STOR 155 Introductory Statistics

Lecture 15: Sampling Distributions for Sample Means
Review

• **Sampling distribution for a sample count:**
  – Binomial experiments
  – Binomial distribution
  – Normal approximation
  – Continuity correction

• **Sampling distribution for a sample proportion:**
  – Exact calculation via binomial distribution
  – Normal approximation *No* continuity correction

• **How about a sample mean?**
Diversification

• A basic principle of investment is that diversification reduces risk.
• That is, buying several securities, (for example stocks), rather than just one, reduces the variability of the return on an investment.
• The following figures show two distributions of returns in 1987.
• Distribution of returns for all 1815 stocks on the NYSE for the entire year 1987.
• The mean return was \(-3.5\%\) and the distribution shows a very wide spread.
• Distribution of returns for all possible portfolios that invested equal amounts in each of 5 stocks in 1987.
• The mean is still –3.5%, but the variability is much less.
The investment example shows that

- Averages are **less variable** than individual observations;
- Averages are closer to ``normal” than individual observations.

**Why?**
• The sampling distribution of $\bar{X}$ for samples of size 10 compared with the distribution of a single observation.
POPULATION
Mean $\mu$
Standard deviation $\sigma$

SRS size $n \rightarrow \bar{x}$

Sampling distribution of $\bar{x}$

Mean $\mu$
SAMPLING DISTRIBUTION OF A SAMPLE MEAN

If a population has the $N(\mu, \sigma)$ distribution, then the sample mean $\bar{x}$ of $n$ independent observations has the $N(\mu, \sigma / \sqrt{n})$ distribution.

CENTRAL LIMIT THEOREM

Draw an SRS of size $n$ from any population with mean $\mu$ and finite standard deviation $\sigma$. When $n$ is large, the sampling distribution of the sample mean $\bar{x}$ is approximately normal:

$$\bar{x} \text{ is approximately } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$
The distributions of $\overline{X}$ for (a) 1 obs. (b) 2 obs. (c) 10 obs. (d) 25 obs. --- correction for `skewness"
• The amount of soda pop in each bottle is normally distributed with a mean of 32.2 ounces and a standard deviation of 0.3 ounces.

• Find the probability that a bottle bought by a customer will contain more than 32 ounces.

• Answer: Let $X$ denote the amount of soda in a bottle.

$$P(X > 32) = P\left( \frac{X - 32.2}{0.3} > \frac{32 - 32.2}{0.3} \right)$$

$$= P(z > -0.67) = 0.7486$$
• Find the probability that a box of four bottles will have a mean of more than 32 ounces of soda per bottle.

• Answer: The random variable here is the mean amount of soda per bottle, which is normally distributed with a mean of 32.2 and standard deviation of $0.3 / 2 = 0.15$. Hence

\[
P(\bar{X} > 32) = P\left(\frac{\bar{X} - 32.2}{0.15} > \frac{32 - 32.2}{0.15}\right)
= P(Z > -1.33) = 0.9082
\]
Weekly Income

- The average weekly income of graduates one year after graduation is $600. Suppose the distribution of weekly income has a standard deviation (SD) of $100.
- What is the probability that 25 randomly selected graduates have an average weekly income of less than $550?
- Answer: According to CLT, $\bar{X}$ approximately has a normal distribution with mean 600 and SD $100 / 5 = 20$.

\[
P(\bar{X} < 550) = P\left( \frac{\bar{X} - 600}{20} < \frac{550 - 600}{20} \right)
= P(Z < -2.5) = 0.0062.
\]
• Suppose we actually found a random sample of 25 graduates, with an average weekly income of $550.
• What can you say about the validity of the claim that the average weekly income is $600?
• Answer:
  – If the population mean is $600, then the probability to have observed a sample mean of $550 is very low (0.0062). The evidence provided by the sample suggests that the assumed average weekly income $600 is unjustified.
  – It’s more reasonable to believe that the population mean is actually smaller than $600. Then a sample mean of $550 becomes more probable.
Sum of independent normal random variables

Fact: If $X$, $Y$ are independent normal random variables and $a$, $b$ are constants, then $aX+bY$ also follows a normal distribution with mean

$$E(aX+bY) = a E(X) + b E(Y)$$

and variance

$$\sigma^2_{aX+bY} = a^2 \sigma^2_X + b^2 \sigma^2_Y$$

Use this fact to solve Problem 5.60 (c) (d) on page 349.
Take Home Message

• Sampling distribution for a sample mean
  – Mean for a sample mean
  – Variance for a sample mean
  – Central Limit Theorem
  – Normal distribution calculation