Week 8 notes

ATTENTION: Some definition might differ from your textbook, please follow your textbook.

Some questions for you:

1. What's markdown and R markdown (RMD). (NOT required to know)
2. What's confidence interval (CI)?
3. What does a 95% CI mean?
    - The probability that CI will cover the population mean is 95%.

If a R.V. X follows a normal distribution $N(\mu, \sigma^2)$

Let n be the sample size, and we draw a sample from population as

$X_1, X_2, \cdots, X_n$

We can obtain sample mean $\bar{x} = \frac{X_1+X_2+\cdots+X_n}{n}$ and sample standard error $se = \hat{\sigma}^2 / \sqrt{n}$ ($\hat{\sigma}^2$ is the sample variance). For proportion data, we use estimated standard deviation for sample mean instead, which is given by $se = \frac{\hat{p}(1-\hat{p})}{n}$.

If we have confidence level as p (e.g. 95%), then we define $\alpha = 1 - p$.

We call $[\bar{x} - z_\alpha se, \bar{x} + z_\alpha se]$ as the p-CI for this sample.

Where $z_\alpha = qt(1 - \frac{\alpha}{2}, df = n - 1)$, $qt()$ is quantile for t distribution, df is degree of freedom.

Corollary: The confidence interval is associated with α and sample size n.