Review ANSEWRS – Unit IV – Scatterplots & Regressions

FREE RESPONSE

1. The association between weight and fuel consumption for these automobiles is...
   - strong (or moderately strong),
   - linear (or fairly linear), and
   - positive – as weight of the automobile increases, so does fuel consumption.

   **make sure to write in complete sentences and use context!!!*
   Grading note: Including the word "correlation" in your response does NOT count for "linear" – you must specify that the relationship between the variables is "linear" (or "fairly linear").

2. Although "x" and "y" have a strong ASSOCIATION, the relationship is curved (nonlinear).
   CORRELATION is only useful for describing a LINEAR association.

3. 9.7

4. a) The association between forearm circumference and strength is moderately weak (due to the low outlier), positive, and mostly linear.

   b) The correlation would become stronger (closer to 1.0)

   c) Currently the right side of the least squares line would be "pulled" down by the point at (14, 38).
   Since removing the point would cause the right side of the line to go UP, the slope of the line would become MORE POSITIVE.

5. a) As the number of stork pairs INCREASES, the populations of these countries also INCREASE.

   b) Having children may not (probably will not?) CAUSE the stork population to increase. Even if there is a strong association between these two variables, **correlation does not imply causation**.

6. From left to right: 0.7994, -0.6112, 0.0023, -0.9713

7. a) Yes: The scatterplot of fare vs distance has a linear pattern, and the plot of residuals vs predicted fare* shows no clear curved pattern.
   (Note: If the horizontal axis of the residual plot is confusing to you… the plot of residuals vs "x" will have the same form as the plot of residuals vs "y-hat". This is because "y-hat" is a direct linear transformation of "x")

   b) $\widehat{airfare} = 177.215 + 0.079(distance)$

   c) For each increase of 1 mile in distance, the model predicts an increase of \$0.079 in airfare.
   (or "For each increase of 1 mile in distance, there is a MEAN increase of \$0.079 in airfare")

   d) A zero-mile long flight is predicted to cost \$177.215. This is perhaps due to overhead costs (also a possible example of extrapolation).

   e) 48% of the variability in airfare is accounted for by the model relating distance and airfare.

   f) s_e = \$41.9. The typical (average-ish) amount that observed airfare differs from the predicted airfare for these flights is about \$41.90. *(don't worry, this is NOT on this test… but we'll see it in the Spring!)*

   g) \$185.115. However 100 miles is outside the range of x-values in this set of data. This is an example of extrapolation, and this prediction might not be very reliable since we don't know if the linear trend continues as we go below 300 miles.

8. a) $\widehat{car\ sold} = 14.2973 + 1.3067(\#\ of\ commercials)$

   b) 37.82 *(expected values should NEVER be rounded to whole numbers!)*

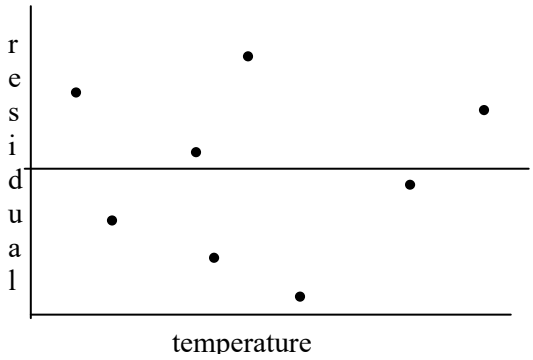
   c) –5.364. This means the dealer sold 5.364 FEWER cars than predicted by the model based on the 10 commercials that the dealer paid for.

   d) OVERestimated

9. There is a curved pattern in the residual plot (and the scatterplot itself is curved!) so a linear equation is not appropriate. The association between age and canal width for these seals is strong, curved, and negative (since canal width DECREASES as age INCREASES).

10. a) r = –.969. The data suggests a strong, negative, linear association between crop duration and crop yield.
    b) $\widehat{yield} = 5.557 - 0.037(duration)$
    c) $r^2 = 0.938$. 93.8% of the variation in crop yield is explained by the linear model with crop duration.
    d) 1.699 tons
    e) 0.153 tons

11. a) $r = 0.972$. This indicates a strong, positive, linear association between temperature and ice cream sales.
    b) $\widehat{ice\ cream\ sales} = -331.839 + 5.77(temperature)$
    c) 0.945; 94.5% of the variability in ice cream sales is accounted for by the model with temperature.
    d) 147.486 pounds

e) 

The residual plot is scattered with no clear curved pattern, which supports the use of a linear regression model for this data set.

f) 222.958 pounds

MULTIPLE CHOICE

12. E
13. B
14. A
15. B
16. B
17. D
18. A
19. A

20. A
21. B
22. B
23. A
24. E

25. B

26. The association between student rating and parent rating for these movies is moderately strong, curved (or non-linear), and positive (generally as student rating increases, so does parent rating).