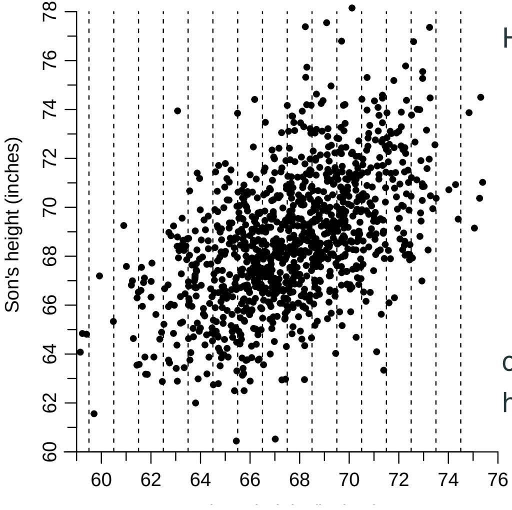
STAT 234 Lecture 23B
Simple Linear Regression Model
Section 12.1

Yibi Huang
Department of Statistics
University of Chicago

Example: Pearson's Father-and-Son Data

Father-son pairs are grouped by father's height, to the nearest inch.



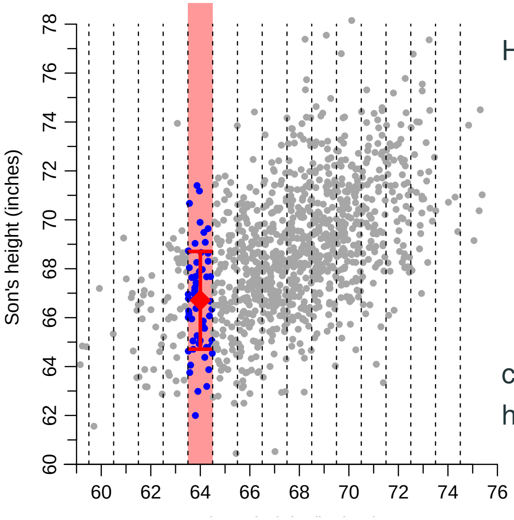
How do the

- conditional mean of son's height (SH),
- conditional SD of SH, and
- conditional distribution of SH?

change with father's height?

Example: Pearson's Father-and-Son Data

Father-son pairs are grouped by father's height, to the nearest inch.



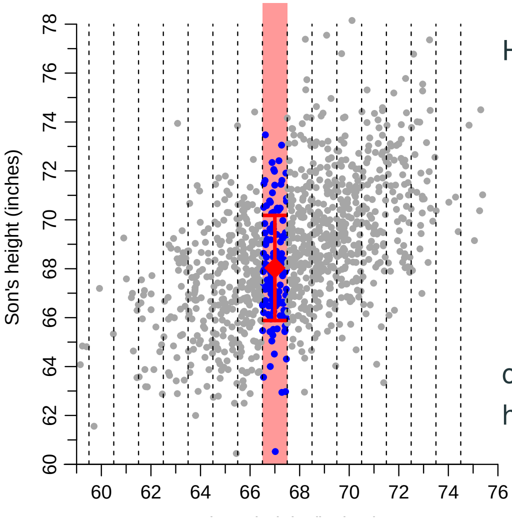
How do the

- conditional mean of son's height (SH),
- conditional SD of SH, and
- conditional distribution of SH?

change with father's height?

Example: Pearson's Father-and-Son Data

Father-son pairs are grouped by father's height, to the nearest inch.



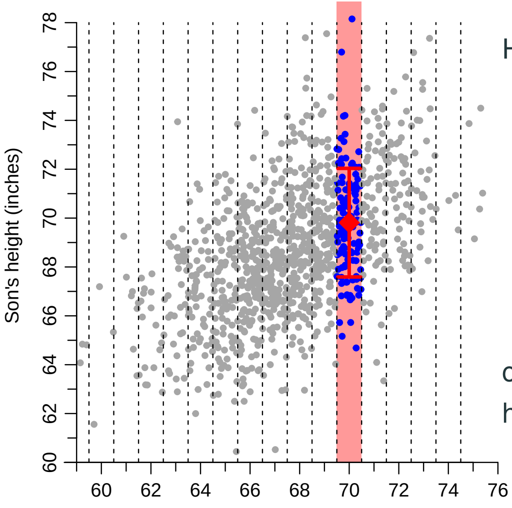
How do the

- conditional mean of son's height (SH),
- conditional SD of SH, and
- conditional distribution of SH?

change with father's height?

Example: Pearson's Father-and-Son Data

Father-son pairs are grouped by father's height, to the nearest inch.



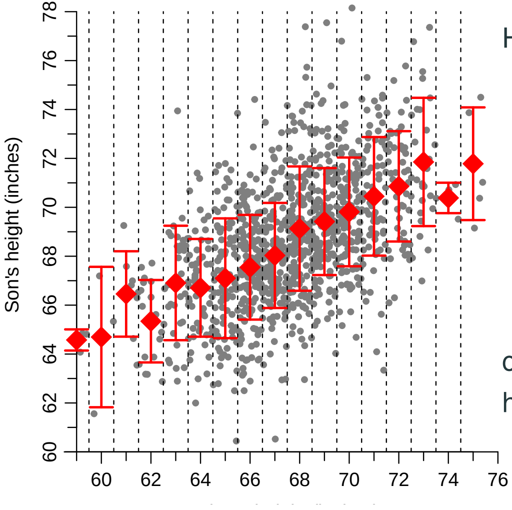
How do the

- conditional mean of son's height (SH),
- conditional SD of SH, and
- conditional distribution of SH?

change with father's height?

Example: Pearson's Father-and-Son Data

Father-son pairs are grouped by father's height, to the nearest inch.

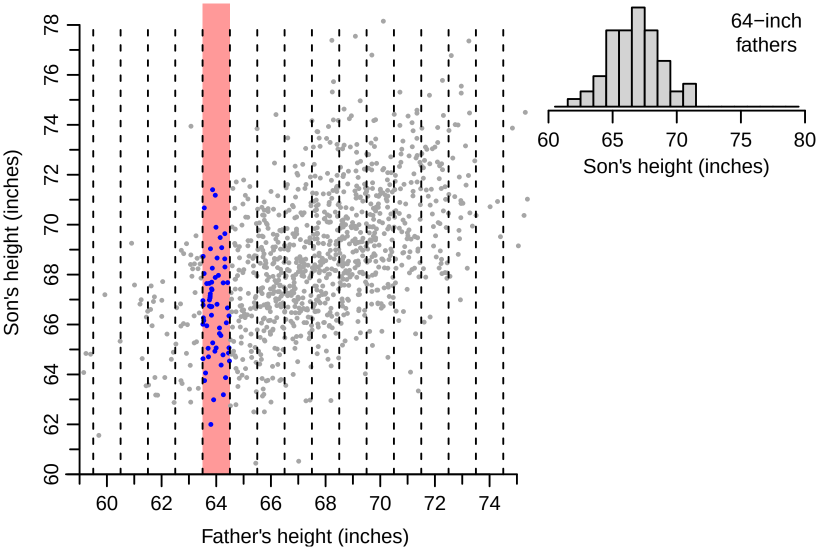


How do the

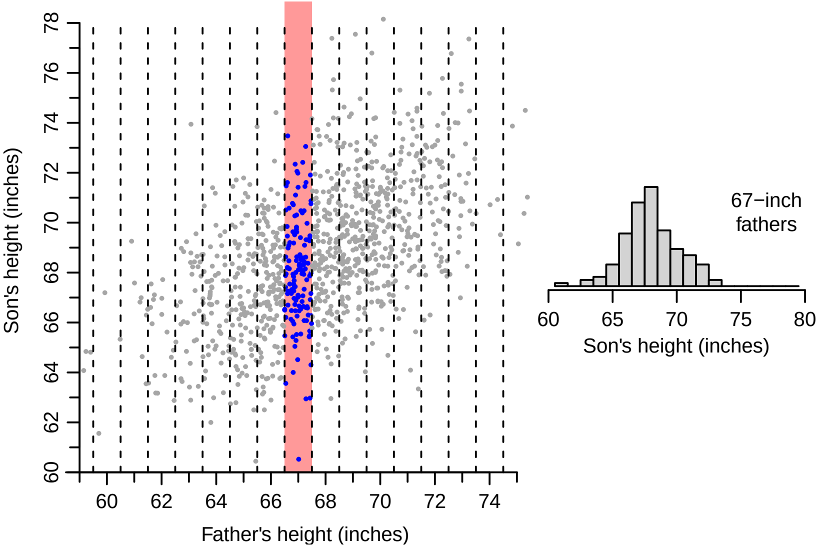
- conditional mean of son's height (SH),
- conditional SD of SH, and
- conditional distribution of SH?

change with father's height?

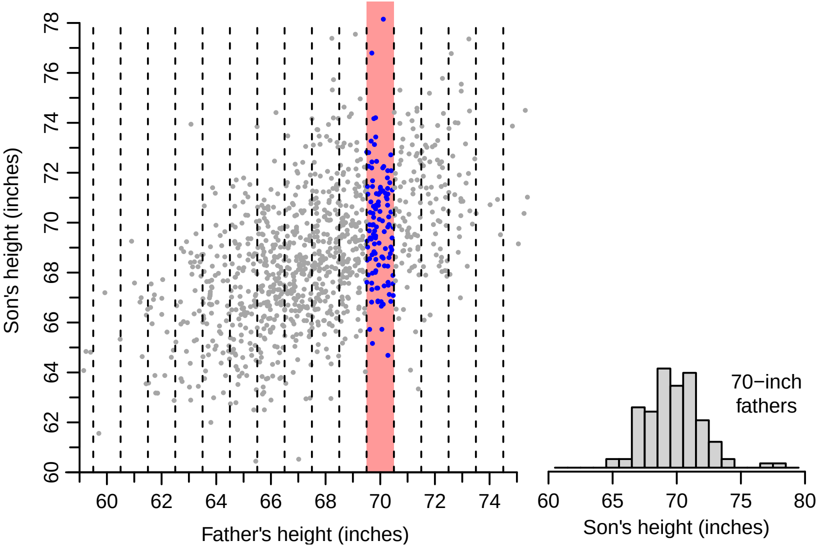
Example: Pearson’s Father-and-Son Data



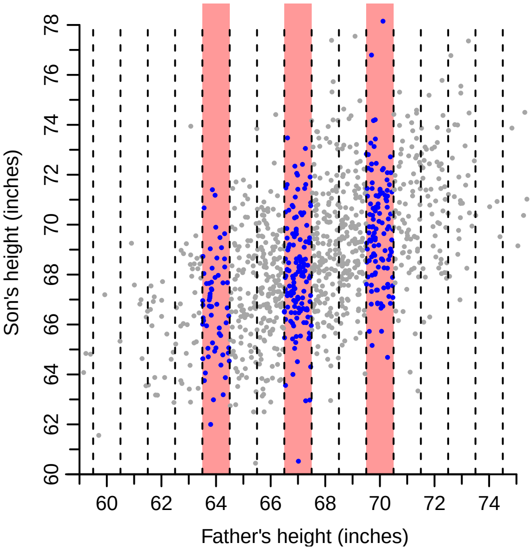
Example: Pearson’s Father-and-Son Data

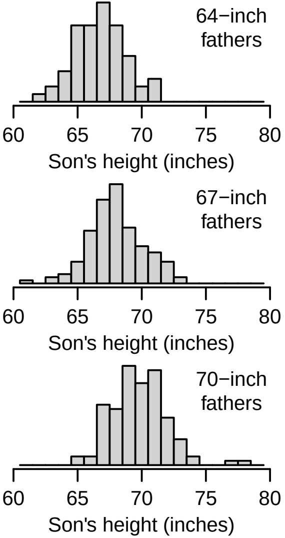


Example: Pearson's Father-and-Son Data



Example: Pearson’s Father-and-Son Data





Simple Linear Regression Model

Pearson's father-and-son data inspire the following assumptions for the simple linear regression (SLR) model:

1. The condition mean of Y given $X = x$is a linear function of x, i.e.,

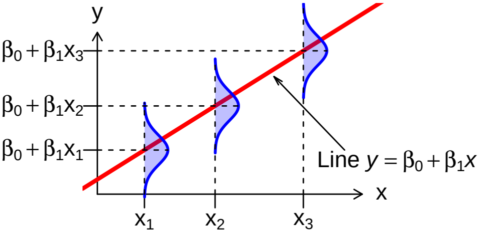
$$\mathrm{E}(Y \mid X = x) = \beta_0 + \beta_1 x$$

2. The conditional variance of Y does not change with x, i.e.,

$$\mathrm{Var}(Y \mid X = x) = \sigma^2 \quad \text{for every } x$$

3. (Optional) The conditional distribution of Y given $X = x$ is normal,

$$(Y|X = x) \sim N(\beta_0 + \beta_1 x,\ \sigma^2).$$



Simple Linear Regression Model

Equivalently, the SLR model asserts the values of X and Y for individuals in a population are related as follows

$$Y = \beta_0 + \beta_1 X + \varepsilon,$$

- the value of ε, called the **error** or the **noise**, varies from observation to observation, follows a normal distribution

$$\varepsilon \sim N(0, \sigma^2)$$

- In the model, the line $y = \beta_0 + \beta_1 x$ is called the **population regression line**.

Data for a Simple Linear Regression Model

Suppose the data comprised of n individuals/cases randomly sampled from a population.

From case i we observe the response y_i and the predictor x_i:

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)$$

The SLR model states that

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

How do we estimate intercept β_0 and the slope β_1?