8.2 and 8.3 - Linear models

(in the context of "best fit lines" for bivariate data)

HW 8.2 #39, 40, 47, 65
HW 8.3 #21, 33, 35, 41, 67, 70

Present Thursday:
8.2 #47c, and 8.3 #41, 70*
(in #70, interpret the slope as a rate of change.)

9:55 Kelly Buc., Timothy, Cory, Kaitlin, Amber, Alyssa Fro.
11:00 Melissa Hla., Samantha, Courtney, Jennifer, Taylor, Elizabeth Lin.

---

pooled st. dev for the mean difference:

$$S = \frac{\left(\frac{S_L + S_R}{2}\right)}{\sqrt{n}}$$

where n is the number of "drops" conducted per hand.

Confidence interval: $\bar{x} \pm z^*(S)$

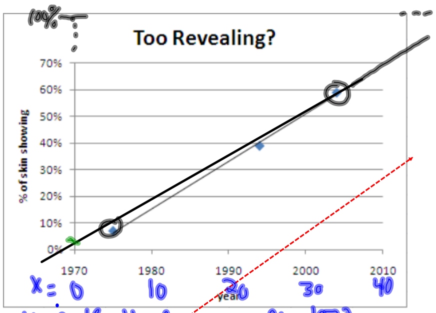
Statistics and Probability 8.SP

1. Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.
2. Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.
3. Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept. *For example, in a linear model for a biology experiment, interpret a slope of 1.5 cm/hr as meaning that an additional hour of sunlight each day is associated with an additional 1.5 cm in mature plant height.*
4. Understand that patterns of association can also be seen in bivariate categorical data by displaying frequencies and relative frequencies in a two-way table. Construct and interpret a two-way table summarizing data on two categorical variables collected from the same subjects. Use relative frequencies calculated for rows or columns to describe possible association between the two variables. *For example, collect data from students in your class on whether or not they have a curfew on school nights and whether or not they have assigned chores at home. Is there evidence that those who have a curfew also tend to have chores?*

Linear Models

Interesting research: In 2004, researchers at Odeon Cinemas determined that celebrities at award shows expose an average of 59% of their skin. That's up from 39% in 1994, and up from 7% in the 1970s. ?

1975



Let 0 represent the year 1970.

x is the # of years after 1970.

Interesting research: In 2004, researchers at Odeon Cinemas determined that celebrities at award shows expose an average of 59% of their skin. That's up from 39% in 1994, and up from 7% in the ~~1970s?~~ 1975

Find a *linear* equation $y = mx + b$ to model this relationship. (We refer to this equation as a "linear model" for the relationship.)

1. Choose two points to compute the slope m. (Choose thoughtfully!)

$$m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{59-7}{34-5} = \frac{52}{29} = 1.793\ldots$$

2. Using the value for m and any appropriate point (x, y), use algebra to solve the equation $y = mx+b$ for b.

$(34, 59)$ $(5, 7)$ (x, y)

$$y = 1.79x + b$$
$$7 = 1.79(5) + b$$
$$7 = 8.95 + b$$
$$-8.95 \quad -8.95$$
$$-1.95 = b$$

$$y = 1.79x + b$$
$$59 = 1.79(34) + b$$
$$59 = 60.86 + b$$
$$-1.86 = b$$

$$-1.9 = b$$

Interesting research: In 2004, researchers at Odeon Cinemas determined that celebrities at award shows expose an average of 59% of their skin. That's up from 39% in 1994, and up from 7% in the 1970s.

3. If the trend continues at this rate, the researchers say, movie stars will be completely naked by... what year??

$$y^{\%} = 1.79x^{(yrs)} - 1.9^{\%}$$
$$100^{\%} = 1.79x - 1.9$$
$$101.9 = 1.79x \Rightarrow \frac{101.9}{1.79} = x \approx 57 \text{ years after 1970}$$

So in 2027, they will be nude!?

This shows the danger of extrapolating beyond the available data.

4. Interpret the slope as a rate of change.

$$m = 1.79 \, \%/\text{yr} = \frac{\Delta y}{\Delta x} \, \frac{\%}{\text{yr}}$$

For each additional year after 1970, the % skin showing increases by ~1.79%.

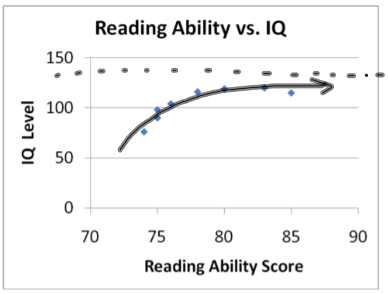
5. Interpret the y-intercept; is it meaningful in this case?

$b = -1.9^{\%}$ in year 1970 $(x=0)$.

Not meaningful in this context.

Linear Model, or Nonlinear Model?

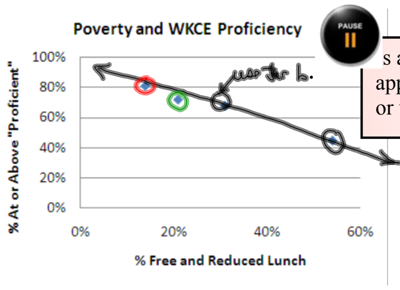
A linear model is not always appropriate. For instance, consider the (hypothetical) relationship between reading ability and IQ shown here.



Finding a Linear Model

Here are data on SES and WKCE from four Wisconsin schools.

x Percent Free & Reduced Lunch:	31	54	21	14
y Percent Above Proficient on WKCE:	68	45	72	81



s a linear model appropriate here? Why or why not?

Finding a Linear Model

Here are data on SES and WKCE from four Wisconsin schools.

x	Percent Free & Reduced Lunch:	31	54	21	14
y	Percent Above Proficient on WKCE:	68	45	72	81

1. Find a linear model for this data (choose points thoughtfully).

$y = mx + b$

$m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{68-45}{31-54} = \boxed{-1}$

$y = (-1)x + b$

$68 = (-1)(31) + b \rightarrow \boxed{b = 99}$

$y = -x + 99$

2. Interpret the slope as a *rate of change*.
"For every 1 point increase in the $\underbrace{\text{FRL rate}}_{\text{x-value}}$, the percent proficient..." decreases by 1 percentage point.

Finding a Linear Model

Here are data on SES and WKCE from four Wisconsin schools.

Percent Free & Reduced Lunch:	31	54	21	14
Percent Above Proficient on WKCE:	68	45	72	81

3. Predict the proficiency level at a school with a 40% FRL rate.

(Un)related question: should teacher compensation be tied to student achievement on standardized tests?

4. Predict the WKCE proficiency level in an affluent district where no students qualify for a free or reduced lunch.
(Is the y-intercept meaningful in this case?)

Statistics and Probability **8.SP**

3. Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept. *For example, in a linear model for a biology experiment, interpret a slope of 1.5 cm/hr as meaning that an additional hour of sunlight each day is associated with an additional 1.5 cm in mature plant height.*

Each student is given a bean sprout to care for, and each decides how many hours per day to let their plant out of the dark closet where they are kept. Plants are measured once at the end of 8 weeks.

A linear model obtained from the class data might be:
$y = 1.5x + 4.5$,
where y is the final height in centimeters and x is the number of hours of daylight the plant received per day.

Interpret the slope, and interpret the y-intercept.

Statistics and Probability **8.SP**

3. Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept. *For example, in a linear model for a biology experiment, interpret a slope of 1.5 cm/hr as meaning that an additional hour of sunlight each day is associated with an additional 1.5 cm in mature plant height.*

Each student is given a bean seed to plant and care for. The bean plants are placed near a window and are measured once each week for 8 weeks.

A linear model obtained from the class data might be:
$y = 3x - 2.4$,
where y is the height in centimeters after x weeks of growth.

Interpret the slope, and interpret the y-intercept.