 * Thickness$

b. $\hat{Thickness} = 2.807 + 0.365 * Speed$

c. $\hat{Speed} = 7.12 + 3.49 * Thickness$

d. $\hat{Thickness} = 2.807 + 7.12 * Speed$

e. $\hat{Thickness} = 7.12 + 2.807 * Speed$

## Question 6

Correct
10 points out of 10

### Question text

What is the range of possible values for the correlation coefficient?

Select one:

a. 0 to 1

b. -1 to 1

c. -infinity to infinity

d. 0 to 100%

e. -2 to +2

## Question 7

Incorrect
0 points out of 10

### Question text

The Summary of Fit table, ANOVA table, and residual-by-predicted plot for a multiple linear regression model are shown. What can you conclude about this model?

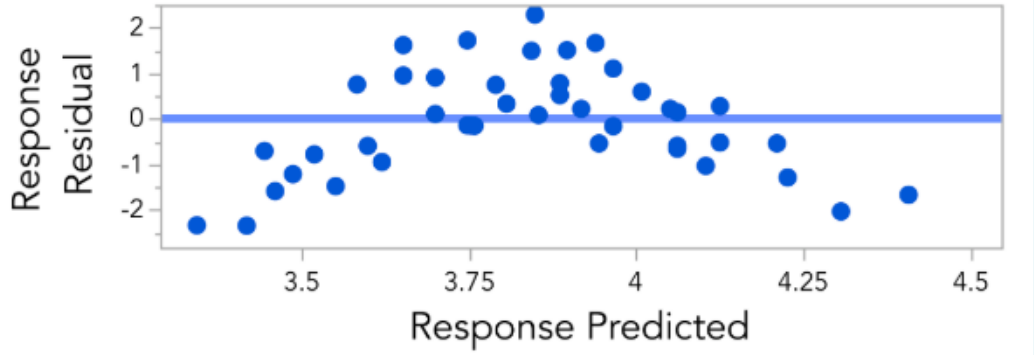
**Summary of Fit**

| | |
|---|---|
| RSquare | 0.041435 |
| RSquare Adj | 0.021464 |
| Root Mean Square Error | 1.188042 |
| Mean of Response | 3.846667 |
| Observations (or Sum Wgts) | 50 |

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 1 | 2.928502 | 2.92850 | 2.0748 |
| Error | 48 | 67.749275 | 1.41144 | Prob > F |
| C. Total | 49 | 70.677778 | | 0.1562 |

**Residual by Predicted Plot**




Select one:

- a. There is no obvious non-random pattern in the residual plot.
- b. The model explains less than 10% of the variation in the response.
- c. The model explains more than 90% of the variation in the response.
- d. The model explains a significant amount of variation in the response.

## Question 8

Incorrect
0 points out of 10

### Question text

The confusion matrix for a logistic regression model is shown. What is the error rate?

**Confusion Matrix**

Training

| Actual Outcome | Predicted Count Fail | Pass |
|---|---|---|
| Fail | 22 | 4 |
| Pass | 2 | 67 |

Select one:

a. 22%

b. 2.1%

c. 4.2%

d. 6.0%

e. 6.3%

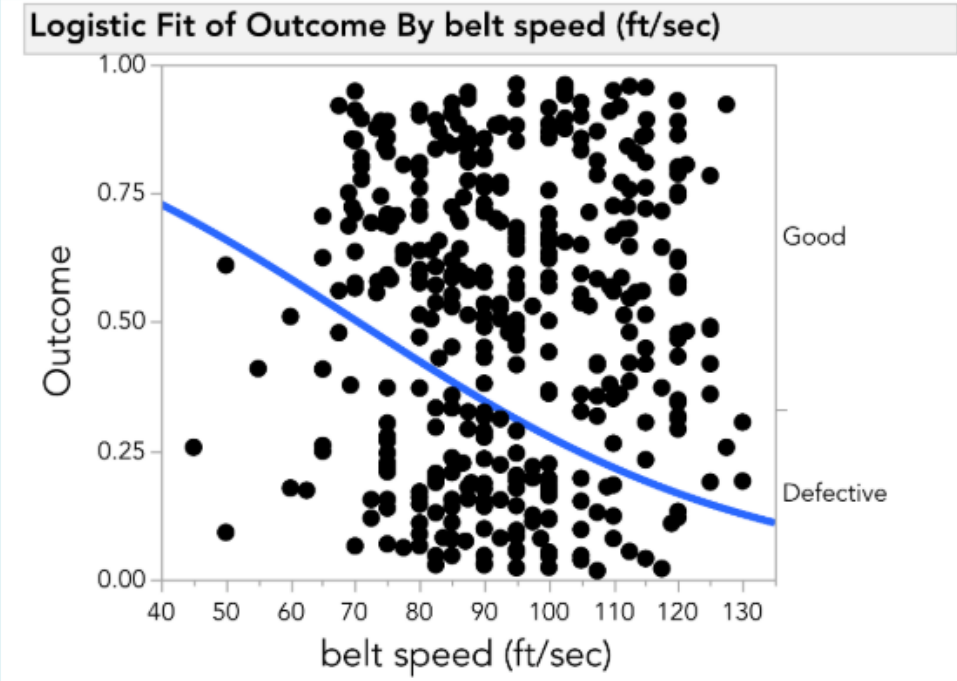
## Question 9

Correct

10 points out of 10

### Question text

You fit a simple logistic model for **Outcome** (**Good** or **Defective**) as a function of **belt speed**. You model the probability that **Outcome** = **Defective**. The logistic curve for the model and the Parameter Estimates table are shown. Which of the following are correct?




**Parameter Estimates**

| Term | Estimate | Std Error | ChiSquare | Prob>ChiSq |
|---|---|---|---|---|
| Intercept | 2.28267685 | 0.662228 | 11.88 | 0.0006* |
| belt speed (ft/sec) | -0.0323879 | 0.0071636 | 20.44 | <.0001* |

For log odds of Defective/Good

Select one:

a. As the value of **belt speed** increases, the probability of **Defective** increases.

b. As the value of **belt speed** increases, the probability of **Good** increases.

c. **Belt speed** is not a significant predictor of **Outcome**.

d. There is no relationship between **belt speed** and the probability of **Defective**.

## Question **10**

Correct

10 points out of 10

### Question text

You fit two multiple regression models to the same data, Model A and Model B. The Summary of Fit tables for the two models are shown. Which of the following statements is correct?

| Model A - Summary of Fit | |
|---|---|
| RSquare | 0.644249 |
| RSquare Adj | 0.636838 |
| Root Mean Square Error | 1.38959 |
| Mean of Response | 19.10668 |
| Observations (or Sum Wgts) | 50 |

| Model B - Summary of Fit | |
|---|---|
| RSquare | 0.844423 |
| RSquare Adj | 0.830594 |
| Root Mean Square Error | 0.949075 |
| Mean of Response | 19.10668 |
| Observations (or Sum Wgts) | 50 |

Select one:

- a. The RSquare Adj for Model B is better than the RSquare Adj for Model A.
- b. The mean response for Model B is higher than the mean for Model A.
- c. Based on the RSquare Adj and RMSE, Model A is better than Model B.
- d. The RMSE for Model A is better than the RMSE for Model B.

Finish review