Regression Examples

Marius Ionescu

09/29/2011

Example

Example

We study IQ versus Math SAT score. Our group has an average IQ of 100 with a SD of 15, and obtained an average SAT of 550 with SD of 80. We calculated the correlation coefficient to be $r = .6$ and found that the scatterplot was football-shaped.

- If a student scores a 150 on the IQ test, what do you estimate for their SAT score?
- If a student scores 710 on SAT, what do we estimate for their IQ?

RMS error and SD

Fact

Suppose that x and y are normally distributed and linearly correlated (they form a football shaped date). Then

- *For each x, the strip above x is normally distributed.*
- *The average is the predicted value of y*
- *SD equals RMS error.*

Example

Example

Suppose that the average height of men is 68 inches with SD$=$ 2.7, and the average weight of women is 63 inches with SD$=$ 2.5. Assume that the correlation between the height of husbands and wifes is 0.25 and assume that the data is normally distribute.

- What percentage of women are over 68 inches?
- What percentage of women married to a men of height 72" are over 68"?
- What percentile of women married to a men of height 72" is 68"?

Example

Suppose that in this class the average of the first midterm will be 72, SD= 20, the average of the final exam will be 75, SD= 10, and the scores have a correlation coefficient of $r = 0.6$.

- Of all those who score in the 68^{th}%ile on the midterm, what is the expected outcome on the final (with error estimates)?
- Of those who score with 90^{th}%ile on the midterm, what percentile is expected for the final.
- Of those who score 75 on midterm what is the percentile we expect for the final?
- If you score 58 on the midterm what do you expect for the score on the final (with error estimates)
- If you score 70.8 on the final what is your expected score on the midterm?

Example

Example

Suppose that the average of violent crimes during night in a city is 235, SD= 34, the average night temperature is 60, SD= 7, and the correlation between crimes and temperature is 0.5.

- If the temperature is 74, how many violent crimes are predicted tonight?
- If there were 200 crimes last night, how hot was it?
- What is the RMS error for these predictions?
- If tonight is hotter that 68% of all nights, what is the predicted percentile for crimes?

Nonlinear relationships

Fact

What happens if the relationship between the variables is not linear?

- *For exponential relationships*

$$y = Ae^{Bx}$$

we can use linear regression with $\ln y$ *and* x*:*

$$\ln y = \ln A + \ln e^{Bx} = \ln A + Bx.$$

- *For parabolic data we can try a general quadratic:*

$$y = Ax^2 + Bx + C,$$

where we have two variables: x^2 *and* x*.*