Week 2: Discussion

David Diez

Assignments/Labs in general :: please include both first and last name along with your section number (8am or 9am works).

Lab 1 Recap :: some comments:

- Recall that the data is from *samples*, not populations. (There were more than 750 and 1500 people in CA in 1880 and 1990, respectively.) We have to be cautious to only look at proportions, not counts, for trends.
- Example on why proportions is important and not counts. The number of supporters of Barack Obama or Hillary Clinton in polls taken over successive days could shoot from 500 to 1500. This doesn't mean he/she has 3 times as much support – we have to look at the proportions (because the sample sizes may have been 1000 and 3000, meaning the proportion didn't appear to change).
- Be certain to describe what is happening and why.
- Only need to turn in 'on your own' sections.
- All labs must be typed.
- Printing on paper that already has one side used is okay with me.

Example with mean and standard deviation :: Find the mean and standard deviation of $\{0, 1, 4, 4, 4, 5\}$. Mean: 3. Standard deviation: 2.

Boxplots ::

- We define an outlier to be a value that falls outside of $Q_1 - 1.5 * IQR$ and $Q_3 + 1.5 * IQR$. Note: this definition is arbitrary and is not perfect.
- Collapsed boxplots occur when there is very little data or the first or third quantile equals the median. This will look very odd. Ex, boxplot of the example above has its 5 number summary as $(0, 1, 4, 4, 5)$, corresponding to the $(min, Q_1, median, Q_3, max)$.

Independence :: Two variables or events are independent if there is no association between them. That is, knowing the state or outcome of one variable or event gives no useful information in determining the other variable/event state/outcome. An example of two variables that are **not** independent are education level and income. If we know an individual has a high level of education, we would think her income would also be higher (this isn't true in all cases, of course, but it is 'on the average').

Independence in a table :: The following data was presented by Ben Franklin and regards small pox inoculation. Is survival independent of inoculation? Column or row proportions are an easy and straightforward method for checking independence in a table.

Inoculated	Lived	Died	
Yes	1601	6	1607
No	3856	1002	4858
	5457	1008	6465

If the two variables are independent, then we would expect the same proportion of people who were inoculated to live as who lived that were not innoculated. This is just a way to say, if they are independent, then the chance of surviving should be the same regardless of whether an individual was inoculated or not. The row proportions for living are $\frac{1601}{1607} = 0.996$ and $3856/4858 = 0.794$. Because these chances for survival as not close (especially with so many samples), the inoculation and survival are not independent. If you lived in 1775, would you rather be inoculated or not?

Review quiz :: Run over each problem briefly and some in detail.

p123, #18 :: IQ test scores are normally distributed with mean 100 and standard deviation 16. (a) Draw the model for the IQ scores and label the 68-95-99.7 areas. (b) skip! (c) About what percent of people have IQ scores above 116? (d) About what percent of people have scores between 68 and 84? (e) Skip!

p125, #37 (tangent) :: Angus weights are normal with mean 1152 and standard deviation 84.
(Q1) How many st. dev.'s from the mean would a steer weighing 1000 pounds be?
(Q2) Which would be more unusual, a steer weighing 1000 pounds, or one weighing 1250 pounds?
(Q3) What is the cutoff for the 90th percentile for the steer?
(Q4) What percent of steer are smaller than 1100 pounds?
(Q5) What percent of steer are between 1000 pounds and 1100 pounds?

Discuss how to solve problems using the normal model:
data $(x, \mu, \sigma) \Leftrightarrow Z \Leftrightarrow$ tail probability.