There are actually 2 different reg lines in any scatter plot: one for predicting y from x, & a different line for predicting x from y. novel: if somebody switches the role of x & y in the middle of a problem, watch out.
final is open-book, open notes; 5 problems, in no particular order:

- correlation & regression
- probability models for (residuals, problem sums)
- 2 indep samples (continuous data)
- 1-sample problem with 0/1 data (Spock)
- 2 sample paired comparison

somewhere in 1 of these problems:
you will be asked either to do inference or to say why doing inference would be inappropriate

\[ y = 66.75 + (0.25 \frac{\text{in}}{\text{yr}})(x) \]

let: for a guy with high school ed, 
\[ x = 12 + y = 66.75 + (0.25 \frac{\text{in}}{\text{yr}})(12 \text{yr}) \]

= 69.75 in
\[ y = 6.75 \pi + (0.25 \frac{\text{in}}{\text{yr}})(16 \text{ yrs}) \]

\[ = 70.75 \text{ in} (1 \text{ inch taller}) \]

Q: What does this mean? \[ \frac{\text{d}y}{\text{d}x} = \frac{0.25}{0.25} \]

Going back to college, seeing if the guy gets taller.

Longitudinal question but data cross-sectional.

A1: Slope come out pos \( \rightarrow \) corr. is pos \( \rightarrow \)

PCF: Income (income \( \uparrow \) height \( \uparrow \) (nutrition))

A2: If we look at 2 groups of men who differ in age by 1 yr of ed.

we can expect them to differ in age by 0.25 in. in ht. \( (b) + \text{cm due to } \)

better nutrition or child (might get a raise with elevator shoes, but simpler explanation of this corr. is PCF (nutrition))
Sample observed people
1 = F
0 = M
possible p

N = huge
mean p = 53%

1s
4 0s

gender

mean p = 53%

N = 350

mean p = ?

51.9%
56%

E\text{IID}(\bar{p}) = p

SE\text{IID}(\bar{p}) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.53 \times 0.47)}{350}} = 0.0227 = 2.27%

actual % = \frac{102}{350} = 29%

29% - 53% = 2.7% - 8.8%

change (\text{F}) = 0%
3. It always makes sense to think about whether a difference is large in practical terms.

\[
\frac{12.5\% - 9.8\%}{9.8\%} = \frac{2.7\%}{9.8\%} = 0.28
\]
i.e. U.S. had 28% more elderly people in 1990 than in 1970 (huge practical difference)

1970

1 = old

imaginary data

\[ N = \left[ \begin{array}{c} \frac{1}{2} \times 10^5 \\ \text{millions} \end{array} \right] \]

\[ \text{mean } p = 9.8\% \]

\[ \frac{1}{2} \times \frac{1}{0.5} \text{ million} \]

\[ \text{mean } p = 9.8\% \]

\[ \text{pop - sample so inference (stats) is silly} \]