Class Notes

this time: regression
next time: Final Exam Review (4pm-6pm)
HW 6 due Mon at Final Review
-At final review (Mon 4pm-6pm) Draper will tell exactly what will be on the final

ECS 18.1

\[ SE(\hat{y}) = \sigma \sqrt{1-r^2} \]

is only accurate within data range
as far as \( x \) is concerned, not necessarily accurate beyond.

Here in ECS18 for a guy who is 70.5 inches tall, we predict his weight to be \( \hat{y} = 167 \) lb, give or take

\[ SE(\hat{y}) = \sigma \sqrt{1-r^2} = 25.16 \sqrt{1-(0.76)^2} = 9.1 \text{ lb} \]

- regression prediction not very accurate because correlation

95% CI: 167 \pm 2(9.1) = 167 \pm 18.2 \text{ lb} : 95% of guys with height 70.5 in will have weight from 148.8 to 185.2 lb.

Ep 134 - CS 19.1 - Health Examination Survey

- there are two ways to collect correlation and regression data
  1) Cross sectional: take a snapshot of lots of people at one moment in time (This is what they did)
  2) Longitudinal: follow people along and get data at multiple time points

- It is hard to draw valid longitudinal inference from cross sectional data because of PCBs as cross sectional data is typically observational
Here, the 70 year olds in 1960 were born in 1890, and the 25 year olds in 1960 were born in 1935. The difference in childhood nutrition in childhood (PCF)

C187-1417 - Interpreting Corr and Regr results

\[ y = \hat{a} + \hat{b}x \]

- 2 other ways to gather Corr and regression data:
  1) controlled experiment ( longitudinal )
     - control X
     - see how y changes as we change x
  2) observational study (cross sectional)
     - just take a crosssectional snapshot of x & y
     and see how they seem to be related

A: So no, men should NOT expect to get 3 inches shorter after age 30. We were being tricked by the Age-Peiod Cohort effect.

- just because there is correlation, there may not be cause-effect.
\[ \text{Variable} \quad \text{mean} \quad \text{SD} \]
\[
\begin{array}{lcr}
\text{ed level (x)} & 14 \text{ yrs} & 3 \text{ yrs} \\
\text{Income (y)} & $8000 & $3000 \\
\end{array}
\]

\[ n = 144 \quad r = 0.4 \]

Let's predict income from ed level:

\[
\hat{y} = \hat{b}_0 + \hat{b}_1 x = r \frac{\sigma_y}{\sigma_x} = (+0.4) \frac{5000}{3 \text{ yrs}} = \$400/\text{yr of education}
\]

Q: If a woman took a leave of absence and go to school for a year, could her income go up $400?

A: We cannot draw this longitudinal inference from this cross-sectional data, so no.

Q: What does the $400/year of education actually mean?

A: Associated with each extra year of education, there is an increase of $400 in income, on average.