AP Statistics
Chapter 7 – 9 Review

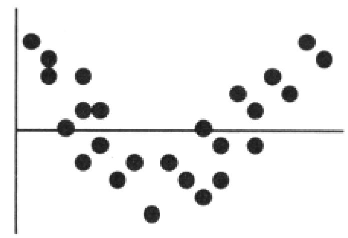
Name: [illegible] Per: ∞

MULTIPLE CHOICE Select the letter that corresponds to the **best** answer.

A 1. The correlation between *X* and *Y* is r = 0.35. If we double each *X* value, increase each *Y* value by 0.20, and interchange the variables (put *X* on the *Y-axis* and vice versa), the new correlation

A) is 0.35 B) is 0.50 C) is 0.70 D) is 0.90 E) cannot be determined

E 2. The residuals plot for a linear model is shown. Which is true?



A) The linear model is okay because approximately the same number of points are above the line as below it.
B) The linear model is okay because the association between the two variables is fairly strong.
C) The linear model is no good because the correlation is near 0.
D) The linear model is no good because some residuals are large.
E) The linear model is no good because of the curve in the residuals.

B 3. All but one of the statements below contains a mistake. Which one could be true?

A) The correlation between height and weight is 0.568 inch~~es per pound~~. X
B) The correlation between weight and length of foot is 0.488. ✓
C) The correlation between the ~~breed~~ of a dog and its weight is 0.435. X
D) The correlation between ~~gender~~ and age is -0.171.

C 4. A correlation of zero between two quantitative variables means that

A) We have done something wrong in our calculation of *r*.
B) There is no association between the two variables.
C) There is no linear association between the two variables.
D) Re-expressing the data will guarantee a linear association between the two variables.
E) None of the above.

C 5. A residual plot is useful because

I. it will help us to see whether our model is appropriate
II. it might show a pattern in the data that was hard to see in the original scatterplot.
III. it will clearly identify influential points.

A) I only B) II only C) I and II only D) I and III only E) I, II, and III

A 6. When using midterm exam scores to predict a student's final grade in a class, one would prefer to have a

A) positive residual, because that means the student's final grade is higher than we would predict with the model.
B) positive residual, because that means the student's final grade is lower than we would predict with the model.
C) residual equal to zero because that means the student's final grade is exactly what we would predict with the model.
D) negative residual, because that means the student's final grade is higher than we would predict with the model.

B 7. Which one of the following statements is true?

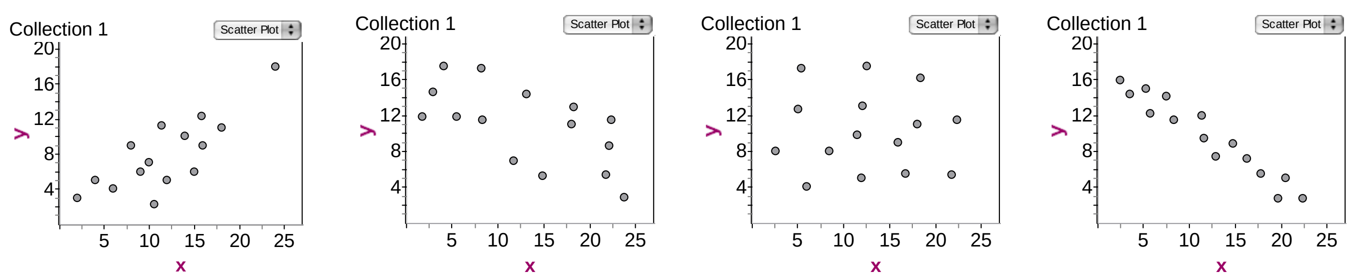
A) Values of *r* near zero indicate a ~~strong~~ linear relationship.
B) The correlation can be strongly affected by a few outlying observations. ✓
C) Changing the measurement units of x and y may affect the correlation between x and y. nope
D) Strong correlation means that there is a definite ~~cause-and-effect~~ relationship between x and y
E) Correlation changes when the x and y variables are reversed. nope

E 8. Which of the following associations is likely to have a negative correlation?

A) Number of hours devoted to studying for a final exam and a student's grade on the final ✗
B) A teacher's salary and the number of years teaching experience that the teacher has ✗
C) The age of an automobile and the number of miles an automobile has ✗
D) The number of children in a family and the weekly amount of money spent on food ✗
E) The speed a car travels and the time required to travel a given distance on a flat deserted road ✓
more speed, less time to travel

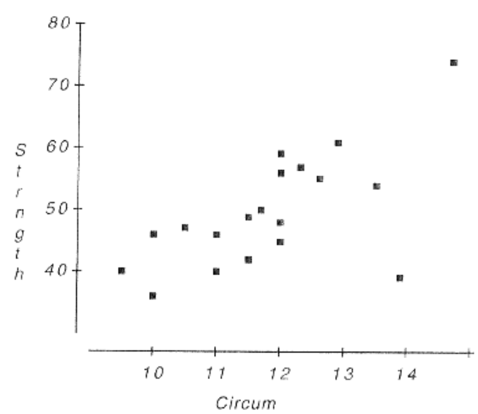
9. For the following scatterplots, write the appropriate correlation coefficient underneath each plot.

-0.6112 ✓ 0 .7994 ✓ -0.9713 ✓ 0.2005 0.0023 ✓ ~~-1~~



0.7994 -0.6112 0.0023 -0.9713

10. Researchers investigating the association between the size and strength of muscles measured the forearm circumference (in inches) of 20 teenage boys. Then they measured the strength of the boys' grips (in lbs). Their data are plotted at right.



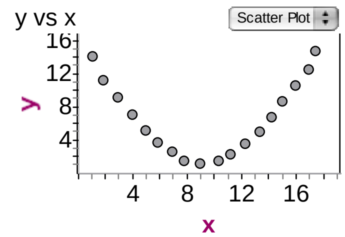
a) Describe the *association* between forearm circumference and strength.

There is a moderate (or weak), positive, linear association between forearm circumference and strength.
(the association is moderate – not strong – due to the outlier)

b) If the point at the lower right corner (at about 14" and 38 lbs.) were removed, would the correlation become stronger, weaker, or remain about the same?

The correlation would become STRONGER (closer to 1).

11. The scatterplot shows a relationship between x and y that results in a correlation coefficient of r = 0.024. Explain why r = 0.024 in this situation even though there appears to be a strong relationship between the x and y variables.



The association between x and y is NON-linear.
Correlation is only useful for describing LINEAR association.

12. A student who has created a linear model is disappointed to find that their R^2 value is a very low 13%.

a) Does this mean that a linear model is not appropriate? Explain.

Not necessarily. Although the linear model may **not** have a tremendously strong correlation (r = 0.361), a linear model may still be appropriate, as long as the scatterplot looks fairly linear and as long as there is no clear pattern in the residual plot (it just wouldn't be a very strong association).

R^2 values only tell strength of association – NOT appropriateness of the linear model.

b) Does this model allow the student to make accurate predictions? Explain.

No. Since the r-value is only 0.361, this indicates that the residual values would be relatively large, and there is a good chance that most predictions made from the linear model would be pretty far off from the actual (observed) values.

13. **Storks** Data show that there is a positive association between the population of 17 European countries and the number of stork pairs in those countries.

a) Briefly explain what "positive association" means in this context.

As the number of stork pairs increases, so does the population of that European country (or vice-versa).

b) Wildlife advocates want the stork population to grow, so they approach the governments of these countries to encourage their citizens to have children. As a statistician, what do you think of this plan? Explain briefly.

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**.

14. **Car commercials** A car dealer investigated the association between the number of TV commercials he ran each week and the number of cars he sold the following weekend. He found the correlation to be r = 0.56. During the time he collected the data he ran an average of 12.4 commercials a week with a standard deviation of 1.8, and sold an average of 30.5 cars with a standard deviation of 4.2. Next weekend he is planning a sale, hoping to sell 40 cars.

a) Write an equation of the linear model to estimate the number of cars he might sell on a certain weekend based on the number of TV commercials run that week. Define any variable used in this equation or state the equation *in context*. (Show your work as well as your formulas... but only if you want any credit!)

x = # of commercials that week

y = # of cars sold that weekend

$\text{slope} = r \cdot \frac{S_y}{S_x}$

$= 0.56 \times \frac{4.2}{1.8}$

$= 1.3067$

$y\text{-int} = \bar{y} - \text{slope}(\bar{x})$

$= 30.5 - (1.3067)(12.4)$

$= 14.2973$

$$\boxed{\widehat{\#\text{ of cars sold}} = 14.2973 + 1.3067\,(\#\text{ of TV commercials run that week})}$$

b) If the car dealer decides to pay for 18 TV commercials, how many cars might he expect to sell the following weekend?

$\widehat{\text{cars sold}} = 14.2973 + 1.3067(18)$

$= 37.8179$

$\boxed{\text{about 38 cars}}$

c) Let us suppose that in a particular week, the car dealer paid for 10 commercials and sold 22 cars. Calculate the residual for the number of cars sold that weekend, and interpret this value in context.

Actual cars sold = 22

$\widehat{\text{cars sold}} = 14.2973 + 1.3067(10)$

$= 27.364$

Residual = observed − predicted

$e = y - \hat{y}$

$= 22 - 27.364$

$= \boxed{-5.364}$

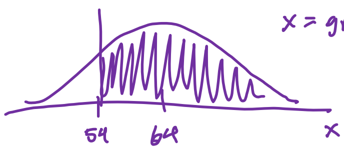
This means the dealer sold 5.364 FEWER cars that the model predicted, given that the dealer paid for 10 commercials.

15. Professor Rogers has found that the grades on the nursing final exam are normally distributed with a mean of 64 and standard deviation of 11.

a) If the passing grade is 54, what percent of the class will pass (will make greater than 54)?

$N(64, 11)$

x = grades on nursing final exam



$P(x > 54)$

$= P(z > -0.91)$

$= 1 - 0.1814$

≈ 0.8186

$\boxed{\text{about } 82\% \text{ of the class}}$

$z = \frac{x - \mu}{\sigma} = \frac{54 - 64}{11}$

$= -0.909090\ldots$

b) If Professor Rogers wants only 85% of the class to pass, what should the passing grade be?

find cutoff for bottom 15%

$z \approx -1.04 = \frac{x - 64}{11}$

$\boxed{x = 52.56}$

16. There is a linear relationship between posted speed limit and the average number of accidents. A least squares fit of some data collected by the department of traffic control gives the model

$$\hat{y} = -10.91 + 0.72x$$

where x is the posted speed limit and $\hat{y}$ is the estimated number of accidents.

a) What is the estimated increase in accidents that corresponds to an increase of 20 mph?

$0.72(20)$

$= \boxed{14.4 \text{ accidents}}$

b) The department of traffic control reported 23 accidents when the posted speed limit was 50 mph. Does the least squares model overestimate or underestimate the number of accidents? Show your thinking.

$\hat{y} = -10.91 + 0.72(50)$

$= 25.09$ ← the model OVEREstimates the # of accidents.

17. Landslides are common events in tree-growing regions of the Pacific Northwest, so their effect on timber growth is of special concern to foresters. The following is information on clear-cut growth, with age of the tree (years) used to predict the 5-year height growth (cm). A scatterplot, a residual plot, and the computer output from a regression analysis are shown:

Predictor	**Coefficient**	**St Dev**	**T**	**P**
Age	21.3	3.45	4.1	0.04
Constant	9.5	5.6	2.07	0.01
S = 51.6	R-Sq = 67%	R-Sq(adj) = 64.2%		

a) Is a linear model appropriate to summarize this data? Explain.

YES. 1) Scatterplot is fairly linear

2) Residual plot has no clear pattern.

b) State the least squares regression line equation that summarizes the relationship between the age of trees in years and the height of the tree. Define any variables used in this equation or state the equation in context.

$$\widehat{\text{tree height}} = 9.5 + 21.3(\text{age})$$

c) Interpret the meaning of the slope of the regression line in context.

For each increase of 1 year in tree age, the model predicts an increase of 21.3 cm in tree height.

d) If meaningful, interpret the y-intercept. If not meaningful, explain why not.

A zero-year old tree is 9.5 cm tall?
(Not sure if that is meaningful...?)

e) What is the percent of variation height of the trees that can be explained by the linear association of tree age in years and height of the trees?

67%

f) Interpret s_e in this context.

51.6 cm is the typical distance between predicted tree height and actual tree height.

g) Predict the height of trees that are 30 years old. Comment on your prediction.

648.5 cm... but 30 years is outside of our data.
this is extrapolation - we cannot assume that this prediction is reliable.