Lecture 1 - Economics 113

Professor Alan Spearot

September 25th, 2009

Summary measures

Preliminaries

- Σ is shorthand for addition
- Suppose x_i is the *ith* observation.

$\Longrightarrow$

$$x_1 + x_2 + x_3 = \sum_{i=1}^{3} x_i$$

- If a is a constant

$\Longrightarrow$

$$\sum_{i=1}^{3} a = 3a$$

- Mixed example:

$\Longrightarrow$

$$\sum_{i=1}^{3} ax_i = a \sum_{i=1}^{3} x_i$$

Summary measures

Preliminaries

- b a constant
- n observations
- Simplify $\sum_{i=1}^{n}(a+bx_i)$
- Group terms!

$\Longrightarrow$

$$\begin{aligned}\sum_{i=1}^{n}(a+bx_i) &= a+bx_1+a+bx_2+...+a+bx_n \\ &= a+a+a....bx_1+bx_2+...bx_n \\ &= an+b\sum_{i=1}^{n}x_i\end{aligned}$$

Summary measures

Preliminaries (cont.)

- Let $\{x_i, y_i\}$ be paired observations
- k another constant.
- Try this one.

$$\sum_{i=1}^{n} \left(ak + bx_i + c\left(x_i y_i\right)\right) = ???$$

- Treat $x_i y_i$ as any other random variable:

$$\sum_{i=1}^{n} \left(ak + bx_i + c\left(x_i y_i\right)\right) = nak + \sum_{i=1}^{n} bx_i + \sum_{i=1}^{n} \left(cx_i y_i\right)$$

- Pull out the constants:

$$nak + \sum_{i=1}^{n} bx_i + \sum_{i=1}^{n} \left(cx_i y_i\right) = nak + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} \left(x_i y_i\right)$$

Summary measures

- Standard notation

	Estimated from the sample	**Parameter from the population**
Mean	$\widehat{\mu}_x$	μ_x
Variance	$\widehat{\sigma}_x^2$	σ_x^2
Standard Deviation	$\widehat{\sigma}_x$	σ_x
Size	n	N

- We estimate parameters to describe a population using the sample.
- The "hats" represent statistics estimated using the sample.
- What properties should these statistics have?

Summary measures

An example

- We want to estimate the height of all UCSC students.
- Is this expensive?
- Where should we go to sample?
  1. Bus stop?
  2. Bay tree?
  3. Bar?
  4. Office hours?
- Are there problems with these sampling points?

$\Rightarrow$ *Must argue that students at these points represent a random sample of the population.*

Summary measures

Statistics that describe distributions

- Measures of central tendency
  1. Mean
  2. Median
  3. Mode
- Measures of variation
  1. Range
  2. Variance
  3. Standard Deviation
- Measures of shape
  1. "Skew"

Summary measures

Measures of central tendency

- **Mean**
  - Most common
  - Central tendency

$$\widehat{\mu}_x = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- **Median**
  - Mid-point of the data.
- To calculate:
  1. Order the data.
  2. Calculate $(n+1)/2$

     Take the value at $(n+1)/2$, or the average of the two closest (if $(n+1)/2$ is not whole)

Summary measures

Measures of central tendency (cont.)

- **Mode**
  - The value that occurs most often
- Question: which measure(s) of central tendency are not affected by extreme values?

  $\Rightarrow$ *Median and Mode*
- US income distribution?

Summary measures

Measures of variation

- **Range**

  A measure of data dispersion, though not used for many applications

  1. Identify the largest observation
  2. Identify the smallest observation
  3. Take the difference

Summary measures

Measures of variation (cont.)

- **Sample Variance**
  - The most commonly used measure of dispersion.
  - Summarizes the how far a typical observation is from the mean

$$\widehat{\sigma}_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \widehat{\mu}_x)^2$$

- Why do we divide by $n-1$ instead of n?

$\Rightarrow$ *We used one "piece" of information to calculate* $\widehat{\mu}_x$

Summary measures

Measures of variation (cont.)

- **Sample Standard deviation**

$$\widehat{\sigma}_x = \sqrt{\widehat{\sigma}_x^2}$$

- This is more desirable than the sample variance. Why?

$\Rightarrow$ On the same scale as $\widehat{\mu}_x$.