AP Statistics Unit 5 – Quick Notes

1. Sampling Distributions Basics

Statistic: A number from a sample (e.g., $\bar{x}$, $\hat{p}$, $\bar{x}_1 - \bar{x}_2$, $\hat{p}_1 - \hat{p}_2$).
Parameter: A number describing the population (e.g., μ, p, $\mu_1 - \mu_2$, $p_1 - p_2$).

Sampling distribution: The distribution of a statistic over many samples from the same population(s).

Unbiased Estimators:

$$E[\bar{x}] = \mu, \quad E[\hat{p}] = p$$

$$E[\bar{x}_1 - \bar{x}_2] = \mu_1 - \mu_2, \quad E[\hat{p}_1 - \hat{p}_2] = p_1 - p_2$$

2. Mean (Center) of the Sampling Distribution

$$\mu_{\bar{x}} = \mu, \quad \mu_{\hat{p}} = p$$

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2, \quad \mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$$

3. Standard Deviation (Spread)

Single-sample:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}, \quad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Two-sample:

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Effect of n: Larger $n \Rightarrow$ smaller spread.

4. Shape and Normality Conditions

For Means ($\bar{x}$):

- Random sample
- Independence: $n \leq 0.1N$ if without replacement
- Normality: population normal (any n) OR $n \geq 30$ (CLT)

For Proportions ($\hat{p}$):

- Random sample
- Independence: $n \leq 0.1N$
- Success–failure: $np \geq 10$ and $n(1-p) \geq 10$

For Difference of Means ($\bar{x}_1 - \bar{x}_2$):

- Random samples from each population
- Independence: within and between samples ($n_1 \leq 0.1N_1$, $n_2 \leq 0.1N_2$)
- Normality: both populations normal (any n) OR both $n \geq 30$

For Difference of Proportions ($\hat{p}_1 - \hat{p}_2$):

- Random samples from each population
- Independence: within and between samples ($n_1 \leq 0.1N_1$, $n_2 \leq 0.1N_2$)
- Success–failure:
  - $n_1 p_1 \geq 10$, $n_1(1-p_1) \geq 10$
  - $n_2 p_2 \geq 10$, $n_2(1-p_2) \geq 10$

5. Central Limit Theorem (CLT)

If n is large enough, the sampling distribution of $\bar{x}$ is approximately normal regardless of population shape. Same logic applies to $\bar{x}_1 - \bar{x}_2$ if both n_1 and n_2 are large enough.

6. z-Score Formulas

Single-sample:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}, \quad z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

Two-sample:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

7. Interpretation Tips

- Large n reduces spread, making differences easier to detect.
- Always check all normality conditions before using z.
- In two-sample cases, verify independence *between* samples.

8. Common Pitfalls

- Confusing population distribution with sampling distribution.
- Forgetting success–failure checks for proportions.
- Using single-sample formulas for difference problems.