

Data 8, Lab 9

Skewness, Normality, and Sample Means

Hubert Luo

Fall 2019

8 November 2019

Agenda

1. Skewness
2. Variability
3. Chebyshev's Bounds
4. Standard Units
5. Normal Distribution
6. Central Limit Theorem
7. Distribution of Sample Means

Skewness

- Left skew
 - Long left tail
 - $\text{Mean} < \text{Median}$
- Right skew
 - Long right tail
 - $\text{Mean} > \text{Median}$

Variability

- Variance: How spread out is the data?
- Standard Deviation: Square root of the variance
 - Same unit as the data
 - The larger the SD, the more spread out the data is

Chebyshev's Bounds

- Regardless of the distribution, the proportion of values in the range "average $\pm z$ SDs" is at least $1 - 1/z^2$

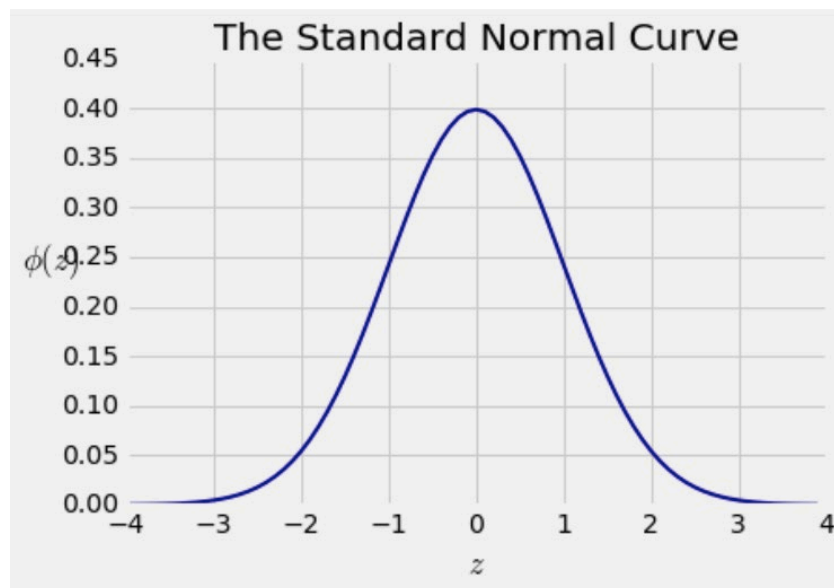
Range	Proportion
average ± 2 SDs	at least $1 - 1/4$ (75%)
average ± 3 SDs	at least $1 - 1/9$ (88.888...%)
average ± 4 SDs	at least $1 - 1/16$ (93.75%)
average ± 5 SDs	at least $1 - 1/25$ (96%)

Standard Units

- Standard Unit: Number of SD's above or below average
- Allows us to easily compare different distributions and units
- $Z = (\text{value} - \text{average}) / \text{SD}$
- Average of standard units is always 0
- SD of standard units is always 1

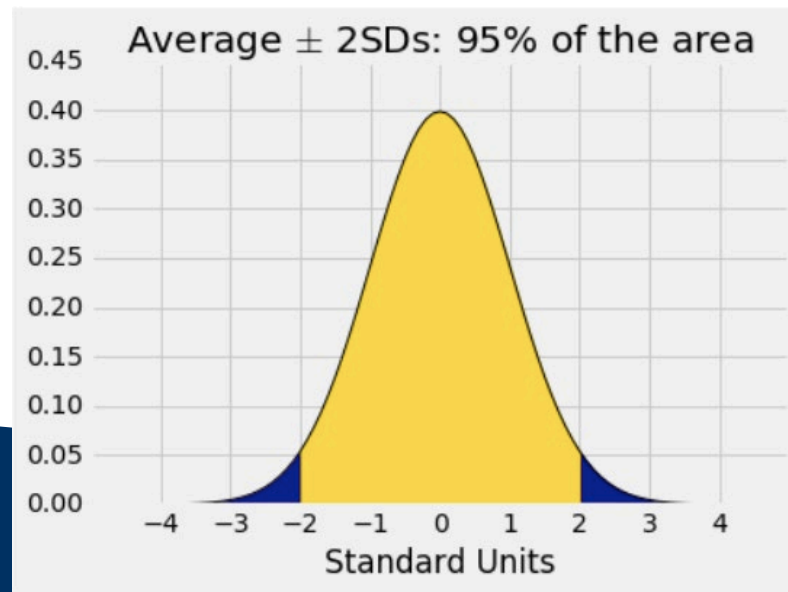
Normal Distribution

- An extremely common distribution in statistics, shaped like a bell curve
- Most of the data is within a few SD's of the mean



Normal Distribution (cont'd)

Range	All Distributions	Normal
average \pm 1 SDs	at least 0%	68%
average \pm 2 SDs	at least 75%	95%
average \pm 3 SDs	at least 88.9%	99.7%



Central Limit Theorem

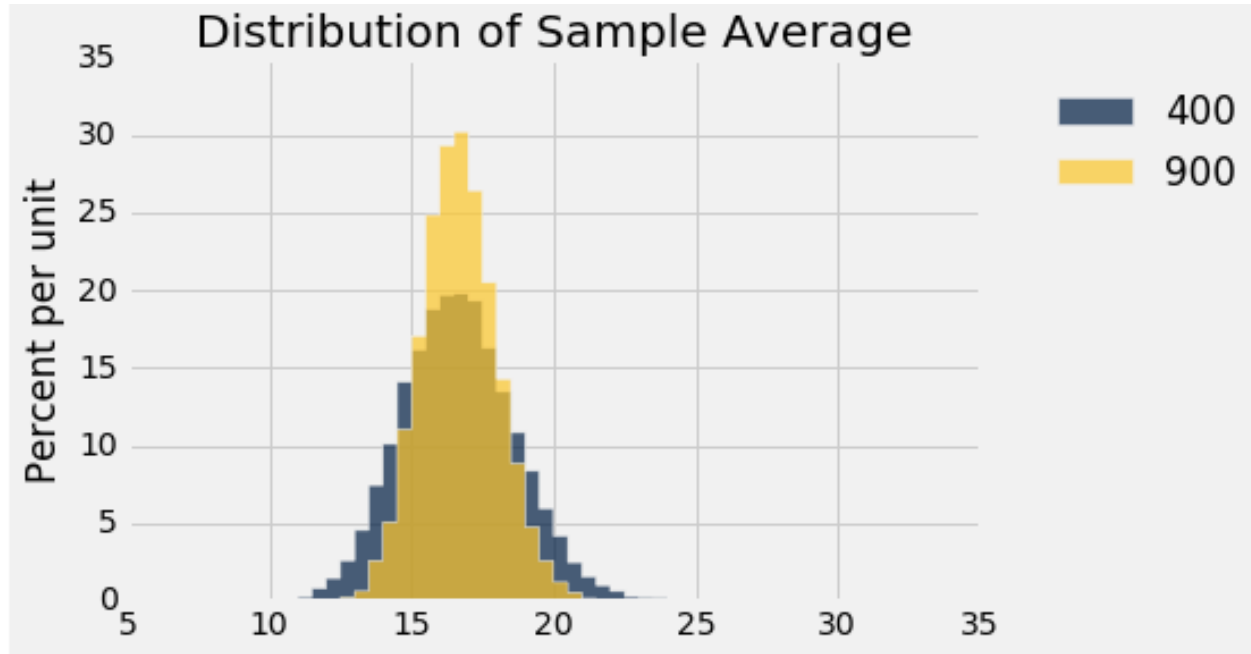
- If the sample is large and drawn at random with replacement
- Regardless of the distribution of the population, the distribution of the sample sum or average is **roughly normal**
- Distribution of the sample sum/average:
 - Many possible random samples of the same size
 - Distribution is based on the sum/average of different samples

Distribution of Sample Mean

- As the sample size increases, the sample mean is more likely to be closer to the population mean
- As a result, the distribution of sample means will have lower SD – a “narrower bell shape” when the sample size increases

$$SD \text{ of Sample Means} = \frac{\text{Population SD}}{\sqrt{\text{Sample Size}}}$$

Distribution of Sample Mean



Announcements

- Checkpoint 2 of Project 2 due today (11/8)
- Homework 10 due next Thursday (11/14)
- No class next Monday (11/11)