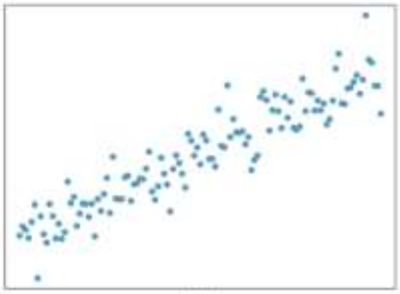
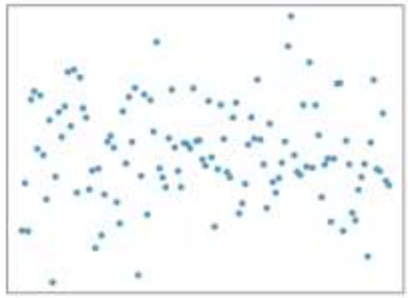
Chapter (11) practice questions

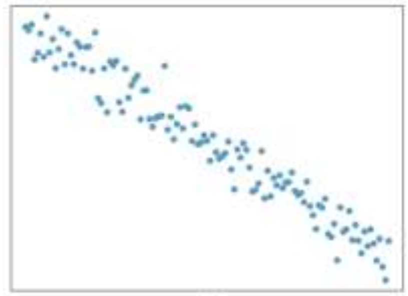
Q1 For each of the following plots, identify the strength of the relationship (e.g. weak, strong) in the data.



A

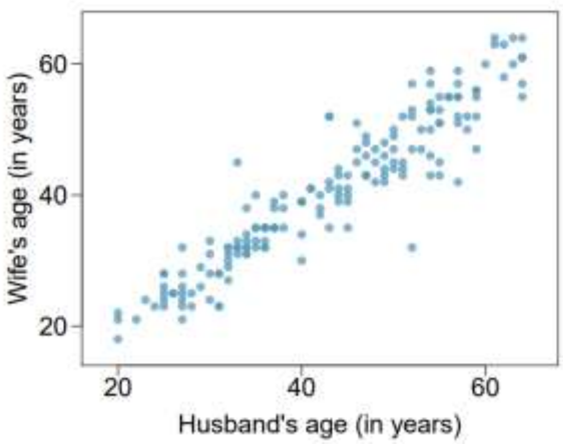


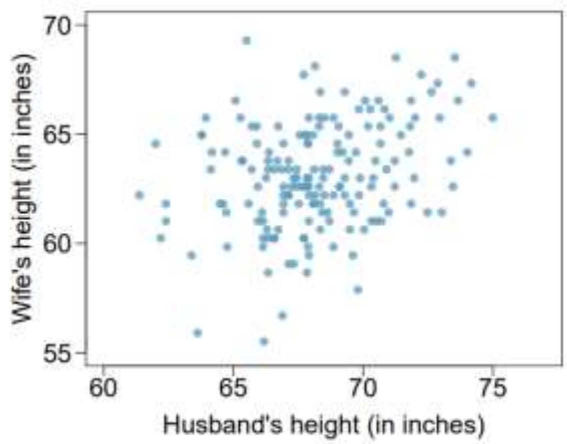
B



C

Q2 The Great Britain Office of Population Census and Surveys once collected data on a random sample of 170 married couples in Britain, recording the age (in years) and heights (converted here to inches) of the husbands and wives. The scatterplot on the left shows the wife's age plotted against her husband's age, and the plot on the right shows wife's height plotted against husband's height.

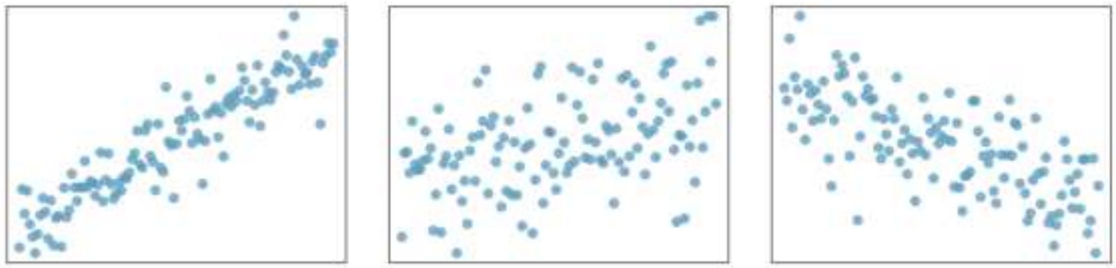




(a) Describe the relationship between husbands' and wives' ages.
(b) Describe the relationship between husbands' and wives' heights.
(c) Which plot shows a stronger correlation? Explain your reasoning.
(d) Data on heights were originally collected in centimeters, and then converted to inches. Does this conversion affect the correlation between husbands' and wives' heights?

Q3 Match each correlation to the corresponding scatterplot.

a) R = −0.7 (b) R = 0.06 (c) R = 0.92



A B C

Q4 Compute the correlation coefficient for the data.

Company	Cars x (in ten thousands)	Revenue y (in billions)
A	63.0	7.0
B	29.0	3.9
C	20.8	2.1
D	19.1	2.8
E	13.4	1.4
F	8.5	1.5

Q5 A study in Bogalusa, Louisiana, and Brooks County, Texas, explored strategies for modifying cardiovascular risk factors in children. The correlation between LDL cholesterol and ponderal index (a measure of obesity) was 0.28 for 122 Caucasian boys and 0.14 for 62 African American boys, highlighting potential modifiable variables.

A) Test the population correlation is different from zero for the Caucasian boys. Use $\alpha = 0.01$.
B) Construct 90% confidence interval for the correlation coefficient of the Caucasian boys.

C) Test the population correlation is different from 0.45 for the African American boys. Find the value of the test statistics.

D) Calculate the p-value for part (C). Use $\alpha = 0.01$.

Q6 The following statistics are taken from an article by Burch relating cigarette smoking to lung cancer. The article presents data relating mortality from lung cancer to average cigarette consumption (lb/person) for 12 females in England and Wales over a 40-year period. The data are given as follows:

$\sum X_i Y_i = -4.125, \sum Xi = 2.38, \sum X_i^2 = 1.310, \quad \sum Yi = -15.55, \sum Y_i^2 = 30.708$

Calculate the correlation coefficient and describe the relationship.

Q7 True or False:

1. A correlation coefficient of +1 indicates a perfect positive linear relationship between two variables.
2. A correlation coefficient of -1 indicates a perfect negative linear relationship between two variables.
3. Correlation implies causation; if two variables are strongly correlated, one must cause the other.
4. A correlation coefficient of 0 indicates no linear relationship between two variables.
5. The correlation coefficient is sensitive to outliers, meaning that extreme values can significantly impact its value.
6. If the correlation coefficient is close to zero, it implies that there is a strong linear relationship between the variables.
7. The correlation coefficient can only measure linear relationships between variables.
8. A positive correlation implies that as one variable increases, the other variable decreases.
9. The correlation coefficient is a measure of the strength and direction of a linear relationship between two variables.
10. Correlation coefficients can only range between -1 and +1, inclusive.

Q8 Suppose the correlation coefficient between weight is 0.78 for the 30 sets of identical twins and 0.50 for the 42 sets of fraternal twins.

A) Test for whether the true correlation coefficients differ from zero for the identical twins, use $\alpha = 0.01$.
B) The z transformation of fraternal twins' correlation.
C) Test for whether the true correlation coefficients differ from 0.80 for the fraternal twins, use $\alpha = 0.10$. Report p-value.
D) Establish 90% confidence interval for the correlation coefficient for fraternal twins.

Q9 Choose the correct answer:

A) The strength of the linear relationship between two quantitative variables is determined by:
   a) Sxy b) Lxy c) r d) Sxx e) Syy

B) To test the significance of r, a ___________ test is used:
   a) χ^2 b) t c) F d) None

C) The test of significance for r has ________ degrees of freedom:
   a) 1 b) n c) n-1 d)n-2

D) We can use _______ to plot the linear relationship between two numeric variables:
   a) Box plot b) scatter plot c) Histogram d) Bar graph.

E) The range of r is _____ to ___ :
   a) -1, +1 b) 0, 1 c) -1,0 d) -2, 2

F) The sign of r and _____ will always be the same:
   a) Lxy b) Sxy c) Lxx d) Lyy e) a+b

G) If r = 0.67, then the z transformation of r =
   a) 0.78 b) 0.56 c) 0.34 d) 0.811

H) The sample correlation coefficient r can then be used as an estimator of ρ if the following assumptions are valid:
   I. The variables x and y are linearly related.
   II. The variables are constant variables.
   III. The two variables have a bivariate normal distribution.
   a) I only b) II only c) I and II d) I and III

Q10 Gas Tax and Fuel Use The data below indicate the state gas tax in cents per gallon and the fuel use per registered vehicle (in gallons). Given that:

Sxy = -198.64, S_x^2= 10.367 and S_y^2 =28211.9.

Is there a significant relationship between these two variables?

Tax	21.5	23	18	24.5	26.4	19
Usage	1062	631	920	686	736	684

Chapter (11) solutions:-

Q1 a) strong positive

b) weak.

c) strong negative.

Q2

a) strong positive.

b) weak postive.

c) the one to the left

d) no, since correlation is unit-free.

Q3 A → (c)

B → (b)

C → (a)

Q4

Company	Cars x (in 10,000s)	Revenue y (in billions)	xy	x^2	y^2
A	63.0	7.0	441.00	3969.00	49.00
B	29.0	3.9	113.10	841.00	15.21
C	20.8	2.1	43.68	432.64	4.41
D	19.1	2.8	53.48	364.81	7.84
E	13.4	1.4	18.76	179.56	1.96
F	8.5	1.5	12.75	72.25	2.25
	$\Sigma x = 153.8$	$\Sigma y = 18.7$	$\Sigma xy = 682.77$	$\Sigma x^2 = 5859.26$	$\Sigma y^2 = 80.67$

$$\lambda_{xy} = \Sigma x \cdot y - \frac{(\Sigma x)(\Sigma y)}{n}$$

$$= 682.77 - \frac{(153.8)(18.7)}{6} = 203.43$$

$$\lambda_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 5859.26 - \frac{(153.8)^2}{6} = 1916.85.$$

$$\lambda_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 80.67 - \frac{(18.7)^2}{6} = 22.39.$$

$$r = \frac{\lambda_{xy}}{\sqrt{\lambda_{xx} \cdot \lambda_{yy}}} = \frac{203.43}{\sqrt{1916.85\,(22.39)}} = 0.982$$

∴ The correlation coeff. suggests a strong relationship between # of cars a rental agency has & It's annual revenue

Q5

Caucasian

$r = 0.28$

$n = 122$

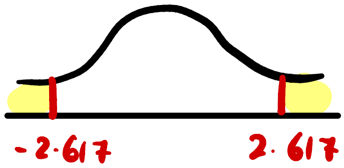
African

$r = 0.14$

$n = 62$

a) $H_0: \rho = 0$ vs $H_1: \rho \neq 0$.

$$t = r \cdot \sqrt{\frac{n-2}{1-r^2}} = 0.28 \cdot \sqrt{\frac{122-2}{1-(0.28)^2}} = 3.195$$



$d.f = 122 - 2 = 120.$

$\alpha = 0.01 \rightarrow \alpha/2 = 0.005 \Rightarrow t_{0.995}^{(120)} = 2.617$

$\therefore$ we rej H_0.

b) $$Z = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right) = \frac{1}{2} \ln\left(\frac{1+0.28}{1-0.28}\right) = 0.288$$ [or by tables]

$1-\alpha = 0.90 \rightarrow \alpha = 0.10 \rightarrow \alpha/2 = 0.05 \longrightarrow 0.90 + 0.05 = 0.9500$

$Z_{\alpha/2} = 1.96$

$$Z_1 = Z - \frac{Z_{\alpha/2}}{\sqrt{n-3}} = 0.288 - \frac{1.96}{\sqrt{122-3}} = 0.108$$

$$Z_2 = Z + \frac{Z_{\alpha/2}}{\sqrt{n-3}} = 0.228 + \frac{1.96}{\sqrt{122-3}} = 0.468$$

$$\rho_1 = \frac{e^{2Z_1}-1}{e^{2Z_1}+1} = \frac{e^{2(0.108)}-1}{e^{2(0.108)}+1} = 0.108$$

$$\rho_2 = \frac{e^{2Z_2}-1}{e^{2Z_2}+1} = \frac{e^{2(0.468)}-1}{e^{2(0.468)}+1} = 0.437.$$

A 90% CI for $\rho \Rightarrow (0.108, 0.437)$

c) $H_0: \rho = 0.45$ vs $H_1: \rho \neq 0.45$

$$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) = \frac{1}{2}\ln\left(\frac{1+0.14}{1-0.14}\right) = 0.141$$

$$z_0 = \frac{1}{2}\ln\left(\frac{1+\rho_0}{1-\rho_0}\right) = \frac{1}{2}\ln\left(\frac{1+0.45}{1-0.45}\right) = 0.485$$

$$\lambda = (z - z_0)\cdot\sqrt{n-3} = (0.141 - 0.485)\cdot\sqrt{62-3} = -2.64$$

d) $P(Z < -2.64) = P(Z > 2.64) = 1 - P(Z < 2.64)$
$= 1 - 0.9959 = 0.0041$

P-value $= 2 \times 0.0041 = 0.0082 < 0.01 \rightarrow$ rej H_0

Q6 $$l_{xy} = \sum x \cdot y - \frac{(\sum x)(\sum y)}{n} = -4.125 - \frac{(2.38)\cdot(-15.55)}{12} = -1.04$$

$$l_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 1.310 - \frac{(2.38)^2}{12} = 0.838$$

$$l_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 30.708 - \frac{(-15.55)^2}{12} = 10.558$$

$$r = \frac{l_{xy}}{\sqrt{l_{xx}\cdot l_{yy}}} = \frac{-1.04}{\sqrt{0.838(10.558)}} = -0.35$$ weak negative.

Q7

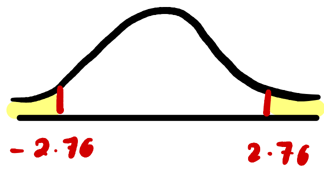
1) True
2) True
3) False, correlation doesn't imply causation.
4) True
5) True
6) False
7) True
8) False
9) True
10) True.

Q_8 identical : $r = 0.78$, $n = 30$

Fraternal : $r = 0.50$, $n = 42$

a) $H_0: \rho = 0$ vs $H_1: \rho \neq 0$.

$$t = r \cdot \sqrt{\frac{n-2}{1-r^2}} = 0.78 \cdot \sqrt{\frac{30-2}{1-(0.78)^2}} = 6.596.$$



$$d.f = 30-2 = 28$$

$$\alpha = 0.01 \rightarrow \alpha/2 = 0.005 \Rightarrow t_{0.995}^{(28)} = 2.763$$

$\therefore$ we rej H_0.

b) $$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) = \frac{1}{2}\ln\left(\frac{1+0.5}{1-0.5}\right) = 0.549$$

c) $H_0: \rho = 0.80$ vs $H_1: \rho \neq 0.80$

$$z = 0.549$$

$$z_0 = \frac{1}{2}\ln\left(\frac{1+\rho_0}{1-\rho_0}\right) = \frac{1}{2}\ln\left(\frac{1+0.80}{1-0.80}\right) = 1.099$$

$$\lambda = (z - z_0)\cdot\sqrt{n-3} = (0.549 - 1.099)\cdot\sqrt{42-3} = -3.43$$

$$P(Z < -3.43) = 0.0003 \rightarrow \text{p-value} = 2 \times 0.0003 = 0.0006$$

p-value < 0.10 $\rightarrow$ we rej H_0.

d) $z = 0.549$, $z_{\alpha/2} = 1.64$

$$z_1 = z - \frac{z_{\alpha/2}}{\sqrt{n-3}} = 0.549 - \frac{1.64}{\sqrt{42-3}} = 0.286$$

$$z_2 = z + \frac{z_{\alpha/2}}{\sqrt{n-3}} = 0.549 + \frac{1.64}{\sqrt{42-3}} = 0.812$$

$$\rho_1 = \frac{e^{2z_1}-1}{e^{2z_1}+1} = \frac{e^{2(0.286)}-1}{e^{2(0.286)}+1} = 0.278$$

$$\Rightarrow (0.278,\ 0.671).$$

$$\rho_2 = \frac{e^{2z_2}-1}{e^{2z_2}+1} = \frac{e^{2(0.812)}-1}{e^{2(0.812)}+1} = 0.671$$

Q9

question	answer	question	answer
A	(c)	E	(a)
B	(b)	F	(e)
C	(d)	G	(d)
D	(b)	H	(d)

Q10

$$S_x = \sqrt{10.367} = 3.22$$

$$S_y = \sqrt{28211.9} = 167.96$$

$$r = \frac{S_{xy}}{S_x \cdot S_y} = \frac{-198.64}{3.22(167.96)} = -0.367$$

$\therefore$ weak Negative corr.