1 · *Explanation and evidence*

Due Monday, January 26, 2015

(1) Assessing causal claims

Read the article entitled "Contraceptive Used in Africa May Double Risk of H.I.V.," published in the *New York Times* on October 3, 2011.[1]

If you are not an online subscriber to the New York Times, and have run out of your free page views this month, you may retrieve the article through UT's LexisNexis Academic subscription, available through the UT library website.

Write a summary of the article, no more than one single-spaced page in length, that addresses the following questions:

1. What the primary causal claim made by the researchers who conducted the original study?
2. Are there any issues of confounding and endogeneity that the authors of the original study would have had to consider in making this causal claim? In your estimation, does the author of the newspaper article do a satisfactory job of discussing these issues?
3. Do you see anything else in the newspaper article, whether positive or negative, worth commenting on?

[1] `http://www.nytimes.com/2011/10/04/health/04hiv.html`

(2) The sample mean

Suppose we have a data set with N observations $y_1, \ldots, y_N$, and we want to summarize these N numbers with a single number $\hat{\theta}$ that represents the center of the data set. A standard choice is the sample mean: $\hat{\theta} = \bar{y}$. In this question, you'll provide a justification of this choice.

Consider the following function of θ:

$$f(\theta) = \sum_{i=1}^{N} (y_i - \theta)^2 \, .$$

This function represents the total "error" for a particular choice for θ. The quantity $(y_i - \theta)$ is the error in using θ to approximate the ith data point. The function $f(\theta)$ just squares these quantities (so that positive and negative errors count equally) and sums them up to measure the total approximation error. Show mathematically that $f(\theta)$ is minimized when you choose θ to be the sample mean:

$$\hat{\theta} = \bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i \, .$$