LINEAR REGRESSION

TWO CONTINUOUS VARIABLES

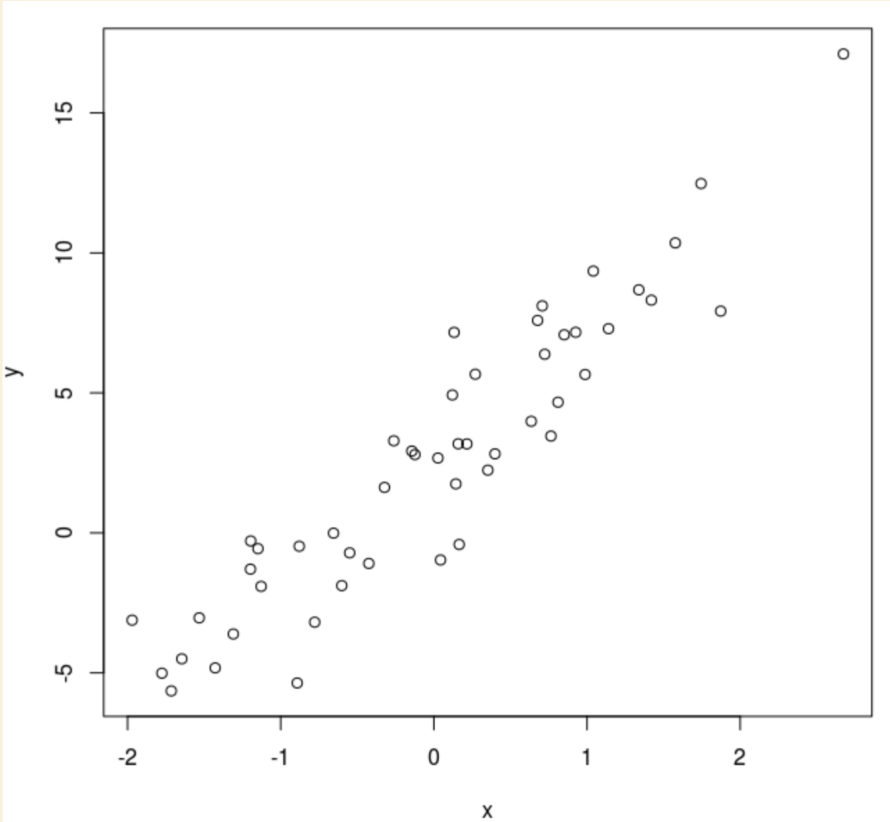
BRIAN STOCK

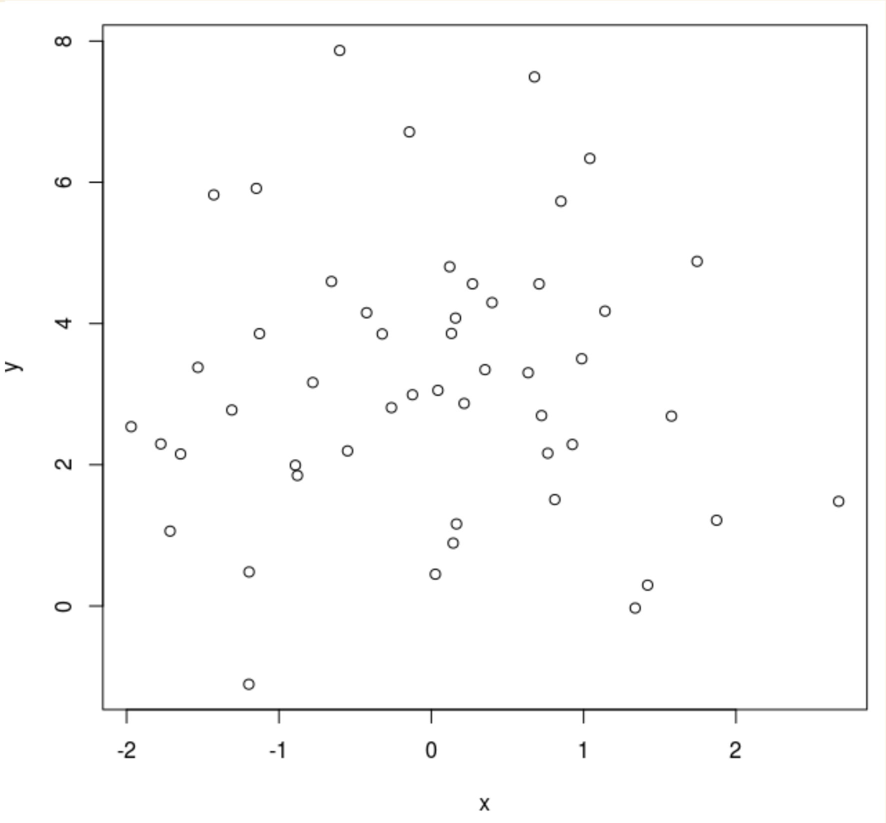
03.27.17

LINEAR REGRESSION

- Predicts *Y* from *X* (both continuous)

$$Y = \alpha + \beta X$$



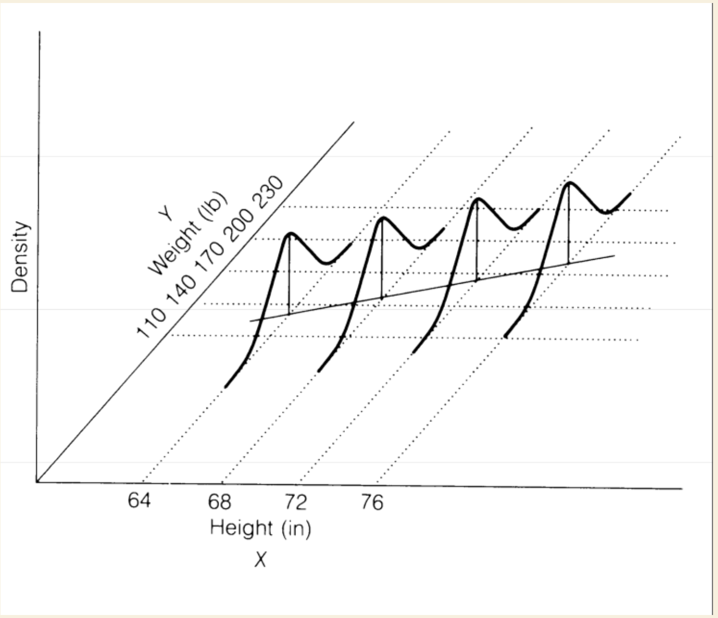


REGRESSION ASSUMES…

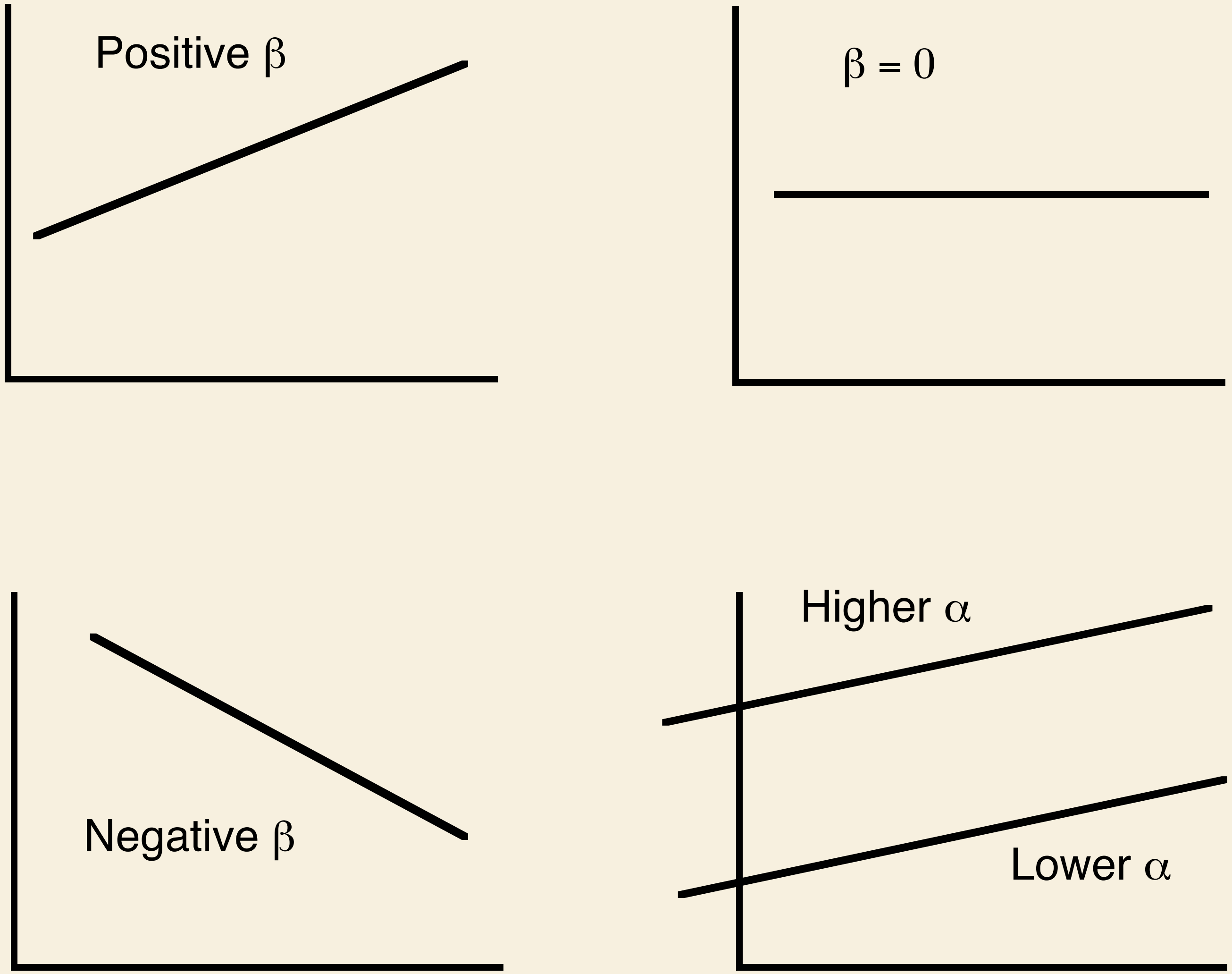
1. Random sample
2. The relationship between *X* and *Y* can be described by a **line:** $Y = \alpha + \beta X$

REGRESSION ASSUMES…

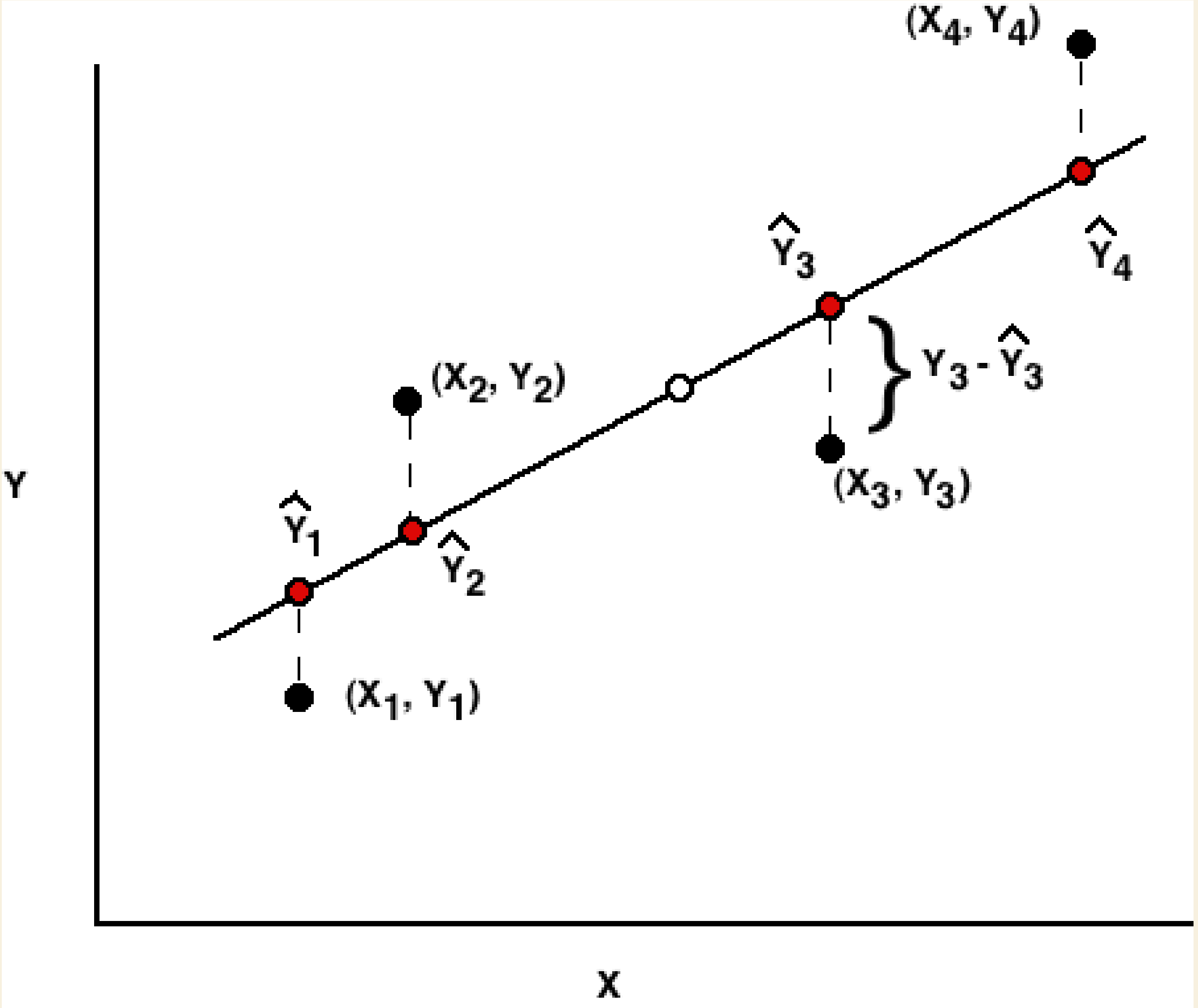
1. Random sample
2. The relationship between *X* and *Y* can be described by a **line:** $Y = \alpha + \beta X$
3. *Y* is normally distributed with equal variance for all values of *X*



2 PARAMETERS $Y = \alpha + \beta X$



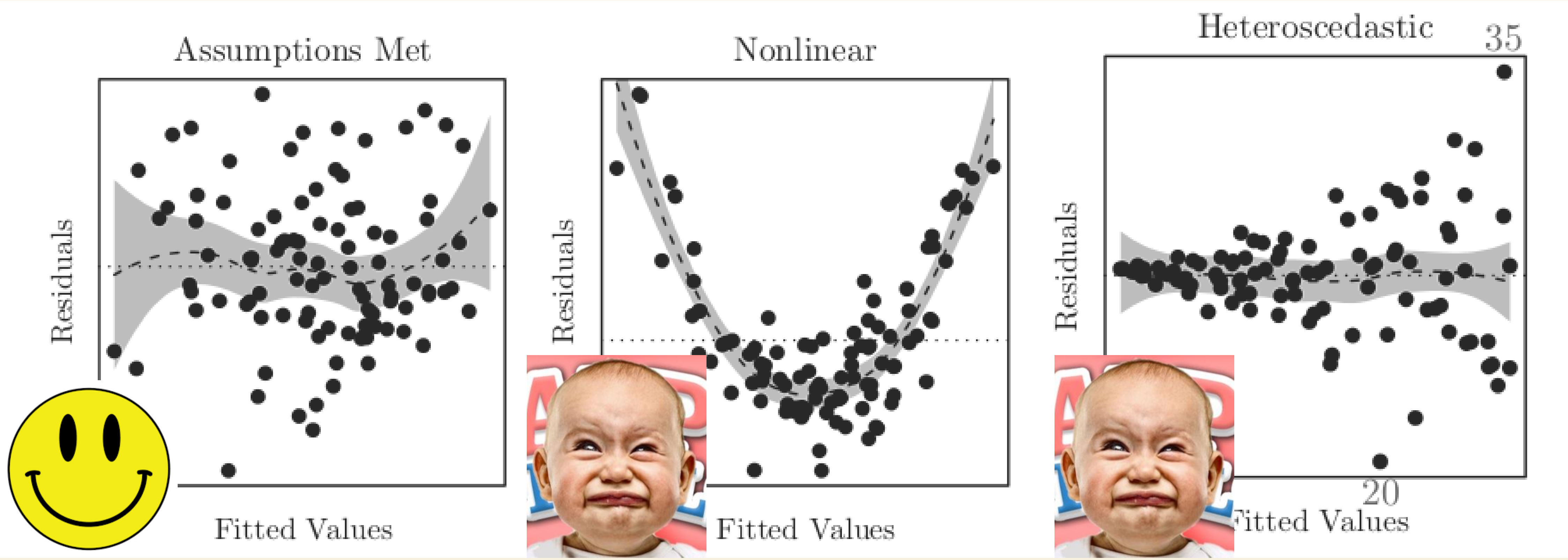
“RESIDUAL”



$$Y_i - \hat{Y}_i$$

RESIDUAL PLOTS

1. Normal (no pattern)
2. Equal variance



FINDING THE "LEAST SQUARES" REGRESSION LINE

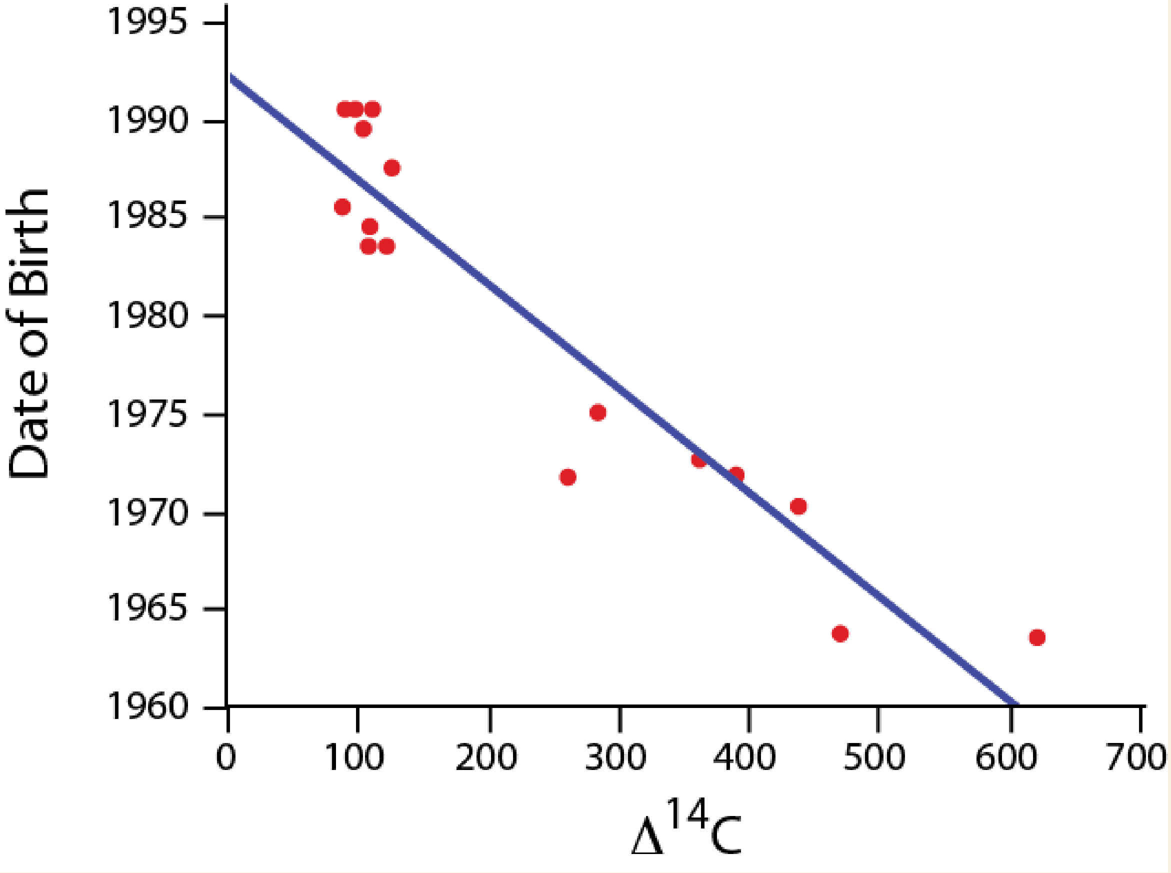
Minimize: $$SS_{residual} = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$

Your computer does this for you…

EX: AGE ~ TOOTH RADIOACTIVITY

Many above ground nuclear bomb tests in the '50s and '60s may have left a radioactive signal in developing teeth.

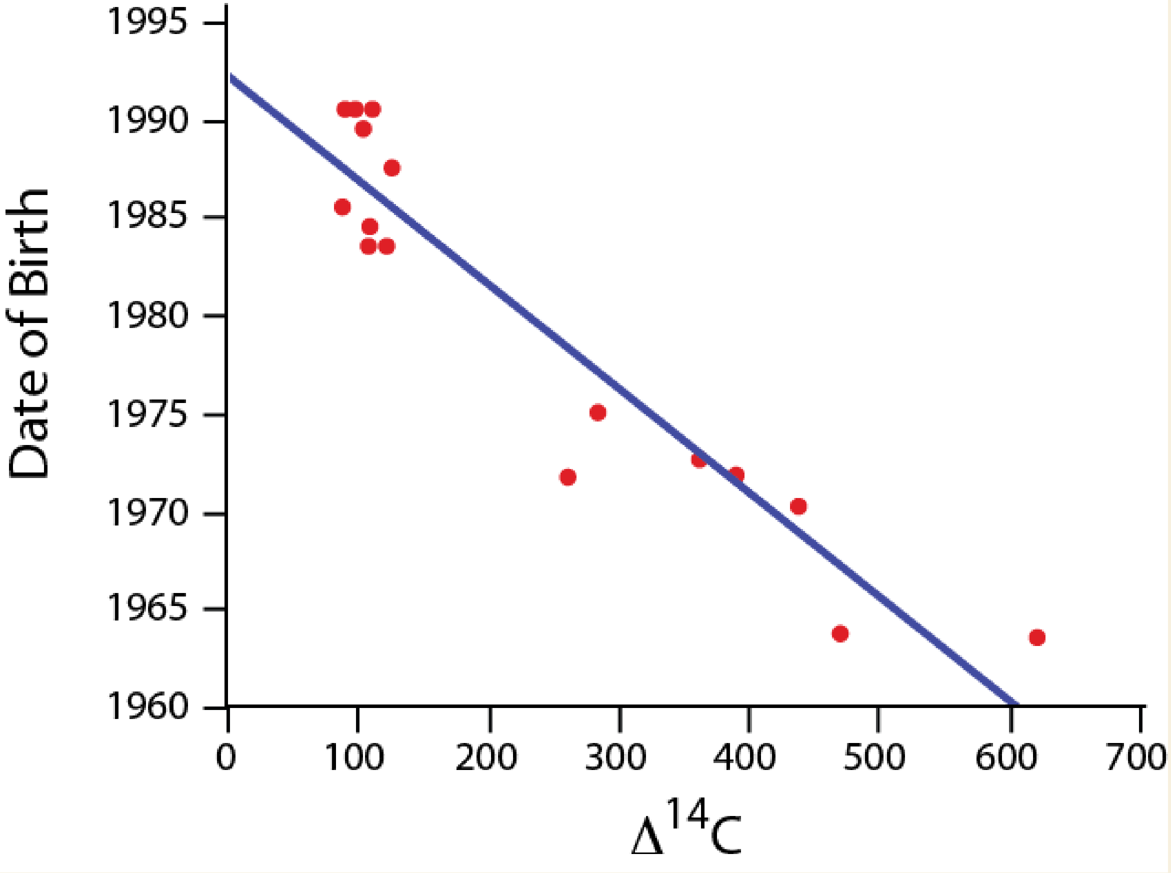
Is it possible to predict a person's age based on dental C^{14}?



Data from 1965 to present from Spalding et al. 2005. Forensics: age written in teeth by nuclear tests. ***Nature*** *437: 333–334.*

EX: AGE ~ TOOTH RADIOACTIVITY

$$\hat{Y} = 1992.2 - 0.053X$$



Data from 1965 to present from Spalding et al. 2005. Forensics: age written in teeth by nuclear tests. ***Nature*** *437: 333–334.*

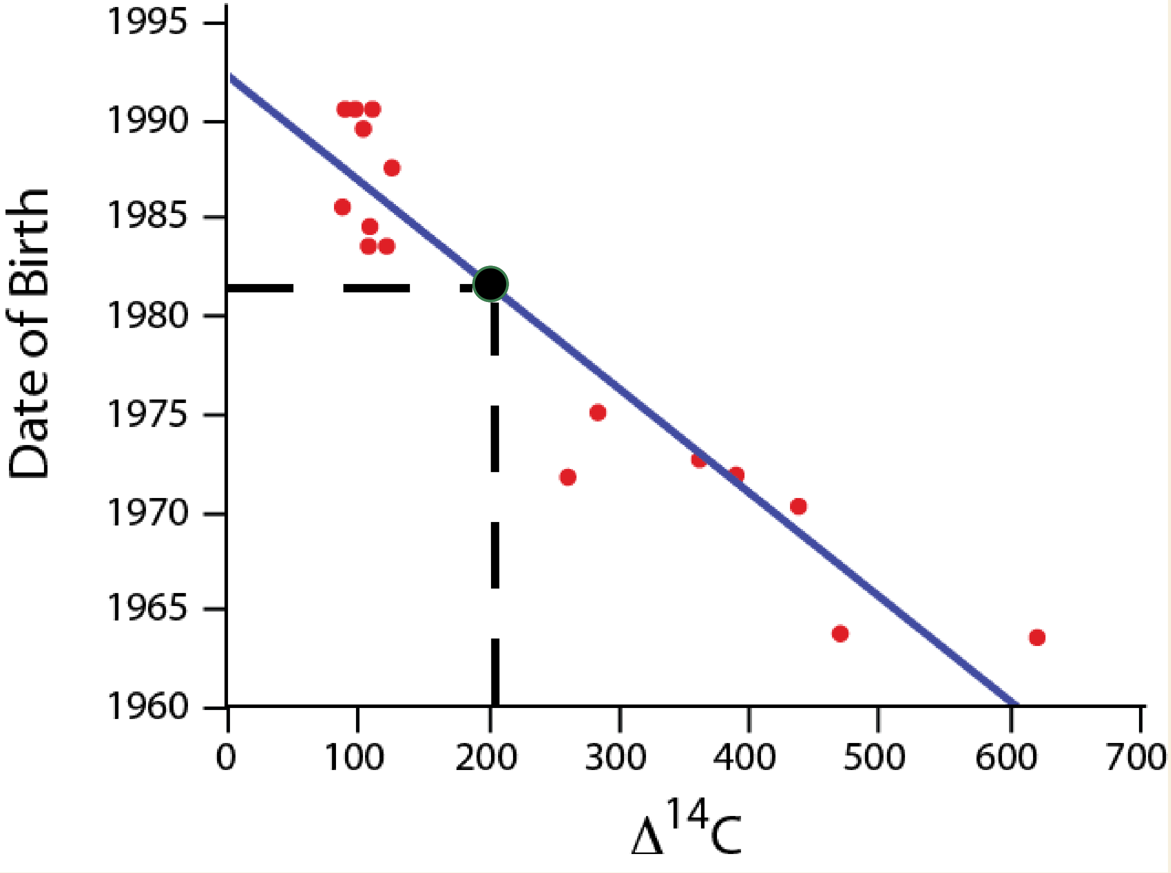
PREDICTING *Y* FROM *X*

If a cadaver has a tooth with $\Delta^{14}C = 200$, what does the linear model predict its year of birth to be?

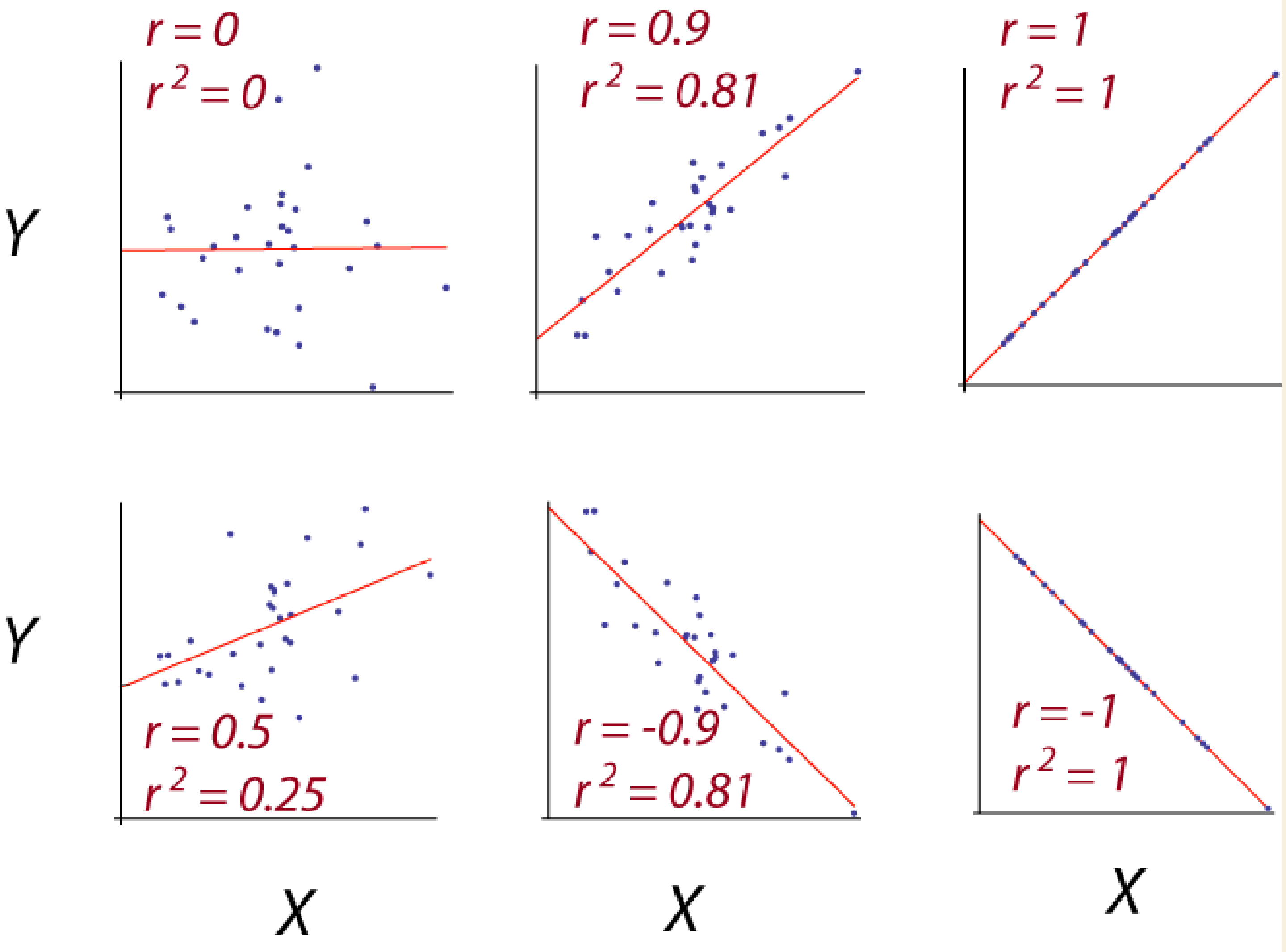
$$\hat{Y} = 1992.2 - 0.053X$$

$$= 1992.2 - 0.053(200)$$

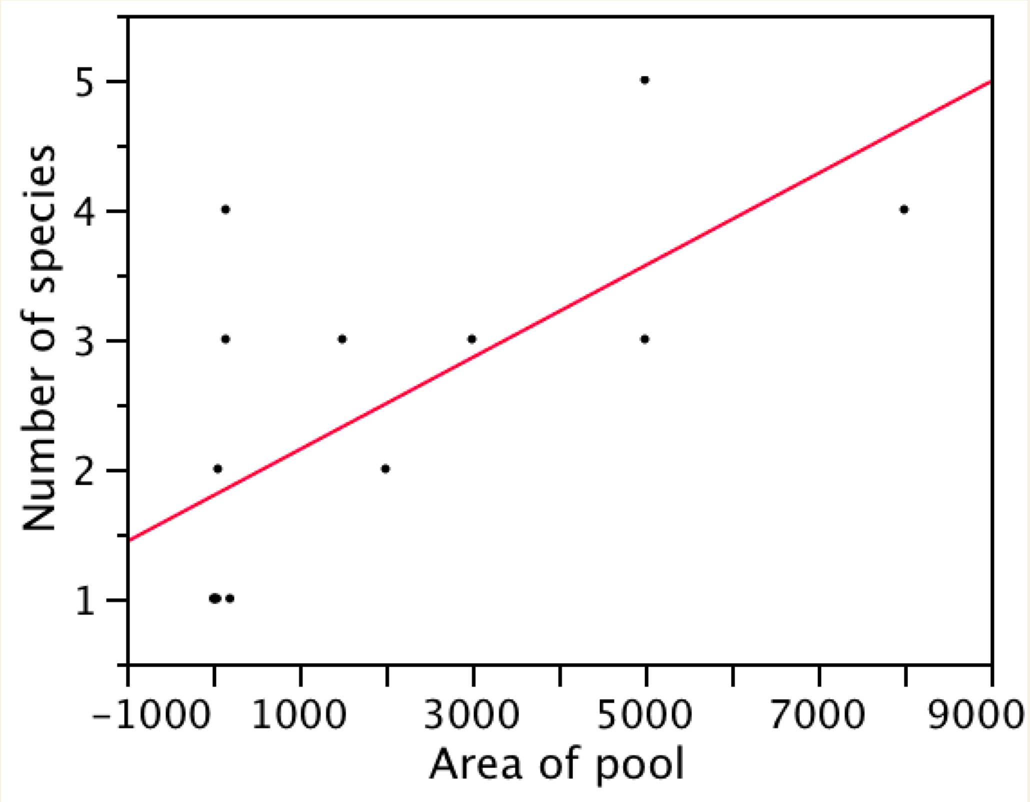
$$= 1981.6$$



R^2 predicts the amount of variance in *Y* explained by the linear model



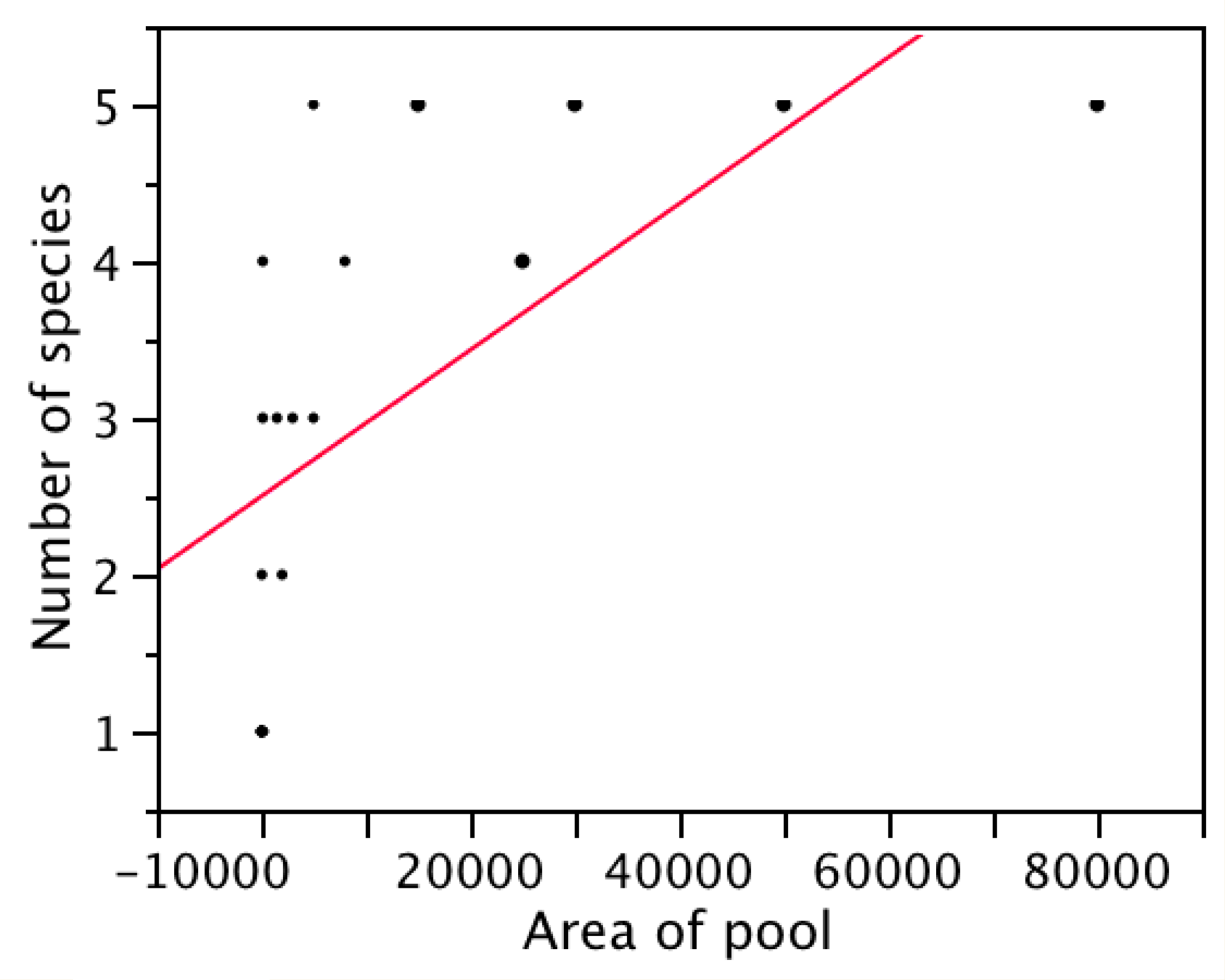
CAUTION: IT IS UNWISE TO EXTRAPOLATE BEYOND THE RANGE OF THE DATA.



Number of species of fish as predicted by the area of a desert pool

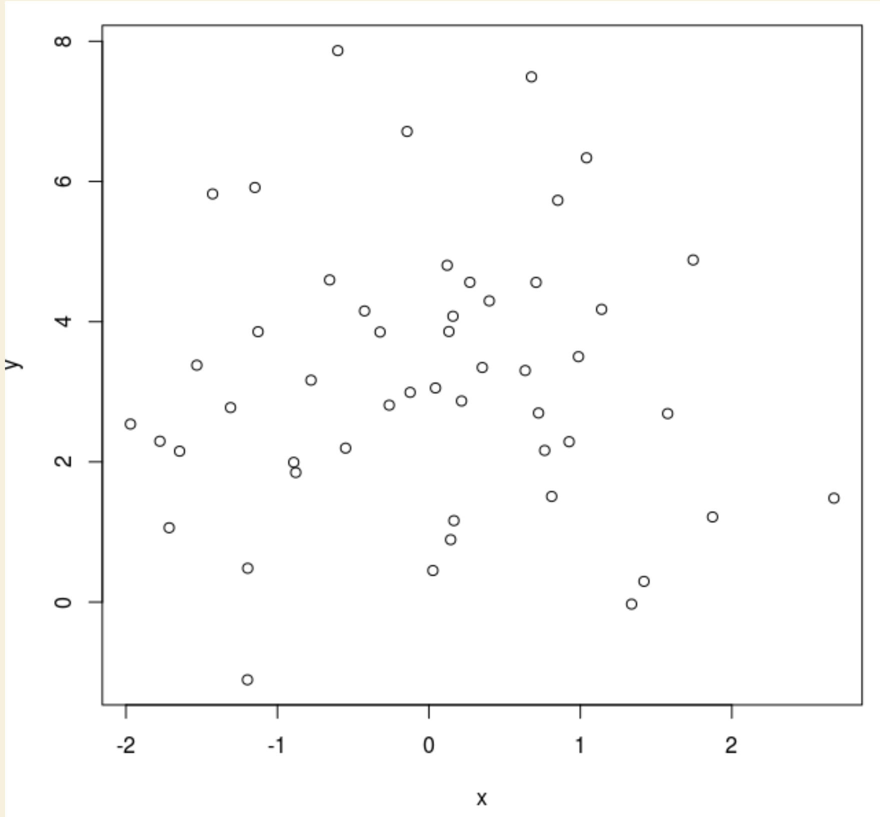
If we were to extrapolate to ask how many species might be in a pool of 50000m^2, we would guess about 20.

MORE DATA ON FISH IN DESERT POOLS

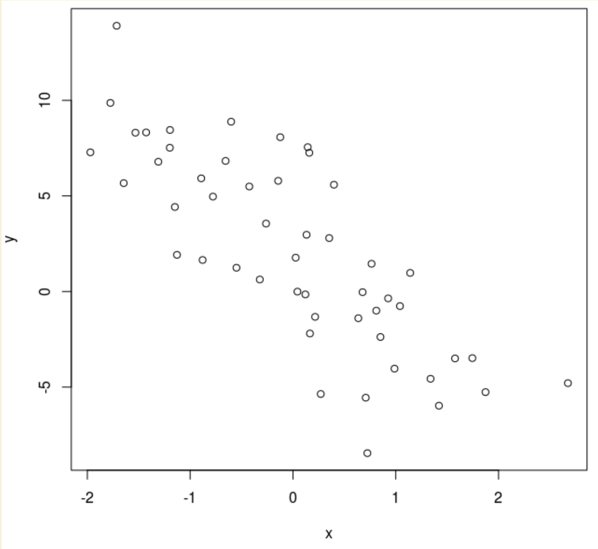


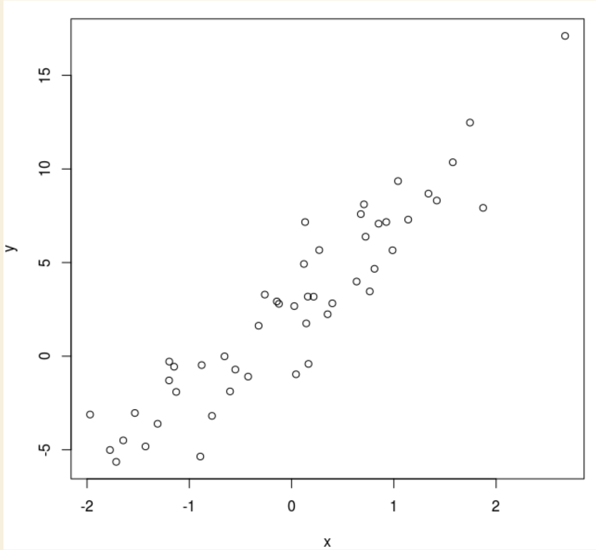
REGRESSION HYPOTHESES

$H_0: \beta = 0$



$H_A: \beta \neq 0$





HYPOTHESIS TEST: β

H_0: $\beta = 0$

H_A: $\beta \neq 0$

$$t = \frac{b - b_0}{SE_b}$$

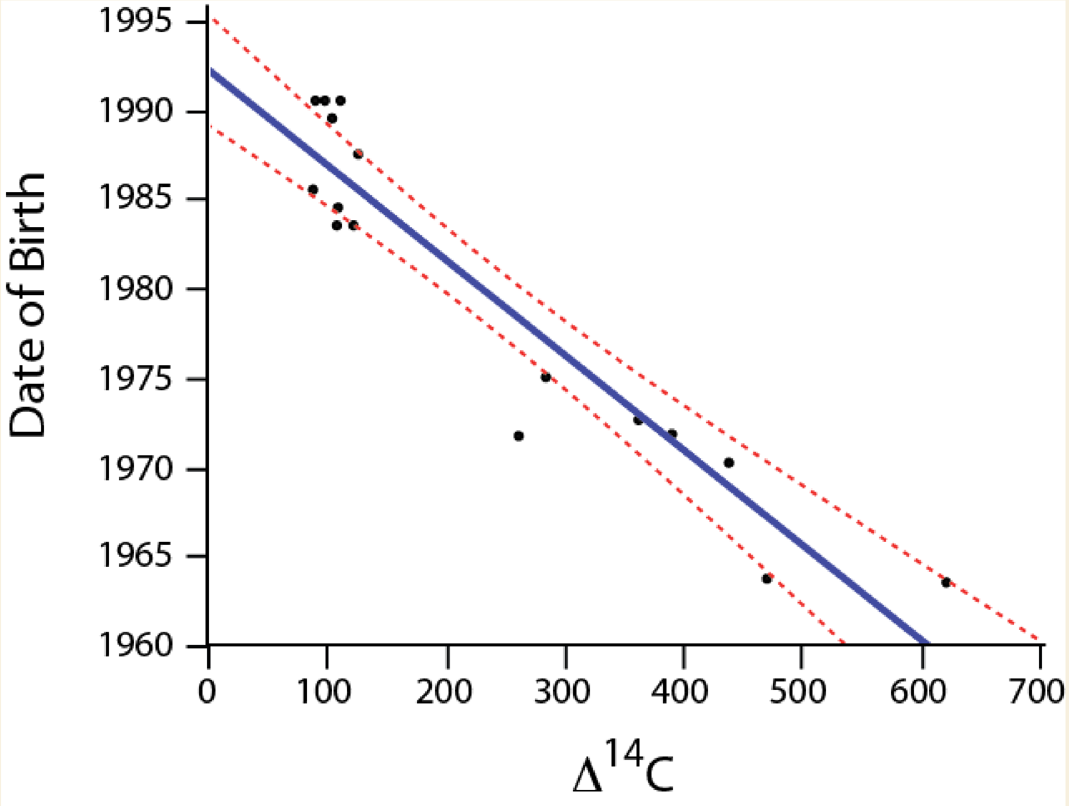
$$t = \frac{-0.053 - 0}{0.004} = 13.25$$

$t_{0.0001(2),14} = \pm 5.36$

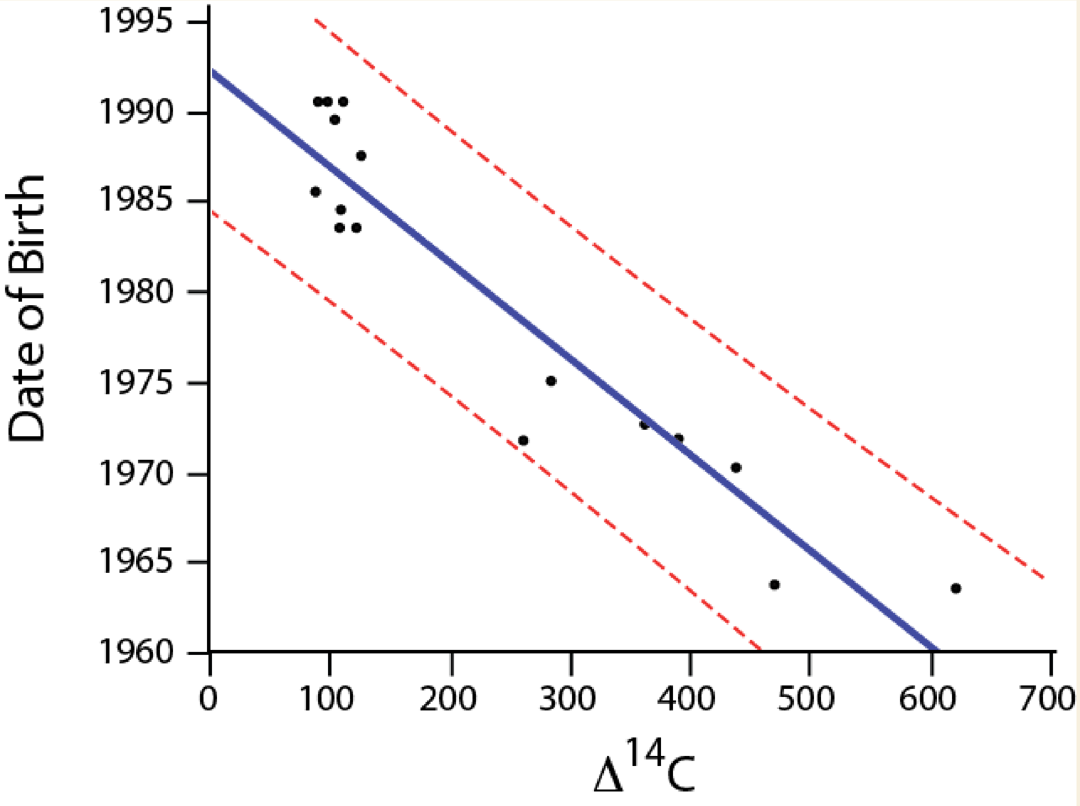
So we can reject H_0, $P<0.0001$

CONFIDENCE INTERVALS FOR PREDICTIONS OF:

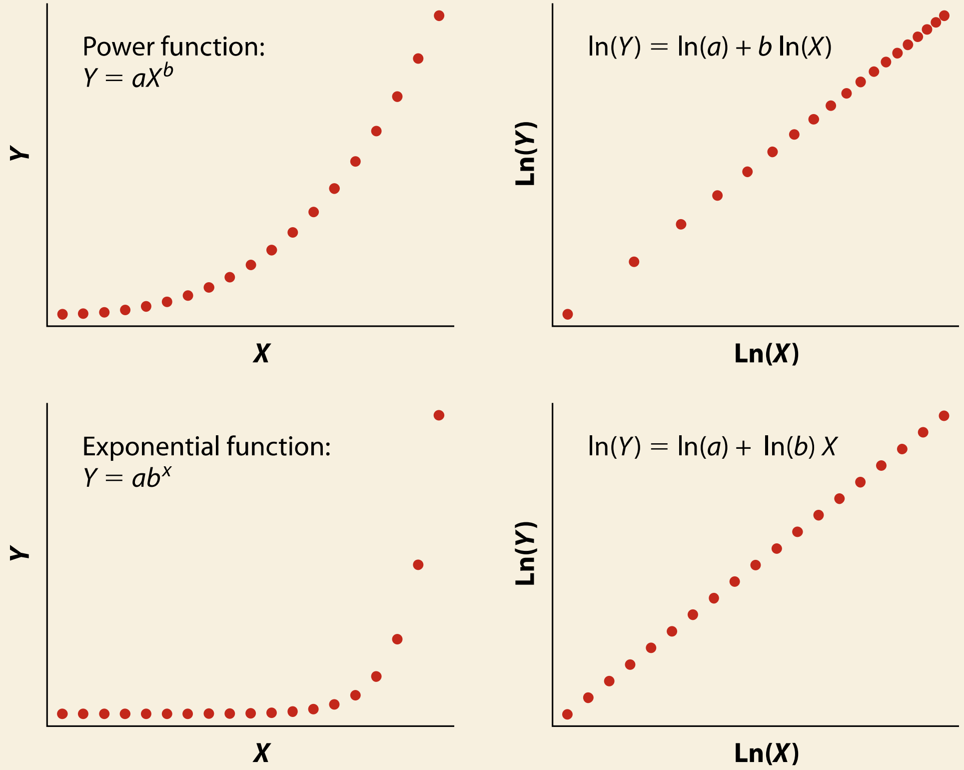
MEAN *Y*



INDIVIDUAL *Y*



TRANSFORMATIONS



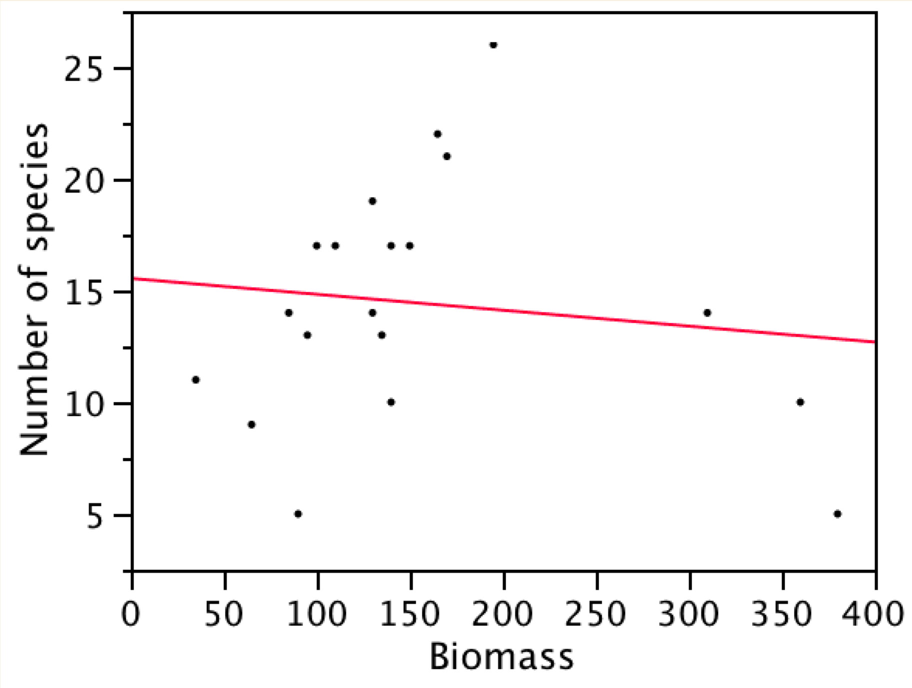
If $Y = aX^b$ then $\ln Y = \ln a + b\ln X$.

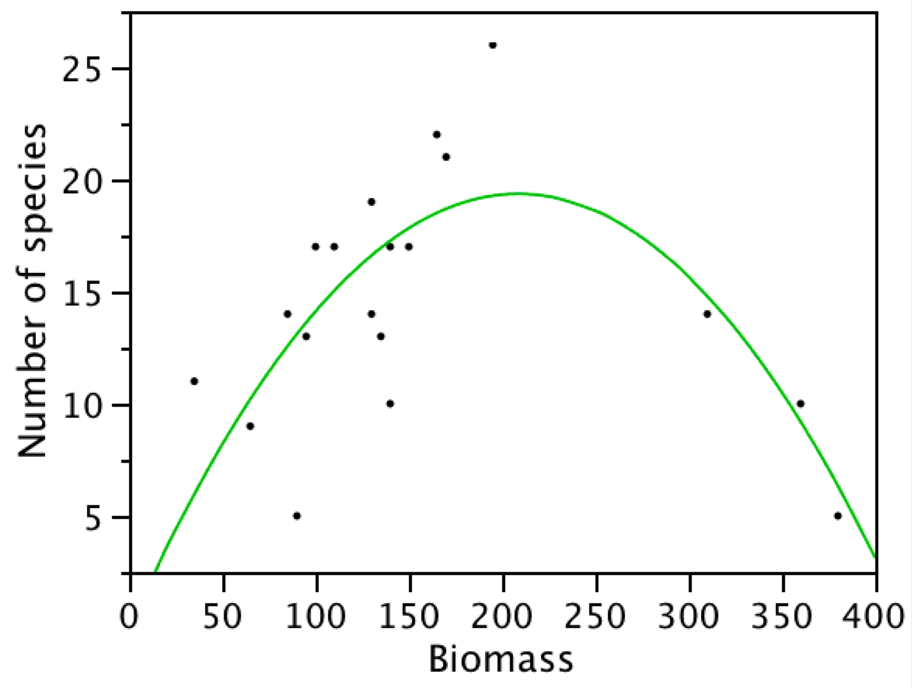
If $Y = ab^X$ then $\ln Y = \ln a + X\ln b$.

If $Y = a + \dfrac{b}{X}$ then set $X' = \dfrac{1}{X}$, and calculate $Y = a + bX'$.

All of the equations on the right have the form Y=a+bX.

POLYNOMIAL REGRESSION





Number of species = 0.046 + 0.185 Biomass - 0.00044 Biomass^2