Lesson 3. Confidence Intervals – Part 1

1 Populations vs. samples

Example 1. Suppose researchers are interested in estimating the average BMI (body mass index) of American men over the age of 18. The researchers obtain a random sample of 422 men and calculate a mean BMI of 28.0 kg/m^2. Further suppose the standard deviation of all American men BMIs is known to be 5.4 kg/m^2.

	Definition	**For Example 1**
population	All individuals in the group of interest.	
parameter	Numerical characteristic of the population distribution. Fixed value. Does not depend on data.	
sample	A collection of independent, identically distributed (i.i.d) r.v.s.	
data	An observed sample; actual numbers.	
statistic (**estimator**)	A function of the random variables in the sample. Used to estimate a parameter.	
observed statistic (**estimate**)	The numerical value of a statistic, calculated with data values.	

- A **p**arameter describes a **p**opulation, while a **s**tatistic describes a **s**ample

- A **simple random sample (SRS)** is a subset of the population, chosen randomly such that each individual has the same chance of being chosen, and all subsets of the same size have the same chance of being chosen
- Examples of statistics:
  - Sample mean:
  - Sample standard deviation:

2 Confidence intervals for population mean

- We can use the sample mean to get a **point estimate** of the population mean, but...
- Usually, we would like to also provide an **interval estimate**, a range of plausible values that includes a margin of error
- The most general form of an interval estimate is
- The margin of error is composed of the **critical value** times a **standard error (SE)** term:

2.1 CI when population variance (σ^2) is known

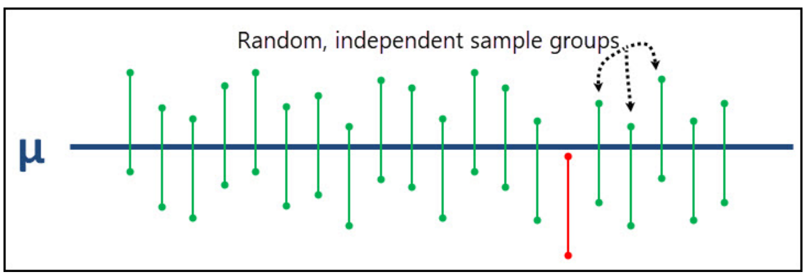
- Suppose $x_1, \ldots, x_n$ is data from a simple random sample from a population with unknown mean μ and known variance σ^2
- The $(1-\alpha)100\%$ **CI for the population mean** is

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Example 2.

a. Based on Example 1, calculate a 95% confidence interval for the mean BMI of all American men. Note that $z_{0.025} = 1.96$.
b. Interpret your interval from part a.

2.2 What does it mean, mathematically, to be "95% confident"?



- "95% confidence" means that if we were to repeatedly take samples of size n and construct the corresponding confidence intervals, 95% of the intervals would contain the true population mean μ
- The probability that the process of forming a CI will capture μ is 0.95
- It is NOT correct to say that the probability we captured μ with our particular CI estimate is 0.95

2.3 CI when population variance (σ^2) is not known

- Now suppose $x_1, \ldots, x_n$ is data from a simple random sample from a population with unknown mean μ and unknown variance σ^2
- The $(1-\alpha)100\%$ **CI for the population mean** is

$$\bar{x} \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$$

- Why do we switch to the t distribution?

When σ^2 is known:

- Central Limit Theorem says $\overline{X} \sim N(\mu, \sigma^2)$

$$\Rightarrow \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$\Rightarrow$ critical value from $N(0,1)$

When σ^2 is unknown:

- It can be shown that

$$\frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t(df = n-1)$$

$\Rightarrow$ critical value from $t(df = n-1)$

2.4 Technical conditions to check

- Two things must be met for the above CI formulas to be appropriate:

1. Data must be from a $\boxed{}$
2. Either $\boxed{}$ or $\boxed{}$