Lecture 4: Poisson Approximation to Binomial Distribution; Measures of Center and Variability for Data (Sample);

Chapter 2

No Lab this week, but…

- Questions in Lab# 2 are related to this week's topics…

- Hw#2 is due by 5pm, next Monday

Poisson Approximation for the Binomial Distribution

- For Binomial Distribution with large *n*, calculating the mass function is pretty nasty
- Good news:

  when $n \to \infty, \quad \pi \to 0, \quad n\pi \to$ a constant λ

  $\text{Binomial}(n, \pi) \to \text{Poisson}(\lambda)$, i.e.:

$$\frac{n!}{x!(n-x)!}\pi^x(1-\pi)^{n-x} \to \frac{e^{-\lambda}\lambda^x}{x!}$$

- So for those nasty "large" Binomials ($n \geq 100$) and for small π (usually ≤ 0.01), we can use a Poisson with $\lambda = n\pi$ (≤ 20) to approximate it!

Example

Suppose 1 in 5000 light bulbs are defective. Let X denote the number of defective light bulbs in a group of size 10000. What is the chance that at least 3 of them is defective?

Check the appropriateness for Poisson approximation:

n ≥100? Yes. $\pi \leq 0.01$? Yes. $n\pi \leq 20$? Yes.

- For $\lambda = n\pi = 2$,

(proportion with $x \geq 3$) = 1 – (proportion with $x < 3$)

$$= 1 - [p(0) + p(1) + p(2)] = \ldots$$

Density (for Continuous) and Mass (for Discrete) functions

- tell you the “chance/proportion/probability” that a variable takes a certain value
  - Need to know the distribution expression

- both used to rigorously describe populations or processes
  - How to know which distribution is applicable?

→See Chapter 2
- Numerical measures for both samples and populations
- **Bring Your Calculator from now on…**

2.1
Measures of Center (Data)

- The sample mean – arithmetic average
- From a sample of *n* observations, x_1, x_2,… x_n, the mean is given by:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum x_i$$

Examples

- Scores for 10 students are:

80 85 81 87 78 82 80 83 85 86

- So, $\sum x_i = 80 + 85 + \ldots + 86 = 827$

$$\bar{x} = \frac{1}{10} \sum x_i = 82.7$$

Means: not resistant to outliers…

- Scores for 11 students are:

80 85 81 87 78 82 80 83 85 86 2

- So, $\sum x_i = 80 + 85 + \ldots + 86 + 2 = 829$

$$\bar{x} = \frac{1}{11}\sum x_i = 75.4$$

- What does this say about the mean?

Measures of Center (Data)

- The sample median – midpoint
- From a sample of *n* observations, $x_1, x_2, \ldots x_n$, the median is given by

1. Order the observations from smallest to largest

$$2.\ \tilde{x} = \begin{cases} (\frac{n+1}{2})\text{th value on the ordered list,} & \text{when } n \text{ is odd} \\ \text{average of } (\frac{n}{2})\text{th and } (\frac{n}{2}+1)\text{th value on the ordered list,} & \text{when } n \text{ is even} \end{cases}$$

- Intuitively, it is the middle observation (in an ordered list)

Examples

- Scores for 10 students are:

80 85 81 87 78 82 80 83 85 86

Step 1: Reordered observations:

78 80 80 81 82 83 85 85 86 87

Step 2: n=10, an even number. so take the average of 5^{th} and 6^{th} observation (in the sorted list)

$$\tilde{x} = \frac{(82 + 83)}{2} = 82.5$$

Median: a more resistant measure of center

- Scores for 11 students are:

80 85 81 87 78 82 80 83 85 86 2

- Reordered,

2 78 80 80 81 82 83 85 85 86 87

$$\frac{n+1}{2} = 6th \text{ position,} \qquad \tilde{x} = 82$$

- What does this say about the median?

Trimmed Means (page 62)

- Rank the observations from smallest to largest, then trim off a percentage from both ends of the data before taking the mean
- So for our example with 11 students:

  2 78 80 80 81 82 83 85 85 86 87

- We could trim off say 1/11 or 9% from each end (just one value)

  78 80 80 81 82 83 85 85 86

- Trimmed mean is 82.222

Review

Up to Now:

- Measure of center of Data (Sample)
  –Sample mean
  –Sample median
  –Trimmed means

Later,

- Measure of variability for Data (Sample)
  –Sample variance
  –Standard deviation

And: Measure of Center for Distributions
 - Population Mean/ Expected Value;
 - Population Median, for continuous distributions.

- *how to measure* ***variability for distributions*** *(Population)*
- ***graphically display*** *both the center and variability of Data (Sample);*

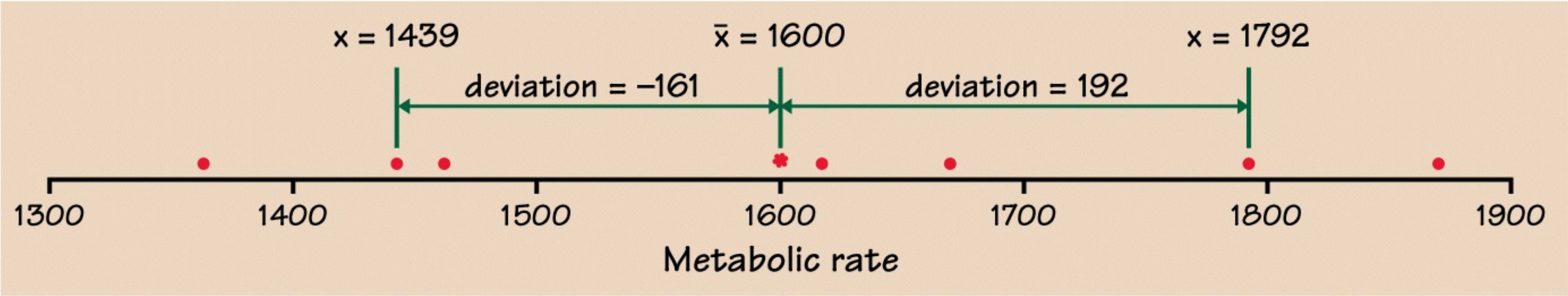
2.2
Standard Deviation for Data

- Deviation : $x_i - \bar{x}$
- Variance : s^2

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n-1} = \frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2$$

- Standard Deviation : s

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2}$$



DATA: 1792 1666 1362 1614 1460 1867 1439
Mean = 1600

- Find the deviations from the mean:
  - Deviation1 = 1792 – 1600 = 192
  - Deviation2 = 1666 – 1600 = 66
  - …
  - Deviation7 = 1439 – 1600 = -161
- Square the deviations.
- Add them up and divide the sum by n-1 = 6, this gives you s^2. n-1: degrees of freedom.
- Take square root: Standard Deviation = s = 189.24

Measures of Variability (Data)

- The sample variance, s^2
  - From a sample of n observations, x_1, x_2,…x_n, the sample variance is given by

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1} = \frac{S_{xx}}{n-1}, \text{ where } S_{xx} = \sum (x_i - \overline{x})^2$$

- Why divide by $n - 1$? From the degrees of freedom
- The sample standard deviation, s
  - Just take the square root of the variance

$$s = \sqrt{s^2}$$

Example

- Scores for 10 students are:

80 85 81 87 78 82 80 83 85 86

- Calculations on board…

  Mean = 82.7,

- OR just plug into calculator $s=2.983287$

After Class …

- Start Hw#2 now
- Review sections 2.1 and 2.2, especially Pg 63 – 68, 74 – 77
- Read section 2.3, self-reading Pg95