

60s

12.15

do these from 11.50am
until 12.15pm

Questions 18 to 23 refer to the following information.

The researcher believes that the engine size of cars with small to moderate sized engines (under 2500cc) could be used to predict the weight of a car. The results of a linear regression analysis using SPSS and associated plots are shown in Figure 9, Table 14 and Figure 10 (all given below).

Scatter Plot of Wt versus Eng (Eng less than 2500cc)

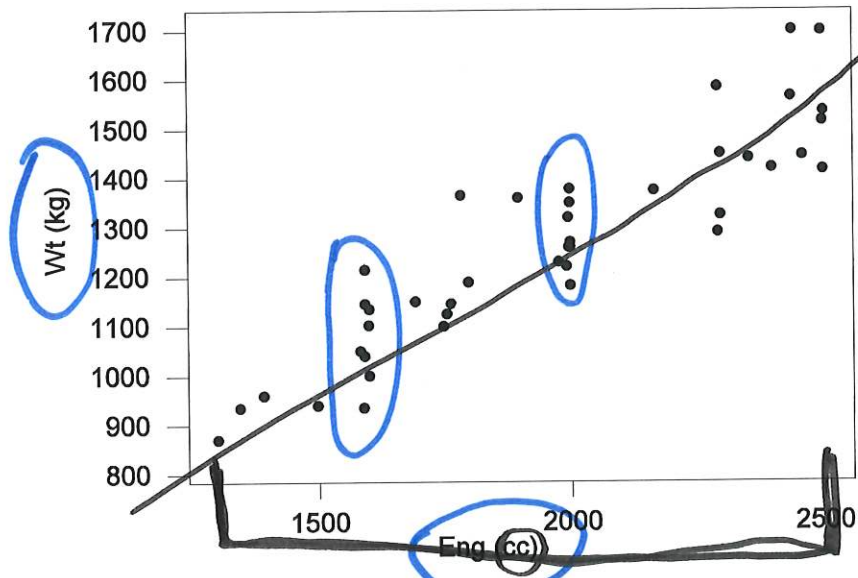


Figure 9: Scatter plot of weight versus engine size for cars with engines smaller than 2500cc

Regression

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	235.41	73.68		3.19	.003
	Eng	.52594	0.03710	.862	14.18	.000

a. Dependent Variable: Wt (kg)

Table 14: SPSS output, linear regression analysis of the relationship between weight and engine size

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\Rightarrow \hat{y} = 235.41 + .52594x$$

$$\Rightarrow \text{pred. weight} = 235.41 + .52594 \times \text{Eng}$$

↑
av.

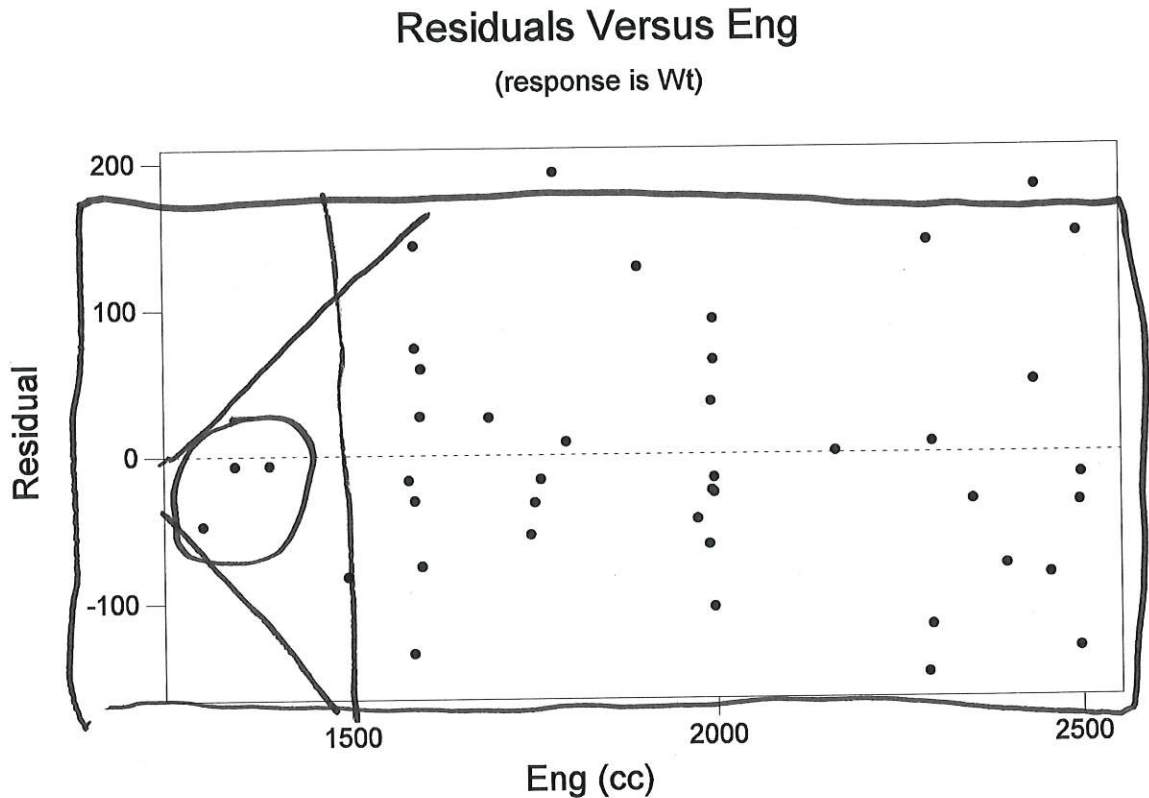


Figure 10: Scatter plot of residuals versus engine size for cars with engines smaller than 2500cc

18. One of the cars in the sample has an engine size of 1590cc and a weight of 1215kg. If a new car has an engine size of 1590cc, the regression equation predicts the car's weight to be approximately:

- (1) 1215kg
- (2) 1826kg
- (3) 836kg
- (4) 1321kg
- (5) 1072kg

$$\hat{y} = 235.41 + .52594 \times 1590$$

$$= 1071.6546$$

cf Q14

19. Another of the cars in the sample has an engine size of 1497cc and a weight of 940kg. Based on the regression equation, the residual for this car is approximately:

- (1) -83kg
- (2) 83kg
- (3) 767kg
- (4) 1023kg
- (5) -767kg

$$\text{res} = \text{obs} - \text{exp}$$

$$= y - \hat{y}$$

$$= 940 - 1023 = -83$$

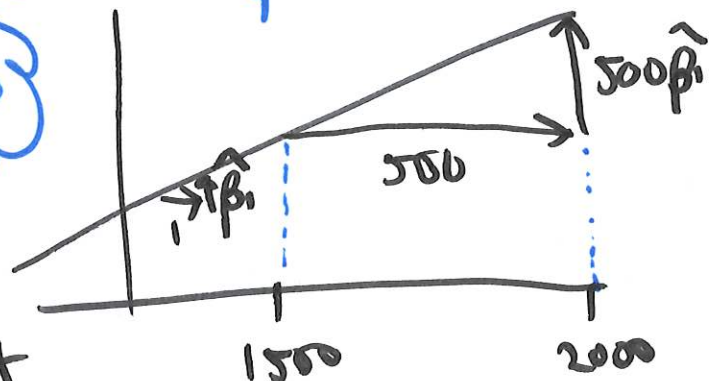
cf Q15

$$\hat{y} = 235.41 + .52594 \times 1497$$

$$= 1022.74218$$

17. Suppose that the engine sizes of two cars differ by 500cc. The regression equation predicts that the difference in the weights of these two cars will be:

- (1) 498kg
(2) 139kg
(3) 263kg
(4) 117.5kg
(5) 504kg



$$500 \times .52594 = 262.97$$

18. In a test for no linear relationship between engine size and weight the hypotheses are:

- (1) $H_0: \beta_0 \neq 0$ and $H_1: \beta_0 = 0$
(2) $H_0: \hat{\beta}_0 = 0$ and $H_1: \hat{\beta}_0 \neq 0$
(3) $H_0: \hat{\beta}_1 = 0$ and $H_1: \hat{\beta}_1 \neq 0$
(4) $H_0: \beta_1 = 0$ and $H_1: \beta_1 \neq 0$
(5) $H_0: \beta_0 = 0$ and $H_1: \beta_0 \neq 0$

Slope, β_1

19. You may need to refer to Figure 9 and Figure 2 to help answer this question. Which **one** of the following statements about this linear regression analysis is **false**?

T (1) It is reasonable to assume that the error terms have a constant underlying standard deviation.

F (2) It would be difficult to have faith in a 95% prediction interval for an engine size of 2150cc ^{not} because there are so few observations with a similar engine size. ^{even though}

T (3) Engine size is a numeric variable and weight is a continuous random variable. ^{numeric}

T (4) It would be unwise to use this data to predict the weight of a car with a 3000cc engine. ^{max 2500cc (1300-2500 is the}

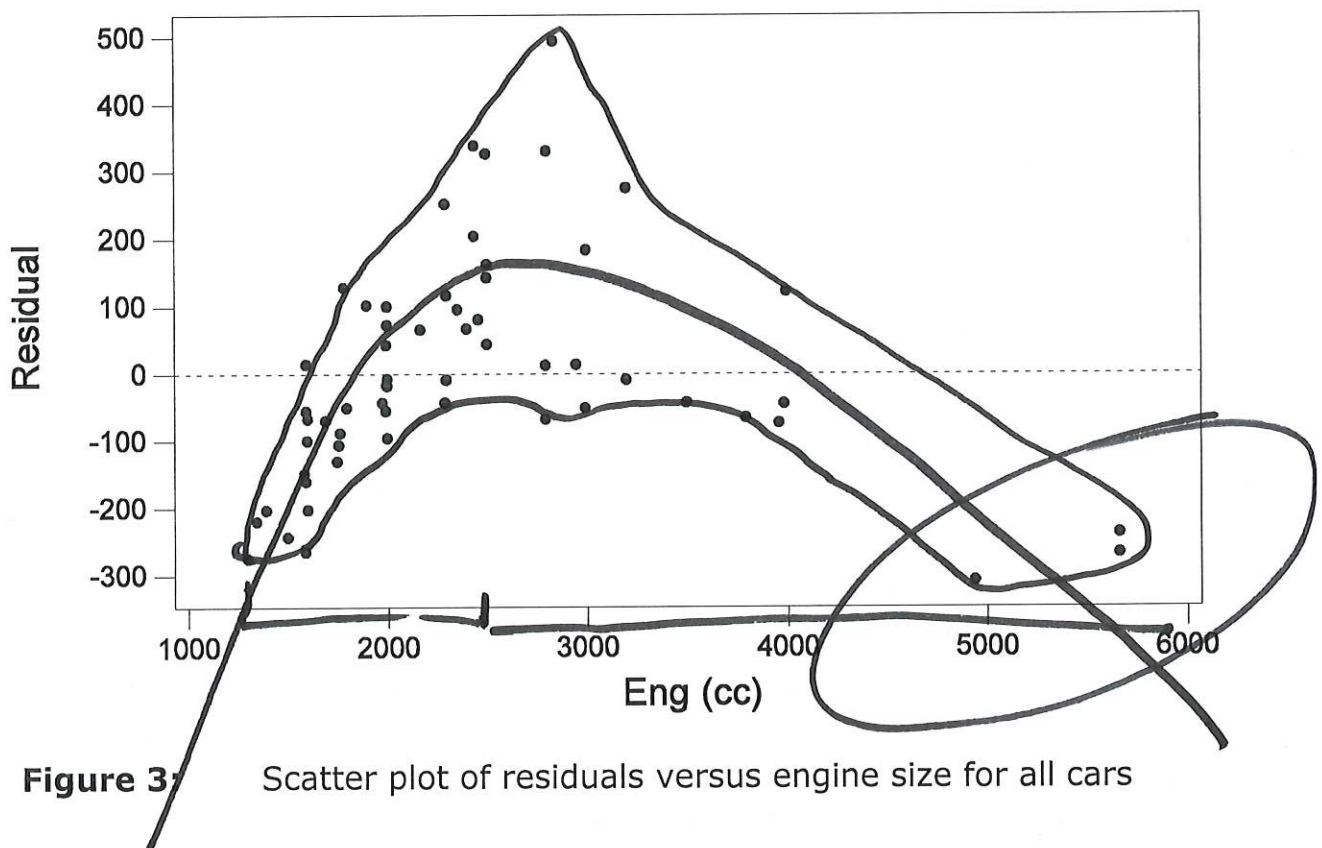
T (5) It is believable that the error terms are Normally distributed with a mean of zero. ^{x range here!}

20. The researcher also used data from a sample of 60 cars (irrespective of engine size, recall there were 43 cars previously) to investigate if engine size could be used to predict the weight of a car. The residual plot in Figure 3 (given below) was produced as part of this investigation. The most useful information provided by this plot about this linear regression model is that:

- ~~(1)~~ the errors are not independent. *can't be checked!*
- ? (2) the errors are not Normally distributed. *outliers*
- ? (3) the relationship between weight and engine size is not linear.
- ? (4) the mean of the errors is not equal to zero.
- ~~(5)~~ all of the assumptions underlying this regression model are satisfied.

Residuals Versus Eng

(response is Wt)



Questions 24 to 32 refer to the following information.

Various measurements were made on 312 African elephants, all with known ages up to 15 years.

Three of the variables recorded were:

Sex — Male
— Female

Height Shoulder height of the elephant in centimetres

Age Age of the elephant in years

continuous
discrete

Assume that these 312 elephants are a random sample of some larger population of the same breed of elephant in the same age range.

A scatter plot of **Height** against **Age** for the male elephants aged between 5 and 15 years is shown in Figure 5 below.

A simple linear regression analysis was carried out on the data from male elephants aged between 5 and 15 years to determine the expected height from age for elephants like these ones. A residual plot of this analysis is shown in Figure 6, page 24, and other regression output is shown in Tables 3 and 4, page 24.

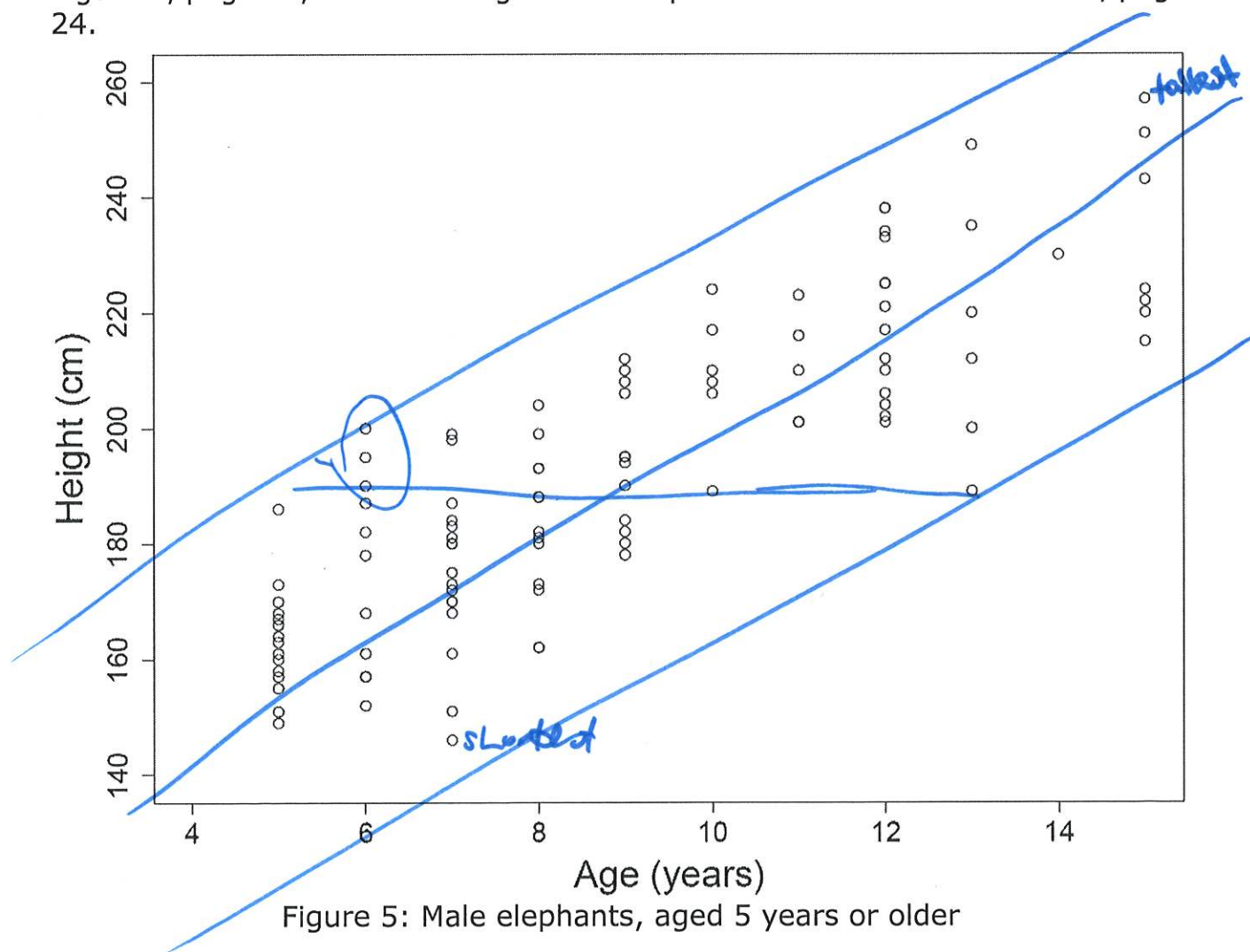


Figure 5: Male elephants, aged 5 years or older

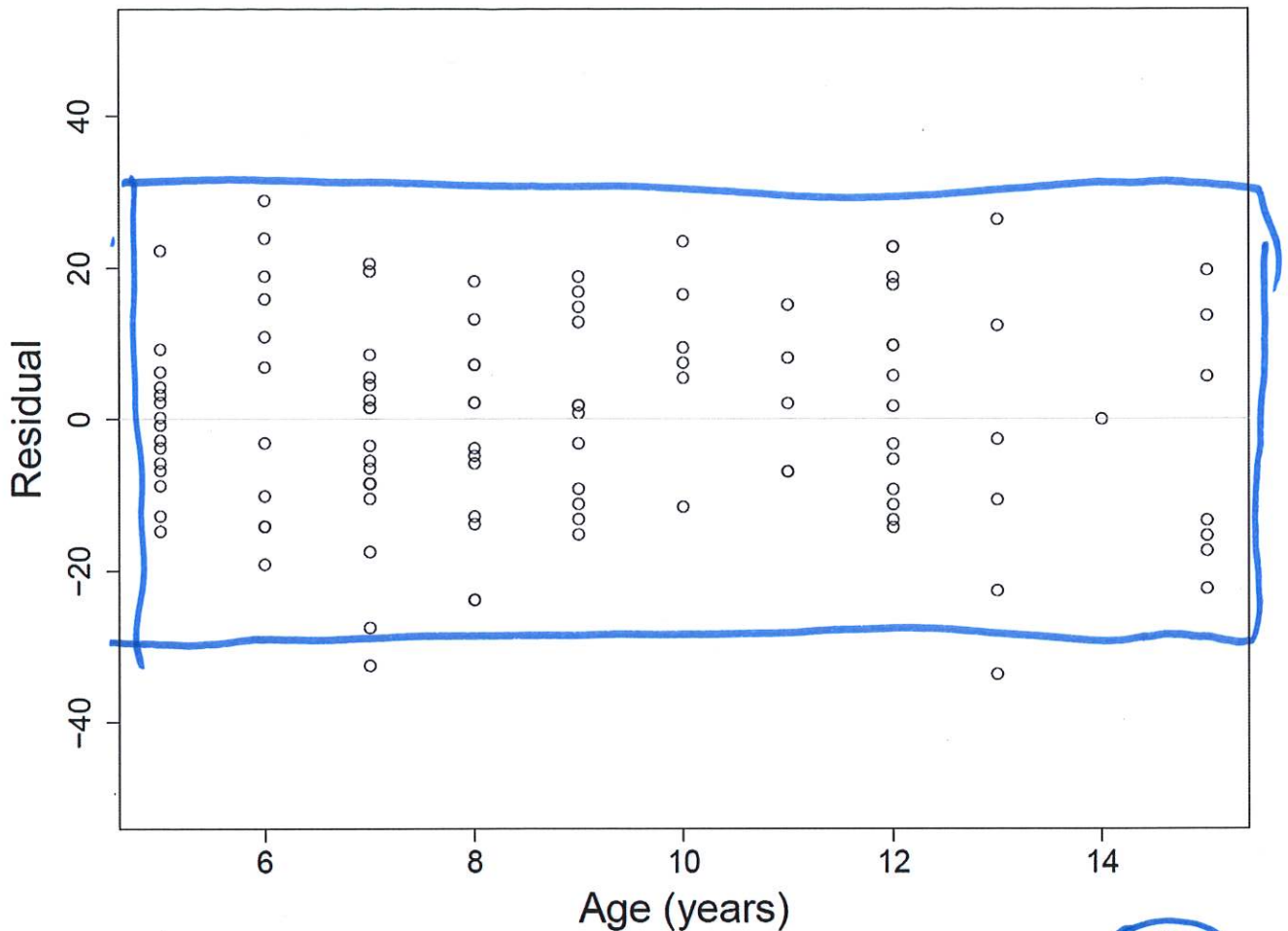


Figure 6: Residual plot for regression of Height against Age (Q27)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	127.005	4.184		30.355	.000	118.710	135.301
	Age	7.357	.447	.848	16.459	.000	6.471	8.243

a. Dependent Variable: Height

Table 3: Simple linear regression output

Sex	Age	Height	LCMI_1	UMCI_1	LCI_1	UICI_1
M	5	161	159.46522	168.11615	136.17266	191.40870
M	6	168	167.48694	174.80854	143.62598	198.66949
M	7	180	175.39612	181.61347	151.05103	205.95856
M	9	193	190.59199	195.84582	165.81551	220.62230
M	12	214	211.47749	219.10266	187.74772	242.83243

Table 4: Simple linear regression prediction output

(Data courtesy of Dr Sam Ferreira, Conservation Ecology Research Unit, University of Pretoria, South Africa.)

Questions 24 and 25 refer to the scatter plot in Figure 5, page 23.

24. Which **one** of the following statements about these male elephants aged between 5 and 15 years is **false**?

- T (1) For any pair of these elephants, the older elephant is not necessarily the taller.
- F (2) This plot shows that by the age of 12 years, elephants like these appear to have, on average, stopped growing in height.
- T (3) This plot suggests, on average, a reasonably constant increase in height from one year to the next for elephants like these.
- T (4) Some of the 6-year-old elephants are taller than at least one of the 13-year-old elephants.
- T (5) The shortest elephant is not in the youngest year group, but the tallest elephant is in the oldest year group.

25. Which **one** of the following statements **best** describes the linear relationship between **Height** and **Age**?

The sample correlation coefficient, r , is closest to:

(1) 0.95

(2) ~~0.35~~

(3) ~~0.85~~

(4) 0.85

(5) 0.35

st. +ve linear assoc.!

26. Based on Figures 5 and 6, pages 23 and 24, which **one** of the following statements is **true**?

T (1) None of the features in these plots raises any concerns about conducting a simple linear regression analysis on these data.

F (2) The residual plot confirms that the errors are related for male elephants like these and so we ~~should~~ be concerned about conducting a simple linear regression analysis on these data.

F (3) The plots clearly suggest an ~~increase in spread~~ of the errors with an increase in age for male elephants like these and so we should be concerned about conducting a simple linear regression analysis on these data.

f (4) The plots provide enough of a hint of the existence of two groups, (5–10 year-olds and 11–15 year-olds) to give us concerns about conducting a simple linear regression analysis on these data.

f (5) There is a strong suggestion in the plots of a non-linear relationship for male elephants like these and it is strong enough for us to be concerned about using a linear model on these data.

• scatterplot shows st. linear trend with reasonably constant scatter ✓

• residual plot shows horizontal band of patternless scatter ✓

Questions 27 to 32 refer to the regression output in Tables 3 and 4 on page 24 and assume that the simple linear regression model is valid.
(Note that this assumption may not be true.)

27. The equation for the least squares regression line for this analysis is:

- (1) Expected **Height** = 127.005 + 7.357 × **Age**
- (2) Expected **Height** = 127.005 + 4.184 × **Age**
- (3) Expected **Height** = 7.357 + 0.447 × **Age**
- (4) Expected **Height** = 7.357 + 0.848 × **Age**
- (5) Expected **Height** = 7.357 + 127.005 × **Age**

see pg 24
for colour-
coding...

28. Under this regression analysis, we would expect the height of an 8-year-old male elephant in this population to be approximately:

- (1) 135 cm
- (2) 158 cm
- (3) 178 cm
- (4) 186 cm
- (5) 173 cm

$$\hat{y} = 127.005 + 7.357 \times 8 = 185.861$$

95% CI for β_1 !

29. Which **one** of the following statements is the **best** interpretation of the confidence interval (6.471, 8.243) given in Table 3, page 24?

With 95% confidence, we estimate that for male elephants in this population aged between 5 and 15 years:

- F (1) the y-intercept of the true regression line is somewhere between 6.471 and 8.243.
- F (2) there is, on average, an increase in age of somewhere between 6.471 and 8.243 years associated with each additional centimetre increase in height.
- F (3) there is, on average, an increase in height of 1.772 cm for each additional year in age. *with an MC of .886 → 8.243 - 6.471!*
- T (4) the slope of the true regression line is somewhere between 6.471 and 8.243.
- T (5) there is, on average, an increase in height of somewhere between 6.471 cm and 8.243 cm for each additional year in age.

Questions 30 and 31 refer to the t -test for no linear relationship between **Height** and **Age**.

30. In a test for no linear relationship between **Height** and **Age**, the hypotheses are:

- (1) $H_0: \beta_0 \neq 0$ $H_1: \beta_0 \neq 0$
 (2) $H_0: \beta_1 = 0$ $H_1: \beta_1 \neq 0$
 (3) $H_0: \beta_1 = 0$ $H_1: \beta_1 \neq 0$
 (4) $H_0: \beta_1 \neq 0$ $H_1: \beta_1 \neq 0$
 (5) $H_0: \beta_0 = 0$ $H_1: \beta_0 \neq 0$

31. Which **one** of the following is a **correct** interpretation of the P -value for the test for no linear relationship?

- (1) We have evidence of a very strong positive linear relationship between **Height** and **Age**.
 (2) We have very strong evidence of no linear relationship between **Height** and **Age**.
 (3) We have very strong evidence of a linear relationship between **Height** and **Age**.
 (4) We have evidence of no linear relationship between **Height** and **Age**.
 (5) We have very strong evidence of a very strong positive linear relationship between **Height** and **Age**.

32. Referring to Table 4, page 24, which **one** of the following statements is **true**?

With 95% confidence, we estimate that:

- (1) the mean height of all 9-year-old male elephants in this population is 193.21891 cm.
 (2) the mean height of all 7-year-old male elephants in this population is somewhere between 181 cm and 206 cm.
 (3) a 12-year-old male elephant in this population is somewhere between 211 cm and 219 cm in height.
 (4) a 5-year-old male elephant in this population is 163.79068 cm in height.
 (5) a 6-year-old male elephant in this population is somewhere between 144 cm and 199 cm in height.