MATH 10

INTRODUCTORY STATISTICS

Ramesh Yapalparvi

Week 3

- **Chapter 5 – Probability**
- **Chapter 7 – Normal Distribution**
- **Chapter 8 – Advanced Graphs**

- **Chapter 9 – Sampling Distributions** ← **today's lecture**

Sampling distributions of the mean and p. Difference between means.
Central Limit Theorem.

Chapter 9, Section 2 - Introduction

- Inferential statistics.
- Have population, and a variable of interest, X.
- Take a simple random sample, calculate estimators for certain aspects of the population distribution of X.

- E.g. sample mean = estimator for population mean.
- E.g. estimator of the sample variance.

- We will now quantify how “good” these estimators are.

Chapter 9, Section 6 – Sampling Distribution of the Mean

- The sample mean, of a simple random sample A with size n, is a random variable.
- If you collect another simple random sample B with size n, it is likely to have a different sample mean.
- If X is a random variable that represents the mean of a sample of size n, then X has a distribution.

- The distribution of X is the sampling distribution of the mean (of a sample of size n).
- This distribution has mean $\mu_M = \mu$, where μ is the population mean.
- This distribution has variance $\sigma_M^2 = \frac{\sigma^2}{n}$, where σ^2 is the population variance.

Chapter 9, Section 6 – Sampling Distribution of the Mean

- Sampling Distribution of the Mean has,

$$\mu_M = \mu \qquad\qquad \sigma_M^2 = \frac{\sigma^2}{n}$$

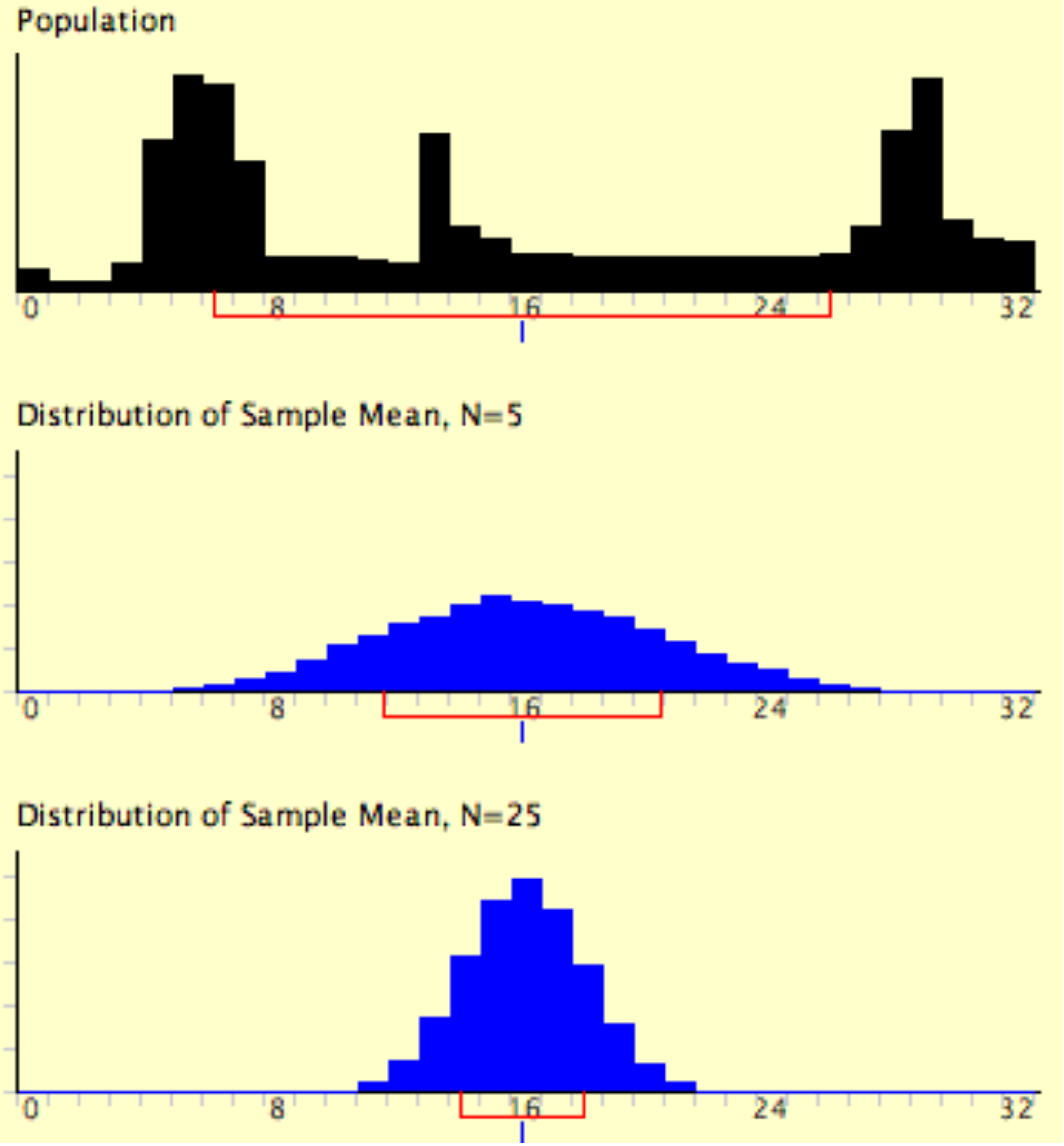
- Standard error, $\sigma_M = \frac{\sigma}{\sqrt{n}}$.

- **Central Limit Theorem!!!** ᕕ(ᐛ)ᕗ

If the population has finite mean μ, and finite non-zero variance σ^2, then the sampling distribution of the mean becomes better approximated by a normal distribution N(μ, $\frac{\sigma^2}{n}$), as sample size n increases.

Chapter 9, Section 6 – Sampling Distribution of the Mean

- Central limit theorem works for **any** distribution with finite mean and finite non-zero variance.



Chapter 9, Section 7 – Difference Between Means

- Finally, we can use statistics to compare two populations.
- Suppose you have two simple random samples with size n_1 and n_2.
- Samples from population 1 and 2 respectively.

- Calculate their sample means M_1 and M_2.
- The difference has a sampling distribution with mean $\mu_{M_1-M_2} = \mu_1 - \mu_2$.

Chapter 9, Section 7 – Difference Between Means

- The difference has a sampling distribution with mean $\mu_{M_1-M_2} = \mu_1 - \mu_2$.
- And variance $\sigma^2_{M_1-M_2} = \sigma^2_{M_1} + \sigma^2_{M_1}$.

- $\sigma^2_{M_i} = \frac{\sigma^2}{n_i}$, which is variance of the sampling distribution of M_i.
- Since the sample means are independent (as random variables), the variance sum law was used to derive the variance.

- $\sigma^2_{M_1-M_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

Chapter 9, Section 7 – Difference Between Means

- The difference has a sampling distribution with mean $\mu_{M_1-M_2} = \mu_1 - \mu_2$.
- And variance $\sigma^2_{M_1-M_2} = \sigma^2_{M_1} + \sigma^2_{M_1} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$.
- Standard error $\sigma_{M_1-M_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.
- This becomes much easier if the sample sizes and population variances are equal.

Public Service Announcement

- We are skipping "Chapter 9, Section 8, Sampling Distribution of r".
- This chapter is about the sampling distribution of the correlation coefficient.
- Not usually taught at Math 10 level.
- So we're nuking it from orbit (it's the only way to be sure).

Chapter 9, Section 9 – Sampling Distribution of p

- Population with N individuals. A proportion P of them are of type A, and the rest are of type B.
- E.g. Type A = those who voted for candidate A, and type B = those who voted for candidate B.

- Take a simple random sample of size n.
- You can see this sample as an experiment with n trials and probability of "success" P.

Chapter 9, Section 9 – Sampling Distribution of p

- Take a simple random sample of size n.
- You can see this sample as an experiment with n trials and probability of "success" P.

- The Binomial distribution modelling the distribution of the number of "successes" in this sample would have mean $\mathrm{n}P$.

- Let p be the proportion of type A ("successes") in your sample.
- This p has sampling distribution with mean P.

Chapter 9, Section 9 – Sampling Distribution of p

- Let p be the proportion of type A (“successes”) in your sample.
- This p has sampling distribution with mean P.

- The standard deviation of the Binomial distribution modeling our sample is $\sqrt{nP(1-P})$.
- Divide by n to get the standard error of p to be $\sigma_p = \sqrt{\frac{P(1-P}{n}}$.

- The sampling distribution is approximately normally distributed for large n.