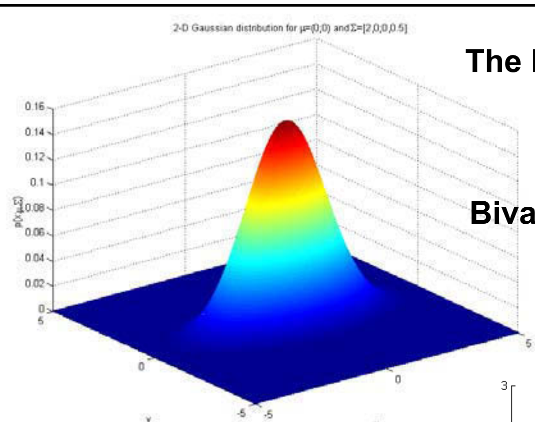
STATISTICS FOR BRAIN AND COGNITIVE SCIENCES

9.07

Lecture 13 Addendum: The Meaning of Regression

Emery N. Brown

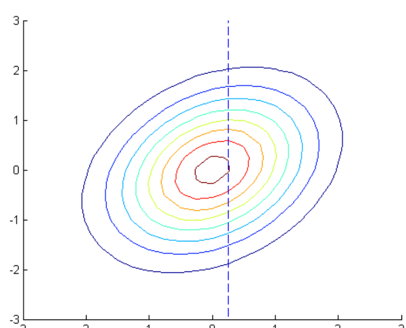
December 7, 2016



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

$$\Pr(B \mid A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

$$f(y \mid x) = \frac{f(x,y)}{f(y)}$$

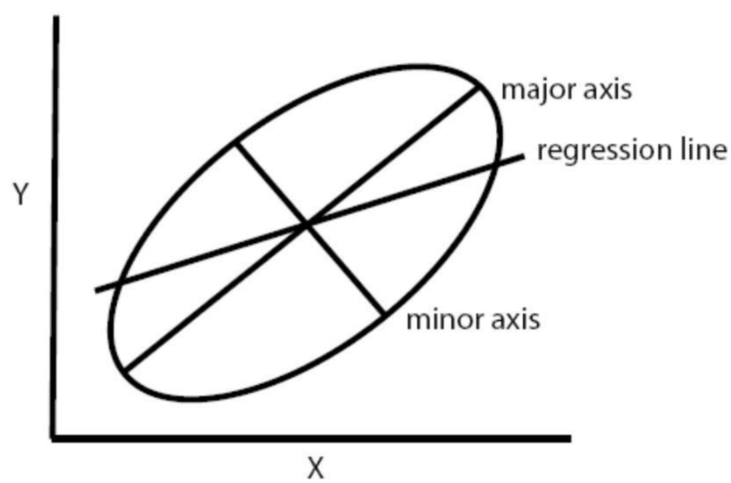
Conditional Expectation: Theoretical Regression Line

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

Conditional Variance

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

The Geometry of the Regression Line



The Meaning of Regression

Conditional Expectation: Theoretical Regression Line

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

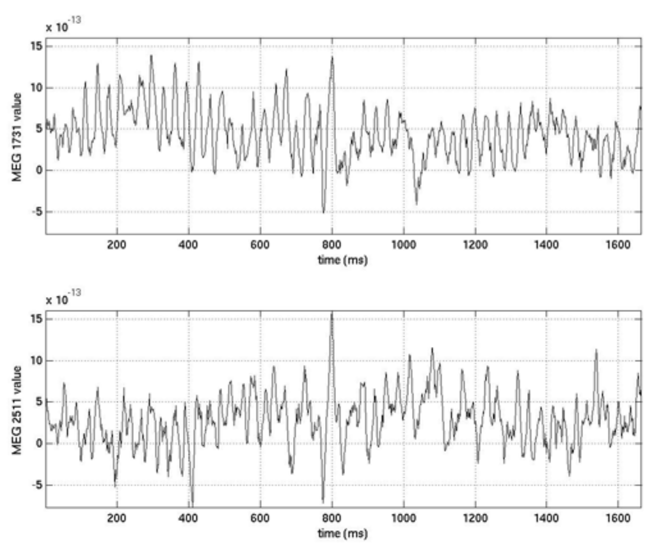
The Empirical Regression Line Variance

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

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

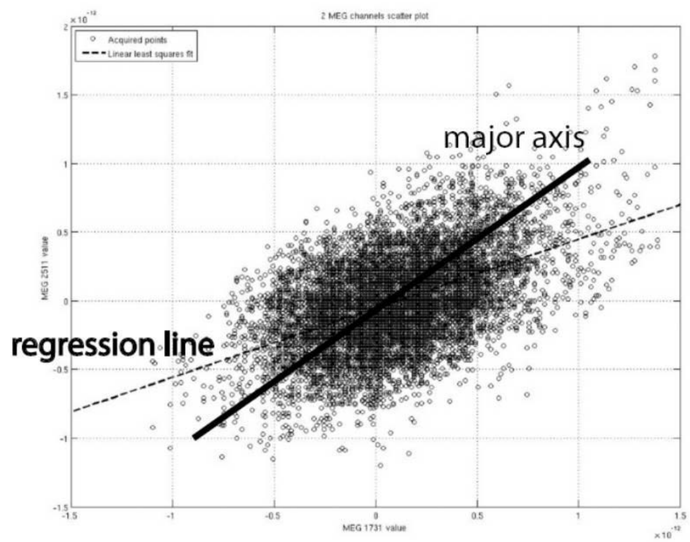
Y

X



Time (ms)

The Meaning of Regression

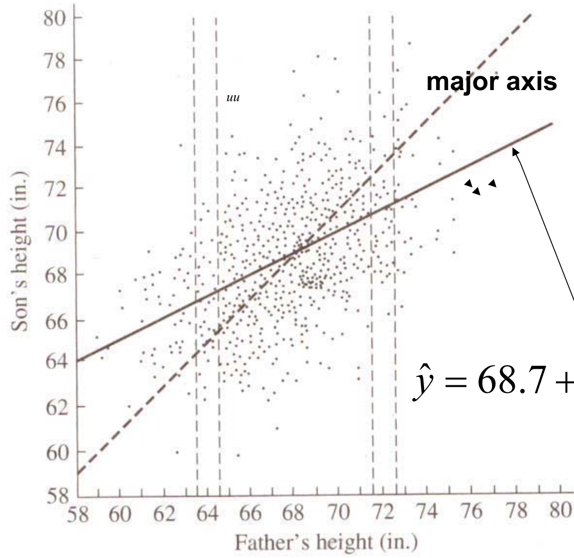


The Meaning of Regression

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



$$\bar{x} = 67.7$$
$$\hat{\sigma}_x = 2.74$$
$$\bar{y} = 68.7$$
$$\hat{\sigma}_y = 2.81$$
$$\hat{\rho} = 0.5$$

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

Rice (2007)

$$x = 72 \Rightarrow \hat{y} = 70.9$$
$$x = 64 \Rightarrow \hat{y} = 66.8$$

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.