Key

4 - 26 - 10 PREP Open Book

1-13. From the data table below, calculate requested quantities.

x	y	x^2	y^2	xy
1	3	1	9	3
4	9	16	81	36
3	7	9	49	21
5	10	25	100	50
—	—	—	—	—
3.25	7.25	12.75	59.75	27.5

ck #2 $x+y$: 4, 13, 10, 15; AVG 10.5

1. Average of list 4x + 1. $\overline{4x+1} = 4\bar{x}+1 = 4(3.25)+1 = 14$

a) 13 b) 15 c) 14 d) 16 e) 15

2. Average of list x + y $\overline{x+y} = \bar{x}+\bar{y} = 3.25+7.25 = 10.5$

a) 12 b) 11 c) 9 d) 10.5 e) 13.5

3. $\hat{\sigma}_y$ (equal to $\sqrt{\overline{y^2} - \bar{y}^2}$) is 2.681. What is $\hat{\sigma}_x$? $\hat{\sigma}_x = \sqrt{\frac{\sum_1 (x-\bar{x})^2}{n}} = \sqrt{\overline{x^2}-\bar{x}^2}$

$= \sqrt{12.75-3.25^2} = 1.479$

a) 1.643 b) 1.425 c) 1.479 d) 2.105 e) 1.886

4. S_y (equal to $\sqrt{\frac{n}{n-1}}\ \hat{\sigma}_y$) is 3.096. What is S_x? $s_x = \sqrt{\frac{n}{n-1}}\ \hat{\sigma}_x = \sqrt{\sum_1 (x-\bar{x})^2/(n-1)}$

$= \sqrt{4/3}\ 1.479\cdots = 1.707$

a) 1.90 b) 1.65 c) 1.71 d) 2.43 e) 2.18

5. $S_{2x+4} = s_{2x} = |2|\, s_x = 2 s_x$

a) 2 S_x b) 2 S_x + 4 c) S_x d) 4 S_x e) 4 S_x + 4

6. r[x, y] (equal to $\frac{\overline{xy}-\overline{x}\,\overline{y}}{\hat{\sigma}_x \hat{\sigma}_y}$) $= \frac{27.5 - 3.25\ 7.25}{1.479\ 2.681} = 0.9930\ldots$

a) 0.84 b) 0.99 c) 0.80 d) 0.72 e) 0.91

7. r[2x+4, 9y-8] $= r[x, y]$ Since $2(9) > 0$.

a) 18 r[x, y] b) 18 r[x, y] + 12 c) r[x, y] + 12 d) r[x, y] e) r[x, y] + 50

8. b_1 (slope of sample L.S. line) is

a) S_y/S_x b) r[x, y] S_y/S_x c) r[x, y] d) $\sqrt{1-r[x,\ y]^2}\ S_y/S_x$

9. Usual estimate of slope β_1 (of population L.S. line) is $b_1 = r\ s_y/s_x = 1.8$

a) 1.800 b) 1.048 c) 0.886 d) 0.773 e) 0.993

10. SE(b_1) = $\sqrt{\frac{1-r^2}{n-2}\ \frac{b_1}{r}}$

a) 1.9426 b) 2.131 c) 0.0834 d) 0.1510 e) 0.0007

11. Applicable df if population is 2D normal (or if, more generally, the normal errors model applies)

a) 4 b) 2 c) 3 d) 5 e) 1 $df = n-2 = 4-2 = 2$

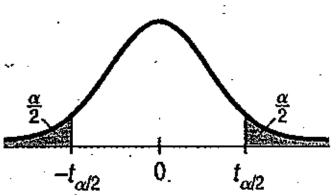
12. Applicable t-score for 95% CI for β_1 95% CONFIDENCE, $df = 2$

a) 12.706 b) 4.303 c) 3.182 d) 2.776 e) 2.571

13. 95% CI for β_1. $b_1 \pm t\ SE(b_1) = 1.8 \pm 4.303\ 0.1510$

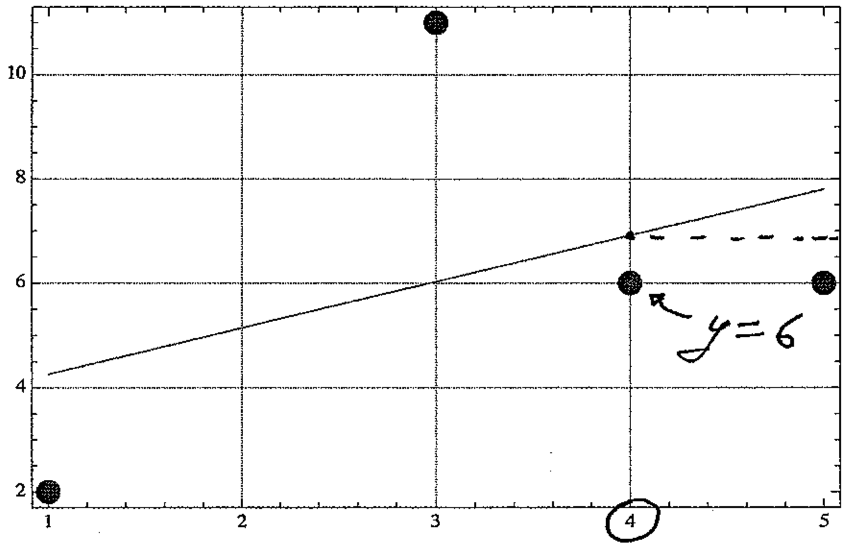
a) {0.826341, 1.85352} b) {0.182634, 2.20535}
c) {0.634182, 2.35205} d) {0.218263, 3.00535} (e) {1.5025, 2.4498}

Two-tail probability		0.20	0.10	0.05	0.02	0.01	
One-tail probability		0.10	0.05	0.025	0.01	0.005	
Table T	df						df
Values of t_α	1	3.078	6.314	12.706	31.821	63.657	1
→	2	1.886	2.920	4.303	6.965	9.925	2
	3	1.638	2.353	3.182	4.541	5.841	3
	4	1.533	2.132	2.776	3.747	4.604	4
	5	1.476	2.015	2.571	3.365	4.032	5
	6	1.440	1.943	2.447	3.143	3.707	6
	7	1.415	1.895	2.365	2.998	3.499	7
	8	1.397	1.860	2.306	2.896	3.355	8
	9	1.383	1.833	2.262	2.821	3.250	9
	∞	1.282	1.645	1.960	2.326	2.576	∞
Confidence levels		80%	90%	95%	98%	99%	



Two tails

14-16. Answer questions about the data with its L.S. fit.



$\hat{y} \sim 7.8$ (BY EYE)
CALL IT 8

14. Residual $y - \hat{y}$ for the point with x = 4. $y - \hat{y} \sim 6 - 7.8 \sim -2$ CLOSEST

CLOSEST TO

a) -1 b) 2 c) 3 d) -2 e) 1

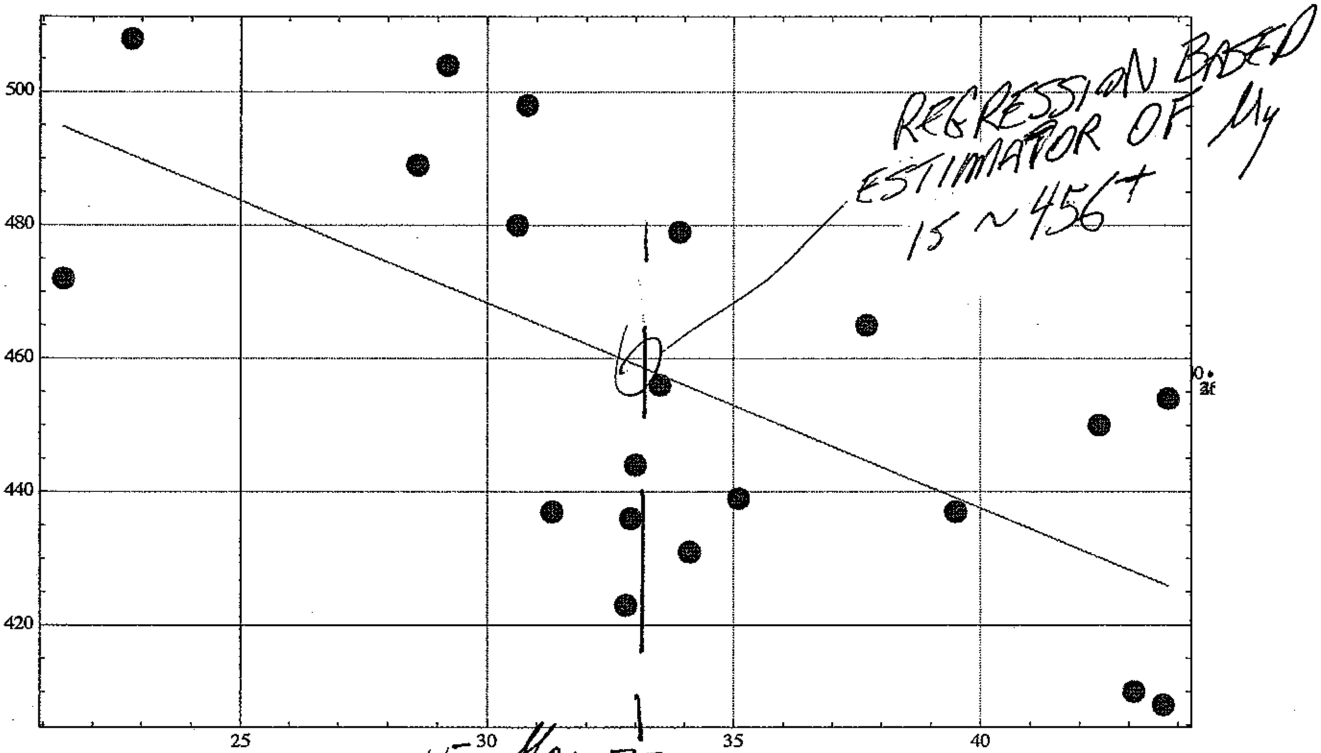
15. b_1 (closest value, by eye) $\frac{RISE}{RUN} = \frac{RISE \sim 4 \text{ TO } 8}{RUN = 1 \text{ TO } 5}$ ie $\frac{8-4}{5-1} \sim 1$

a) 1.7 b) 2.6 c) 1.0 d) 0.5 e) 1.8

16. Contribution of the point (4, 6) to the sum of squares of residuals.

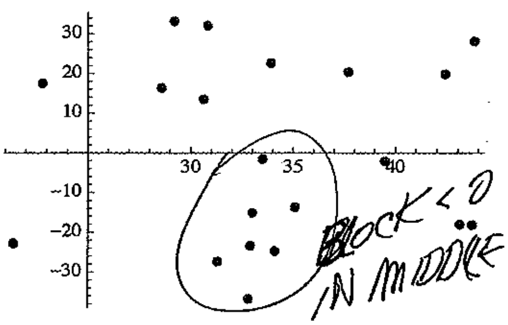
a) 1 b) 4 c) 9 d) -1 e) -3 $(y-\hat{y})^2 \sim 4$

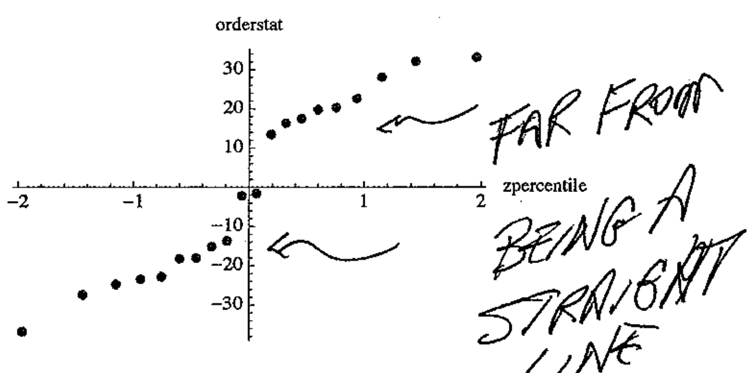
17-20. Here is a data plot
x = time at table
y = calories consumed
together with regression statistics. Also shown are a plot of the residuals versus x (left plot) and a normal probability plot of the residuals (right plot).



regrstats[x_, y_] := {mean[x], mean[y], s[x], s[y], r[x, y], b_1}

34.01	456	6.31647	29.9403	-0.649167	-3.07707





17. Would you agree that the plots strongly confirm normal errors?

a) only left does b) both do c) only right does d) neither

FRACTION

18. What ~~percentage~~ of the sample variance of y is explained by L.S. fit of y on x?

a) 1-(-0.65) b) $(-0.65)^2$ c) $\sqrt{1-(-0.65)^2}$ $r = -0.649 \cdot \sim -0.65$

d) $\sqrt{1-(-3.08)^2}$ e) 0.65 ANS. $r^2 (\times 100\%)$

19. If the population mean is **known** to be $\mu_x = 33$ then the regression-based estimate of μ_x is (from the least squares plot) around ~456 SEE PLOT ABOVE.

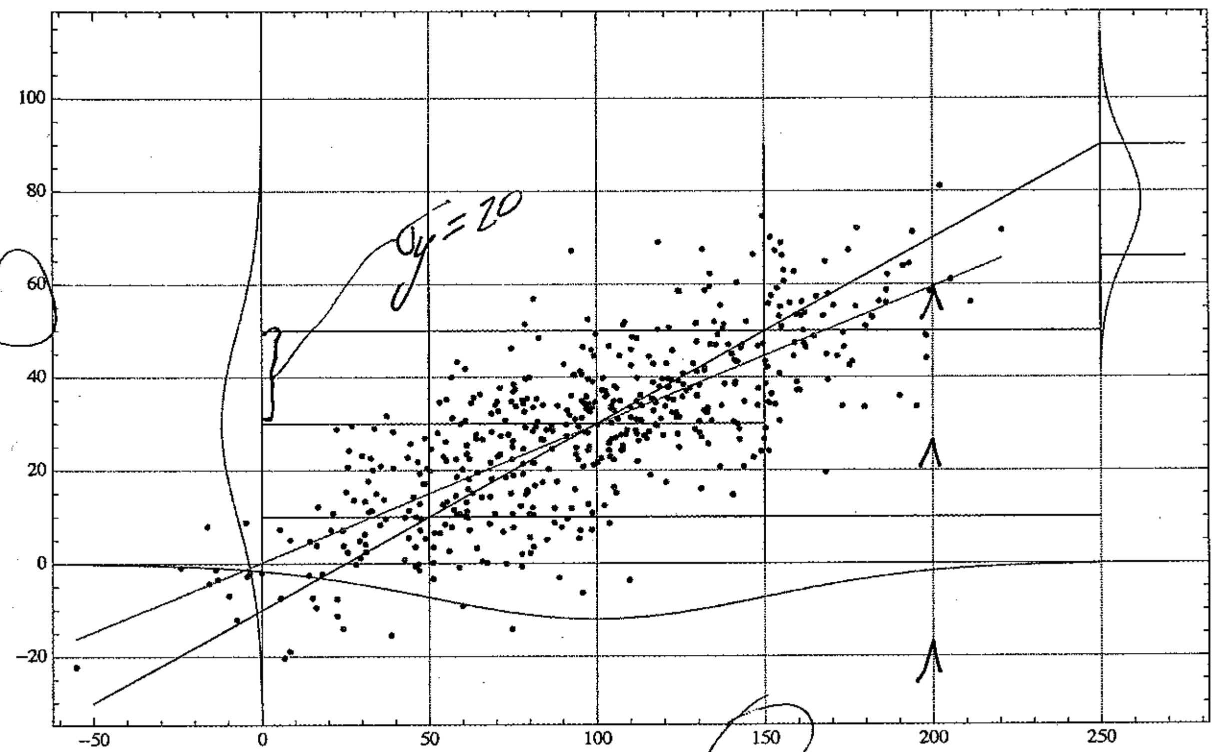
a) 456 b) 430 c) 440 d) 468 e) 424

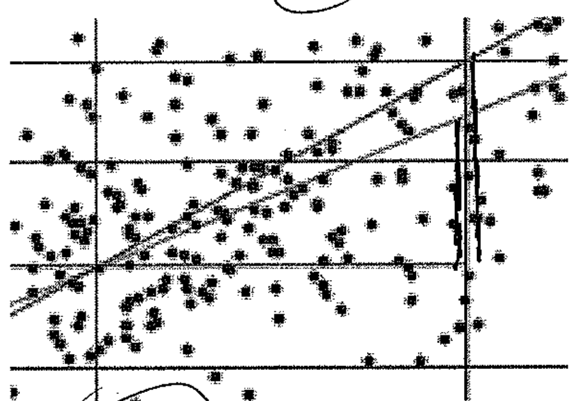
20. If the population mean is **known** to be $\mu_x = 33$ then the SE of the regression-based estimate of μ_x is $s_y/\sqrt{n}$ multiplied by

a) $\sqrt{1-(-3.08)^2}$ b) 0.65 c) 1-(-0.65) $\sqrt{1-r^2}$

d) $(-0.65)^2$ e) $\sqrt{1-(-0.65)^2}$

21-23. Read off the requested **population** quantities from the **population** fit show below.





21. ρ (CLOSEST TO)

$r = \frac{\text{RISE OF L.S.}}{\text{RISE OF NAIVE}}$

~ 0.75

BY EYE

a) 0.5 b) 0.9 c) 0.2 d) 0.4 e) 0.75

22. $\mu_X + \sigma_X$

a) 130 b) 180 c) 150 d) 100 e) 200 SEE PLOT ABOVE

23. L.S. prediction of y for x = 200.

a) 70 b) 40 c) 30 d) 60 e) 50 SEE PLOT ABOVE

24. Mean of all population y with x = 200. SAME AS #23 FOR NORMAL

a) 60 b) 50 c) 70 d) 40 e) 30

25. SD of all y with x = 200.

a) $20\sqrt{1-\rho}$ b) 20 c) $\sqrt{1-\rho}$ d) $\sqrt{1-\rho^2}$ e) $20\sqrt{1-\rho^2}$ $\sigma_y = 20$ (SEE PLOT)

26-32. (a) True or (b) False?

26. For a sample of many pairs (x, y) from a 2D normal distribution the plot of vertical strip averages of y for each value x is roughly a straight line. TRUE (a). GALTON'S DISCOVERY (HTS.).

27. For a sample of n = 2 pairs (x, y) from a 2D normal population a t-based CI for the population slope β_1 applies with df = 1. FALSE (b).

28. Galton's plot of pairs x = parental sweet pea size, y = filial sweet pea size for EQUAL #S OF parent seeds selected at x = 0, ±1, ±2, ± 3 units of s_x from $\overline{x}$, the plot appeared elliptical. FALSE (b)

29. The Least Squares line is always the same as the regression line. FALSE (b).

30. For a large enough sample size n from any given population of (x, y), the sample least squares line will most probably fall very close to the population least squares line. FALSE (b)

31. For a large enough sample size n, the joint distribution of ($\overline{x}$, $\overline{y}$) will be approximately 2D normal even though the population may not be 2D normal. Moreover, the least squares line for that 2D normal will be the same as the population least squares line. TRUE (a).

32. Galton's regression effect implies that, owing to the 2D normality of the plot, sons of fathers scoring 2 standard deviatioons above average height for fathers will on average be only (2 r) standard deviations above average height for sons. TRUE (a).