1) In lab, your resistor is labeled "ARCOL HS25 8R F." Through investigation, determine the temperature coefficient of this resistor. How is the temperature coefficient related to a change in resistance? Estimate the variability in resistance in your experiment.

From online store, mouser.com, $\boxed{\text{Temp Coeff: } \pm 100 \frac{ppm}{°C}}$

- The temp coefficient of resistance describes how much the resistance changes as the temperature changes.
- A value of ±150 ppm/°C means the resistors Ω will vary by ±.000100 Ω for every change in °C

Temp ranged from 70°F → 140°F
21°C → 60°C

@ T = 60°C, our resistor has a variability of $\boxed{\pm 0.006\ °C}$

2) Show that for a set of data falling on a straight line of slope C, the magnitude of *r* is 1.

Assume b = 0
y = Cx + b

$$y_i = x_i C \qquad C = \frac{y_i}{x_i} = \frac{\bar{y}}{\bar{x}}$$

(3) $$r^2 = C^2 \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \dots}{(Cx_1 - C\bar{x})^2 + (Cx_2 - C\bar{x})^2 + \dots}$$

(4) $$r^2 = C^2 \frac{x_1^2 - 2x_1\bar{x} + \bar{x}^2 + x_2^2 - 2x_2\bar{x} + \bar{x}^2 + \dots}{C^2x_1^2 - 2C^2x_1\bar{x} + C^2\bar{x}^2 + C^2x_2^2 - 2C^2x_2\bar{x} + \bar{x}^2 + \dots}$$

(5) $$r^2 = \frac{C^2(x_1^2 + x_2^2 + \dots) - 2\bar{x}(x_1 + x_2 + \dots)}{C^2(x_1^2 + x_2^2 + \dots) - 2\bar{x}(x_1 + x_2 + \dots)} = 1$$

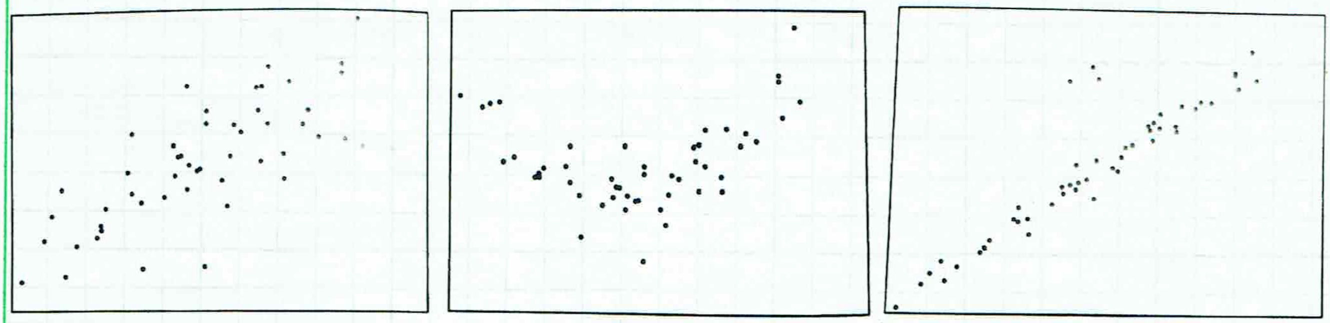
(6) $\boxed{r = 1}$

(1) $$r = \hat{\beta}_1 \frac{S_x}{S_y} = C\frac{S_x}{S_y} = \frac{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} C$$

(2) $$r^2 = C^2 \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + (x_3-\bar{x})^2 + \dots}{(y_1-\bar{y})^2 + (y_2-\bar{y})^2 + (y_3-\bar{y})^2 + \dots}$$

$$\frac{\bar{y}}{\bar{x}} = \frac{(y_1 + y_2 + \dots)/\cancel{n}}{(x_1 + x_2 + \dots)/\cancel{n}} = \frac{\Sigma y_i}{\Sigma x_i} = \frac{\Sigma C x_i}{\Sigma x_i} = C = \frac{\bar{y}}{\bar{x}}$$

7.1.3. For each of the following scatterplots, state whether the correlation coefficient is an appropriate summary, and explain briefly.



(*a*) (*b*) (*c*)

(a) → Yes, there is a linear relationship. Finding the coefficient would be appropriate.

(b) This relationship is correlated, but not linearly. Also, the symmetry would create an adverse coefficient. → NO

(c) While there is a linear correlation, the outliers would make the coefficient innaccurate. → NO

7.1.4. True or false, and explain briefly:

a. If the correlation coefficient is positive, then above-average values of one variable are associated with above-average values of the other.

b. If the correlation coefficient is negative, then below-average values of one variable are associated with below-average values of the other.

c. If y is usually less than x, then the correlation between y and x will be negative.

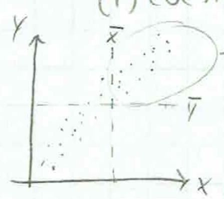
9/9

a) TRUE

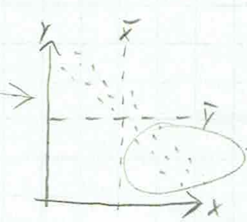
b) FALSE

c) FALSE

(+) coefficient

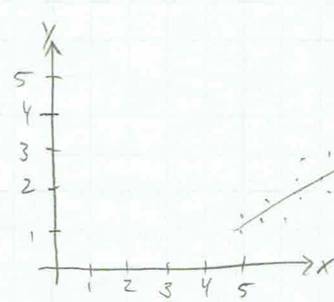


above avg. x are associated w/ above avg y



(-) coefficient

below avg y values are associated w/ above avg x values



(+) coefficient

You can have a (+) coefficient and still have y values usually < x

7.1.9. Tire pressure (in kPa) was measured for the right and left front tires on a sample of 10 automobiles. Assume that the tire pressures follow a bivariate normal distribution.

$$r = \frac{1}{9}\sum_{i=1}^{10}\left[\left(\frac{L_1(i) - mean(L_1)}{stddev(L_1)}\right)\left(\frac{L_2(i) - mean(L_2)}{stddev(L_2)}\right)\right]$$

$L_1 = List\ 1$
$L_2 = List\ 2$

List 1 Right Tire Pressure	List 2 Left Tire Pressure
184	185
206	203
193	200
227	213
193	196
218	221
213	216
194	198
178	180
207	210
AVG: 201.3	202.2

$n=10$

$r = 0.9307$

$W = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) = 1.6636$

$(1.6636) - (1.96)\sqrt{\frac{1}{7}} < \mu_W < (1.6636) + (1.96)\sqrt{\frac{1}{7}}$

$0.92277 < \mu_W < 2.40439$

$\rho = \frac{e^{2\mu_W} - 1}{e^{2\mu_W} + 1} \Rightarrow 0.7272 \Rightarrow 0.9838$

a. Find a 95% confidence interval for ρ, the population correlation between the pressure in the right tire and the pressure in the left tire.

(a) $\boxed{.7272 < \rho < .9838 \quad 95\%\ CI}$

b. Can you conclude that $\rho > 0.9$?

c. Can you conclude that $\rho > 0$? (For parts b and c, take $\alpha = 0.05$ such that you may base your conclusions on your confidence interval alone instead of using hypothesis testing)

(b) $\boxed{\text{No}}$ our range includes $\rho < 0.9$

(c) $\boxed{\text{Yes}}$ our range is all above 0

7.2.3. A least-squares line is fit to a set of points. If the total sum of squares is $\sum(y_i - \bar{y})^2 = 9615$, and the error sum of squares is $\sum(y_i - \hat{y}_i)^2 = 1450$, compute the coefficient of determination r^2.

$$r^2 = \frac{\text{regression } \Sigma \text{ square}}{\text{Total } \Sigma \text{ square}} = \frac{\text{Total} - \text{error}}{\text{total}}$$

$$r^2 = \frac{9615 - 1450}{9615}$$

$$\boxed{r^2 = 0.8492}$$

7.2.**8.** The following table presents shear strengths (in kN/mm) and weld diameters (in mm) for a sample of spot welds.

Diameter	Strength
4.2	51
4.4	54
4.6	69
4.8	81
5.0	75
5.2	79
5.4	89
5.6	101
5.8	98
6.0	102

a. Construct a scatterplot of strength (y) versus diameter (x). Verify that a linear model is appropriate.

b. Compute the least-squares line for predicting strength from diameter.

c. Compute the fitted value and the residual for each point.

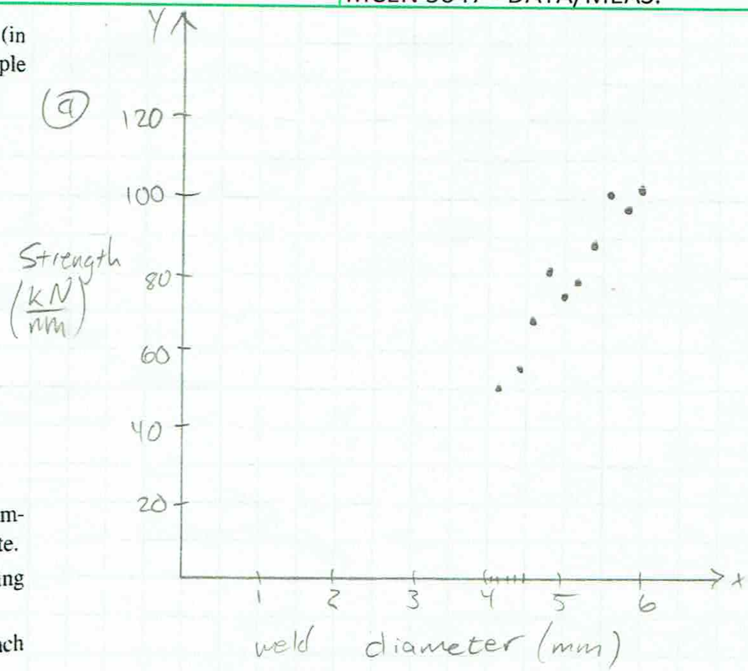
d. If the diameter is increased by 0.3 mm, by how much would you predict the strength to increase or decrease?

e. Predict the strength for a diameter of 5.5 mm.

f. Can the least-squares line be used to predict the strength for a diameter of 8 mm? If so, predict the strength. If not, explain why not.

g. For what diameter would you predict a strength of 95 kN/mm?

(a)



$\boxed{\text{Linear model appropriate}}$

(b) Diameter → list 1 → $\bar{x} = 5.1$, $s_x = .606$

Strength → list 2 → $\bar{y} = 79.9$, $s_y = 18.242$

$$\hat{\beta}_1 = \frac{\sum_{x=1}^{10} \left[(L_1(x) - mean(L_1))(L_2(x) - mean(L_2))\right]}{\sum_{x=1}^{10} (L_1(x) - mean(L_1))^2}$$

$\hat{\beta}_1 = 28.94$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = mean(L_2) - 28.94\, mean(L_1)$

$\hat{\beta}_0 = -67.69$

(b) $\boxed{\hat{y} = -67.69 + 28.94x}$ → Fitted value, plug into L_3

(c)

Fitted Value $\hat{y}$	53.9	59.6	65.4	71.2	77.0	82.8	88.6	94.4	100.2	105.9
residual e	-2.85	-5.64	3.57	9.78	-2.01	-3.79	0.42	6.63	-2.16	-3.95

← $e = y_i - \hat{y}_i = L_2 - L_3 = L_4$

(d) $\hat{y}(4.0) = -67.69 + 28.94(4.0) = 48.07$

$\hat{y}(4.3) = -67.69 + 28.94(4.3) = 56.75$

$\hat{y}(4.3) - \hat{y}(4.0) = \boxed{8.68 \frac{kN}{mm}}$

(e) $\hat{y}(5.5) = -67.69 + 28.94(5.5) = 91.48$

$\boxed{\hat{y}(5.5) = 91.48 \frac{kN}{mm}}$

7.2.8 cont.

(F) Can the least squares line be used to predict the strength for a dia. of 8mm? If so, predict the strength. If not, explain why not.

$\boxed{\text{No}}$ → An 8mm diameter is outside the range of our data. We shouldn't extrapolate outside this range b/c linear relationships often are only true for a certain range.

(g) $\hat{y} = 95 = -67.69 + 28.94(x)$

$\boxed{x = 5.62 \text{ mm}}$

7.2.15. A simple random sample of 100 men aged 25–34 averaged 70 inches in height, and had a standard deviation of 3 inches. Their incomes averaged $34,900 and had a standard deviation of $17,200. Fill in the blank: From the least-squares line, we would predict that the income of a man 70 inches tall would be ________.

i. less than $34,900.
ii. greater than $34,900.
iii. equal to $34,900.
iv. We cannot tell unless we know the correlation.

$n = 100$

$\bar{x} = 70$ $\qquad$ $\bar{y} = 34,900$

$s_x = 3$ $\qquad$ $s_y = 17,200$

$\hat{\beta}_1 = r$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$\hat{y}(x=70) = (\bar{y} - \cancel{\hat{\beta}_1 \bar{x}}) + \cancel{\hat{\beta}_1 \bar{x}}$

$\hat{y}(x=70) = \bar{y} = 34,900$

$\boxed{iii}$

7.4.3. To determine the effect of temperature on the yield of a certain chemical process, the process is run 24 times at various temperatures. The temperature (in °C) and the yield (expressed as a percentage of a theoretical maximum) for each run are given in the following table. The results are presented in the order in which they were run, from earliest to latest.

Order	Temp	Yield	Order	Temp	Yield	Order	Temp	Yield
1	30	49.2	9	25	59.3	17	34	65.9
2	32	55.3	10	38	64.5	18	43	75.2
3	35	53.4	11	39	68.2	19	34	69.5
4	39	59.9	12	30	53.0	20	41	80.8
5	31	51.4	13	30	58.3	21	36	78.6
6	27	52.1	14	39	64.3	22	37	77.2
7	33	60.2	15	40	71.6	23	42	80.3
8	34	60.5	16	44	73.0	24	28	69.5

$n = 24$

a. Compute the least-squares line for predicting yield (y) from temperature (x).
b. Plot the residuals versus the fitted values. Does the linear model seem appropriate? Explain.
c. Plot the residuals versus the order in which the observations were made. Is there a trend in the residuals over time? Does the linear model seem appropriate? Explain.

Temp $\rightarrow L_1 \rightarrow x$ $\quad \bar{x} = 35.04$
Yield $\rightarrow L_2 \rightarrow y$ $\quad \bar{y} = 64.22$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (L_1(x) - \text{mean}(L_1))(L_2(x) - \text{mean}(L_2))}{\sum_{i=1}^{n} (L_1(x) - \text{mean}(L_1))^2}$$

$\hat{\beta}_0 = \text{mean}(L_2) - \hat{\beta}_1 \cdot \text{mean}(L_1) = 21.373$ $\quad \hat{\beta}_1 = 1.223$

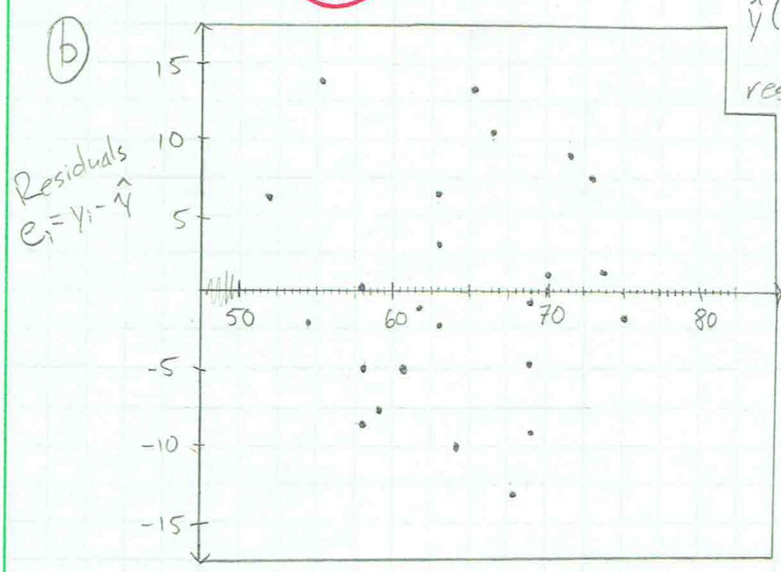
(a) $\boxed{\hat{y} = 1.223x + 21.373}$ -1

(b)

$\hat{y}(L_1) \rightarrow$ list 3

residuals $= L_2 - L_3 \rightarrow$ list 4

Plot($x = L_3$, $y = L_4$)

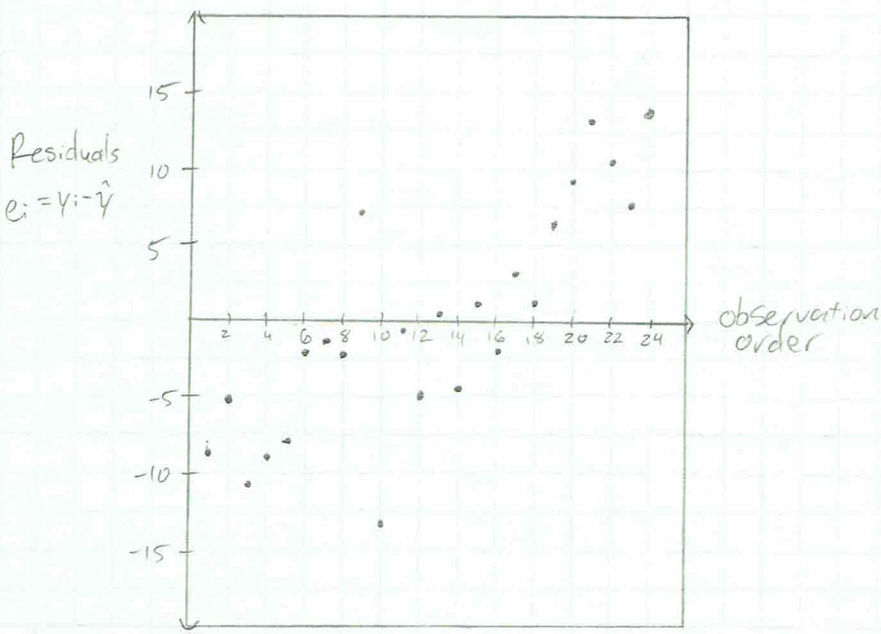


8/9

(c)

(c) 7.4.3 cont. ↗ L4

Plot the residuals vs. the order in which the observations were made. Is there a trend in the residuals over time? Does the linear model seem appropriate? Explain.



- There is a positive trend in the Residuals vs. time. This makes the linear mode innappropriate b/c there are more confounding variables not being accounted for.

7.4.5. Good forecasting and control of preconstruction activities leads to more efficient use of time and resources in highway construction projects. Data on construction costs (in \$1000s) and person-hours of labor required on several projects are presented in the following table and are taken from the article "Forecasting Engineering Manpower Requirements for Highway Preconstruction Activities" (K. Persad, J. O'Connor, and K. Varghese, *Journal of Management Engineering*, 1995:41–47). Each value represents an average of several projects, and two outliers have been deleted.

Person-Hours (x)	Cost (y)	Person-Hours (x)	Cost (y)
939	251	1069	355
5796	4690	6945	5253
289	124	4159	1177
283	294	1266	802
138	138	1481	945
2698	1385	4716	2327
663	345		

$n = 13$

a. Compute the least-squares line for predicting y from x.
b. Plot the residuals versus the fitted values. Does the model seem appropriate?
c. Compute the least-squares line for predicting $\ln y$ from $\ln x$.
d. Plot the residuals versus the fitted values. Does the model seem appropriate?

(x) Person Hours → List 1
(y) Cost → List 2

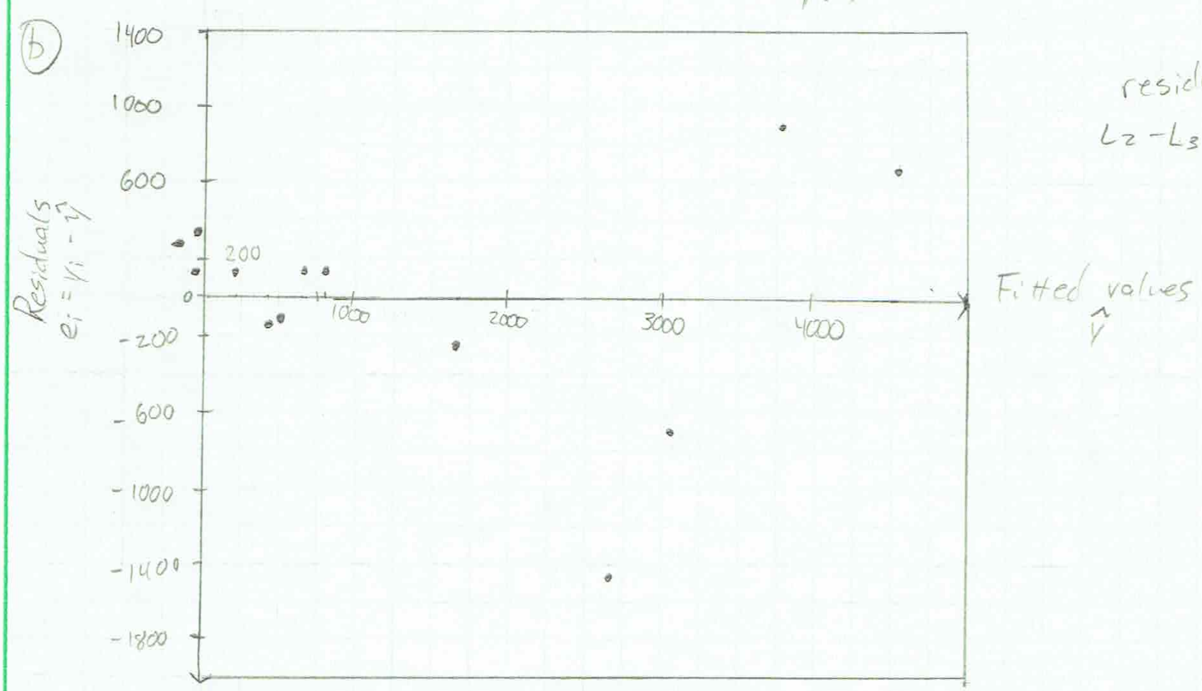
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{13} (L_1(x) - \text{mean}(L_1))(L_2(x) - \text{mean}(L_2))}{\sum_{i=1}^{13} (L_1(x) - \text{mean}(L_1))^2}$$

$\hat{\beta}_1 = 0.6946 \rightarrow \hat{\beta}_0 = \text{mean}(L_2) - \hat{\beta}_1 \,\text{mean}(L_1) = -235.32$

(a) $\boxed{\hat{y} = 0.6946x - 235.32}$

$\hat{y}(L_1) \rightarrow$ list 3

(b)



residuals: $L_2 - L_3 \rightarrow$ list 4

• The model isn't appropriate b/c the vertical spread increases as the fitted value increases → heteroscedastic.

(c)

7.4.5 cont.

(c) Compute the least-squares line for predicting lny from lnx

$\ln(\hat{y}) = \hat{\beta}_1 \ln(x) + \hat{\beta}_0$ $\qquad \ln(x) = \ln(\text{list }1) \longrightarrow \text{list } 3$

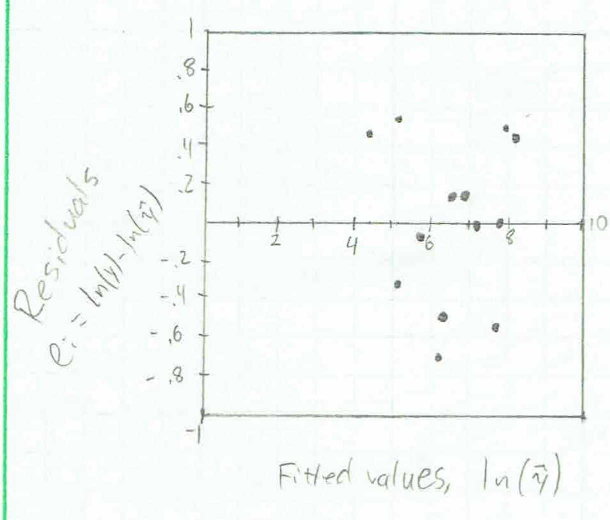
$\ln(\hat{y}) =$ $\qquad \ln(y) = \ln(\text{list }2) \longrightarrow \text{list } 4$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{13} (L_3(x) - \text{mean}(L_3))(L_4(x) - \text{mean}(L_4))}{\sum_{i=1}^{13} (L_3(x) - \text{mean}(L_3))^2}$$

$\hat{\beta}_1 = 0.9255 \rightarrow \beta_0 = \text{mean}(L_4) - \hat{\beta}_1 \,\text{mean}(L_3) = -0.0745$

$\boxed{\ln(\hat{y}) = 0.9255 \cdot \ln(x) - 0.0745}$

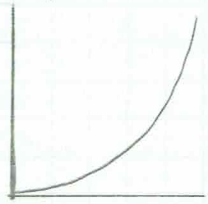
(d) $\ln(\hat{y}(\text{list }3)) \longrightarrow \text{list } 5$ $\qquad$ Residuals: $(\text{list }4) - (\text{list }5) \rightarrow \text{list } 6$



- The model seems appropriate b/c there is no trend or pattern to the residuals vs Fitted values

7.4. **15.** The article "Some Parameters of the Population Biology of Spotted Flounder (*Ciutharus linguatula* Linnaeus, 1758) in Edremit Bay (North Aegean Sea)" (D. Türker, B. Bayhan, et al., *Turkish Journal of Veterinary and Animal Science*, 2005:1013–1018) models the relationship between weight W and length L of spotted flounder as $W = aL^b$ where a and b are constants to be estimated from data. Transform this equation to produce a linear model.

$W = aL^b \xrightarrow{\text{on calc w/ experimental } a,b}$



to make a linear relationship, we need to remove the b from the exponent of L.

$$\ln(W) = \ln(aL^b) = \ln(a) + b\ln(a)$$

$$\boxed{\ln(w) = b\ln(a) + \ln(a)} + \varepsilon$$

?

$$y = m x + b$$