This Time: Final Exam Review

↓

overflow time: Sat. June 4 - 12-2 pm

6-2-05

Dana Morton

Final Idea on Regression p123 #4

variable	mean	SD	
(H) ed level (y)	12 yrs	3 yrs	$r = +0.5$
(W) ed level (x)	12 yrs	3 yrs	

Q1: (W) has 18 yrs of schooling → $\frac{18-12}{3} = 2$ SD's above avg.

- for (W) → we expect her (H) is only $r \cdot 2$ SD's above avg. in ed level.

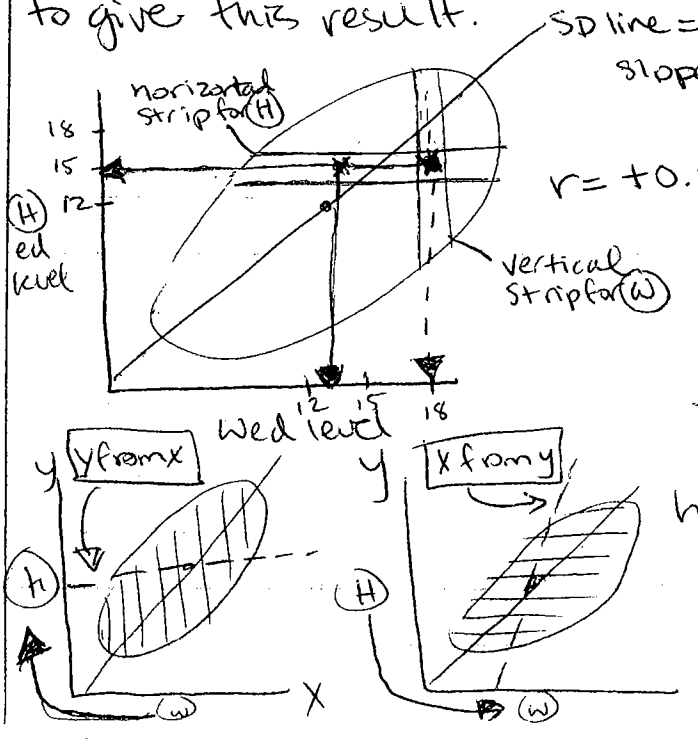
$(0.5)(2\ SD) = 1\ SD \rightarrow$ 15 yrs

Q2: (H) has 15 yrs of schooling → $\frac{15-12}{3} = 1$ SD above avg. for (H).

- we expect his (W) to be only $r \cdot 1\ SD = (0.5)(1\ SD)$ above avg → 1.5 yrs above avg. = 13.5 yrs of schooling predicted for (W).

Q3: What? this doesn't make sense.

- Draw scatter plot to clarify what happened to give this result.



- when predicting y (H) from x (W), use vertical strip.
- But, when predicting x (W) from y (H), we use a horizontal strip.
- The regression line changes!

Key Point

- In any scatter plot, there are actually 2 regression lines that are different.
  - predicting x from y: slope of regression line is greater than slope of SD line; avg of horizontal strips.
  - predicting y from x: slope of regression line is less than slope of SD line; avg of vertical strips.

Moral: If somebody switches the role of x and y in the middle of the problem, watch out!

Final Exam:

Final is open book, open notes.

- 5 problems which consist of (in no particular order):
  - correlation and regression
  - probability models for sums
  - 2 independent samples; continuous data
  - 1 sample problem w/ 0 & 1 data
  - 2 sample paired comparison.

* somewhere in 1 of these problems:
  - will be asked either to do inference or to explain why doing inference would be inappropriate.

#1 on practice final is identical to #5 on midterm #1. (escalator problem)

Practice Problem #5

a) $y = 66.75\text{ in} + (0.25\frac{\text{in}}{\text{yr}})(x)$ ← ed level

↑ ht. for a guy w/ highschool education; $x = 12$