\[ \bar{x} \quad \bar{y} \quad \bar{x}^2 \quad \bar{y}^2 \quad \bar{x}y \quad 577200 \quad 2-16-09 \]

1. \( s_y. = \sqrt{\frac{\sum_{i=1}^{n} (208788 - \bar{y})^2}{n-1}} \)

2. \( r. \) in large \( n \).

3. The fraction of \( s_y^2 \) accounted for by regression on \( x \).

4. The slope of the naive line.

5. The slope of the regression line of \( y \) on \( x \).

6. The slope of the regression line of \( x \) on \( y \) (which would apply if the variables were interchanged).

Sample SD of \( x_1 \ldots x_n = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum x_i^2 - \bar{x}^2}{n-1}} \)

Sample Correlation = \[ \frac{\bar{x}y - \bar{x}\bar{y}}{\sqrt{\bar{x}^2 - \bar{x}^2} \sqrt{\bar{y}^2 - \bar{y}^2}} = \frac{15392 - (34)458}{\sqrt{1194-34^2} \sqrt{208788-458^2}} \]
7. \( r[2x - 4, 6y + 2] \)  
   \( r = 6 \) have same sign

8. \( r[-x + 2, y - 6] = -2 \leq [x, y] \)  
   \( -1, 1 \) opposite signs

9. For an ELLIPTICAL plot having the above averages, the average calories for all subjects having time 36.  
   \( \bar{y} = 456 \)

10. For an ELLIPTICAL plot having the above averages, the best (by least squares) prediction for calories for a student with time 36.  
    \( \hat{y} + (36 - \bar{x}) \) (you work it out)

11. The independent variable.  
    \( \bar{x} = 34 \)  
    \( \bar{y} = 456 \)

12. The dependent variable.  

13-21. \( r[x, y] = 0.9, s_x = 2, s_y = 5, \bar{x} = 22, \bar{y} = 54 \).

13. Determine \( r[y, x] \).

14. For points \((x, y)\) on the regression line determine the numerical value of \( \frac{y - \bar{y}}{x - \bar{x}} \).

15. For \( x = \bar{x} + s_x \) the regression prediction of \( y \) is \( y + (\ ?) s_y \).

16. For \( x = 18 \) the regression prediction of \( y \) is 0.9.

17. Regression predictions (15), (16) are sometimes useful even if the plot is not elliptical. If the plot IS ELLIPTICAL what is the special nature of the plot of vertical strip averages?

\[
\#16. \quad \bar{y} + 2 dy \left( \text{510score} \right) = \bar{y} + 1.9 dy \left( \frac{18 - 22}{2} \right)
\]

\[
\text{nns } \bar{y} + 1.9 dy (-2)
\]

\[
\#17. \quad \text{plot of vert strip avgs} \sim \text{regression line}
\]
\[ r[x, y] = 0.9, \ s_x = 2, \ s_y = 5, \ \bar{x} = 22, \ \bar{y} = 54. \]

18. If the plot is ELLIPTICAL what is the average y-score for all \((x, y)\) pairs with \(x = 18\)? \text{Pt on Reg Line at } x = 18

19. If the plot is ELLIPTICAL what is the standard deviation of y-scores for all \((x, y)\) pairs with \(x = 18\)?

20. If the plot is elliptical, sketch the distribution of \(x\), and the distribution of \(y\).

21. Draw a picture illustrating all of (18), (19), (20).

\[ \#18. \quad \text{G}-\bar{y} = \sum \frac{y}{n} = .9 \left( \frac{5}{2} \right) \text{ Solve for } y \]

\[ \#19. \quad \text{G} \quad \text{SD} \sqrt{1-s^2} \quad \text{G} = \sqrt{1-0.81} (5) \text{ So without } x \text{ you guess} \]

\[ x = 18 \text{ Gues Reg.} \]

\[ \text{REGR Gues} \quad \text{SD} \approx \sqrt{1-0.9^2} (5) \]
22-23. \( r[x, y] = 0.9, s_x = 2, s_y = 5, \bar{x} = 22, \bar{y} = 54. \)

22. Using (14), if I tell you that the population mean of \( x \) is \( \mu_x = 26 \) what is the regression-based estimate for \( \mu_x \)?

\[
\text{Regr Based Est} \quad \text{Insert to Regr.} \quad \text{Est of } \mu_y = \bar{y} + (\bar{x} - \bar{x}) \frac{\bar{y}}{\bar{x}}
\]

23. Give the 95% CI for the estimate (19) if \( n \) is large. (The plot need not be elliptical since the estimator (19) is dependent upon \( \bar{x} \) and \( \bar{y} \) which are approximately jointly normal distributed for large \( n \).) CI is above est \( \pm 1.96 \frac{\text{S.E.} (\hat{B}_0)}{\sqrt{n}} \)
24. For the plot below, sketch the regression line for \( y \) on \( x \) (usual). Also, sketch the regression line for \( x \) on \( y \).
\[ n = \frac{\overline{x} \overline{y} - \overline{x} \overline{y}}{\sqrt{\frac{1}{n} \sum (x_i - \overline{x})^2} \sqrt{\frac{1}{n} \sum (y_i - \overline{y})^2}} \]

\[ n = \frac{5.33 - 1.33}{\sqrt{12 - 3.33^2}} \]

\[ n = \frac{3.33}{\sqrt{12 - 3.33^2}} \]

\[ n = \frac{3.33}{\sqrt{12 - 3.33^2}} \]

25. For the plot just above calculate the slope of regression. Confirm it with what you see in the plot.

\[ \text{slope} = n \frac{\sum xy}{\sum x^2} = \frac{5.33 - 1.33}{3.33} \]

26. From your calculations (25) what is \( s_y \)?

27. Calculate \( s_e \).

28. Verify that \( r^2 = 1 - \frac{s_e^2}{s_y^2} \)

\[ r \text{ works out to } 0.25 \]

\[ r = 0.5 \]

\[ \hat{y} \text{ for } \{0, 0, 4\} = \frac{2}{3} \]

\[ \hat{y} \text{ for } \{1, 1, 0\} = \frac{1}{3} \]

\[ \hat{y} \text{ for } \{0, 0, 4\} = \hat{y} \text{ for } \{1, 1, 0\} = \frac{2}{3} = 1 \]

\[ s_y^2 = 1.1547^2 \]
29. Interestingly, the correlation of $x$ with the fitted values is exactly equal to $|r|$. Verify that it is so in this case.

\[
\begin{align*}
e &= 1 & \hat{y} &= 3 \\
e &= -1 & \hat{y} &= 3
\end{align*}
\]

\[
\begin{align*}
\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \hat{y}) &= \sum_{i=1}^{2} (0, 0, 1)(3, 3, 4) \\
\sum_{i=1}^{n} (x_i - \bar{x})^2 &= \sum_{i=1}^{2} (0, 4, 4) \\
\sum_{i=1}^{n} (y_i - \hat{y})^2 &= \sum_{i=1}^{2} (3, 3, 4)
\end{align*}
\]

\[
\begin{align*}
r_{x, y} &= \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \hat{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \hat{y})^2}} \\
r_{x, y} &= \frac{\sum_{i=1}^{2} (0, 0, 1)(3, 3, 4)}{\sqrt{\sum_{i=1}^{2} (0, 4, 4) \sum_{i=1}^{2} (3, 3, 4)}} \\
r_{x, y} &= \frac{16}{\sqrt{2} \sqrt{12}} \\
r_{x, y} &= \frac{16}{\sqrt{2} \sqrt{12}} \frac{16}{\sqrt{2} \sqrt{12}} \\
r_{x, y} &= \frac{16}{\sqrt{2} \sqrt{12}} \frac{16}{\sqrt{2} \sqrt{12}} \\
\end{align*}
\]

Exactly the same.
30. Draw in the regression of $y$ on $x$. Identify and label $\bar{x}$, $\bar{y}$, $s_x$, $s_y$ (using 68% rule).
- \[ x \quad \bar{y} \quad x^2 \quad \bar{y}^2 \quad xy \quad \text{NEED}\ N \]

Suppose \( N \) LARGE.

1. \( s_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n}} \)

2. \( r = \frac{\sum xy - \bar{x} \bar{y}}{\sqrt{\sum x^2 - \bar{x}^2} \sqrt{\sum y^2 - \bar{y}^2}} \)

3. The fraction of \( s_y^2 \) accounted for by regression on \( x \).

4. The slope of the naive line.

5. The slope of the regression line of \( y \) on \( x \).

6. The slope of the regression line of \( x \) on \( y \) (which would apply if the variables were interchanged).

\[
N = \frac{15392 - 34(456)}{\sqrt{1195 - 34^2} \sqrt{208788 - 456^2}}
\]

\# 3. Ans. \( \hat{N}^2 = 1 - \frac{\text{Resid}^2}{\text{Total}^2} \)

Short ans. \( \text{Resid}^2 = 0.9 \text{ you've explained } 9^2 \sim 81\% \text{ of } \hat{N}^2 \).

\# 6. Regr \( y \) on \( x \) (usual) Regr slope is \( \bar{y}' / \bar{x} \)

Flip to regr \( x \) on \( y \): \( \bar{x}' / \bar{y} \).
7. \( r[2x - 4, 6y + 2] \) 2, 6 HAVE SAME SIGN

8. \( r[-x + 2, y - 6] = -r[x, y] \) -1, 1 OPPOSITE SIGN

9. For an ELLIPTICAL plot having the above averages, the average calories for all subjects having time 36. \( \bar{x} \) so are y

   FOR SUCH \( x = 36 \) \( \bar{y} = 456 \).

10. For an ELLIPTICAL plot having the above averages, the best (by least squares) prediction for calories for a student with time 36. IS PROXIMATELY \( \approx 456 \)

11. The independent variable.

   ALWAYS \( x (= \text{TIME}) \)

12. The dependent variable.

   \( y \) -
13-21. \( r[x, y] = 0.9, s_x = 2, s_y = 5, \bar{x} = 22, \bar{y} = 54. \)

13. Determine \( r[y, x]. \)

14. For points \((x, y)\) on the regression line determine the numerical value of \(\frac{y - \bar{y}}{x - \bar{x}} = n \frac{y}{x} = 0.9 \sqrt{\frac{5}{2}} \). 

15. For \(x = \bar{x} + s_x\) the regression prediction of \(y\) is \(\bar{y} + (\text{?}) s_y.\)

16. For \(x = 18\) the regression prediction of \(y\) is?

17. Regression predictions (15), (16) are sometimes useful even if the plot is not elliptical. If the plot IS ELLIPTICAL what is the special nature of the plot of vertical strip averages?

\[ \#16. \text{Pred } y \text{ for } x = 18 \quad \hat{y} = 0.9 (\frac{5}{2}) \text{ Slope} \]

\[ \text{Pred } y = 54 + (18 - 22), 0.9 \left(\frac{5}{2}\right). \]

\[ \#17. \text{ Plot of Vertical Strip Averages.} \]
\( r[x, y] = 0.9, \ s_x = 2, \ s_y = 5, \ \bar{x} = 22, \ \bar{y} = 54. \)

18. If the plot is ELLIPTICAL what is the average \( y \)-score for all \((x, y)\) pairs with \( x = 18 \)?

19. If the plot is ELLIPTICAL what is the standard deviation of \( y \)-scores for all \((x, y)\) pairs with \( x = 18 \)?

20. If the plot is elliptical, sketch the distribution of \( x \), and the distribution of \( y \).

21. Draw a picture illustrating all of (18), (19), (20).

\[
\#18. \text{ Point on } \bar{x} \text{ at } x=18 \ (15 \ ar{y} + (18-x)^2) \frac{dy}{dx}
\]

\[
\#19. \text{ Eqn. } \sqrt{1-r^2} \ dy = \ dy.
\]

\[
\#20. \text{ If } \sqrt{1-r^2} \ dy < \ dy, \text{ then } \frac{dy}{dx} \text{ is all the same}
\]
22-23. \( r[x, y] = 0.9, s_x = 2, s_y = 5, \bar{x} = 22, \bar{y} = 54. \)

22. Using (14), if I tell you that the population mean of \( x \) is \( \mu_x = 26 \), what is the regression-based estimate for \( \mu_x \)?

23. Give the 95% CI for the estimate (19) if \( n \) is large. (The plot need not be elliptical since the estimator (19) is dependent upon \( \bar{x} \) and \( \bar{y} \) which are approximately jointly normal distributed for large \( n \).)

For above get 95% CI

\[
\bar{y} + (\mu_x - \bar{x})\frac{s_y}{\sqrt{n}} \pm 1.96 \frac{s_y}{\sqrt{n}} \sqrt{1 - r^2}
\]

New estimator

\( \mu \approx \bar{y} \)

\( n \to \infty \)

\( \frac{1}{\sqrt{1 - r^2}} \approx 1 \)

Achieves same
24. For the plot below, sketch the regression line for $y$ on $x$ (usual). Also, sketch the regression line for $x$ on $y$. 

$\text{REGR OF } x \text{ ON } y$ 

$\text{REGR LINE (ALSO LEAST SQ) OF } y \text{ ON } x \text{ (USUAL)}$ 

$\text{DIFFERENT LINES}$
\[ n = \frac{xy - \bar{x}\bar{y}}{\sqrt{(x^2 - \bar{x}^2)(y^2 - \bar{y}^2)}} \]

\[ = \frac{5.33 - 1.33}{\sqrt{5.33 - 1.33^2}} \cdot \frac{3.33}{\sqrt{12 - 3.33^2}} \]

\[ = 0.5 \]

25. For the plot just above calculate the slope of regression. Confirm it with what you see in the plot.

\[ \text{Slope} = \lambda \frac{\hat{y}}{\hat{x}} = \frac{5.33 - 1.33^2}{3.33} \]

26. From your calculations (25) what is \( s_y \)?

\[ \sqrt{12 - 3.33^2} \]

27. Calculate \( s_e \).

\[ e \text{ are the residuals of regression} \]

28. Verify that \( r^2 = 1 - \frac{s_e^2}{s_y^2} \)

\( n \) works out to \( 0.25 \)

\( r = 0.5 \)

\[ e^2 = (\hat{e}^2) \{ -1, 1, 0 \} = \frac{e^2}{(\hat{e}^2)} = 1 \]

\( s_y^2 = 1.154 \)
29. Interestingly, the correlation of \( x \) with the fitted values is exactly equal to \( |r| \). Verify that it is so in this case.

\[
\begin{align*}
e &= 1 & \hat{y} &= 3 \\
e &= -1 & \hat{y} &= 3
\end{align*}
\]

\[
\begin{align*}
r_{x,y} &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \\
r_{x,\hat{y}} &= \frac{\sum (x_i - \bar{x})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (\hat{y}_i - \bar{\hat{y}})^2}} = \frac{16/3 - \frac{4}{3} \frac{10}{3}}{\sqrt{16/3 - (\frac{4}{3})^2} \sqrt{36/3 - (\frac{10}{3})^2}} \\
&= \frac{16/3 - \frac{4}{3} \frac{10}{3}}{\sqrt{16/3 - (\frac{4}{3})^2} \sqrt{36/3 - (\frac{10}{3})^2}} \sqrt{\frac{16/3 - (\frac{4}{3})^2}{36/3 - (\frac{10}{3})^2}} = \frac{16/3 - \frac{4}{3} \frac{10}{3}}{\sqrt{16/3 - (\frac{4}{3})^2} \sqrt{36/3 - (\frac{10}{3})^2}}
\end{align*}
\]

exactly the same
30. Draw in the regression of $y$ on $x$. Identify and label $\bar{x}$, $\bar{y}$, $s_x$, $s_y$ (using 68% rule).