

## Statistics 1211 Spring 2008 HW1

Due in class or in my mailbox before class on Feb.4

### Section 1 (credit/no credit):

From the first lecture handout, the last histogram on the left panel gives the distribution of indurations of tuberculin skin tests of 24 year olds sampled from the population. From the picture, form a guess at the underlying distribution of indurations for both the vaccinated and non-vaccinated populations; ie., .18 · 441 people of the vaccinated people had indurations between 12 and 14 mm. Do your best to make this guess at the underlying distribution, it does not need to be exact. From now on,  $n$  is the total number of counts, so  $n = 441$  for the vaccinated and  $n = 176$  for the not vaccinated populations.

- Compute the mean induration size of both the vaccinated and non-vaccinated populations.
- The mean is one measure of the center of a distribution; the median is another measure, in some ways preferable. From a sample

$$x_1, x_2, \dots, x_n,$$

the median (denoted  $\tilde{x}$ ) is defined by ordering the data in increasing order, and taking the midpoint. Formally, if

$$x_{(1)}, \dots, x_{(n)}$$

are the ordered data points so

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)},$$

which are just a rearrangement of the original  $x_i$ .

If  $n$  is odd,

$$\tilde{x} = x_{(\frac{n+1}{2})}$$

and if  $n$  is even,

$$\tilde{x} = \frac{1}{2} \left( x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right).$$

Calculate the medians for the two populations, vaccinated and not vaccinated.

- A measure of spread of the data is  $s^2$  where

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Calculate  $s^2$  for each distribution.

d) Later, we will use the  $t$  statistic to make formal tests about the center of a distribution. This statistic is defined as

$$t = \frac{\bar{x}}{\sqrt{s^2}}.$$

Calculate the  $t$  statistic for each of the two populations, and think (but no need to answer) about why it is a good measure of a distribution.

**Section 2 (graded) from Devore, 7th edition:** Exercises

2.1.5, 2.2.20, 2.4.55, 2.4.64, 2.5.73, 2.5.84.

**Section 3 (credit/no credit, due with HW2):** Complete the tutorial on R at

[www.stat.columbia.edu/~jsalzman/1211/Rtutorial.html](http://www.stat.columbia.edu/~jsalzman/1211/Rtutorial.html)

When there are points in the tutorial with exercises to complete, do them and hand them in.