Contents

Day 16 **1**

Error and Variability **2**
Important Facts 2

Central Limit Theorem **3**
Example 3
Answers 3

Day 16

Error and Variability

Rounding variability: error due to precision of our machine/scale/etc.

1) Never measure exact, only to some tolerance
   - $\rightarrow$ typically rounding error $\sim (E, -E)$

Example: weight is 150 pounds - reality $\rightarrow 150 \pm U(-0.5, 0, 5)$

2) When making repeated measurements of something, there will be some natural variability, due to many small sources of error. Usually (as long as errors are on the same scale), we can make measurement error of $\sim N(0, \sigma)$

3) Sampling error: error due to only having a sample from the population. Estimate a population mean μ based on a sample mean:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_{\mathrm{i}}$$

What is the distribution of $\bar{X}$ over all possible samples = sampling distribution of $\bar{x}$ which is the sampling mean

Consider simple random sample from a given population, the values $\mathrm{x}_1 \rightarrow \mathrm{x}_\mathrm{n}$ values of some numerical variables. We assume the x_i's are independent and identically distributed random variables.

With theoretical/population mean μ and standard deviation σ, then $\bar{X} = \frac{1}{n} \rightarrow (X_1 + X_2 + ... + X_\mathrm{n})$ is considered a linear combination.

$$E(\bar{X}) = E(\frac{1}{n} \times (X_1 + X_2 + ... + X_\mathrm{n})) = \frac{1}{n}(E(x_1) + E(x_2) + ... + x_\mathrm{n})$$

$$E(\bar{X}) = n\mu \times \frac{1}{n} = \mu$$

Important Facts

- The mean of the sampling distribution of $\bar{X}$ is equal to the population mean μ

$$(Var\bar{X}) = var(\frac{1}{n} \times (x_1 + x_2 + ... + x_\mathrm{n}))$$
$$= (\frac{1}{2})^2 \times [var(x_1 + x_2 + ... + x_\mathrm{n}]$$
$$= (\frac{1}{n})^2 \times [var(x_1) + var(_2) + ... + var(x_\mathrm{n})]$$

- The variance of sampling distribution of $\bar{X}$ is smaller than the population variance by a factor of n. The standard deviation is smaller by a factor of $\sqrt{n}$. Consider a normally distributed population. Theorem: any linear combination of normal random variables is also normally distributed.

$$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

Central Limit Theorem

For a simple random sample (ASRS) of size `n` from a population with finite mean μ and finite standard deviation σ:

When `n` is "large enough", $\bar{x}$ is approximately $\sim N(\mu, \frac{\sigma}{\sqrt{n}})$

What does "large enough" depend on?

- How good the approximation needs to be (robust procedures-approximation just needs to be OK)
- Shape of population distribution
  - Higher skew requires larger `n`
  - Outliers in sample suggest larger `n` is needed

Consider a normally distributed population.

$\bar{X}$ is a linear combination of random variables $X_{\mathrm{i}} \sim N(\mu, \sigma)$

Example

You take a sample size of 64 from a population normally distributed with mean of 82 and standard deviation of 24.

a) Find the sampling distribution of the sample mean $\bar{X}$
b) Middle 95% of values of x are expected to be in what interval.
c) Middle 95% of <u>sample means</u> $\bar{X}$ are expected to be in what interval?

Answers

a) $\bar{X} \sim N(82, \frac{24}{\sqrt{64}}) \sim N(82, 3)$

b) (34, 130)

c) $E[\bar{x}] = \mu = 82$, $SD[\bar{x}] = \frac{\sigma}{\sqrt{n}} = \frac{24}{\sqrt{64}} = 3$

$\mu + 2SD[\bar{x}] = 82 + 6 = 88$

$\mu - 2SD[\bar{x}] = 82 - 6 = 76$

The range between 76 and 88

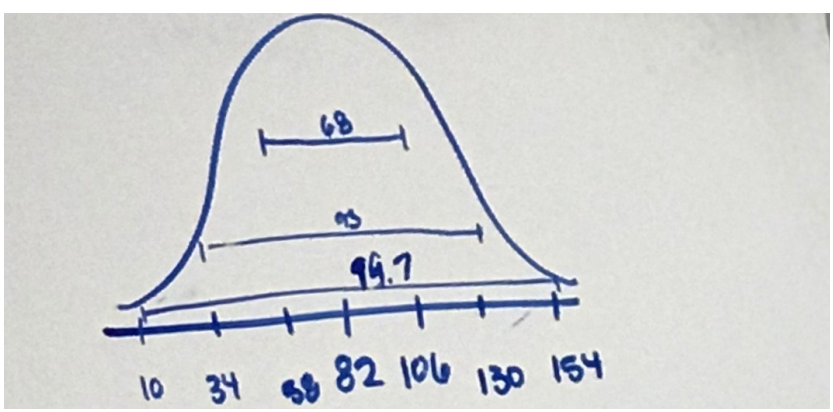

Figure 1: Curve