MATH 1342

Homework 7 (Sections 5.4 – 5.6)

Instructions: Answer all questions through the EMCF tab of casa under the assignment named "Homework 7" before the deadline.

There is no "Submit" button. Your answers will be automatically submitted once the deadline arrives.

Assignments will be graded out of 20 points.

---

1. Section 5.4; Problem 2

    A. 196

    B. 2484.95

    C. -2484.95

    D. 112.9

    E. No Residual Available

2. Section 5.4; Problem 3

    A. Due to the pattern in the residual plot, the LSRL is a good fit for the data.

    B. Due non-linear nature of the residual plot, we can conclude that the LSRL is a not a good fit to the data.

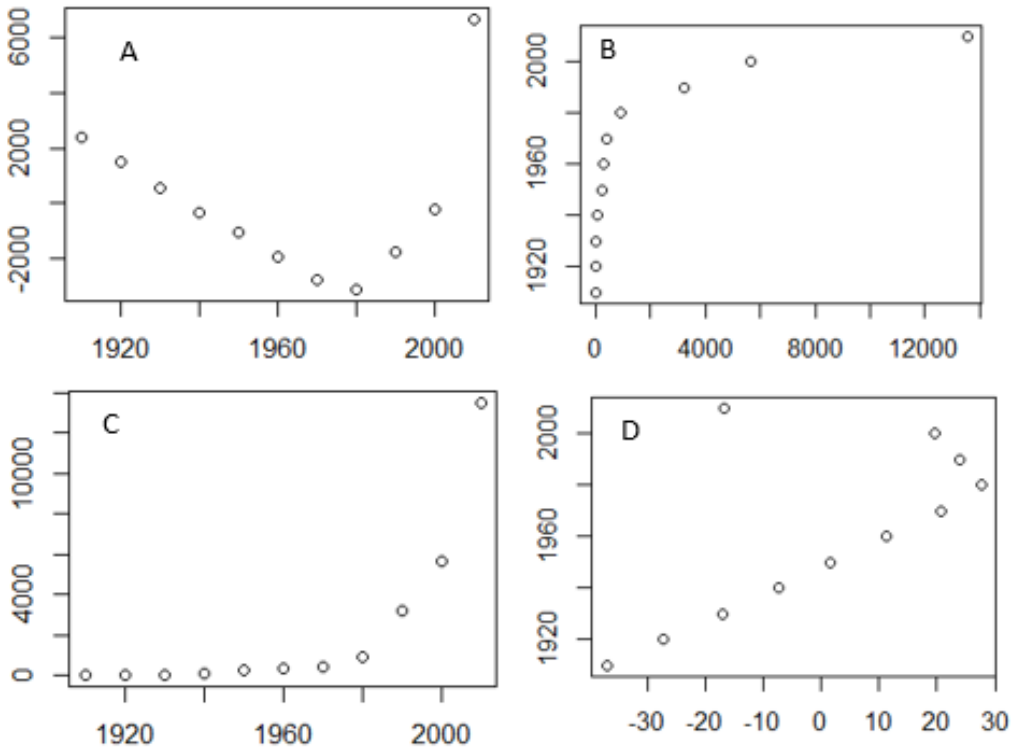
    C. Due to the pattern in the residual plot, the LSRL is not a good fit for the data.

    D. Since, for large values of x, the residual plot is increasing, we know that the LSRL must have a positive slope.

    E. The residual plot cannot be used to determine the accuracy or lack of accuracy in the model that created it.

3. Section 5.5; Problem 2 (a), Scatter Plot

4. Section 5.5; Problem 2 (a), Residual Plot



E. None of the above solutions

5. Section 5.5; Problem 2 (b) (Explanation of Proposed Model)

A. Exponential Model, since the r^2 value is the highest

B. Quadratic Model, since the scatter-plot resembles a U-Shape.

C. Logarithmic Model, since the r^2 value is the highest when calculating cor(year, log(spending))^2.

D. Exponential Model, since the r value is the highest

E. Quadratic Model, since the r^2 value is the highest

6. Section 5.5; Problem 2 (b) (Construction of non-linear model)

Note: Based on the system you use, you may get different formats of your answer. The left and right column answers are equivalent.

A. $\hat{y} = e^{148.925x - 0.07881}$ or $\hat{y} = (1.082 \times 10^{-65})2.1027^x$

B. $\hat{y} = e^{0.07881x - 148.925}$ or $\hat{y} = (2.1027 \times 10^{-65})1.082^x$

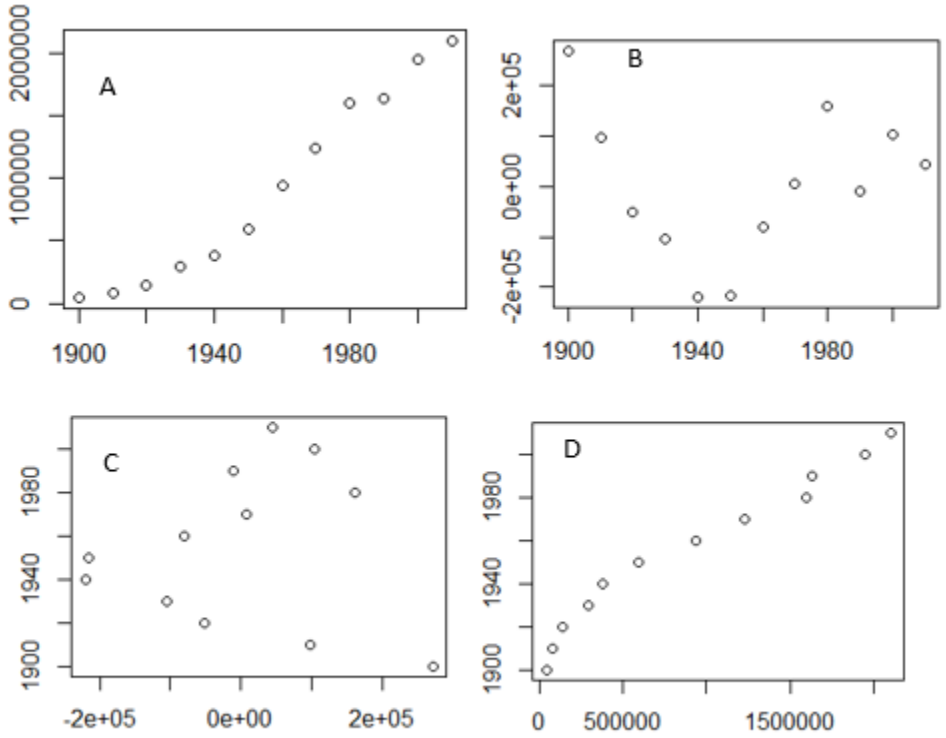
C. $\hat{y} = e^{0.07881x + 148.925}$ or $\hat{y} = (2.1027 \times 10^{-65})0.922^x$

D. $\hat{y} = e^{-0.07881x - 148.925}$ or $\hat{y} = (-2.1027 \times 10^{-65})1.082^x$

E. None of the above models.

7. Section 5.5; Problem 4 (a), Scatter Plot

8. Section 5.5; Problem 4 (a), Residual Plot



E. None of the above solutions

9. Section 5.5; Problem 4 (b) (Proposed Method for Model Selection)

A. Select the model whose generic graph most matches the scatterplot

B. Select the model with the coefficient of determination closest to 1.0

C. Select the model that has a residual plot with the least obvious pattern

D. Any of choices A, B, or C are valid selection criteria.

E. None of the above choices are valid selection criteria.

10. Section 5.5; Problem 4 (b) (Selection of Model)

A. Linear B. Quadratic C. Exponential D. Logarithmic

Refer to the following data, using R Studio data packages, to answer questions 11, 12, 13, and 14.

In R Studio use the data *airquality* to determine the following. *Hint: The data set is already in R studio use the quick reference guide to determine the following.*

[In R Studio, use command(file$column), such as mean(airquality$Temp)]

***Description**: The data gives several measurements of air quality in New York City from 1973.* Format: A data frame with 153 observations.

Temp numeric Temperature (in degrees F))

Wind numeric Wind speed (in mph)

11. Determine the LSRL for determining temperature (Response Variable) as it relates to wind speed (Explanatory Variable).

A. $\hat{y} = -1.23x - 90.13$ B. $\hat{y} = 1.23x - 90.13$

C. $\hat{y} = -90.13x + 1.23$ D. $\hat{y} = 90.13x - 1.23$ E. $\hat{y} = -1.23x + 90.13$

12. Interpret the slope of the LSRL.

A. There is an increase of 1.23 degrees for every 1 mph in wind speed.

B. There is a decrease of 1.23 degrees for every 1 mph in wind speed.

C. There is an increase of 1.23 degrees for every 90.13 mph in wind speed.

D. There is a decrease of 1.23 degrees for every 90.13 mph in wind speed.

E. The wind speed increasing by one mile per hour causes the temperature to drop by 1.23 degrees.

13. Find the values of the correlation coefficient and coefficient of determination.

A. $r = 0.2098$; $r^2 = 0.4580$ B. $r = -0.2098$; $r^2 = 0.4580$

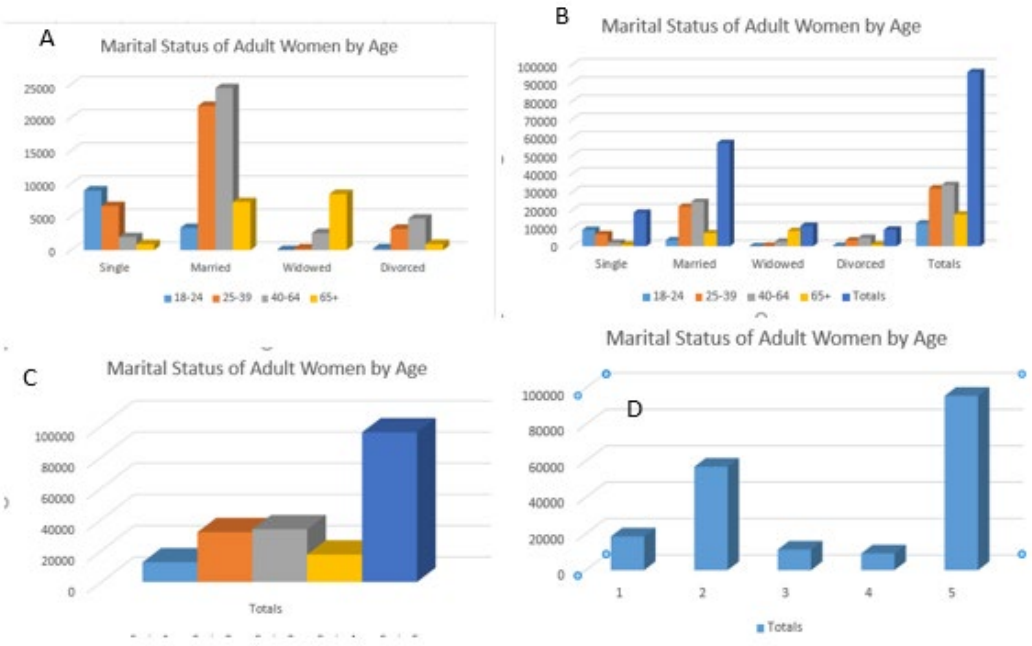
C. $r = -0.4580$; $r^2 = 0.2098$ D. $r = 0.2098$; $r^2 = 0.0440$

E. $r = -0.2098$; r^2 = unreal answer

14. One of the data points indicated a day with a temperature of 84 degrees and a wind speed of 24 mph. Determine the residual of that data point.

A. -23.39 B. 37.19 C. 0.0 D. 23.39 E. 10.81

15. Section 5.6; Problem 4 (b)



16. Section 5.6; Problem 4 (c, d)

A. (c) 0.0940; (d) 0.0347

B. (c) 0.4858; (d) 0.3628

C. (c) 0.7135; (d) 0.0347

D. (c) 0.4848; (d) 0.1011

E. (c) 0.7135; (d) 0.1011

17. In the least-squares regression line, the desired sum of the errors (residuals) should be

A. positive B. negative C. zero D. maximized

18. A prediction of the world's population in the year 2088 is an example of:

A. An outlier B. Seasonality C. Extrapolation D. Correlation

19. An observation that causes the values of the slope and the intercept in the line of best fit to be considerably different from what they would be if the observation were removed from the data set is said to be

A. A causation variable B. Extrapolation C. Influential D. A residual

20. For a set of data: x = (0,1,2,3,4,5,6) and y=(36, 28, 25, 24, 23, 21, 19), is it wise to use a linear regression to extrapolate data for x = 50?

Proposed Solution:

Since the coefficient of determination is 0.8582, the linear model is a reasonably good fit for the data, so extrapolation for any x-value is acceptable.

What is wrong with the proposed solution?

A. As the extrapolation value gets farther away from the known data points, the accuracy diminishes. An x-value of x = 50 is far too distant from the known data to obtain accurate results.

B. It is possible that other models (non-linear models) would have a coefficient of determination closer to 1.0, and therefore be a better model to make predictions from.

C. The coefficient of determination is a measure used to gauge the relative effectiveness of different data models, not as an indication of the accuracy of distant extrapolation.

D. Choices A, B, and C are all valid of the incorrectness of the proposed solution.

E. The proposed solution is correct.