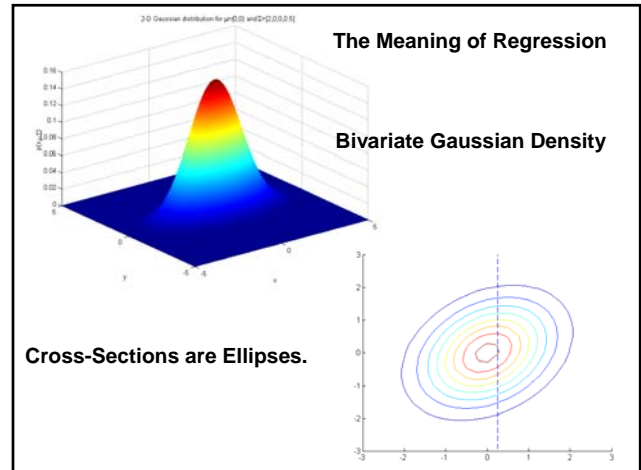**STATISTICS FOR NEUROSCIENCE RESEARCH**

9.073 / HST 460

Class 6: Review of Statistics III: Simple Regression

Emery N. Brown

February 27, 2017

The Meaning of Regression

Bivariate Gaussian Density

Cross-Sections are Ellipses.

Bivariate Gaussian (Joint) Density of (X, Y)

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y(1-\rho^2)^{\frac{1}{2}}} \times$$

$$\exp\{-\frac{1}{2(1-\rho^2)}[\frac{(x-\mu_x)^2}{\sigma_y^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}]\}$$

Conditional Density of Y Given X $\Pr(B\,|\,A) = \frac{\Pr(A\bigcap B)}{\Pr(A)}$
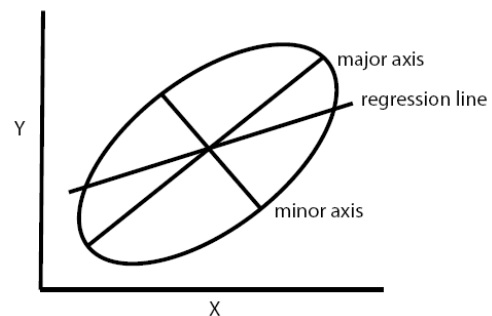
$$f(y\,|\,x) = \frac{f(x,y)}{f(y)}$$

Conditional Expectation: Theoretical Regression Line

$$E(Y\,|\,X = x) = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x)$$

Conditional Variance

$$Var(Y\,|\,X = x) = \sigma_y - \rho\sigma_{xy}\sigma_y$$

The Geometry of the Regression Line

major axis

regression line
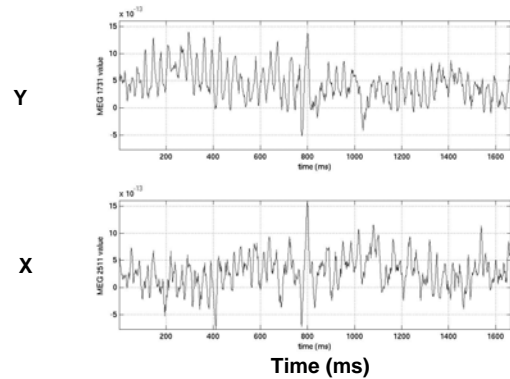
Y

minor axis

X

1

The Meaning of Regression

Conditional Expectation: Theoretical Regression Line

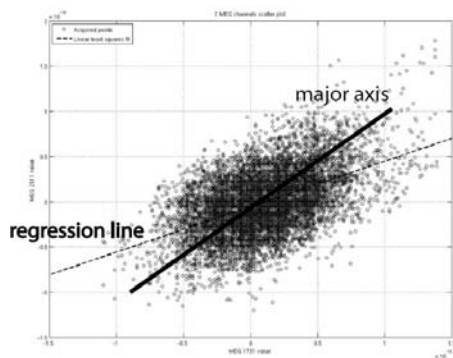$$E(Y \mid X = x) = \mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x)$$

The Empirical Regression Line Variance

$$\hat{y} = \overline{y} + \hat{\rho}\frac{\hat{\sigma}_y}{\hat{\sigma}_x}(x - \overline{x})$$
$$= \overline{y} + \hat{\beta}(x - \overline{x})$$
$$= \overline{y} - \hat{\beta}\overline{x} + \hat{\beta}x$$
$$= \hat{\alpha} + \hat{\beta}x$$

Front (Y) and Back (X) MEG Sensor Background Noise Recordings

Y

X

Time (ms)

## The Meaning of Regression

major axis

regression line

The Meaning of Regression

The geneticist Sir Francis Galton (1822-1911) observed that the sons of fathers who were taller than average tended to be shorter than average and that the sons of fathers who were shorter than average tended to be taller than average. He termed the phenomenon "regression towards mediocrity". More recently it has been termed "regression to the mean".

This relation is given exactly by the regression line.

Heights of 1078 Pairs of Fathers and Sons

major axis
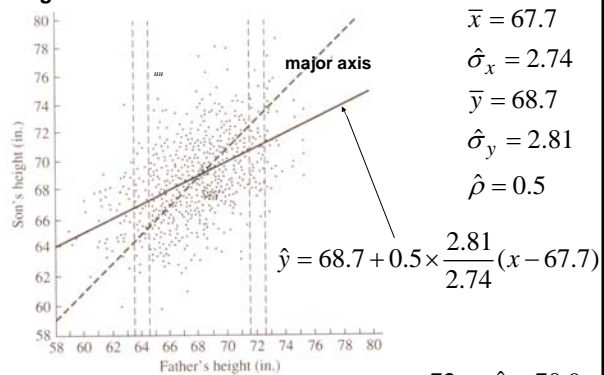
Rice (2007)

$\bar{x} = 67.7$

$\hat{\sigma}_x = 2.74$

$\bar{y} = 68.7$

$\hat{\sigma}_y = 2.81$

$\hat{\rho} = 0.5$

$$\hat{y} = 68.7 + 0.5 \times \frac{2.81}{2.74}(x - 67.7)$$

$x = 72 \Rightarrow \hat{y} = 70.9$

$x = 64 \Rightarrow \hat{y} = 66.8$

Summary

Simple regression is our most basic technique for relating one variable to another.

It is useful to be able to think of simple regression in terms of likelihood, method-of-moments and least squares analyses.

The geometry of the Pythagorean relations carries over to the case of multiple regression.

The concepts discussed here form the basis for our formulations of more complex models.