AP Statistics
Unit 02 REDO – Day 02 Notes

Name_____
Period_____

Transforming Data!

Sometimes our data doesn't appear in a straight line. Since we want to use a least squares regression LINE, we may need to adjust our data. Additionally, r measures the strength of the LINEAR relationship between two variables, which is yet another reason to transform the data. We can use transformations like square roots, cube roots, squaring, cubing, etc., but the two common transformation we will focus on include logs.

LINEAR MODEL	EXPONENTIAL MODEL	POWER MODEL
$\hat{y}$ = a + bx	$\widehat{log(y)}$ = a + bx	$\widehat{log(y)}$ = a + b log(x)
L1 vs. L2	L1 vs. L4	L3 vs. L4

L1 = x
L2 = y
L3 = log(x)
L4 = log(y)

Things to check when comparing models:

1. R^2. We want this to be close to 1. The closer, the better. If you look at r instead, make sure it's close to -1 or +1.

2. The scatterplots. We want to choose the model that looks most linear.

3. The residual plots. If our comparison of R^2 is too close to call, and our plots look linear making it difficult to choose one, check out the residual plot! No pattern is good. Usually one plot has less of a pattern than the others.

EXAMPLE: Below is a table of data representing bacteria growth (in hundreds) after a certain amount of time (in minutes).

Time	Count
1	15
2	19
3	21
4	32
5	36
6	38
7	56
8	60
9	104
10	106
11	142
12	166
13	197
14	211
15	355

Is a linear model a good fit? If not, suggest a better model.

Write the equation for the LSRL of your chosen model.

Calculate and interpret the residual value at 7 minutes.