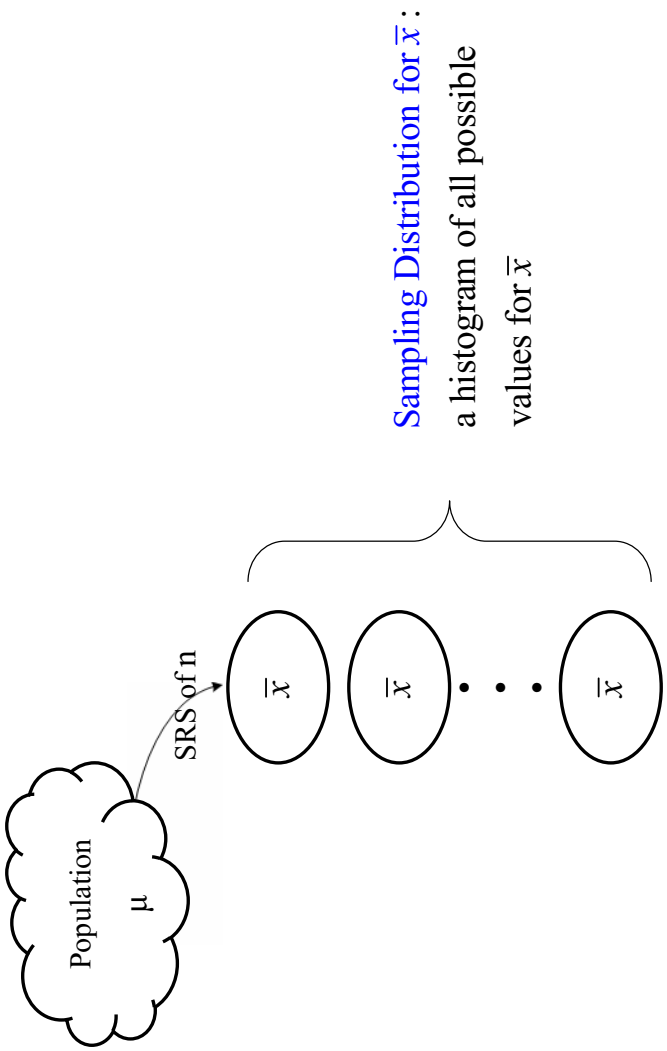


The Sampling Distribution for a Statistic

The distribution of values taken by the statistic in all possible samples of the same size from the same population.

The Sampling Distribution for $\bar{X}$

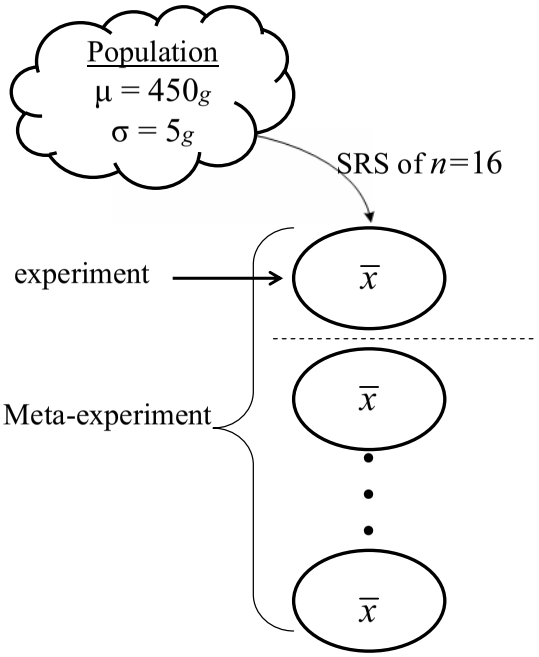
- The mean of the sampling distribution equals the population mean (μ)
- The SD of the sampling distribution is equal to $\frac{\sigma}{\sqrt{n}}$
- If X has a Normal distribution, then the distribution for $\bar{X}$ is exactly Normal.

The Central Limit Theorem (CLT)

For a SRS of size n from a population with mean= μ , SD= σ , and *any shape*, the sampling distribution for $\bar{X}$ is approximately Normal with mean= μ and SD= $\frac{\sigma}{\sqrt{n}}$.

- $\frac{\sigma}{\sqrt{n}}$ is the Standard Deviation of $\bar{X}$

Let X be a R.V. denoting the weight of a chocolate bar in *grams*.



	Distribution for X	*Sampling Distribution* for $\bar{X}$
mean	$450g$	$450g$
SD	$5g$	$5/\sqrt{16}=1.25g$

What is the probability that a sample average will be less than $448g$?

The distribution of heights, X, at a college is normal with

μ = 65 inches and σ = 3 inches

1) Find the probability that a randomly selected individual is between 62 and 67 inches

2) Find the probability that a randomly selected individual is more than 62 inches tall.

A Simple Random Sample of n = 6 is selected from the population.

3) Describe the Sampling Distribution for $\bar{X}$ (Shape, Center, Spread)

4) Find the probability that the sample average is between 62 and 67 inches.

5) Find the 90th percentile of the Sampling Distribution for $\bar{X}$.
State in plain terms what this value represents.

6) Find the probability that all six individuals have a height that is more than 62 inches.

Statistical Inference

- drawing conclusions about a population, based on a sample.
- uses properties of the sampling distribution and random sampling.

Example:

Population: GRE results for a new exam format on the quantitative section
Sample: n=300 test scores

	Population	Sampling Dist. for $\bar{X}$
shape	normal?	approx. normal
mean	μ (unknown)	μ
SD	σ=100 (assume known)	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{100}{\sqrt{300}} = 5.8$

~95% of the Sampling Distribution is within $\pm 2 \cdot \sigma_{\bar{X}}$ of μ.

1) In ~95% of the samples of n=300, $\bar{X}$ is within +/- 11.6 pts of μ.

2) In ~95% of the samples of n=300, μ is within +/- 11.6 pts of $\bar{X}$.

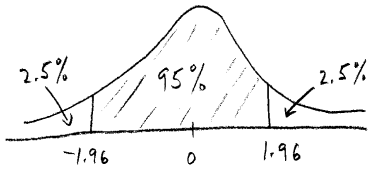
3) In ~95% of the samples of n=300, μ lies between $\bar{X}$ −11.6 *and* $\bar{X}$ +11.6.

4) We are ~95% *confident* that we have one of the samples that gives an interval containing μ.

Confidence Intervals & Margin of Error (CIs & MoE)

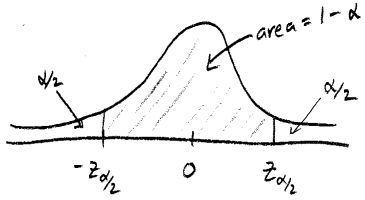
95% CI for μ

$$\bar{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$



100(1- α)% CI for μ

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



Ex: GRE scores (n=300, σ=100, $\bar{x} = 520$)

95% CI for μ: $520 \pm 1.96 \cdot \frac{100}{\sqrt{300}} = 520 \pm 11.37 = (508.63, 531.37)$

90% CI for μ: $520 \pm 1.645 \cdot \frac{100}{\sqrt{300}} = 520 \pm 9.54 = (510.46, 529.54)$

Margin of Error (m):

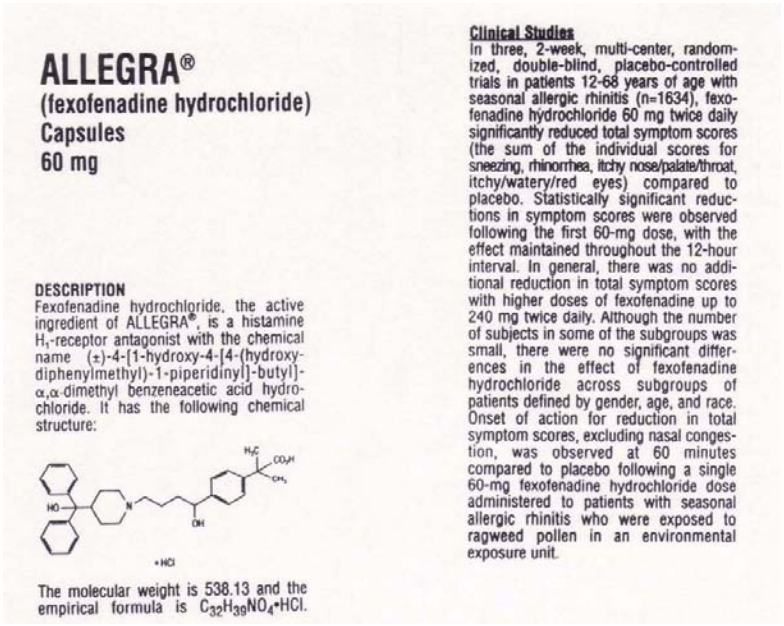
- = ½ of the CI length
- Smaller MoE → more precision
- Decreases if …
  - $z_{\alpha/2}$ gets smaller
  - σ gets smaller
  - n gets larger

Study Design

We need $n = \left(\frac{z_{\alpha/2}\,\sigma}{m} \right)^2$ observations in our sample to have a MoE=m

Ex: (GRE scores) How large a sample is needed for a MoE of 10 points at the 95% level?

$$n = \left(\frac{1.96 \cdot 100}{10} \right)^2 = 384.16 \approx 385$$



ALLEGRA®
(fexofenadine hydrochloride)
Capsules
60 mg

DESCRIPTION
Fexofenadine hydrochloride, the active ingredient of ALLEGRA®, is a histamine H_1-receptor antagonist with the chemical name (±)-4-[1-hydroxy-4-[4-(hydroxydiphenylmethyl)-1-piperidinyl]-butyl]-α,α-dimethyl benzeneacetic acid hydrochloride. It has the following chemical structure:

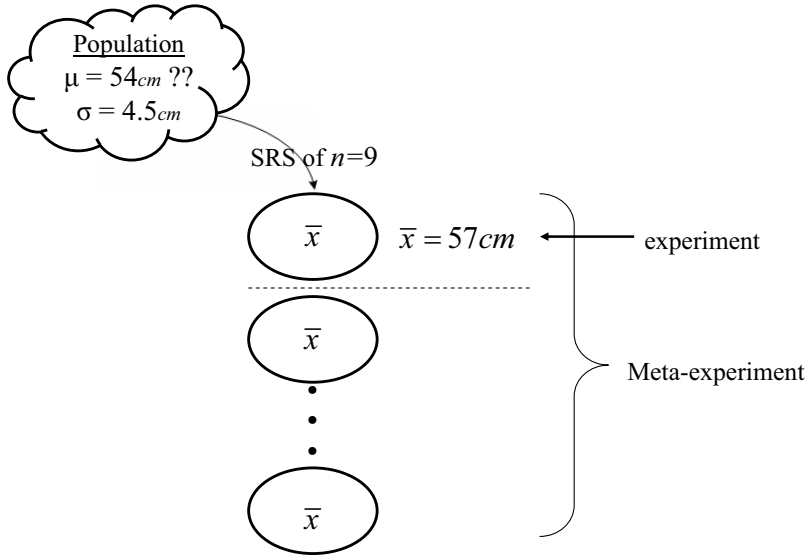
The molecular weight is 538.13 and the empirical formula is $C_{32}H_{39}NO_4$•HCl.

Clinical Studies
In three, 2-week, multi-center, randomized, double-blind, placebo-controlled trials in patients 12-68 years of age with seasonal allergic rhinitis (n=1634), fexofenadine hydrochloride 60 mg twice daily significantly reduced total symptom scores (the sum of the individual scores for sneezing, rhinorrhea, itchy nose/palate/throat, itchy/watery/red eyes) compared to placebo. Statistically significant reductions in symptom scores were observed following the first 60-mg dose, with the effect maintained throughout the 12-hour interval. In general, there was no additional reduction in total symptom scores with higher doses of fexofenadine up to 240 mg twice daily. Although the number of subjects in some of the subgroups was small, there were no significant differences in the effect of fexofenadine hydrochloride across subgroups of patients defined by gender, age, and race. Onset of action for reduction in total symptom scores, excluding nasal congestion, was observed at 60 minutes compared to placebo following a single 60-mg fexofenadine hydrochloride dose administered to patients with seasonal allergic rhinitis who were exposed to ragweed pollen in an environmental exposure unit.

1. Identify the response variable. Is it quantitative or categorical?
2. Identify the blocking variables and 2 levels for each one.
3. Identify the explanatory variable.
4. If the population SD for the total symptom scores was σ = 50 units, how many subjects would be necessary to have a margin of error equal to 2 units at the 90% confidence level?

   How many would be needed at the 98% confidence level?

Let X be a R.V. denoting the length of a fish in *cm*.

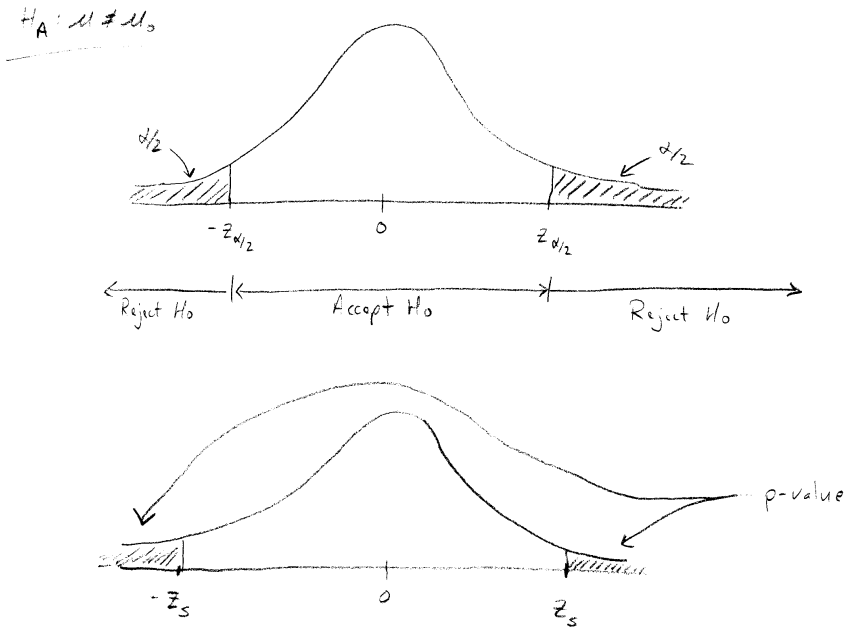


Does $\bar{x} = 57$ give strong evidence that μ>54 *cm*?

How likely are we to get $\bar{x} = 57$ if μ=54 *cm*?

$H_o : \mu = \mu_0$ $\qquad z_s = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}$

		Rejection Region	p-value
1)	$H_a : \mu > \mu_0$	$z_s \geq z_\alpha$	$p-value = P(Z \geq z_s)$
2)	$H_a : \mu < \mu_0$	$z_s \leq -z_\alpha$	$p-value = P(Z \leq z_s)$
3)	$H_a : \mu \neq \mu_0$	$\lvert z_s \rvert \geq z_{\alpha/2}$	$p-value = 2 \cdot P(Z \geq \lvert z_s \rvert)$



Conditions for a Valid 1-Sample CI and level α Test

- The data are a SRS from the population.
- If *n* is small, the population must be approximately normal.
- The population must be large enough (at least 10 x larger than the sample size).

If these conditions are not met, then the *actual* confidence level and/or level of significance are different from what we claim.

→ *actual* vs. *nominal* level of significance (l.o.s.)

"*observed* level of significance" vs. *p-value*

Hypothesis Testing

- Null Hypothesis (H_o): The population mean is 54 cm.
- Alternative Hypothesis (H_A): The population mean is greater than 54 cm.

p-value

- Measures the strength of the sample evidence against H_o
- A small p-value gives strong evidence against H_o
- Definition:

  The probability, computed assuming that H_o is true, of a sample result ($\bar{X}$) as extreme or more extreme than the one from our sample.

Rule of Thumb for the significance of p-values

- If the p-value is less than .05, then our results are statistically significant at the .05 level

H_o: $\mu=54\ cm$ $\qquad \sigma_{\bar{x}} = \frac{4.5\ cm}{\sqrt{9}} = 1.5\ cm$

H_a: $\mu>54$ $(\mu_a = 58)$ $\qquad \alpha = .05$

4 Steps for finding the Power in a test of hypotheses

1) Write the RR for H_o in terms of z-scores: $z_s \geq 1.645$

2) Write the RR for H_o in terms of $\bar{X}$: $\frac{\bar{X} - 54}{1.5} \geq 1.645 \quad \rightarrow \quad \bar{X} \geq 56.47$

3) Find the probability of a Type II error if $\mu=58$ $\qquad \beta(58) = P(\text{Accept } H_o \mid H_a \text{ is true } [\mu=58])$

4) Power = 1 – P(Type II Error): $\qquad PWR(58) = 1 - \beta(58)$

H_o: $\mu = 40$ mpg
H_A: $\mu < 40$

Population Standard Deviation: $\sigma = 6$ mpg

Significance Level: $\alpha = .01$

Sample Results: A SRS of $n = 16$ gives $\bar{X} = 36.7$

1) Write the rejection rule (RR) for H_o in terms of z-scores.

2) Write the rejection rule (RR) for H_o in terms of $\bar{X}$.

3) Find the probability of a Type II error if $\mu=38$ $\left[i.e.,\ \beta(38)\right]$

a) Find the sample z-score (z_s).

b) State a conclusion for the test at the $\alpha = .01$ level.

c) Find the p-value.

Study Design

The number of observations needed to detect a true difference $\Delta = \mu_A - \mu_o$
at the α level of significance with power=1-β is

- 1-sided alternative: $n = \frac{\sigma^2}{\Delta^2}\left(z_\alpha + z_\beta\right)^2$
- 2-sided alternative: $n = \frac{\sigma^2}{\Delta^2}\left(z_{\alpha/2} + z_\beta\right)^2$

Ex: Find the sample size needed to detect a 2 mpg difference at the α=.01 level with 80% power.

H_o: $\mu = 40$ mpg
H_A: $\mu < 40$
Population Standard Deviation: $\sigma = 6$ mpg