Regression Test Review Sheet

Least-Squares Regression Model:

$\hat{y} = ax + b$ (calculator) $\hat{y} = \beta_0 + \beta_1 x$

where a or β0 is the slope (a change in one unit of x produces this average amount of change in y: a statistical (vs algebraic) model

b or β1 is the intercept (may or may not have any meaning in the context of the problem)

Regression:

After entering x-data into L1 and y-data into L2, use STAT → CALC →LinReg for the calculator to calculate the least squares linear regression line.

Y1 (VARS → Y-VARS→FUNC) stores the regression line for graphing

DiagnosticOn must have been run from the CATALOG to get r and r^2 values.

***Linear Correlation Coefficient*: *r*, tells us the how likely is a *line* a good fit to the data. *r* values go between -1 (*perfect negative sloped line fit*) to 1 (*perfect positive sloped line fit*). An *r* value near zero means a line is not a good fit (not that there is no correlation).**

***Coefficient of Determination* (r^2) – measures the percentage of total variation in the response variable that is explained by the least-squares regression line.**

Conditions: LINER

- **Linear: Examine the scatterplot to check that the overall pattern is roughly linear. Look for curved patterns in the residual plot. Check to see that the residuals center on the “residual = 0” line at each *x*-value in the residual plot.**

- **Independent: Look at how the data were produced. Random sampling and random assignment help ensure the independence of individual observations. If sampling is done without replacement, remember to check that the population is at least 10 times as large as the sample (10% condition).**

- **Normal: Make a stemplot, histogram, or Normal probability plot of the *residuals* and check for clear skewness or other major departures from Normality.**

- **Equal variance: Look at the scatter of the residuals above and below the “residual = 0” line in the residual plot. The amount of scatter should be roughly the same from the smallest to the largest x-value.**

- **Random: See if the data were produced by random sampling or a randomized experiment.**

Regression Test:

After entering the x-data into L1 and y-data into L2, use STAT → TESTS → LinRegTest for the calculator to run the least squares linear regression test on the slope of the regression line.

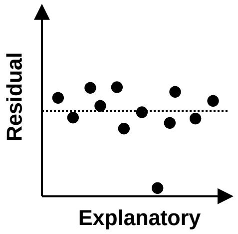
Residuals Residual is the *Error* from the model and is = Observed – Predicted

Least squares regression line minimizes the squares of the residuals

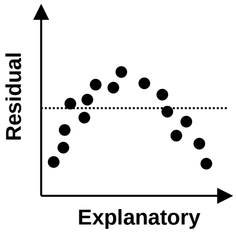
Residual Plots: Linear model is a good fit *if no patterns exists* (no curves) and *variance is constant* (no horn effects in residual plots)

RESID is a list variable formed by your calculator when doing LinReg on calculator
2nd STAT access the list of list variables

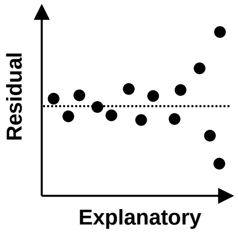
A.



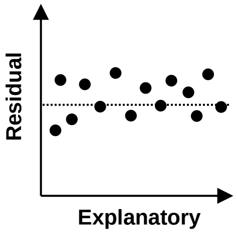
B.



C.



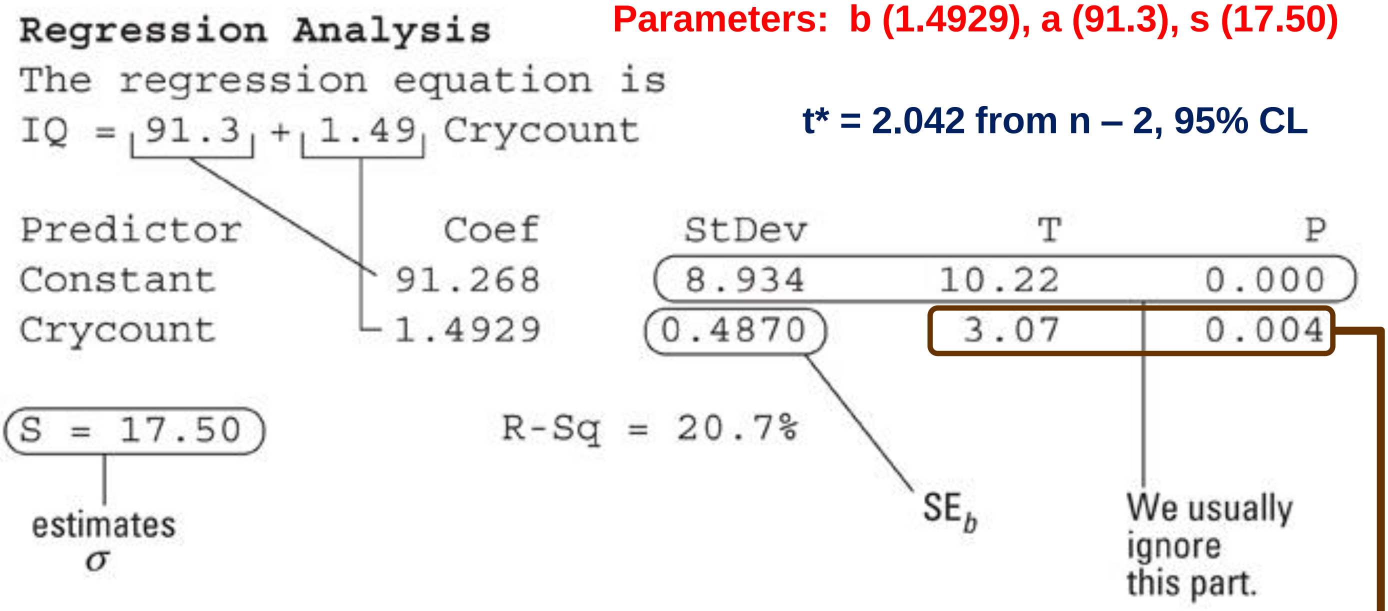
D.



Maybe	No	Line a good fit??	No	Yes
Outlier	Pattern in the residuals	Issues	Variance is non constant (horn effect)	Small variance around 0

Parameters: b (1.4929), a (91.3), s (17.50)

t* = 2.042 from n – 2, 95% CL



```
Regression Analysis
The regression equation is
IQ = 91.3 + 1.49 Crycount

Predictor        Coef       StDev          T          P
Constant       91.268       8.934      10.22      0.000
Crycount       1.4929      0.4870       3.07      0.004

S = 17.50             R-Sq = 20.7%
```

$$CI = PE \pm MOE = 1.4929 \pm (2.042)(0.4870)$$
$$= 1.4929 \pm 0.9944$$
$$[0.4985, 2.4873]$$

Since 0 is not in the interval, then we would conclude that β ≠ 0

Since the p-value for the slope of the regression line is less than our normal $\alpha = 0.05$ value, we would reject H0: $\beta = 0$ and say we have evidence that the slope is not zero.