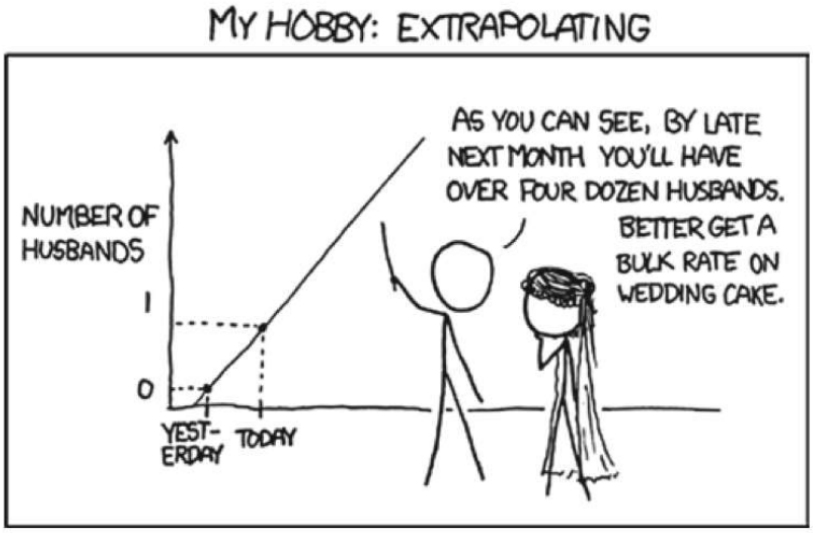
Extrapolation: Reaching Beyond the Data

- Linear models give a predicted value for each case in the data.
- We cannot assume that a linear relationship in the data exists beyond the range of the data.
- The farther the new *x* value is from the mean in *x*, the less trust we should place in the predicted value.
- Once we venture into new *x* territory, such a prediction is called an extrapolation.

Extrapolation (cont.)

- Extrapolations are dubious because they require the additional—and very questionable — assumption that nothing about the relationship between *x* and *y* changes even at extreme values of *x*.
- Extrapolations can get you into deep trouble. You're better off not making extrapolations.

Extreme Extrapolation!

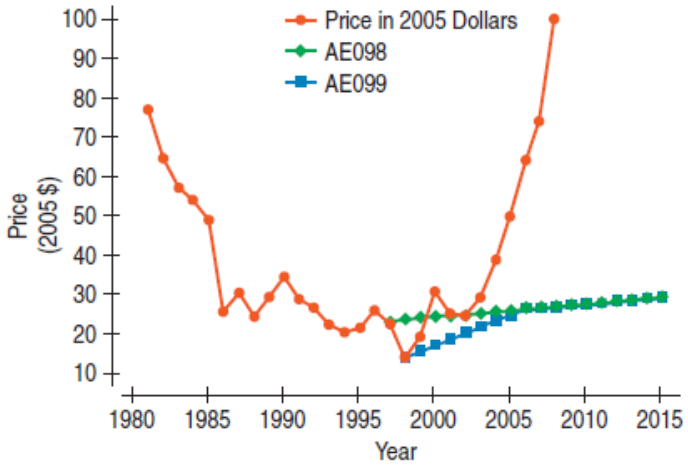


Text pg. 179

Extrapolating from current trends is so tempting that even professional forecasters make this mistake, and sometimes the errors are striking.

In the period from 1982 to 1998 oil prices went down so much that by 1998, prices (adjusted for inflation) were the lowest they'd been since before World War II. Of course, these decreases clearly couldn't continue, or oil would be free by now. The Energy Information Administration offered two *different* 20-year forecasts for oil prices after 1998, and both called for relatively modest increases in oil prices. So, how accurate have these forecasts been? Here's a timeplot of the EIA's predictions and the actual prices (in 2005 dollars).

Oops! They seemed to have missed the sharp run-up in oil prices in the early 2000s.



Homework:

Textbook Pg. 189 #32

Do not draw the scatterplot. Just look at it in the calculator.

Describing slope and y-intercept *in context*

The best fit line is only a model of predicted results based on the data provided. When describing the slope and y-intercept in context be careful how you word it.

Example: $\widehat{Weight} = -149 + 4.4Height.$

Describe the y-intercept and slope:

It does not make sense to say that a person with height of 0" would weight -149 pounds. It is not totally accurate to say that a 1" increase in height causes a 4.4 pound increase in height. (Linear equation does not prove causation, nor is it an absolute truth). It is just a model based on a set of data.

$$\widehat{Calories} = 89.5 + 2.50\,Sugar.$$

Describe the slope and y-intercept.

The intercept of our model predicts that a zero-sugar cereal will have on average 89.5g ~~of sugar~~.

The slope of our model equation suggests that cereals have about 2.50* more calories per additional gram of sugar.

* Actually, 1 gram of sugar equals to 3.90 calories, but our model shows otherwise.

Homework:

Making Sense of Linear Models (finish)

Textbook pg.185-187 # 1 - 5, ~~17~~ 32 [a) look at scatterplot on calc)

Resid = actual − predicted

*Name*_____ *Statistics Chapter 7: Making Sense of Linear Models*

Context	Linear Model Describe slope and intercept in context	Calculate and explain residual for the given data
1. The price (in thousands of dollars) and size (in thousands of square feet) of houses	**Model:** $\widehat{Price} = -3.117 + 94.454Size$ **Slope:** Per our model, each additional 1,000 square feet of house, would result in a price increase of $94,454. **Intercept:** Per our model, a house with 0 sq. ft. would cost −$3,117 (which doesn't make sense)	(2, 195) The model predicts a price of 3.117 + 94.454(2) = $192,025 for a 2,000 sq ft house. This prediction is 195,000 − 192,025 = $2,925 too low.
2. The number of wildfires (in thousands) and years since 1982	**Model:** $\widehat{Fires} = -3.4556year + 153.837$ **Slope:** **Intercept:**	(20, 92)
3. Fiber (in grams) and potassium content (in milligrams) of servings of breakfast cereals.	**Model:** $\widehat{Potassium} = 38 + 27Fiber$ **Slope:** **Intercept:**	(3, 88)
4.	**Model:** **Slope**: For each additional unit of horsepower, a car's mileage is expected to decrease by 0.084 mpg **Intercept**: A car with no horsepower is predicted to have gas mileage of 46.87 mpg. (This is not practical!!)	(150, 36)
5. Predicting the length of an athlete's long jump using the length of his/her high jump (both in meters).	**Model:** **Slope**: 4.20053 in context means… **Intercept**: 1.10541 in context means……	(2.2, 7.1)