

AP Statistics

Linear Regression - Review Problems 2

- 1) The association between a family's weekly income and the amount they spend on restaurant meals is found to be linear with $r = 0.40$.

(a) Answer true or false:

Families tend to spend about 40% of their incomes on restaurant meals.

The line of best fit passes through 40% of the data points.

When income increases by \$100 per month, families tend to spend an additional \$40 on restaurant meals. (This would be true if Slope was 0.40.)

If a family's income increases, then they will spend more in restaurants. (not necessarily)

In general, families with higher incomes spend more in restaurants.

40% of the variability in amount spent in restaurants is explained by differences in family income. (This would be true if R^2 was 0.40.)

- (b) If a family's weekly income is 2 standard deviations above the average family weekly income, the predicted amount this family spends on restaurant meals is 0.8 standard deviations above the mean.

$$\hat{z}_y = r \cdot z_x = (0.4)(2) = 0.8$$

- (c) If a family's weekly income is 1.5 standard deviations below the average family weekly income, the predicted amount this family spends on restaurant meals is 0.6 standard deviations below the mean.

$$\hat{z}_y = r \cdot z_x = (0.4)(-1.5) = -0.6$$

- (d) What are the units for the correlation coefficient?

Correlation coefficients are unitless.

- (e) If we change "weekly income" to "yearly income" (by multiplying by 52 weeks/year), how will the correlation change?

It won't change. Correlation is not affected by changes in the center or scale of either variable.

- 2) A regression analysis of company profits and the amount of money the company spent on advertising found $R^2 = 0.85$.

Answer true or false:

This model can correctly predict the profit for 85% of companies.

On average, companies spend about 85% of their profits on advertising.

Advertising spending explains about 85% of profits.

Differences in advertising spending explain about 85% of the variability in profits.

The R^2 value of 85% shows that this linear model is appropriate. (use a residuals plot)

The R^2 value of 85% shows that this linear model will be fairly reliable.

- 3) A consumer group creates a linear model to estimate the cost of a standard TV (in \$) based on the screen size (in inches). Which is the most likely value of the slope of the line of best fit?

(A) 0.15 (B) 1.5 (C) 15 (D) 150 (E) 1500

If you increase the screen size by 10 inches, the cost is most likely to increase by \$150. $\frac{150}{10} = 15$

- 4) Education research consistently shows that students from wealthier families tend to have higher SAT scores. The slope of the line that predicts SAT score from family income (in thousands of dollars) is 6.25, and the correlation between the variables is 0.48. Then the slope of the line that predicts family income (in thousands of dollars) from SAT score is

(A) 0.037 (B) 0.16 (C) 3.00 (D) 6.25 (E) 13.02

$$b_1 = r \frac{S_y}{S_x}$$

$$6.25 = (0.48) \frac{S_{SAT}}{S_{inc.}}$$

$$\frac{S_{SAT}}{S_{inc.}} = 13.020833$$

$$\frac{S_{inc.}}{S_{SAT}} = \frac{1}{13.020833} = 0.0768$$

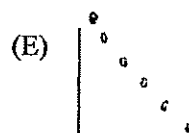
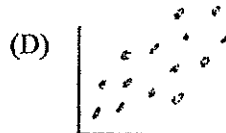
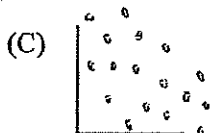
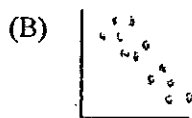
$$\text{new slope: } b_1 = r \frac{S_{inc.}}{S_{SAT}}$$

$$= (0.48)(0.0768) = \boxed{0.0369}$$

- 5) Studies have shown that nations with a higher number of television-sets per person have higher life expectancies. Does this mean that we can lengthen the lives of people in Botswana by shipping them TV sets? Explain.

Shipping TV's to Botswana will probably not lengthen the people's lives. Just because 2 variables are associated does not mean that one causes the other to happen. There may be other factors, such as a lurking variable of national wealth.

- 6) Match each of the following scatterplots with its correlation:



C $r = -0.3$

D $r = 0.5$

B $r = -0.7$

A $r = 0.9$

E $r = 0.99$