Power Regression

Model math functions to fit our data

- Linear Regression

$$\hat{y} = a + bx$$

- Quadratic Regression

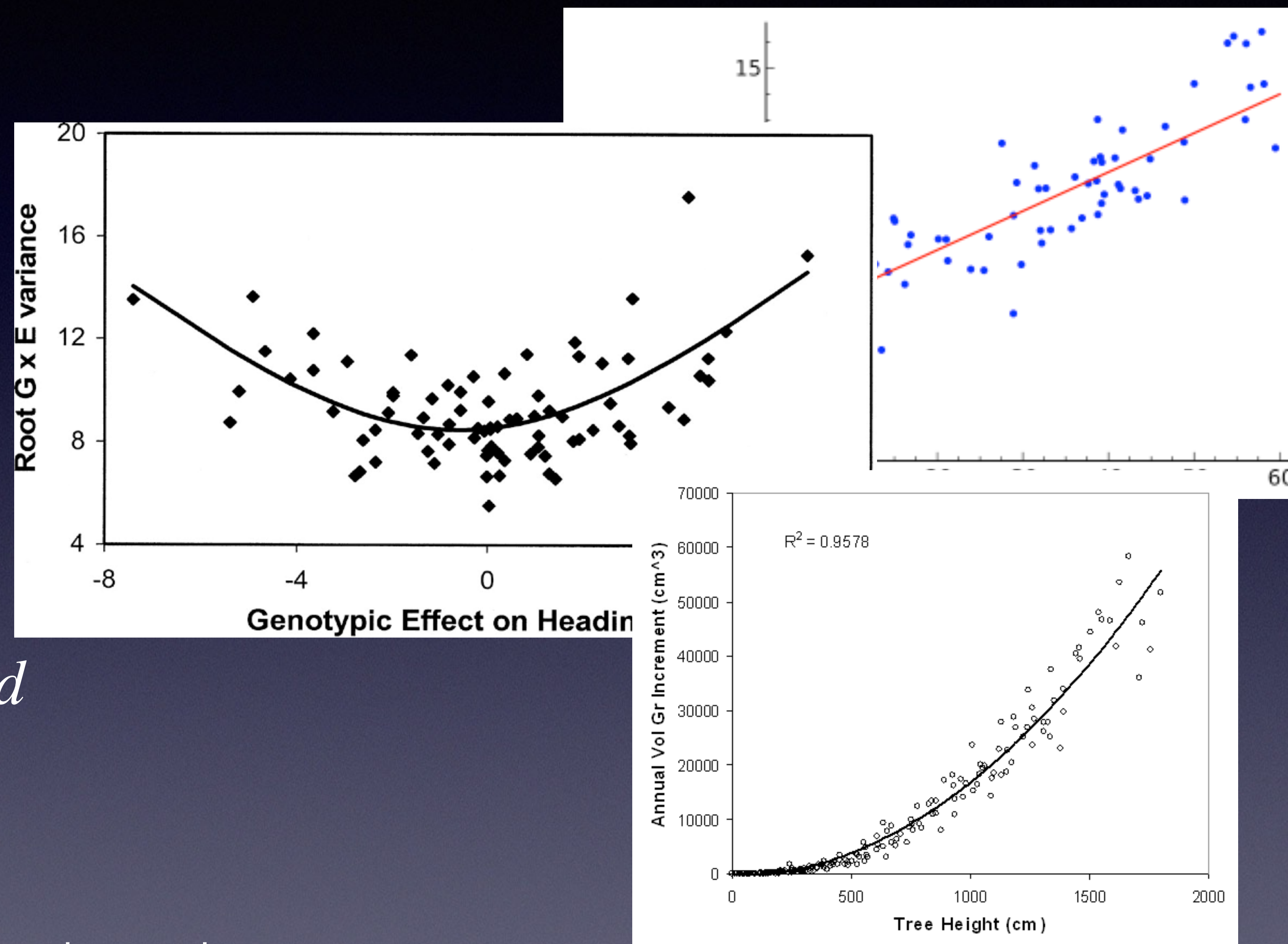
$$\hat{y} = ax^2 + bx + c$$

- Cubic Regression

$$\hat{y} = ax^3 + bx^2 + cx + d$$

- **Power** Regression

$$\hat{y} = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$$



You will choose the best fitting model and use that model to predict.

Power Regression

Model math functions to fit our data

- Linear Regression

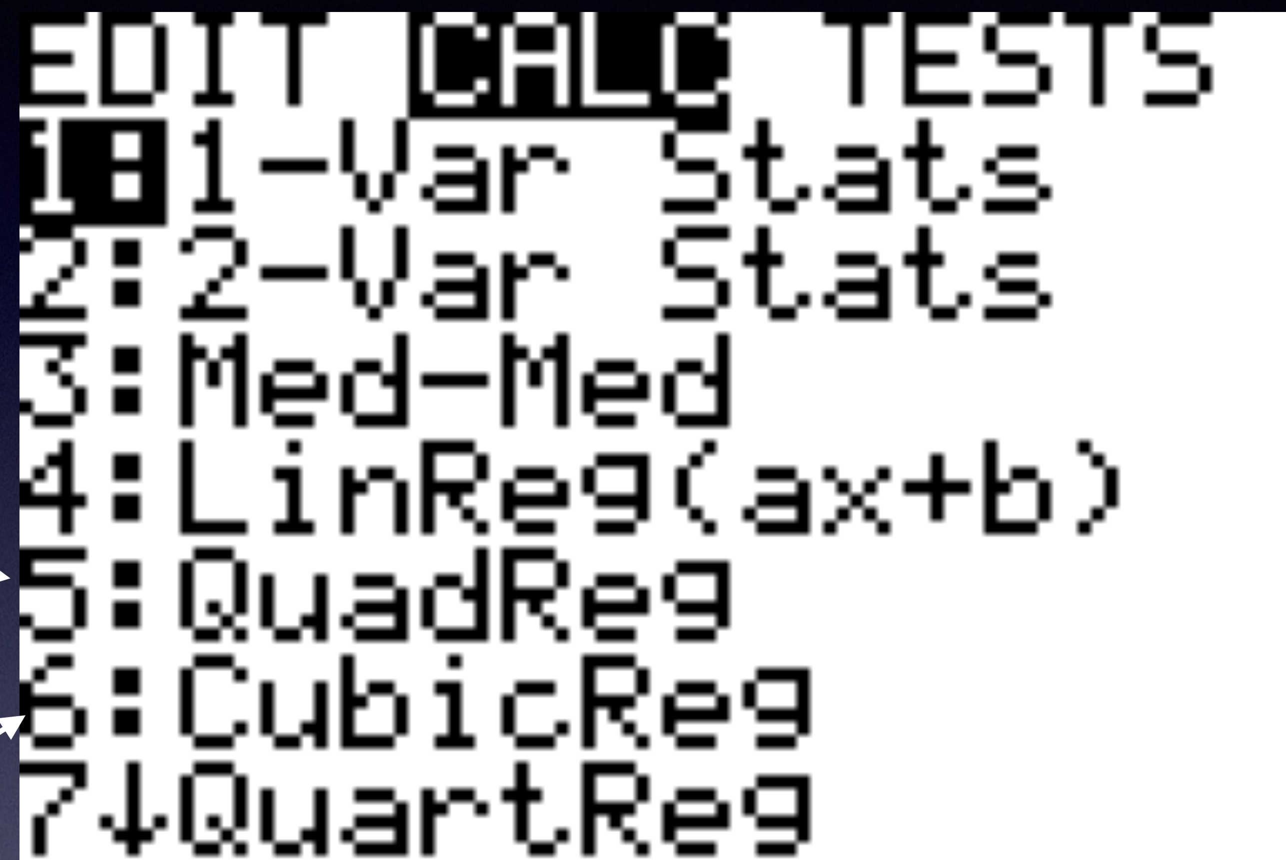
$$\hat{y} = a + bx$$

- Quadratic Regression

$$\hat{y} = ax^2 + bx + c$$

- Cubic Regression

$$\hat{y} = ax^3 + bx^2 + cx + d$$



- **Power** Regression

$$\hat{y} = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$$

You will choose the best fitting model and use that model to predict.

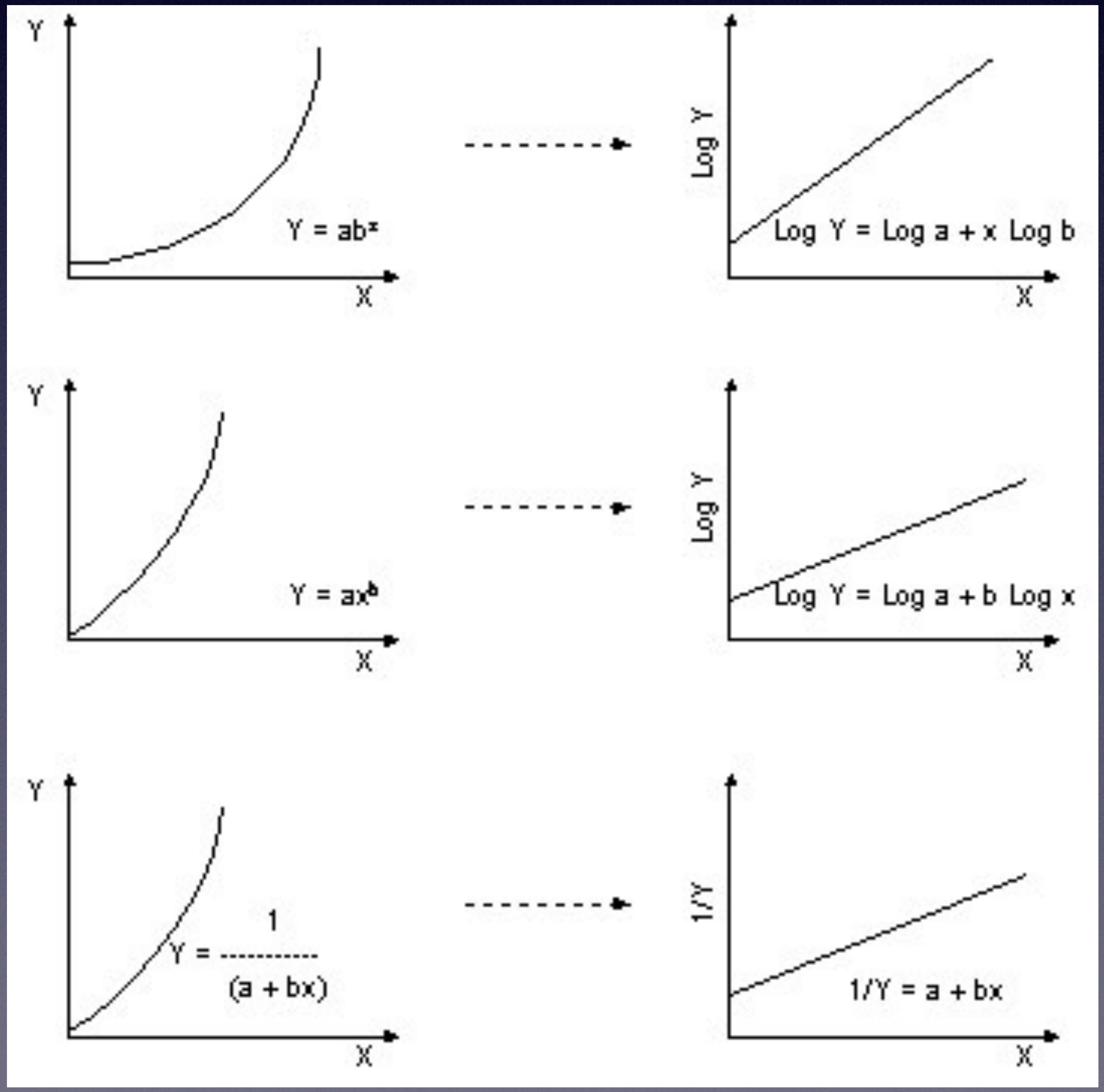
Is this model a good fit?

Four factors go into your decision -

1. Does the scatterplot look like the math function in question?
2. r^2 - should be close to +1 for a good fit
3. Residual plot - should be scattered for a good fit
4. s_e - should be small for a good fit...you can only tell if s_e is small if there is more than one MINTAB output to compare s_e's

Transferring to Get a Linear Model

When a graph of *y* vs. *x* does not appear linear, either *y* or *x* or both may be transformed (for example, by taking the log or raising to a power) in order to get a linear graph.



Predicting with a Transformed Model

When using transformed models to make predictions, first substitute *x* into the equation, then perform the inverse operation to get the predicted *y* value.

Example: Bacteria are allowed to grow in a petri dish. The number of bacteria is recorded each hour. This data results in an exponential graph with the following transformed linear model.

$$\log \hat{y} = 1.457 + 0.239x$$

y = number of bacteria
x = time in hours

Predict the number of bacteria after 6 hours.

Solution:

$$\log \hat{y} = 1.457 + 0.239x$$

$$\log \hat{y}\big|_{x=6} = 1.457 + 0.239(6) = 2.891$$

$$10^{\log \hat{y}|_{x=6}} = 10^{2.891}$$

$$\hat{y}\big|_{x=6} = 778 \text{ bacteria}$$

1. Suppose that the scatterplot of log*X* and log*Y* shows a strong positive correlation close to 1. Which of the following is true?

I. The variables *X* and *Y* also have a correlation close to 1.
II. A scatterplot of the variables *X* and *Y* shows a strong nonlinear pattern.
III. The residual plot of the variables *X* and *Y* shows a random pattern.

(a) I only (b) II only (c) III only (d) I and II (e) I, II, and III

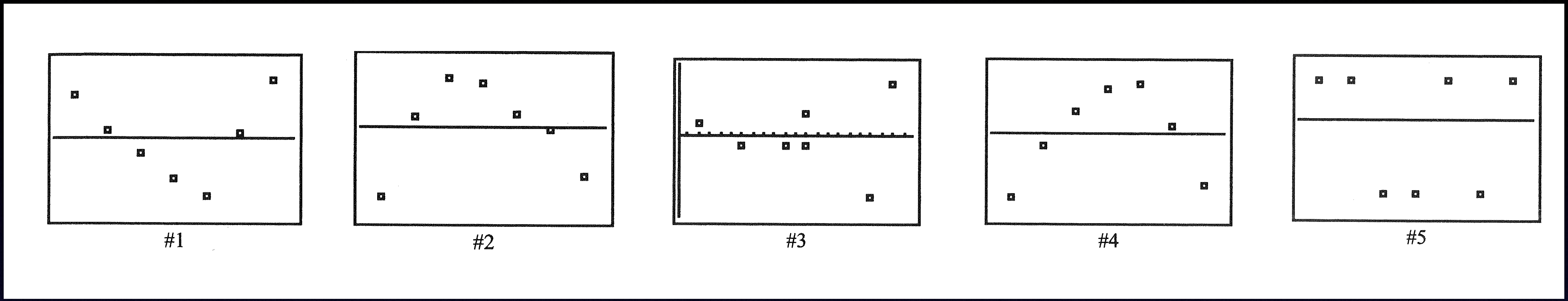
Answer: B

2. Using least-squares regression, I determine that the logarithm (base 10) of the population of a country is approximately described by the equation
Log (population) = -13.5 + 0.01(year)

Based on this equation, the population of the country in the year 2000 should be about
(a) 6.5
(b) 665
(c) 2,000,000
(d)3,162,277
(e) None of the above

Answer: D

3. A researcher made a scatterplot from some previously collected data. The data was clearly nonlinear in shape. The researcher then tried a variety of transformations on the data in an attempt to linearize the results. The residual plot for each is shown below.



Which of the transformations was best at linearizing the data?

(a) #1 (b) #2 (c) #3 (d) #4 (e) #5

Answer: C

4. A residual:

(a)is the amount of variation explained by the least-squares regression line of y on x.
(b)is how much an observed y value differs from a predicted y value.
(c)predicts how well x explains y.
(d)is the total variation of the data points.
(e)should be smaller than the mean of y.

Answer: B

5. Which of the following would indicate the strongest relationship between two variables?

(a) $r = 0.35$ (b) $r = -.28$ (c) $r = .21$ (d) $r^2 = .01$ (e) $r^2 = .23$

Answer: E

6. The coefficient of determination, r^2, between two variables is computed to be 81%. Which of the following statements must be true?

(a) Large values of the explanatory variable correspond with large values of the response variable.
(b) Large values of the explanatory variable correspond with small values of the response variable.
(c) A cause and effect relationship exists between the explanatory and response variables.
(d) There is a strong, positive, linear relationship between the explanatory and response variables.
(e) Approximately 81% of the variability in the response variable is explained by regression on the explanatory variable.

Answer: E

7. If the model for the relationship between the score on the AP Statistics Exam (*x*) and the number of hours spent preparing for the test (*y*) was $\log \hat{y} = 0.1 + 1.9 \log x$, determine the residual if a student studied 9 hours and earned an 85.

(a) 6.53
(b) 3.14
(c) 15.23
(d) 0
(e) -4.86

Answer: B