Point estimator

Confidence interval

If we want to estimate the average height of people in the world, what would it be?

$$\bar{X}$$

Similarly in the Skittles problem, if we want to estimate the average proportion of orange candies, we can use

$$\hat{p}$$

They are point estimators

Definition:

A **point estimator** is a statistic that provides **an estimate** of a population parameter. Ideally, a point estimate is our “best guess” at the value of an unknown parameter.

If we want to estimate the average height of people in the world, what would it be?

$$\bar{X}$$

Similarly in the Skittles problem, if we want to estimate the average proportion of orange candies, we can use

$$\hat{p}$$

They are point estimators

Definition:

A **point estimator** is a statistic that provides an estimate of a population parameter. Ideally, a point estimate is our **"best guess"** at the value of an unknown parameter.

We learned that an ideal point estimator will have no bias and low variability.

Variability is always present when calculating statistics from different samples. Repeated sampling could yield different results.

We want to know the average height of people in the world.

- One sample mean is 168cm ☹

- Use an interval centered at the sample mean. ☺Better

P(the average height is between 148cm and 188cm) = ?

Confidence Interval !!!

Def. Confidence interval for a population characteristic is an interval of plausible values for the characteristic.

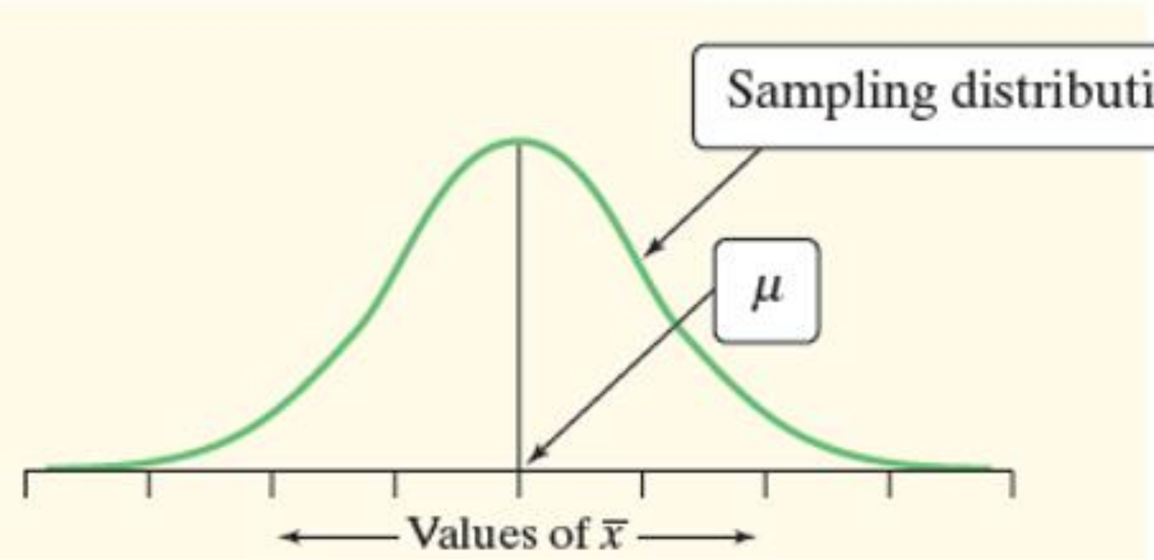
It is constructed so that, with a chosen <u>degree of confidence</u>, the actual value of the population characteristic will be between the lower and the upper endpoints of the interval.

FRQ:
We are C% confident that the interval from _____ to _____ captures the actual value of the [population parameter in context].

<u>The degree of confidence = the confidence level</u>

C% -> confidence level

What is the 90% confidence interval?



$$P(x_l \leq \bar{X} \leq x_r) = 90\%$$
How to find x_l and x_r?

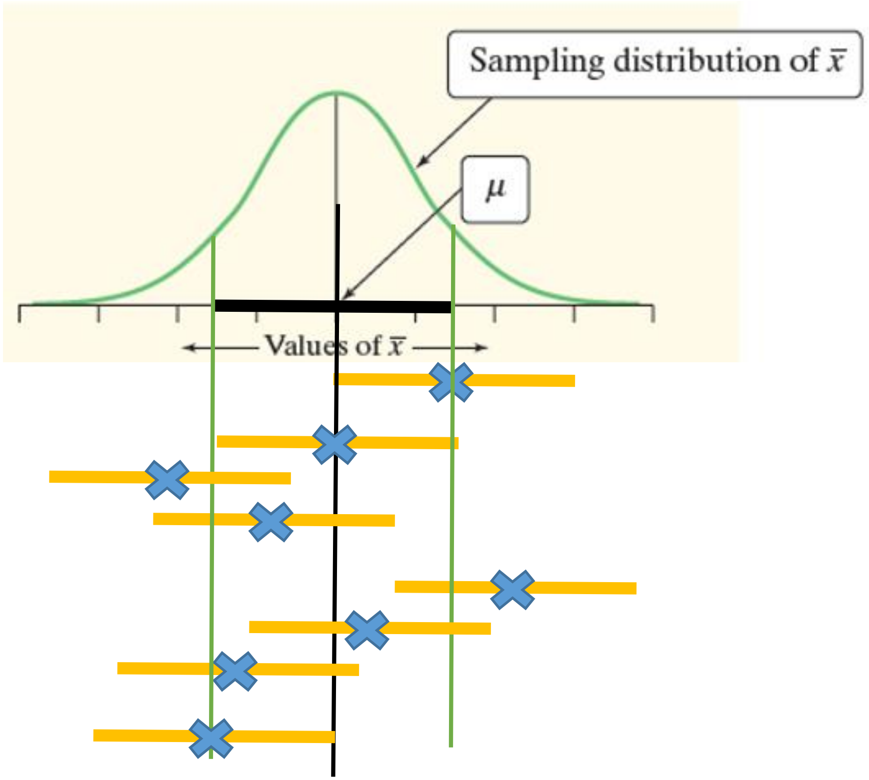
$$P(x_l \leq \bar{X} \leq x_r)$$
$$= P\left(\frac{x_l - \mu}{\sigma/\sqrt{n}} \leq \bar{X} \leq \frac{x_r - \mu}{\sigma/\sqrt{n}}\right)$$
$$= P(z_{0.05} \leq Z \leq z_{0.95}) = 90\%$$

$$x_l = \mu + z_{0.05} \cdot \frac{\sigma}{\sqrt{n}} = \mu - z_{0.95} \cdot \frac{\sigma}{\sqrt{n}}$$
$$x_r = \mu + z_{0.95} \cdot \frac{\sigma}{\sqrt{n}}$$
$$\mathrm{P}(\bar{X} \in \mu \pm z_{0.95} \cdot \frac{\sigma}{\sqrt{n}}) = 90\%$$

$$P(Z \leq z_{0.05}) = 0.05$$

90% of sample means will fall into the interval

$$[\mu - z_{0.95} \cdot \frac{\sigma}{\sqrt{n}}, \mu + z_{0.95} \cdot \frac{\sigma}{\sqrt{n}}]$$



- If $\bar{X} = \mu + z_{0.95} \cdot \frac{\sigma}{\sqrt{n}}$, the interval would be $[\mu, \mu + 2 \cdot z_{0.95} \cdot \frac{\sigma}{\sqrt{n}}]$ which includes μ.
- If $\bar{X} = \mu - z_{0.95} \cdot \frac{\sigma}{\sqrt{n}}$, the interval would be $[\mu - 2 \cdot z_{0.95} \cdot \frac{\sigma}{\sqrt{n}}, \mu]$ which includes μ.

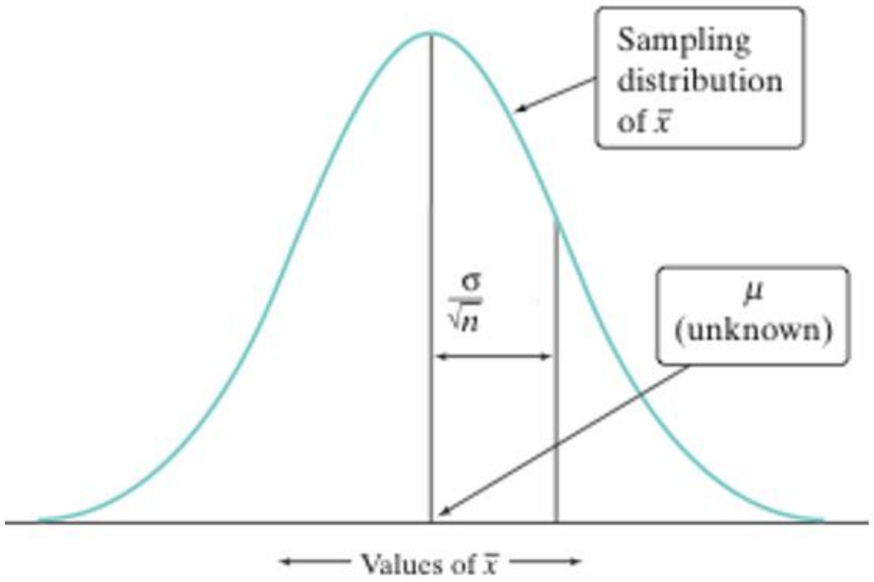
90% of intervals $[\bar{X} - z_{0.95} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{0.95} \cdot \frac{\sigma}{\sqrt{n}}]$

will include the true mean.

Confidence Levels and Confidence Intervals

The confidence level tells us how likely it is that the method we are using will produce an interval that captures the population parameter if we use it many times. The confidence level does not tell us the chance that a particular confidence interval captures the population parameter

Constructing a C.I.



Our estimate came from the sample statistics $\bar{X}$.

Since $\bar{X} \sim \mathrm{N}(\mu, \frac{\sigma^2}{n})$, the 90% confidence interval is

$$[\bar{X} - z_{0.95} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{0.95} \cdot \frac{\sigma}{\sqrt{n}}]$$

Critical value: $z_{0.95}$

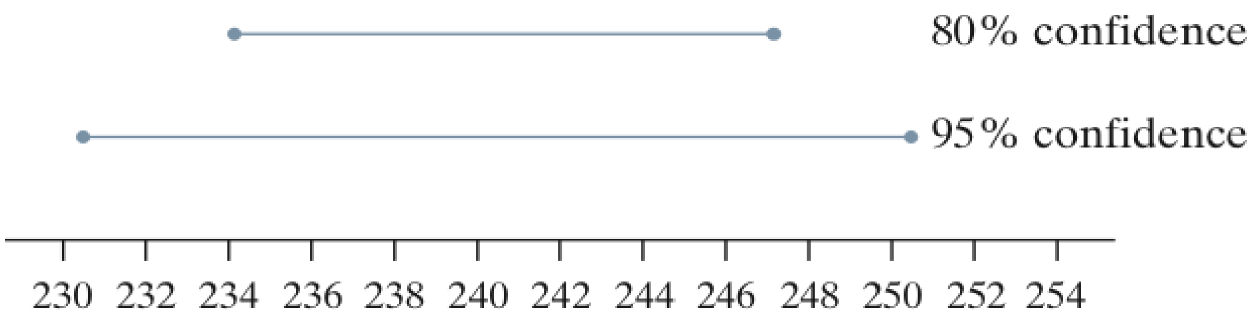
Margin of error (ME): $z_{0.95} \cdot \frac{\sigma}{\sqrt{n}}$

This leads to a more general formula for confidence intervals:

statistic ± (critical value) • (standard deviation of statistic)

Properties of Confidence Intervals:

- The user chooses the confidence level, and the margin of error follows from this choice.
- The critical value depends on the confidence level and the sampling distribution of the statistic.
  - Greater confidence requires a larger critical value
  - The standard deviation of the statistic depends on the sample size *n*



The margin of error gets smaller when:

✓ The confidence level decreases

✓ The sample size *n* increases

Using Confidence Intervals

Before calculating a confidence interval for μ or p there are three important conditions that you should check.

1) **Random**: The data should come from a well-designed random sample or randomized experiment.

2) **Normal:** The sampling distribution of the statistic is approximately Normal.

For means: The sampling distribution is exactly Normal if the population distribution is Normal. When the population distribution is not Normal, then the central limit theorem tells us the sampling distribution will be approximately Normal if n is sufficiently large ($n \geq 30$).

For proportions: We can use the Normal approximation to the sampling distribution as long as $np \geq 10$ and $n(1 - p) \geq 10$.

3) **Independent & 10% condition**: Individual observations are independent. When sampling without replacement, the sample size n should be no more than 10% of the population size N (the *10% condition*) to use our formula for the standard deviation of the statistic.

http://digitalfirst.bfwpub.com/stats_applet/stats_applet_4_ci.html