Stat 201: Introduction to Statistics

Standard 22 – The Central Limit Theorem

Recall: Law of Large Numbers

- (**LLN 1**) – As the sample size increases the sample estimates ($\bar{x}$ or $\hat{p}$) approach the population values (μ or ρ)

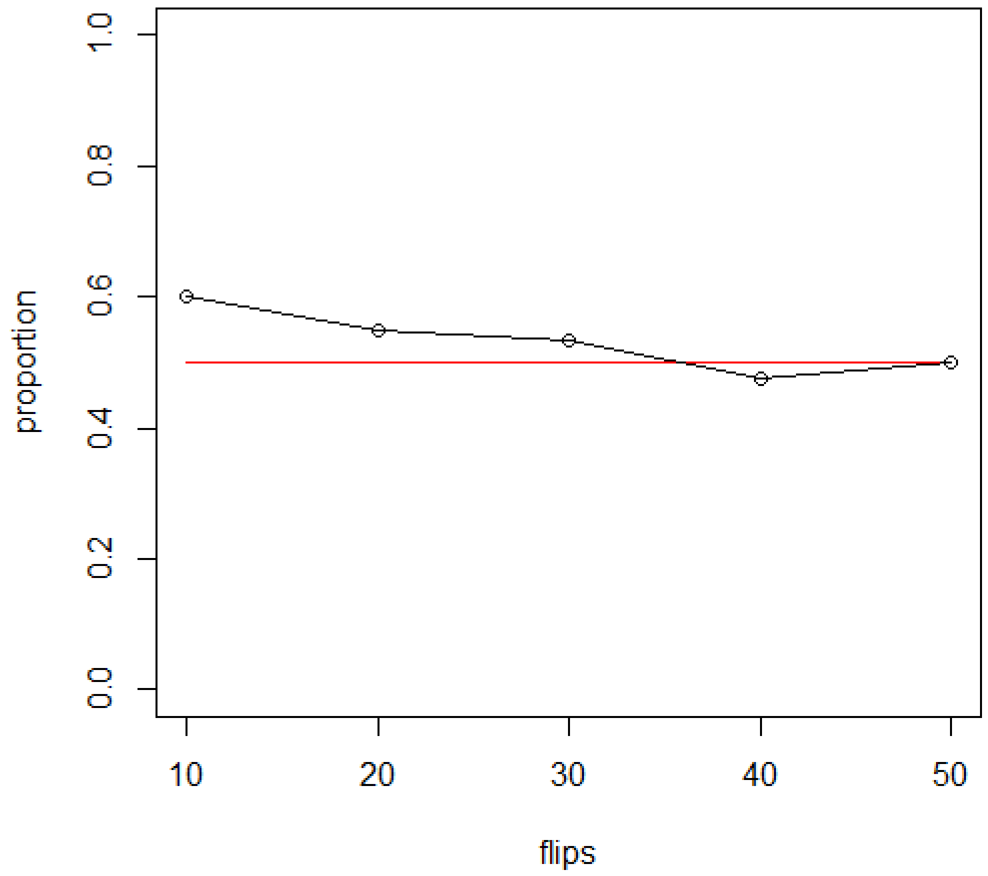
- (**LLN 2**) – As the number of trials increase the **proportion** of occurrences of any given outcome approaches the probability in the long run.

Recall: Law of Large Numbers

- 10 flips: 6 heads were flipped
  - $Total\ proportion\ = \frac{x}{n} = \frac{6}{10} = .60 = 60\%\ heads$
- 10 more flips: 5 heads were flipped
  - $Total\ proportion\ = \frac{x}{n} = \frac{5+6}{10+10} = \frac{11}{20} = .55 = 55\%\ heads$
- 10 more flips: 5 heads were flipped
  - $Total\ proportion\ = \frac{x}{n} = \frac{11+5}{20+10} = \frac{16}{30} = .5333 = 53.33\%\ heads$
- 10 more flips: 3 heads were flipped
  - $Total\ proportion\ = \frac{x}{n} = \frac{16+3}{30+10} = \frac{19}{40} = .475 = 47.5\%\ heads$
- 10 more flips: 6 heads were flipped
  - $Total\ proportion\ = \frac{x}{n} = \frac{19+6}{40+10} = \frac{25}{50} = .5 = 50\%\ heads$

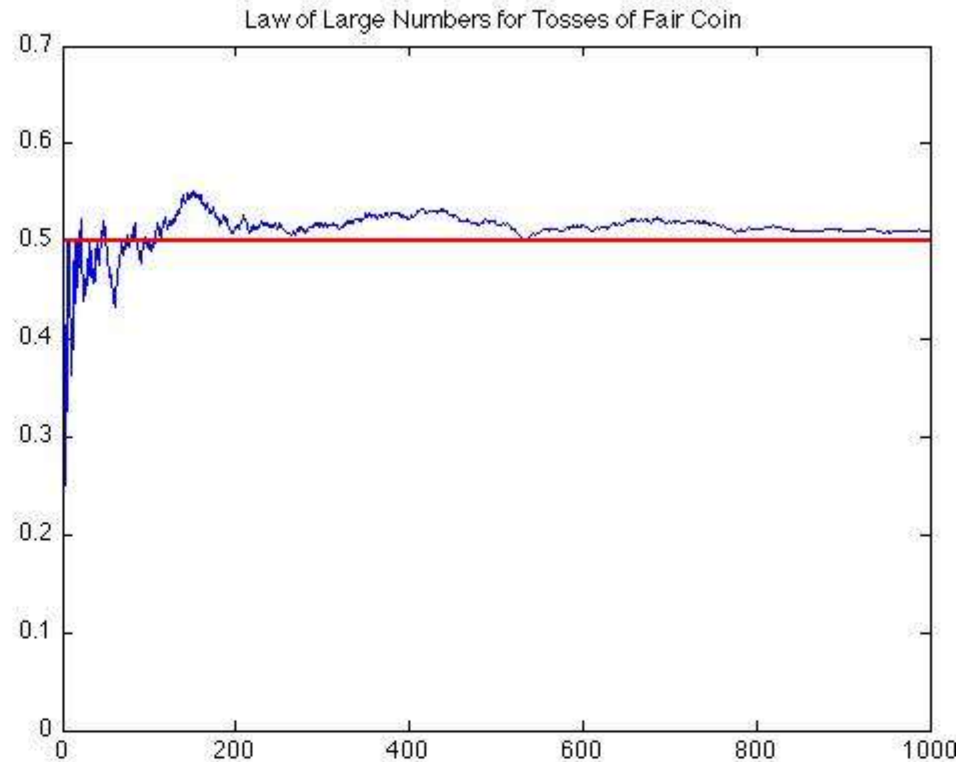
Recall: Law of Large Numbers

- (**LLN**) – As the number of flips increase the **proportion** of heads approaches the **probability** of seeing a heads, P(heads)=.5, which is the red line.



Recall: Law of Large Numbers

- At first the proportion is all over the place – you can see the large spikes in the graph
- Importantly, we see that the proportion of coins that landed on heads levels off and gets closer and closer to 50%, the probability, which is where we expect it to go ‘in the long run!’



Recall: Law of Large Numbers

- In January the Powerball lottery jackpot hit \$800M, which was a record at the time.
- The expected return for each \$2.00 Powerball ticket for this drawing was \$1.06
- What the **Law of Large Numbers** tells us is that we shouldn’t play Powerball. Sure, someone is going to win the jackpot but if you were going to buy tickets every drawing for the rest of eternity it will likely be a losing endeavor.

Recall: Law of Large Numbers

- The idea here is that spending \$2.00 on a Powerball ticket when getting gas for the novelty of playing when the jackpot is so high could turn out well, at least in the short run
- The **Law of Large Numbers** tells us that as the number of independent trials (ticket purchases) increases the average outcome will approach the mean of \$1.06 resulting in an average loss of \$0.96 per ticket.

From *Naked Statistics:* The Central Limit Theorem

"The central limit theorem is the Lebron James of statistics – if Lebron were also a supermodel, a Harvard professor and the winner of the Nobel Peace Prize… and until Lebron James wins as many NBA championships as Michael Jordan (six), the central limit theorem will be far more impressive than he is"

From *Naked Statistics:* The Central Limit Theorem

"The central limit theorem is the "power source" for many of the statistical activates that involve using a sample to make inferences about a large population."

"The core principle underlying the central limit theorem is that a large, properly drawn sample will resemble the population from which it is drawn. Obviously there will be variation from sample to sample, but the probability that any sample will deviate massively from the underlying population is very low."

From *Naked Statistics:* The Central Limit Theorem

"If we have detailed information about a properly drawn sample (mean and standard deviation), we can make strikingly accurate inferences about the population from which that sample was drawn"

This is the avenues we will take for the rest of the semester.

Central Limit Theorem: Proportions

- For random sampling with a **large sample size n, the sampling distribution of the sample proportion** is approximately a normal distribution
  - $n * p \geq 15 \; and \; n * (1 - p) \geq 15$

- Introduction:
  - https://www.youtube.com/watch?v=PujoI1yC1_A

Central Limit Theorem: Means

- For random sampling with a **large sample size n, the sampling distribution of the sample mean** is approximately a normal distribution
  - For us, 30 is close enough to infinity

- Introduction:
  - https://www.youtube.com/watch?v=Pujol1yC1_A

Central Limit Theorem: Means

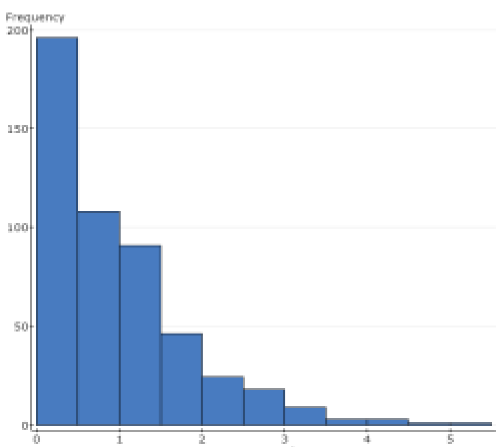
1) For any population the sampling distribution of $\bar{x}$ is bell shaped when the sample size n is large, when n is thirty or more
2) The sampling distribution of $\bar{x}$ is bell-shaped when the population distribution is distribution is bell-shaped, regardless of sample size
3) We do not know the shape of the sampling distribution of $\bar{x}$ if the sample size is small and the population distribution isn’t bell-shaped

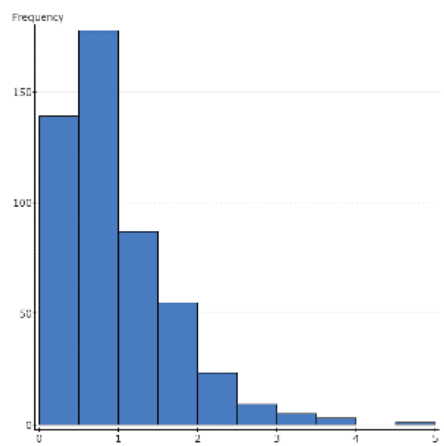
Central Limit Theorem

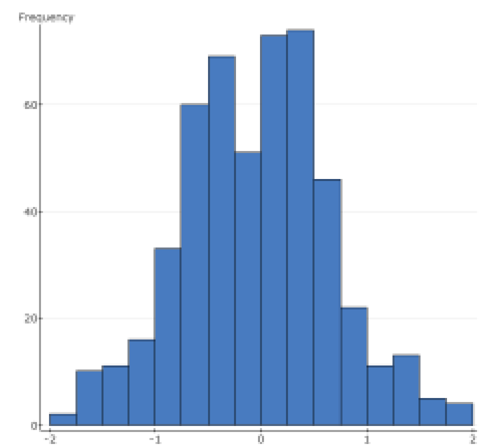
For any population the sampling distribution of $\bar{x}$ is bell shaped when the sample size n is large, when n is thirty or more

Note: for small sample size we can't say this.

Population $\quad\quad$ $\bar{x}$ when n=2 $\quad\quad$ $\bar{x}$ when n=30



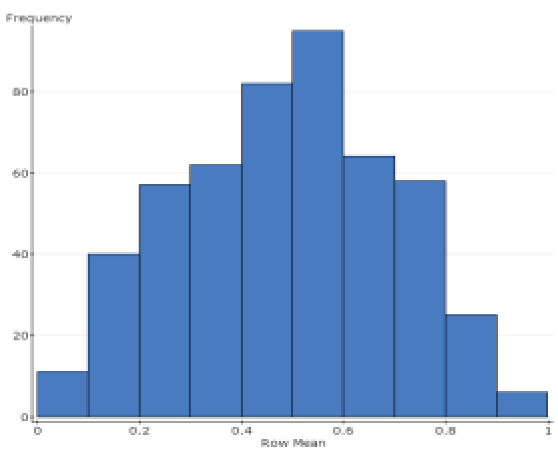




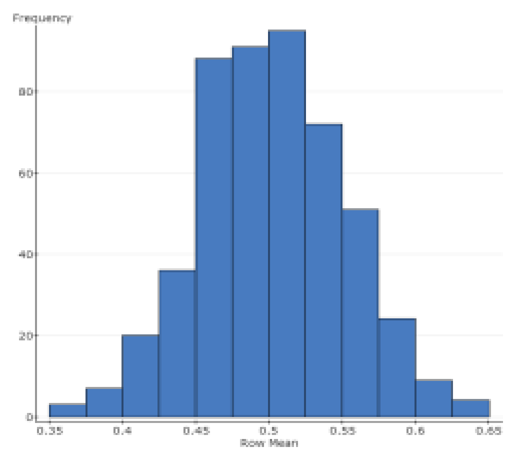
Central Limit Theorem

The sampling distribution of x_{bar} is bell-shaped when the population distribution is distribution is bell-shaped, regardless of sample size

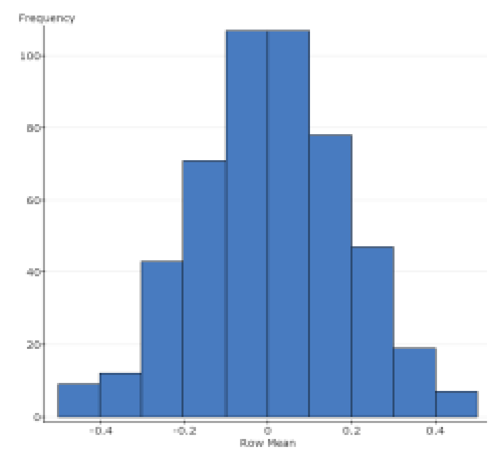
Population



$\bar{x}$ when n=2



$\bar{x}$ when n=30



Central Limit Theorem

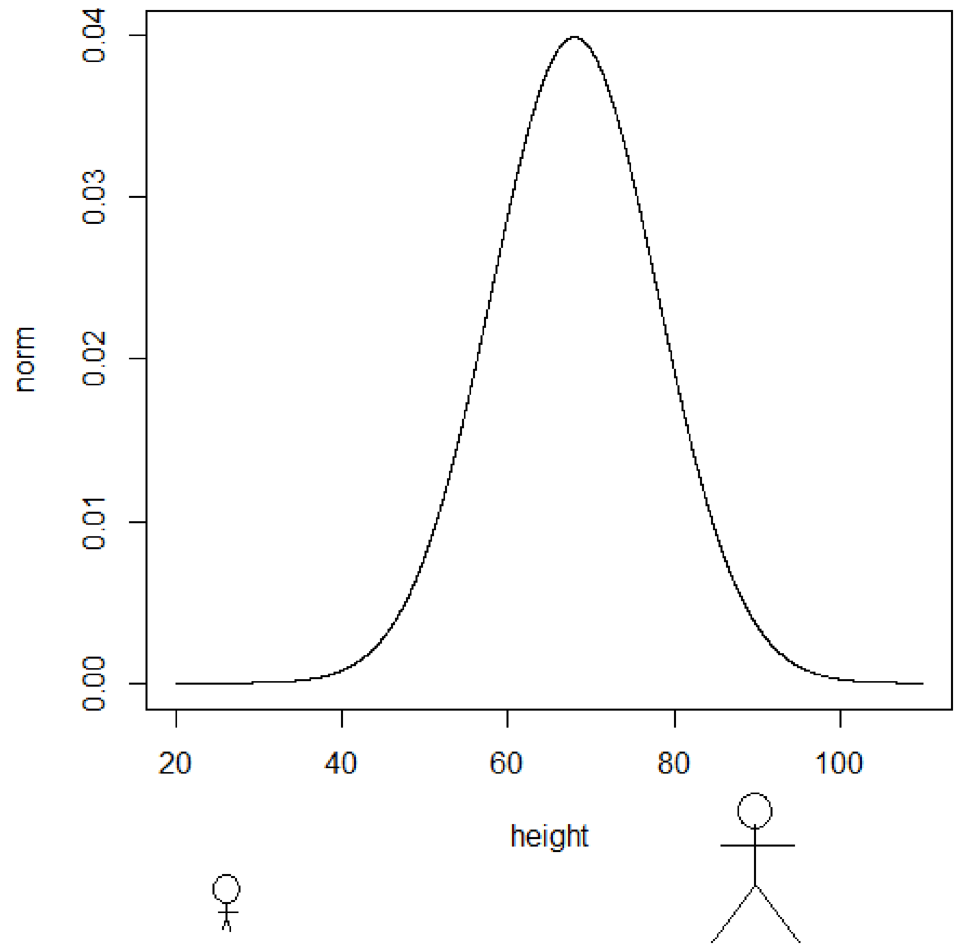
This rule allows us to develop and use powerful methodologies for the rest of the semester!

<u>IT IS VERY</u>

<u>IMPORTANT</u>

Central Limit Theorem

A real life example:



Verne Troyer

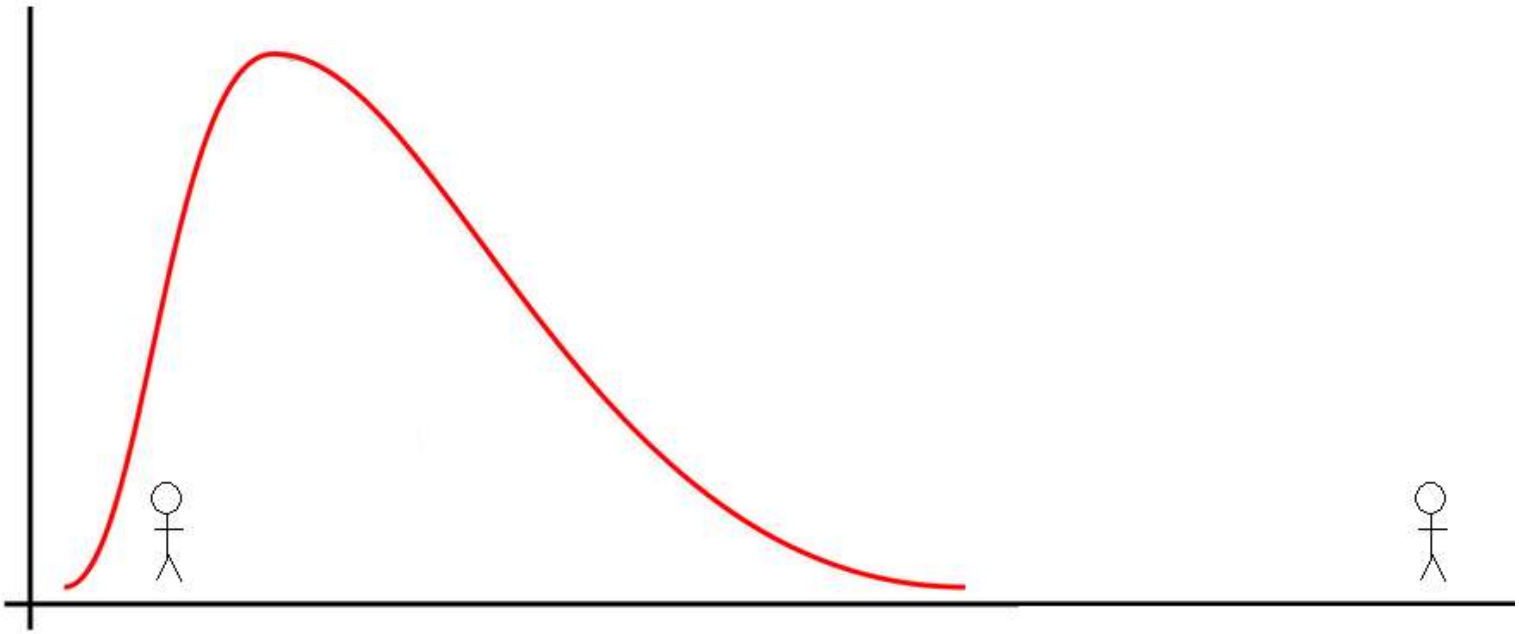
Yao Ming

Central Limit Theorem

- The means in this case are also normally distributed and the sample means will be close to the population mean with some variation.
- Obviously, there will be variation from sample to sample but the probability that any sample will deviate massively from the underlying population is very low – i.e. we wouldn't expect to see sample means too far from the average height.

Central Limit Theorem

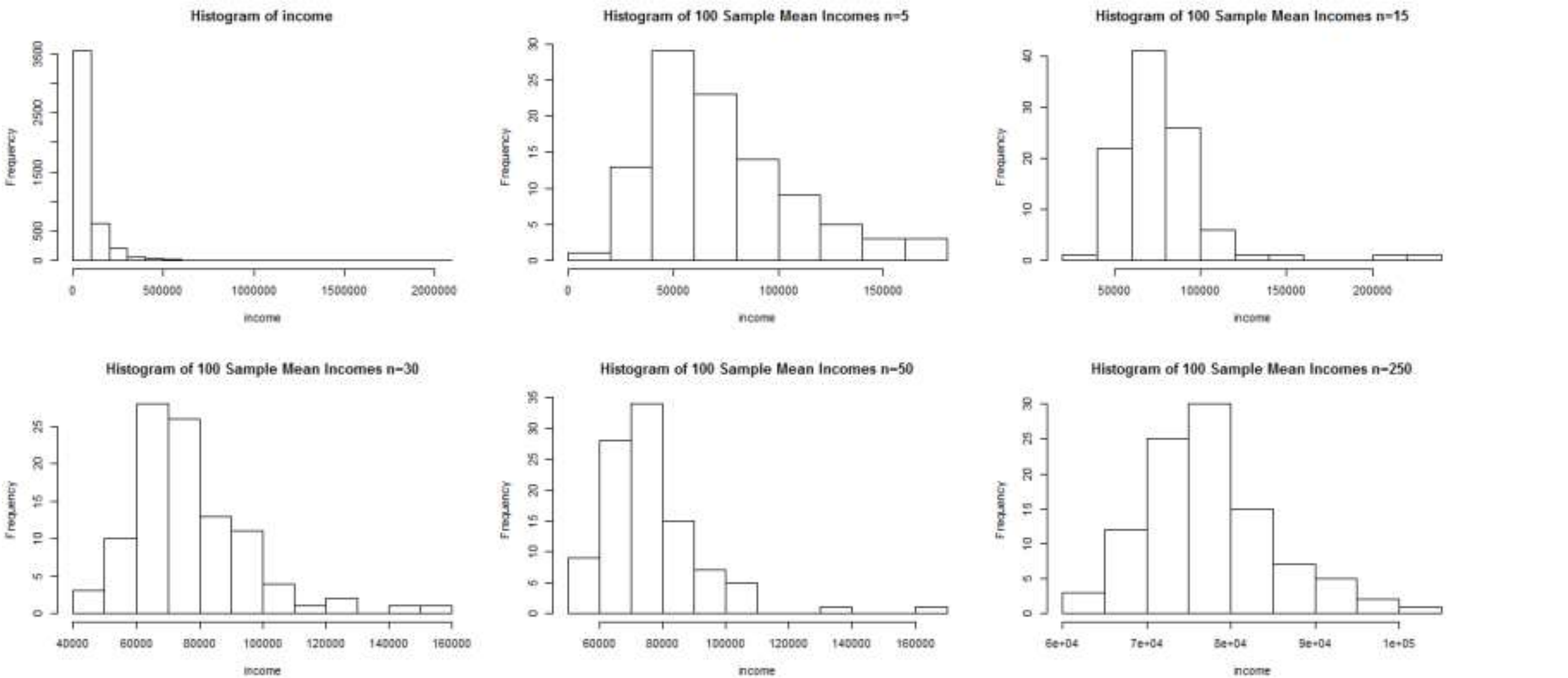
A real life example - income is skewed right



Graduate Students
Like Will

Bill Gates

Central Limit Theorem



Central Limit Theorem

- The means in this case are normally distributed if our sample size is large enough even if the underlying distribution isn't
- Still the sample means will be close to the population mean with some variation.
- Obviously, there will be variation from sample to sample but the probability that any sample will deviate massively from the underlying population is very low – i.e. we wouldn't expect to see sample means too far from the average income.

Video Summaries!*

- Bunnies, Rabbits and the NY times
  - https://www.youtube.com/watch?v=jvoxEYmQHNM