St. ev. against Ho

33\. Suppose we wish to test for no linear relationship between heart rate and temperature. We find the *P-value* is less than 1%. We can **correctly conclude** that the *P-value*:

F (1) indicates strong evidence against the null hypothesis, therefore the data contains strong evidence that a perfect linear relationship exists between heart rate and temperature.

T (2) indicates strong evidence against the null hypothesis, therefore the data contains strong evidence that a linear relationship exists between heart rate and temperature.

F (3) indicates strong evidence against the null hypothesis. However, this tells us nothing about whether or not a linear relationship exists between heart rate and temperature.

F (4) indicates strong evidence against the null hypothesis, therefore the data contains strong evidence that a causal linear relationship exists between heart rate and temperature.

F (5) is very small, therefore the data contain no evidence that a linear relationship exists between heart rate and temperature.

34\. Which **one** of the following statements is **false**?

T (1) A single outlier can have a large influence on the value of the sample correlation coefficient.

T (2) For a least squares regression line, if you add up all the residuals then the total is zero.

F (3) The X-variable is called the independent or explanatory variable and Y-variable is called the dependent or response variable.

T (4) For a simple linear regression, the (average) pattern seen in the scatter plot must be a straight line.

T (5) The two important components of a regression are the average pattern (trend) and the deviation of the observations from that pattern (scatter about the trend).

35\. Which **one** of the following statements is **not** a reason for fitting a linear regression model to the data?

(1) To estimate parameters in a theoretical model.

(2) To make predictions.

(3) To understand a relationship better.

(4) To find the trend line.

(5) To conclusively establish the cause of an effect.

↳ only if well-designed, well-executed experiment!

36. Consider the point labelled "A" in Figure 4.

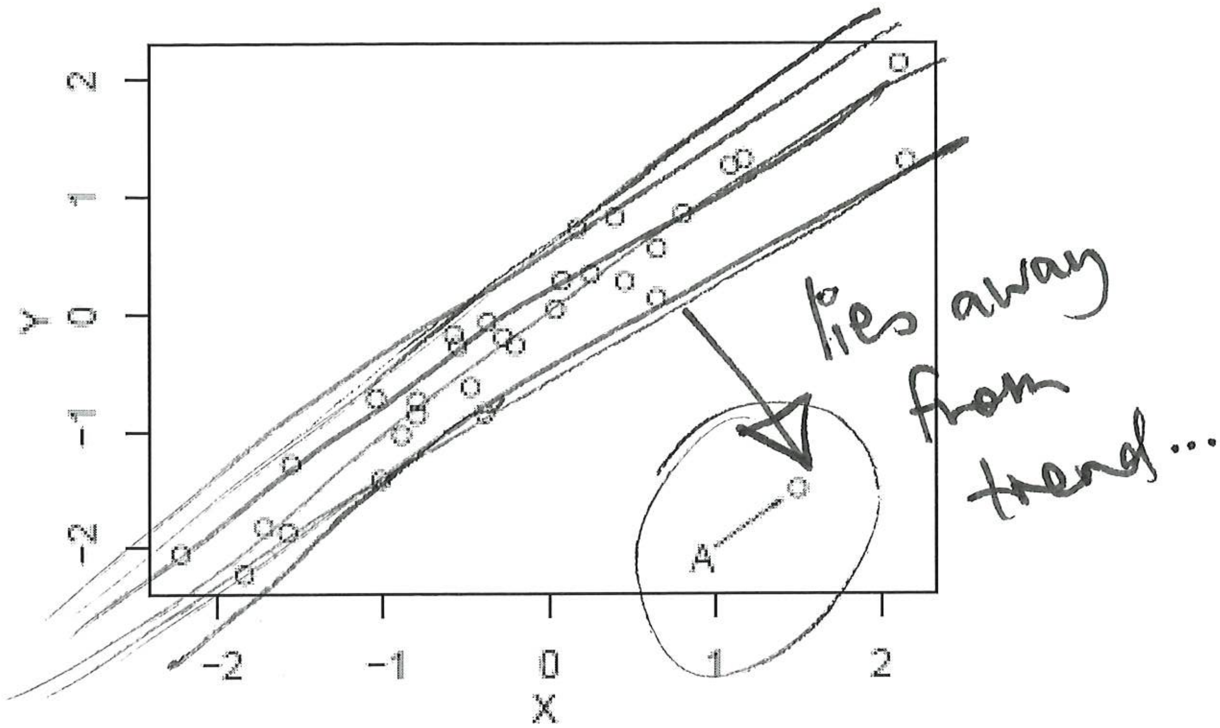


Figure 4: A scatter plot.

Which **one** of the following statements about point "A" is **true**?

F (1) The point is an outlier in X. in range of other x's!

F (2) The point is an outlier in Y. " " " " y's!

F (3) It is impossible to tell whether the point is an outlier without looking at a plot of the residuals.

F (4) The point should be removed from the analysis. (only if mistake!)

T (5) The point is an outlier because it lies much further from the linear trend than the other points.

37. Which **one** of the following assumptions of the simple linear model **cannot** be checked in a residual plot?

(1) ✓ The random errors have a mean of zero.

(2) ✓ The random errors are Normally distributed.

(3) ✓ The random errors have the same standard deviation regardless of the value of x.

(4) ✓ There is a linear relationship between x and E(Y).

(5) ✗ The observations are independent.

38. Which **one** of the following is **not** an assumption of the simple linear regression model?

(1) The random errors have the same standard deviation, regardless of the value of x.

(2) The random errors are all independent.

(3) The ~~random errors~~ observations follow a linear trend.

(4) The random errors have a mean of zero.

(5) The random errors are Normally distributed.

39. Which **one** of the following statements about correlation is **false**?

T (1) Correlation can be used only if both variables are quantitative.

T (2) A scatter plot of the data should be examined before looking at correlation.

T (3) Outliers can deflate the value of the sample correlation coefficient *r*.

T (4) Correlation should not be used when, in a scatter plot, the relationship between two variables appears non-linear.

F (5) Correlation can be used as proof of a causal relationship regardless of how the data were collected. correlation does not prove causation!

40. Which **one** of the following statements about a 95% prediction interval and the corresponding 95% confidence interval for the mean is **true**?

The prediction interval:

(1) can be either narrower or wider than the confidence interval for the mean, depending on the estimated variability of the values for the slope and intercept of the line.

(2) is always narrower than the confidence interval for the mean because it only takes into account the uncertainty about the values of the slope and intercept of the line and not the random scatter about the line.

(3) is always wider than the confidence interval for the mean because as well as taking into account the uncertainty about the values of the slope and intercept of the line, it also takes into account the uncertainty due to the random scatter about the line.

(4) is always narrower than the confidence interval for the mean because it only takes into account the uncertainty about the value of the slope of the line and not the value of the intercept of the line.

(5) is always wider than the confidence interval for the mean because as well as taking into account the uncertainty about the value of the slope of the line, it also takes into account the uncertainty about the value of the intercept of the line.

+ scatter also... X

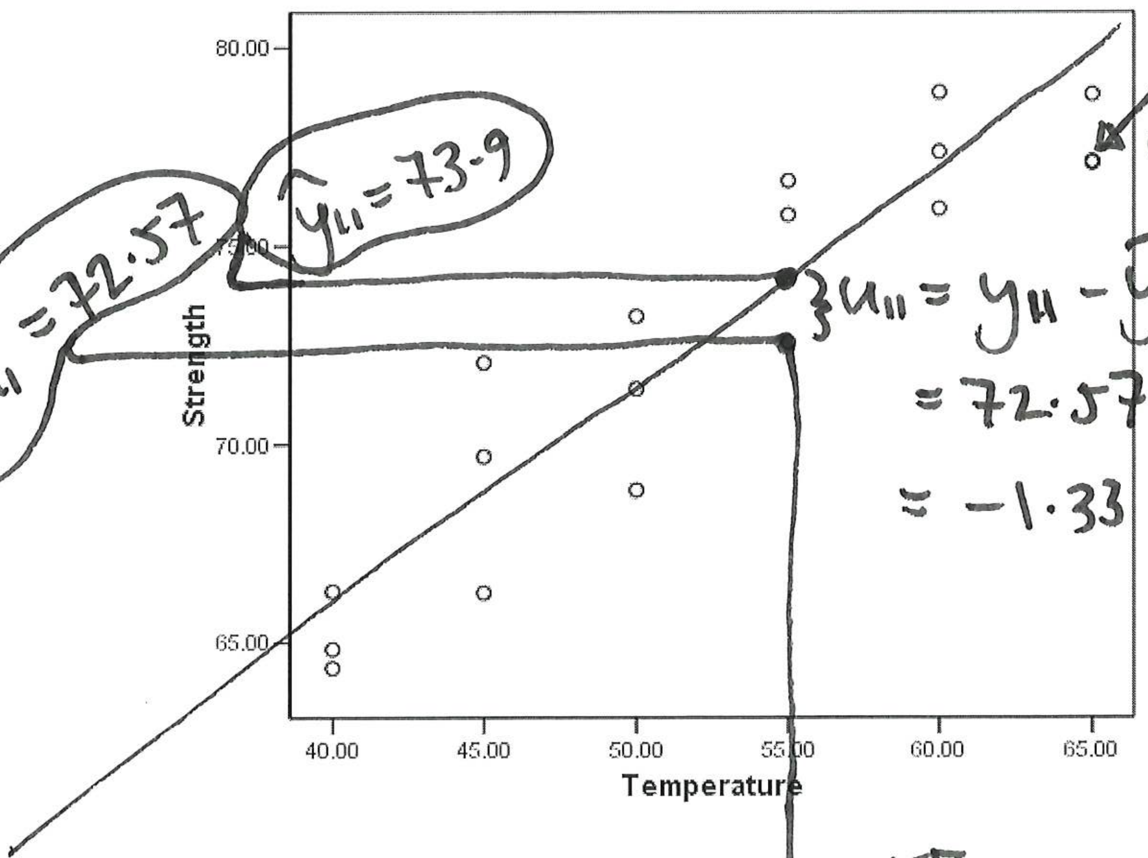
Questions 41 to 49 refer to the following information.

The production of particle boards involves a step in which the boards are baked. A manufacturer of particle boards investigated the effect of the baking temperature (X) on the strength of particle boards (Y). A total of 18 particle boards were baked using 6 different temperatures (3 boards were baked at each temperature) and the strength of these boards was measured. All aspects of the process other than the baking temperature were kept as similar as possible. The assignment of temperatures to boards and the order of production were determined using random processes. The data follows:

	Strength	Temperature (^oC)		Strength	Temperature (^oC)
1	66.30	40	10	75.78	55
2	64.84	40	11	72.57	55
3	64.36	40	12	76.64	55
4	69.70	45	13	78.87	60
5	66.26	45	14	77.37	60
6	72.06	45	15	75.94	60
7	73.23	50	16	78.82	65
8	71.40	50	17	77.13	65
9	68.85	50	18	77.09	65

A scatter plot and some computer output of these data are given below:

Scatterplot of particle board strength against temperature



Regression

$\hat{\beta}_0$

Coefficients(a)

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	45.454	2.846		15.972	.000
	Temperature	.518	.054	.924	9.672	.000

a Dependent Variable: Strength

x, y, $\hat{\beta}_1$

pred. str. = 45.454 + .518 × temp

41. The fitted least squares regression line for these data is:

(1) Predicted Temperature = 45.454 + 0.518 × Strength

(2) Predicted Strength = 2.846 + 0.054 × Temperature

(3) Predicted Strength = 45.454 + 0.518 × Temperature

(4) Predicted Strength = 0.518 + 45.454 × Temperature

(5) Predicted Temperature = 0.518 + 45.454 × Strength

42. For the data in the scatter plot on page 25, which one of the following values would be the closest to the sample correlation coefficient, r?

(1) $r = 0.9$

(2) $r = 0.6$

(3) $r = 0.3$

(4) $r = 0.5$

(5) $r = 0.1$

43. The correct null and alternative hypotheses to test that there is no linear relationship between particle board strength and baking temperature are:

(1) $H_0 : \hat{\beta}_0 = 1$ and $H_1 : \hat{\beta}_0 \neq 1$

(2) $H_0 : \beta_1 = 0$ and $H_1 : \beta_1 \neq 0$

(3) $H_0 : \beta_1 = 1$ and $H_1 : \beta_1 \neq 1$

(4) $H_0 : \beta_0 = 0$ and $H_1 : \beta_0 \neq 0$

(5) $H_0 : \hat{\beta}_1 = 0$ and $H_1 : \hat{\beta}_1 \neq 0$

β_1, slope!

44. To test the hypotheses from the previous question, the test statistic would be:

(1) $t_0 = \dfrac{72.62}{2.846}$

(2) $t_0 = \dfrac{0.518}{2.846}$

(3) $t_0 = \dfrac{45.454}{2.846}$

(4) $t_0 = \dfrac{0.518}{0.054}$

(5) $t_0 = \dfrac{45.454}{0.054}$

45. When the hypothesis test referred to in Questions 43 and 44 is conducted, it is found that there is very strong evidence against the hypothesis of no linear relationship between baking temperature and particle board strength. Which **one** of the following statements is true for this investigation?

(1) The results of this hypothesis test can be taken as evidence that changes in baking temperature *cause* changes in particle board strength since very strong evidence of a relationship always implies causation.

(2) The results of this hypothesis test cannot be taken as evidence that changes in baking temperature *cause* changes in particle board strength since there may be factors other than baking temperature which affect the strength of particle boards.

(3) The results of this hypothesis test cannot be taken as evidence that changes in baking temperature *cause* changes in particle board strength since strong evidence of a relationship does not necessarily mean the relationship is causal.

(4) The results of this hypothesis test cannot be taken as evidence that changes in baking temperature *cause* changes in particle board strength since the scatter plot shows that for each baking temperature there is a substantial amount of variability in the strength of particle boards.

(5) The results of this hypothesis test can be taken as evidence that changes in baking temperature *cause* changes in particle board strength since a random process was used to assign boards to temperatures.

46. The fitted least squares regression line can be used to predict the strength of particle board. Boards baked at a temperature of 55°C have a predicted strength of approximately:

(1) 102.4

(2) 73.9

(3) 13.8

(4) 79.9

(5) 47.0

47. The baking temperature and particle board strength for sample number 11 was 55°C and 72.57 units respectively. Under the fitted least squares line, the value of the residual for this sample is approximately:

(1) 2.30

(2) 0.65

(3) 1.33

(4) −0.65

(5) −1.33

res = obs − exp

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

$= 72.57 - 73.9$

48. The fitted least squares regression line indicates that for each increase of 2.5°C in baking temperature we expect that, on average, the particle board strength will:

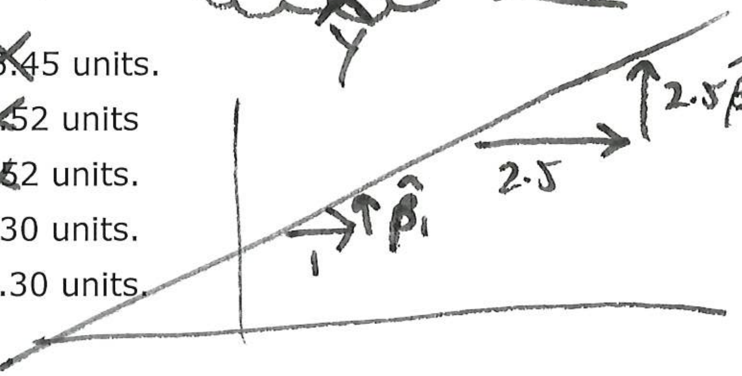
(1) increase by approximately 45.45 units.

(2) decrease by approximately 0.52 units

(3) increase by approximately 0.52 units.

(4) increase by approximately 1.30 units.

(5) decrease by approximately 1.30 units.



2.5×.518=1.295

49. Why is the prediction interval for the strength of particle board baked at a temperature of 55°C more useful than the corresponding point estimate given in question 34? estimate (73.9)

F (1) Because the prediction interval is **only one** of the plausible values of that the strength could be when the baking temperature is 55°C.

(2) The prediction interval is more useful than the point estimate as it estimates (with a certain level of confidence) a range of plausible values the strength could be when the baking temperature is 55°C.

F (3) Because the prediction interval is smaller [wider] than a corresponding confidence interval and therefore can capture the true value more accurately. → 95% of the time!

F (4) The prediction interval is more useful than the point estimate as it always captures the true estimate (with a certain level of confidence) in a range of plausible values the strength could be when the baking temperature is 55°C.

F (5) The prediction interval is not as useful as a corresponding confidence interval. → different purposes, both useful!

(depends on whether you want to estimate average/mean [CI] or next actual/individual value [PI]!)