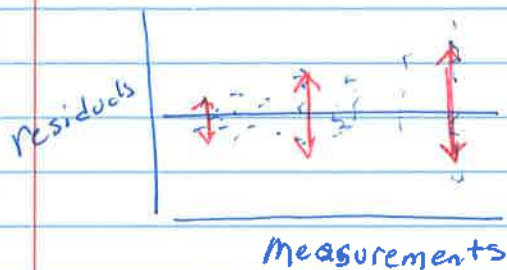


12.11+W

#'s 1, 4, 6, 8, 9, 11, 13, 17, 19

①



NOTICE THE RESIDUAL INCREASE, THIS INDICATES THE VARIANCES OF THE RESIDUALS INCREASES AS LAB MEASUREMENTS INCREASE. THE CONDITION "FOR EQUAL VARIANCES" IS NOT MET

④

CONDITIONS TO CHECK FOR REGRESSION INFERENCE

- UNKNOWN POPULATION STD DEV (σ) - t inference
- RANDOM - THIS WAS A RANDOMIZED EXPERIMENT.
- INDEPENDENT - # OF DRINKS RANDOMLY ASSIGNED
- LINEAR - TO CHECK THIS CONDITION YOU MUST LOOK AT A RESIDUAL PLOT. REVIEWING THE RESIDUAL PLOT, THERE IS NO PATTERN TO THE RESIDUALS CONFIRMING A LINEAR MODEL HAS BEEN CONFIRMED.
- NORMAL - TO CHECK THIS CONDITION YOU CAN LOOK AT A HISTOGRAM OR BOX PLOT OF THE RESIDUALS. REVIEWING THE HISTOGRAM, THE SHAPE OF THE RESIDUALS IS APPROXIMATELY SYMMETRIC WITH NO SKEWNESS OR OUTLIERS
- EQUAL VARIANCES - TO CHECK THIS CONDITION YOU CAN LOOK AT A RESIDUAL PLOT OR SCATTER PLOT. REVIEWING THE RESIDUAL PLOT, IT SHOWS ROUGHLY EQUAL SCATTER FOR ALL X VALUES

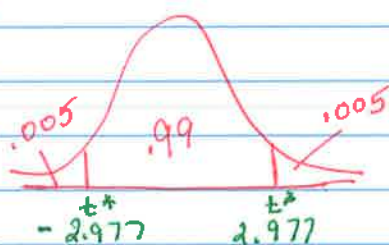
⑥ $\alpha = y\text{intercept}$... THE Y INTERCEPT IS THE STARTING POINT. IN THIS CASE IT WOULD MEASURE THE BAC LEVEL IF NO BEERS WERE DRANK. WE WOULD EXPECT IT TO BE ZERO AND IT IS VERY SMALL. The point estimate is $\alpha^* = .01$ for α

$\beta = \text{slope}$... IN CONTEXT ... It tells us how much the BAC increases, on average, with each additional beer drank. THE ESTIMATE FOR THIS PARAMETER. The point estimate $b = .018$ for β .

σ measures the standard deviation of BAC values about the population regression line. The point estimate is $s = .0204$ for σ .

⑧ (A) $SE_b = .0024$ IF WE REPEATED THE EXPERIMENT MANY TIMES, THE SLOPE OF THE SAMPLE REGRESSION LINE WOULD TYPICALLY VARY BY ABOUT .0024 from the true slope of the population regression line for predicting BAC level from the number of beers consumed.

⑧ (B)



$$df = n - 2 = 16 - 2 = 14$$

$$t^* = \text{invT}(.005, 14) = -2.977$$

$$b = .018$$

$$SE_b = .0024$$

INTERVAL FOR LSRL ← see Green Sheet for CI formula

$$99\% \text{ CI} = b \pm t^* \cdot SE_b$$

$$.018 \pm 2.977 (.0024)$$

$$.018 \pm .0071$$

SE
ME

$(.0109, .0251)$

This is a review of interpreting
in context - CONFIDENCE INTERVAL VS CONFIDENCE
LEVEL

- ⑧ [C] We are 99% confident that the interval from 0.011 to 0.025 captures the true slope of the population regression line for predicting BAC level from the number of beers consumed

Confidence Interval

- [D] If we were to repeat the experiment many times and compute confidence intervals for the regression slope in each case, about 99% of the resulting intervals would contain the slope of the population regression line.

Confidence Level

⑨ TEST: 99% Confidence Interval for β

DEFINE Parameter: β = true slope of the population regression line relating number of stumps to number of clusters of beetle larvae

CONDITIONS: stated conditions were met.

MECHANIS: 99% CI: $b \pm t^* S_b$

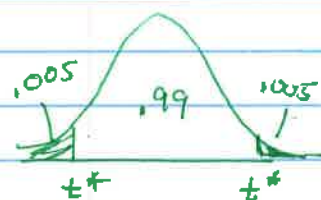
$$11.894 \pm 2.831(1.136)$$
$$11.894 \pm 3.216$$

$$(8.678, 15.11)$$

$$b = 11.894$$

$$S_b = 1.136$$

$$df = 23 - 2 = 21$$



CONCLUDE:

WE ARE 99% confident that the interval from 8.678 to 15.11 captures the true slope of the population regression line for predicting clusters of beetle larvae from the number of stumps

$$t^* = 2.831$$

$$\text{invT}(.005, 21)$$

11 (A)
$$\begin{aligned}\widehat{\text{LARVAE}} &= -1.286 + 11.894 (\text{STUMPS}) \\ &= -1.286 + 11.894 (5) \\ &= 58.184 \text{ larvae clusters}\end{aligned}$$

(B) We would expect our predictions to be off, on average, by about $s = 6.419$ clusters of beetle larvae

13 (A) SCATTER PLOTS - DESCRIBE . STRENGTH, FORM, DIRECTION
DIRECTION - NEGATIVE
FORM - LINEAR
STRENGTH - MODERATE TO WEAK

IN

CONTEXT: THE SCATTER PLOT SUGGESTS A SOMEWHAT WEAK/
MODERATE, NEGATIVE, LINEAR RELATIONSHIP
BETWEEN THE NUMBER OF WEEDS PER METER
AND THE CORN YIELD OF THE PLOTS

OPTION 1

(B) USE MEANINGFUL

VARIABLE NAMES $\rightarrow$
$$\widehat{\text{CORN}} = 166.483 - 1.0987 (\text{WEEDS})$$

OPTION 2

MUST DEFINE

VARIABLES $\rightarrow$

$$\hat{y} = 166.483 - 1.0987x$$

$\hat{y}$ = predicted corn yield
 x = number of weeds per meter

(C) yintercept: (starting point) The yintercept says that if there are no weeds ($x=0$), we would predict a corn yield of about 166.5 bushels

slope: The slope says that for each additional weed per meter, we can expect the average corn yield to decrease by about 1.10 bushels

13D TEST: One-Sample t-test for β $\alpha = .05$

$$H_0: \beta = 0$$

← meaning there is no linear relationship between the 2 variables

$$H_A: \beta < 0$$

where: β is the true slope of the population regression line relating weeds per meter to corn yield.

must define β

CONDITIONS UNKNOWN $\sigma \rightarrow t$ inference

Random - This was a randomized experiment.

Independent - Due to the random assignment, THE OBSERVATIONS can be viewed as independent

LINEAR - the scatter plot (provided) shows a linear relationship but you must check the residual plot to check the linear model was appropriate

RESIDUAL PLOT

CALC COMMANDS

① ENTER DATA IN L1 + L2

② TO POPULATE RESIDUALS

STAT TESTS F:LINREGTEST

③ 2ND STAT PLOT ON

TYPE $\rightarrow$ Scatter

~~X~~ LIST

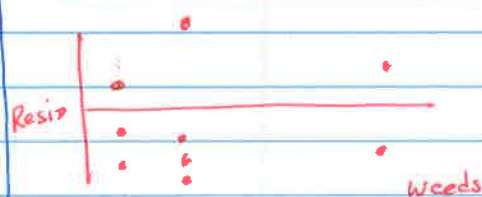
RESID

2ND LIST
8: RESID

~~X~~ LIST L2

④ ZOOM 9:STAT PLOT

SKETCH THE RESIDUAL PLOT



The residual plot shows no pattern therefore the model appears linear

EQUAL VARIANCES - Both the scatter plot (given) and the residual plot above show roughly equal SCATTER FOR ALL X VALUES

You do not need to do both graphs either would be fine

(CONT) $\rightarrow$

13D Continued

Conditions (Continued):

NORMAL: SKETCH EITHER HISTOGRAM OR BOX PLOT
FOR RESIDUALS

HISTOGRAM OF
Residuals

(2ND) (STAT PLOT)

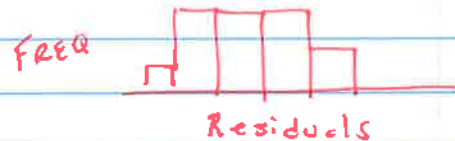
TYPE: $\overline{H_1}$

XLIST: RESID

(Zoom) 9

(2ND) (LIST)
8: RESID

① BASED ON THE HISTOGRAM,
THE RESIDUALS ARE APPROX.
SYMMETRIC WITH NO OUTLIERS



Box Plot of Residuals

(2ND) (STAT PLOT)

TYPE: $\overline{H_1}$

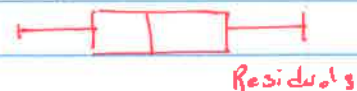
XLIST: RESID

FREQ: 1

Zoom 9

(2ND) (LIST)
8: RESID

OR 2 BASED ON THE BOX PLOT
THE RESIDUALS APPEAR SYMMETRIC
WITH NO OUTLIERS



MECHANICS

$$t = -1.923$$

$$df = 14$$

$p\text{-value} = .037$ (NOTICE: Computer output $p\text{-value}$ is .075 because this is a 2 tail test. For a 1 tail test divide $.075/2 = .0375$)

STAT
TESTS
F: LIN REGTEST
L1
L2
1
40

Conclusion: Since the $p\text{-value}$ (.037) is less than .05, we reject the null hypothesis. And conclude that there is convincing evidence of a negative linear relationship between weeds per meter and corn yield.

17) a) 99% CI $= b \pm t^* SE_b$
 $.79021 \pm 2.977(.07104)$
 $.79021 \pm .21147$

$(.57874, 1.00168)$

From Computer Output

$b = .79021$

$SE_b = .07104$

$df = 16 - 2 = 14$

$t^* = 2.977$

$\text{invT}(.005, 14)$



b) IF ONE METHOD OF MEASURING WEAR GIVES AN INCREASE OF 1 UNIT, WE HOPE THAT THE OTHER WAY OF MEASURING WEAR WOULD ALSO GIVE AN INCREASE OF 1 UNIT, THIS TRANSLATES TO A SLOPE OF 1

c) Since the confidence interval $(.58, 1.00)$ contains 1, it suggests that the slope could plausibly be 1. So we would not reject H_0 and have convincing evidence the slope is 1.

19) (A) TEST: TTEST FOR β $\alpha = .05$

$$H_0: \beta = 0$$

$$H_a: \beta < 0$$

β = the true slope of the population regression line relating wine consumption to heart disease death rate

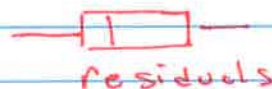
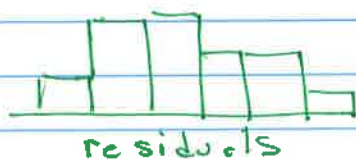
CONDITIONS

Random - the data came from a random sample

Independent - there are more than 10(19) = 190 countries

Linear - the scatter plot shows approx. linear association

NORMAL - looking at a boxplot or histogram shows the residuals are approx. Normal



(STAT) (PLOT) HISTOGRAM

Residual

(2ND) (LIST)
8:RESID

(STAT) (PLOT) Box Plot

Resid

EQUAL VARIANCE: LOOKING AT THE RESIDUAL PLOT SHOWS ROUGHLY EQUAL FOR SCATTER OF X VALUES



(STAT) (PLOT) Scatter plot

XLIST: L1

YLIST: RESID

MECHANICS $DF = 19 - 2 = 17$

$$t = -6.46$$

$$p\text{-value} = 2.6E-6 \approx 0$$

(STAT) (TESTS)

F1: LINREGTTEST

XLIST: L1

YLIST: L2

↵

19A (continued)

Since the p-value is very small we reject H_0 and conclude that there is convincing evidence of a negative linear relationship between wine consumption and death rate due to heart disease.

19B TEST: 95% CI for β

CONDITIONS: CHECK IN TOH (19A)

INTERVAL $(-30.47, -15.46)$

CONCLUDE: WE ARE 95% confident that the interval -30.47 to -15.46 captures the true slope of the population regression line for predicting heart disease death rate from wine consumption

(TEST)

G: LINREGTINT

XLIST: L1

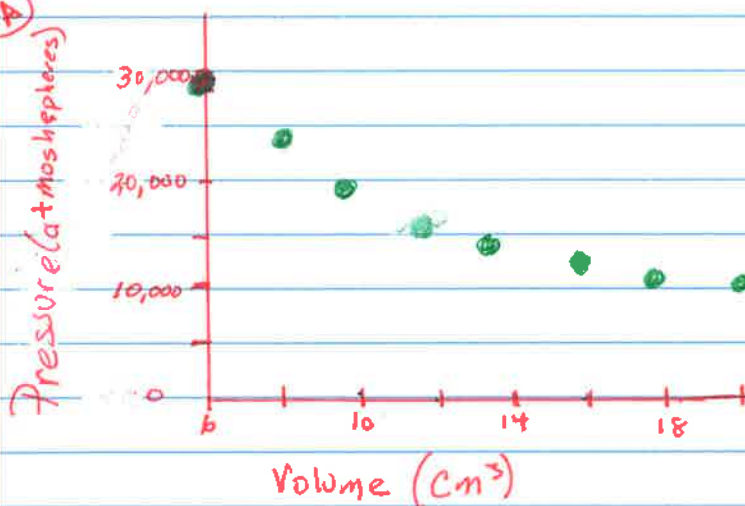
YLIST: L2

FREQ: 1

LEVEL: .95

12.2 HW #15 34 + 35

34 A



Set Window

Xmin: 6 Ymin: 0
Xmax: 22 Ymax: 30,000
Xscl: 2 Yscl: 5,000

IF YOU SEE A GRAPH LIKE THIS THE TRANSFORMATION OF X AND 1/Y OR 1/X AND Y

B
$$\frac{PV}{V} = \frac{k}{V}$$

$$P = \frac{1}{V}(k)$$

EXPLANATORY VARIABLE IS RECIPROCAL OF VOLUME ($\frac{1}{V}$)
RESPONSE VARIABLE IS THE PRESSURE

C
$$\frac{PV}{P} = \frac{k}{P}$$

$$V = \frac{1}{P}(k)$$

EXPLANATORY VARIABLE IS RECIPROCAL OF PRESSURE ($\frac{1}{P}$)
RESPONSE VARIABLE IS THE VOLUME

TRANSFORMATION 1

35 (A) USING $x+y$: $\hat{y} = -.08594 + .21\sqrt{x}$

where y = the period
 x = the length] must define variables

remember $\sqrt{x} = x^{1/2}$

recommend using names !!!

$$\widehat{\text{period}} = -.08594 + .21 (\text{length})^{1/2}$$

TRANSFORMATION 2

$$\widehat{\text{period}^2} = -.155 + .0428 (\text{length})$$

(B) Predict period when length is 80 cm

TRANSFORMATION 1

$$\widehat{\text{period}} = -.08594 + .21\sqrt{80}$$

$$\widehat{\text{period}} = 1.79 \text{ seconds}$$

TRANSFORMATION 2

$$\widehat{\text{period}^2} = -.155 + .0428(80)$$

$$\sqrt{\widehat{\text{period}^2}} = \sqrt{3.269}$$

$$\widehat{\text{period}} = 1.81 \text{ seconds}$$

(C) TRANSFORMATION 1 $S = .046 \rightarrow$ The typical distance that a predicted value of the period will be from the actual value is about 0.046 seconds.

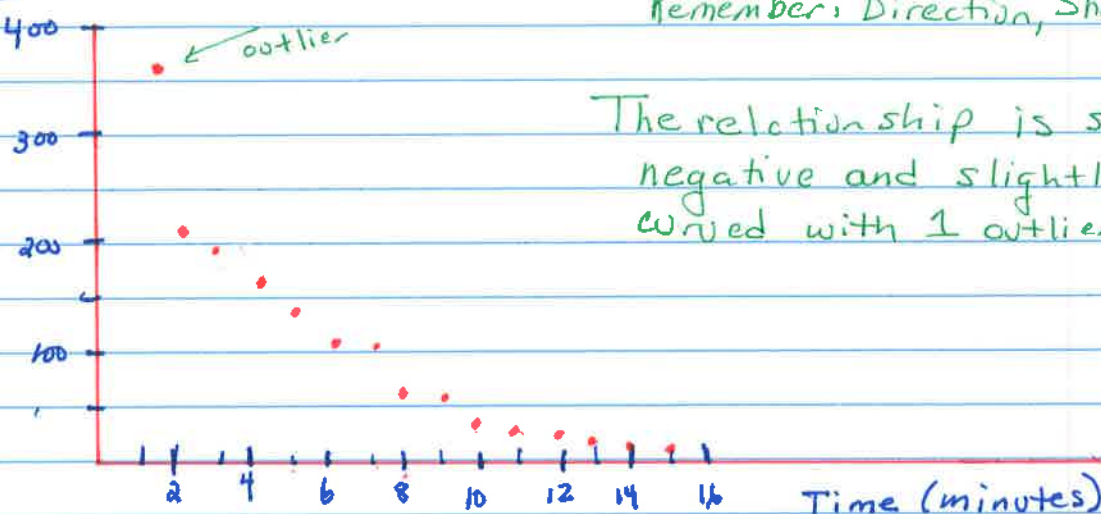
TRANSFORMATION 2 $S^2 = .105 \rightarrow$ The typical distance that the SQUARE OF A PREDICTED VALUE OF THE PERIOD WILL BE FROM SQUARE OF THE ACTUAL VALUE IS ABOUT 0.105 seconds SQUARED

12.2 HW #1's 37, 39, 40, 41, 44

37

Count Surviving Bacteria
(in 000's)

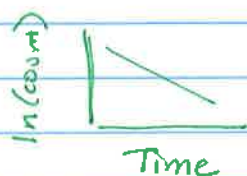
A



Remember: Direction, Shape, Strength

The relationship is strong negative and slightly curved with 1 outlier.

B



Since the graph of the explanatory variable against the natural log of the response variable is fairly linear, an exponential model would be reasonable.

C

The LSAL is $\ln \text{bacteria} = 5.973 - 0.218(\text{time})$

D

predict bacteria when time is 17 minutes

$$\ln \text{bacteria} = 5.973 - 0.218(17)$$

$$e^{\ln \text{bacteria}} = e^{2.267}$$

remember to undo
ln raise both to
a power of e

$$\text{bacteria} = 9.65$$

$$= 965 \text{ bacteria}$$

$$(2ND) (e^x) 2.267$$

Since the residual plot shows a random scatter plot around the value 0, we'd expect this predicted value to be about right

(39)

$$\log y = 1.01 + .72 \log x$$

$$\log \text{brain} = 1.01 + .72 \log (\text{weight})$$

$$\text{weight} = 127 \text{ kg}$$

$$= 1.01 + .72 \cdot \log (127)$$

$$\log \text{brain} = 2.525$$

10

10

To undo log's
Use 10 as the base

$$\text{brain} = 334.97$$

$$10 \wedge 2.525$$

The predicted brain size for Sasquatch
is about 335 grams.

(40)

$$\ln y = -2.00 + 2.42 \ln x$$

$$\ln \text{biomass} = -2.00 + 2.42 \cdot \ln (\text{diameter})$$

$$\text{diameter} = 30 \text{ cm}$$

$$= -2.00 + 2.42 \cdot \ln (30)$$

$$\ln \text{biomass} = 6.231$$

e

e

↑

(2ND)

(e^x)

6.231

$$\text{biomass} = 508.26$$

The predicted total above ground biomass of the
tree is about 508 Kg

- 41 (a) Look at the 2 graphs, the second shows a more linear scatter plot. Therefore the power model would work better with the $\log(\text{abundance})$ versus $\log(\text{body mass})$

(b) $\log(\text{abundance}) = 1.9503 - 1.0481 \cdot \log(\text{body mass})$

(c) $\text{body mass} = 92.5 \text{ Kg}$

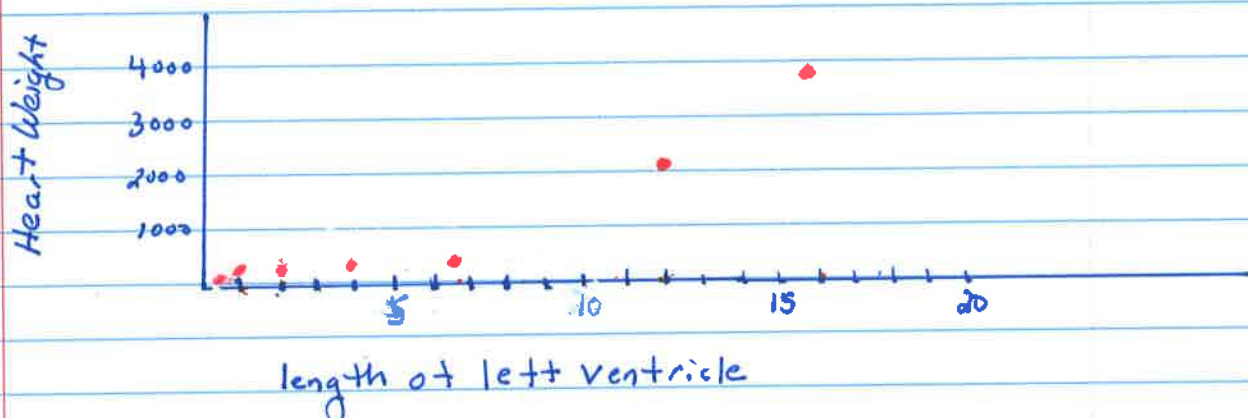
$\log(\text{abundance}) = 1.9503 - 1.0481 \cdot \log(92.5)$

$\log(\text{abundance}) = -0.1104$

$\text{abundance} = .7755 \text{ per } 10,000 \text{ Kg of prey}$

- (d) THE TREND IN THE RESIDUAL PLOT SUGGESTS THAT THE RESIDUAL WOULD BE POSITIVE, WHICH MEANS THAT OUR PREDICTION WOULD BE TOO LOW

44A



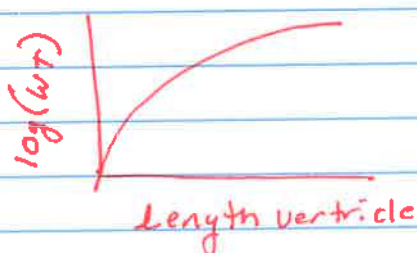
THERE IS A STRONG, POSITIVE, CURVED RELATIONSHIP.

44B

L1	L2	L3	L4
Ventricle	WEIGHT	ln (Ventricle) or * log (Ventricle)	ln (weight) * log (weight)

* book solution uses ln
* here is solution with log

1ST TRY



Scatter plot

XLIST L1

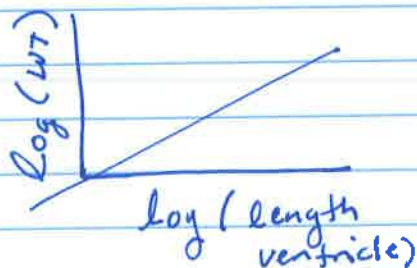
YLIST L4

ZOOM 9

NOTICE

Graph is still curved when taking log of weight.

2ND TRY



Scatter plot

XLIST L3

YLIST L4

ZOOM 9

NOTICE:

Graph is linear when taking log of BOTH weight + length of Ventricle

44C

STAT CALC

8: LIN REG (a+bx)

XLIST [L3]

YLIST [L4]



$$a = -0.1364$$

$$b = 3.139$$

$$r^2 = .9933$$

$$r = .9967$$

LSRL:

$$\widehat{\log(\text{WT})} = -0.136 + 3.139 \cdot \log(\text{ventricle})$$

44D

$$\text{left + ventricle} = 6.8 \text{ cm}$$

$$\widehat{\log(\text{WT})} = -0.136 + 3.139 \cdot \log(6.8)$$

$$\widehat{\log(\text{WT})} = 2.4772$$

$$\widehat{\text{WEIGHT}} = 300.086 \text{ grams}$$