Big Ideas – Time Series Plots

- Look at when data is collected over time
- Overall stability = stable mean + stable variance
- Be careful with your language:
  - There is evidence of increasing mean
  - NOT: The mean is increasing
- If there appears to be a pattern over time, it may not be appropriate to treat the data as a random sample

Big Ideas – QQ-Plots

- Departure from straight line = non-normal data
  - Just need to see a departure, don’t need to say what type
- Make both a histogram and a QQ-plot:
  - Histogram gives general shape of distribution (bell-shaped; skewed; etc.)
  - QQ-plot gives normality of distribution

Big Ideas – Sampling Distribution for the Proportion (See page 74 of notes for details)

- Theory relies on idea of repeated sampling
  - Take all possible random samples of size \( n \) and compute the sample proportion for each sample
  - For a LARGE sample size \( n \), the distribution of the sample proportion will be APPROXIMATELY a NORMAL distribution with a mean of \( p \) and a standard deviation of \( \sqrt{p(1-p)/n} \)
  - Notation:
    \[
    \hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)
    \]

Big Ideas – Sampling Distribution for the Mean

- Theory relies on idea of repeated sampling
  - Take all possible random samples of size \( n \) and compute the sample mean for each sample
  - If the parent population is a normal distribution with mean \( \mu \) and standard deviation \( \sigma \), then for any sample size (small or large), the sample mean will have a NORMAL distribution with a mean of \( \mu \) and a standard deviation of \( \sigma/\sqrt{n} \)
  - Notation:
    \[
    \overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)
    \]

Big Ideas - CLT

- Theory relies on idea of repeated sampling
  - Take all possible random samples of size \( n \) and compute the sample mean for each sample
  - If the parent population is NOT a normal distribution but with mean \( \mu \) and standard deviation \( \sigma \), then for a large sample size, the sample mean will have approximately a NORMAL distribution with a mean of \( \mu \) and a standard deviation of \( \sigma/\sqrt{n} \)
  - Notation:
    \[
    \overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)
    \]

Big Ideas – Confidence Intervals

- Principles for using Confidence Intervals to Guide Decision Making:
  - Principle 1: A value not in a CI can be rejected as possible value of the population parameter.
  - Principle 2: When the CIs for parameters for two different populations do not overlap, it is reasonable to conclude that the parameters for the two populations are different.
**Big Ideas – Confidence Intervals**

- The probability that the true parameter lies in a computed CI is either 0 or 1.
- A 95% Confidence Interval: We are 95% confident that the true parameter value lies inside the CI. OR: The interval provides a range of reasonable values for the population parameter.
- The 95% Confidence Level: If the procedure were repeated many times, we would expect 95% of the resulting CIs to contain the true population parameter.

**Big Ideas – Hypothesis Testing**

- Basic 5 steps to hypothesis testing are the same no matter what the scenario
- Parameters, assumptions, and test statistic change depending on the scenario
  - Result of differences in structure of data

**5 Inference Scenarios**

1. 1-sample inference for pop. proportion \( p \)
2. 1-sample inference for pop. mean \( \mu \)
3. Paired samples inference for a population mean difference \( \mu_0 \)
4. 2 indep. samples inference for the difference between 2 pop. means \( \mu_1 - \mu_2 \)
5. 2 indep. samples inference for the difference between 2 pop. prop. \( p_1 - p_2 \)

**5 Steps in Hypothesis Testing**

1. State \( H_0 \) and \( H_a \)
2. Check assumptions and calculate test statistic
3. Find \( p \)-value
4. Make decision
5. State conclusion

**Step 1 – Basic Truths**

- Use the same population parameter in both \( H_0 \) and \( H_a \)
- Use the same null value in both \( H_0 \) and \( H_a \)
  - The number the researchers want to investigate, or 0 if the parameter involves a difference
- ALWAYS have equality in \( H_0 \) and NEVER have equality in \( H_a \)
  - Sign used in \( H_a \) depends on what the researchers want to investigate
Step 1 – Determining the Scenario

- Good questions to ask yourself:
  - Mean (quantitative data) or proportion (categorical data)?
  - One group or two?
  - If means, two groups: paired or independent?

Step 2 - Assumptions

1. \( p \): The data are a random sample and the sample size is large (\( np \geq 10, n(1-p) \geq 10 \); for large sample z-test)
2. \( \mu \): The data are a random sample from a normal population
3. \( \mu_D \): The differences are a random sample from a normal population of differences
4. \( \mu_1 - \mu_2 \): Each sample is a random sample from a normal population, the 2 samples are independent, and the pop. standard deviations are equal (for pooled version)
5. \( p_1 - p_2 \): Each sample is a random sample and each sample size is large

Example: z-test for \( p \)

1. Ho: \( p = p_0 \) vs. Ha: \( p \neq p_0 \)
   or Ha: \( p > p_0 \)
   or Ha: \( p < p_0 \)
2. Verify assumptions
   - Random sample
   - Large sample: \( np_0 \geq 10 \) and \( n(1-p_0) \geq 10 \)
   and calculate z-statistic

3. Find p-value
   - Assuming Ho is true, the p-value is the probability of observing a z-statistic as extreme or more extreme than what we actually saw
   - Ha: \( p \neq p_0 \) \( \Rightarrow P(Z \leq -z) + P(Z \geq z) \)
   - Ha: \( p > p_0 \) \( \Rightarrow P(Z \geq z) \)
   - Ha: \( p < p_0 \) \( \Rightarrow P(Z \leq z) \)

Example: z-test for \( p \)

4. Decision
   - p-value \( \leq \alpha \) \( \Rightarrow \) Reject Ho
   - p-value > \( \alpha \) \( \Rightarrow \) Fail to Reject Ho

5. Conclusion
   - Reject Ho \( \Rightarrow \) "There is sufficient evidence that Ha is true."
   - Fail to Reject Ho \( \Rightarrow \) "There is insufficient evidence that Ha is true."