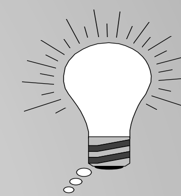
Statistical Test for Population Mean

- In statistical testing, we use deductive reasoning to specify what should happen if the conjecture or null hypothesis is true.
- A study is designed to collect data to challenge this hypothesis.
- We determine if what we expected to happen is supported by the data.
  - If it is not supported, we reject the conjecture.
  - If it cannot be rejected we (tentatively) fail to reject the conjecture.
- We have not proven the conjecture, we have simply not been able to disprove it with these data.

Logic Behind Statistical Tests

Statistical tests are based on the concept of **Proof by Contradiction**.



If P then Q iff **If NOT Q then NOT P**

Analogy with justice system

<u>Court of Law</u>

1. Start with premise that person is **innocent**. (Opposite is **guilty**.)
2. If enough evidence is found to show *beyond reasonable doubt* that person committed crime, reject premise. (Enough evidence that person is guilty.)
3. If not enough evidence found, we don't reject the premise. (Not enough evidence to conclude person guilty.)

<u>Statistical Hypothesis Test</u>

1. Start with **null hypothesis**, *status quo*. (Opposite is **alternative hypothesis**.)
2. If a *significant* amount of evidence is found to refute null hypothesis, reject it. (Enough evidence to conclude alternative is true.)
3. If not enough evidence found, we don't reject the null. (Not enough evidence to disprove null.)

Examples in Testing Logic

CLAIM:	A new variety of turf grass has been developed that is claimed to resist drought better than currently used varieties.
CONJECTURE:	The new variety resists drought no better than currently used varieties.
DEDUCTION:	If the new variety is no better than other varieties (P), then areas planted with the new variety should display the same number of surviving individuals (Q) after a fixed period without water than areas planted with other varieties.
CONCLUSION:	If more surviving individuals are observed for the new varieties than the other varieties (NOT Q), then we conclude the new variety is indeed not the same as the other varieties, but in fact is better (NOT P).

Five Parts of a Statistical Test

1. **Null Hypothesis (H_0:)**
2. **Alternative Hypothesis (H_A:)**
3. **Test Statistic (T.S.)**
   Computed from sample data.
   Sampling distribution is known if the Null Hypothesis is true.
4. **Rejection Region (R.R.)**
   Reject H_0 if the test statistic computed with the sample data is unlikely to come from the sampling distribution under the assumption that the Null Hypothesis is true.
5. **Conclusion:**
   Reject H_0 or Do Not Reject H_0

H_0: $\mu \leq 530$

H_A: $\mu > 530$

Other H_A:
$\mu < 530$
$\mu \neq 530$

T.S. $$z^* = \frac{\bar{y} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

R.R. $z^* > z_\alpha$

Or $z^* < -z_\alpha$

Or $|z^*| > z_{\alpha/2}$

Hypotheses

Research Hypotheses: The thing we are primarily interested in "proving".

$$H_A^1: \quad \mu > \mu_0$$
$$H_A^2: \quad \mu < \mu_0$$
$$H_A^3: \quad \mu \neq \mu_0$$

$$H_A^1: \quad \mu > 1.6\text{m}$$
$$H_A^2: \quad \mu < 1.6\text{m}$$
$$H_A^3: \quad \mu \neq 1.6\text{m}$$

Average height of the class

Null Hypothesis: Things are what they say they are, *status quo*.

$$H_0: \quad \mu = \mu_0$$

$$H_0: \quad \mu = 1.6\text{m}$$

(It's common practice to always write H_0 in this way, even though what is meant is the opposite of H_A in each case.)

Test Statistic

Some function of the data that uses estimates of the parameters we are interested in and whose sampling distribution is known when we assume the null hypothesis is true.

Most good test statistics are constructed using some form of a sample mean.

Why?

The Central Limit Theorem Of Statistics

Developing a Test Statistic for the Population Mean

Hypotheses of interest:

$H_A^1: \quad \mu > \mu_0$

$H_A^2: \quad \mu < \mu_0$

$H_A^3: \quad \mu > \mu_0 \quad \text{or} \quad \mu < \mu_0$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Test Statistic: Sample Mean

Under H_0: the sample mean has a sampling distribution that is normal with mean μ_0.

$$y \sim N(\mu_0, \sigma)$$

$$\bar{y} \sim N\left(\mu_0, \sigma/\sqrt{n}\right)$$

Under H_A: the sample mean has a sampling distribution that is also normal but with a mean μ_1 that is different from μ_0.

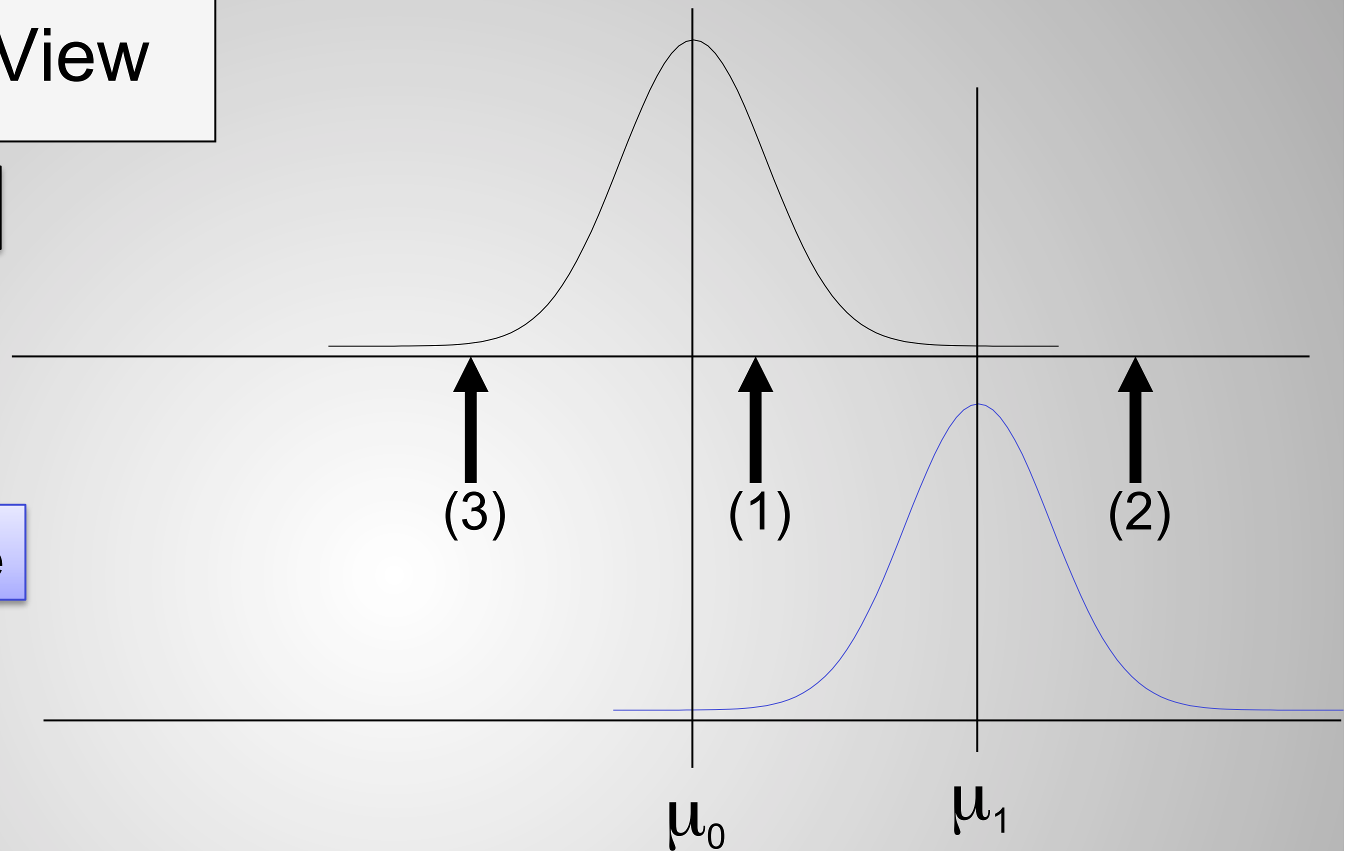
$$y \sim N(\mu_1, \sigma)$$

$$\bar{y} \sim N\left(\mu_1, \sigma/\sqrt{n}\right)$$

Graphical View

H_0: True

$H_A{}^1$: True

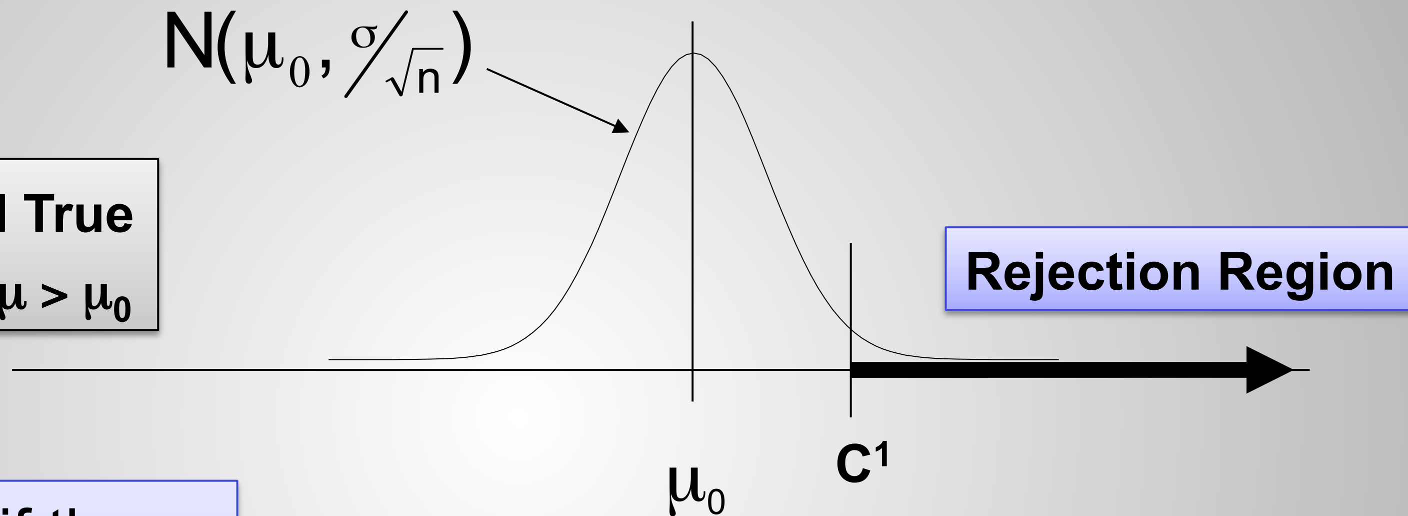


What would you conclude if the sample mean fell in location (1)? How about location (2)? Location (3)? Which is most likely 1, 2, or 3 when H_0 is true?

Rejection Region

H_0: Assumed True

Testing H_A^1: $\mu > \mu_0$



Reject H_0 if the sample mean is in the upper tail of the sampling distribution.

$$\text{If } \bar{y} > C^1 \ \textbf{Reject } H_0$$

How do we determine C^1?

Determining the Critical Value for the Rejection Region

Reject H_0 if the sample mean is larger than "expected".

If H_0 were true, we would expect 95% of sample means to be less than the upper limit of a 90% CI for μ.

$$C^1 = \mu_0 + 1.645 \frac{\sigma}{\sqrt{n}}$$

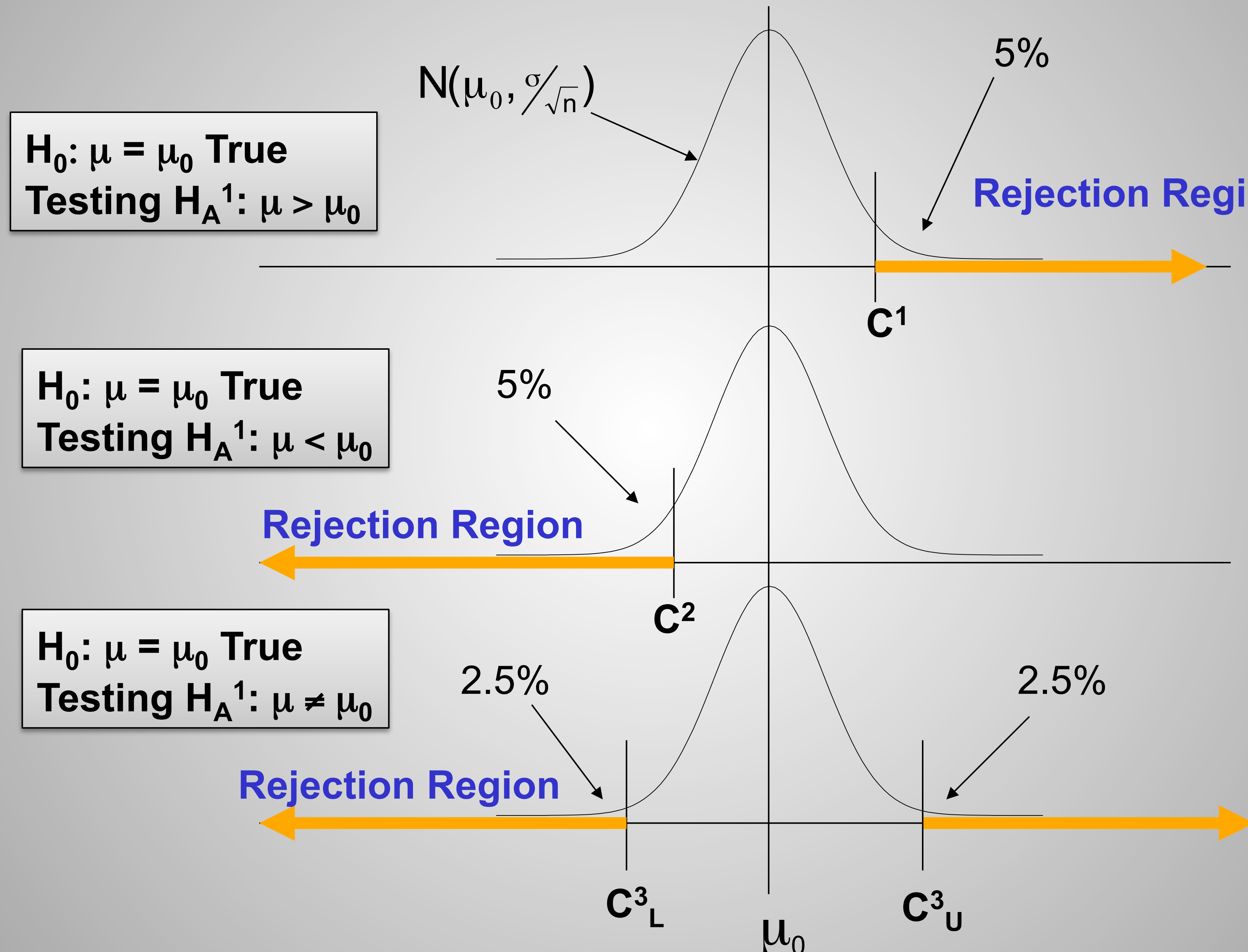
From the standard normal table.

In this case, if we use this critical value, in 5 out of 100 repetitions of the study we would reject H_o incorrectly. That is, we would make an error.

But, suppose $H_A{}^1$ is the true situation, then most sample means will be greater than C^1 and we will be making the correct decision to reject more often.

$$P\left(\frac{\bar{y} - \mu_0}{\sigma / \sqrt{n}} > 1.645 \right) = 0.05$$

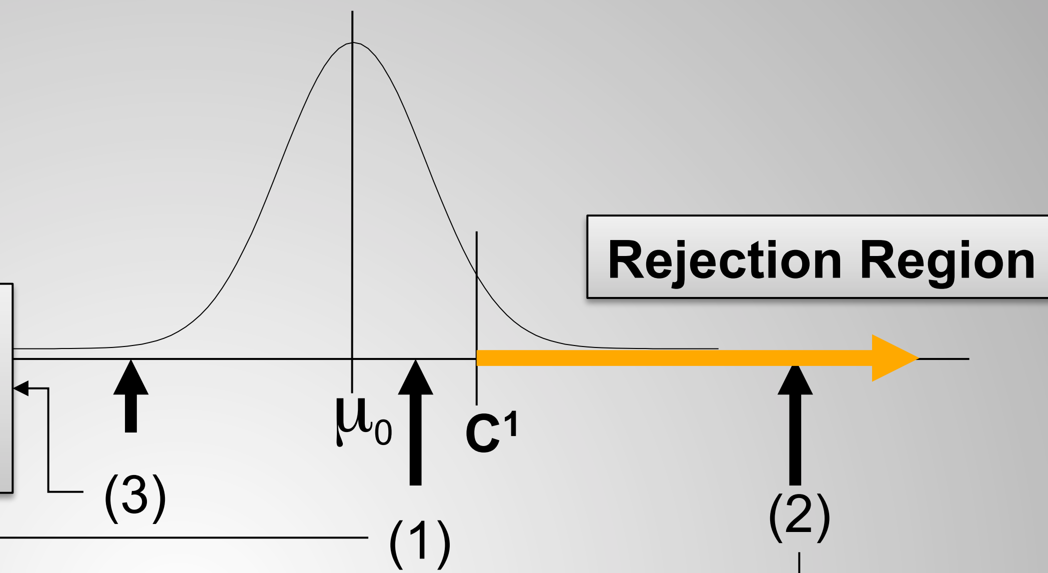
Rejection Regions for Different Alternative Hypotheses



Type I Error

H_0: True

If the sample mean is at location (1) or (3) we make the correct decision.



If the sample mean is at location (2) **and H_0 is actually true**, we make the wrong decision.

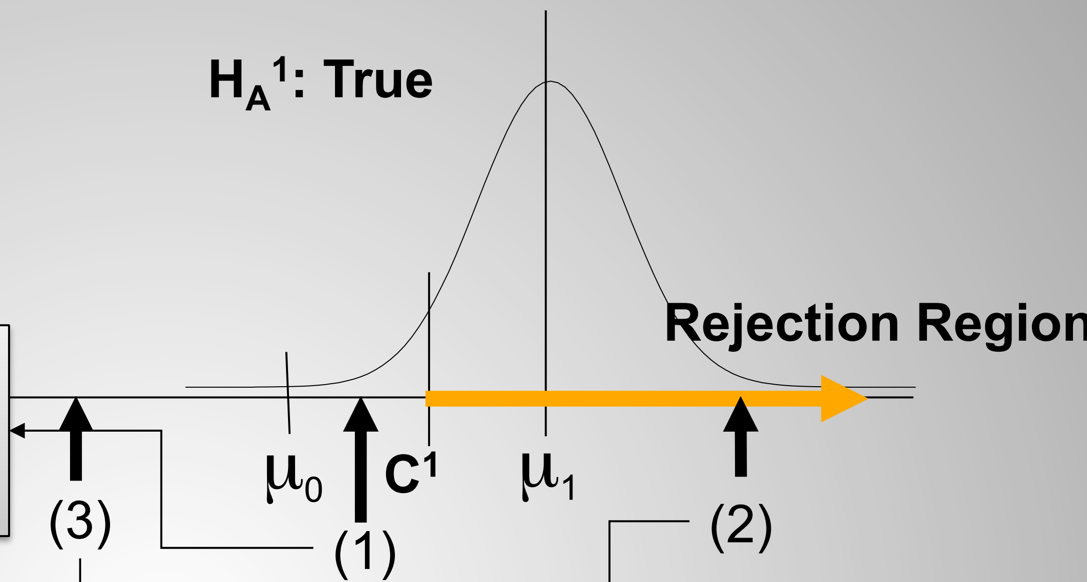
This is called making a **TYPE I error**, and the **probability** of making this error is usually denoted by the Greek letter α.

α = P(Reject H_0 when H_0 is is the true condition)

If $C^1 = \mu_0 + 1.645\dfrac{\sigma}{\sqrt{n}}$ then α = 1/20=5/100 or .05

If $C^1 = \mu_0 + z_\alpha \dfrac{\sigma}{\sqrt{n}}$ then the Type I error is α.

Type II Error



If the sample mean is at location (1) or (3) we make **the wrong decision**.

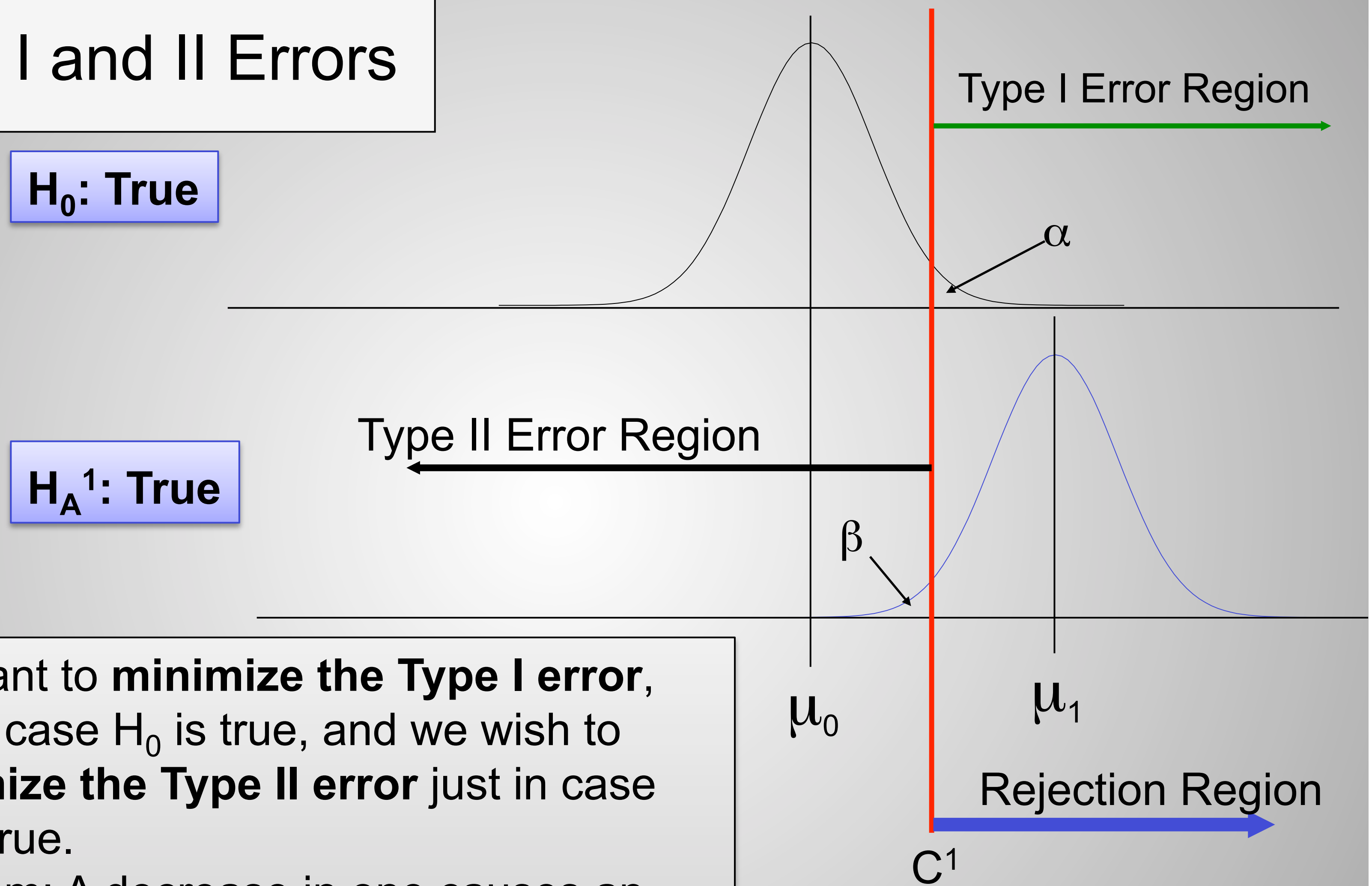
If the sample mean is at location (2) we make the correct decision.

This is called making a **TYPE II error**, and the probability of making this type error is usually denoted by the Greek letter β.

β = P(Do Not Reject H_0 when H_A is the true condition)

We will come back to the Type II error later.

Type I and II Errors



- We want to **minimize the Type I error**, just in case H_0 is true, and we wish to **minimize the Type II error** just in case H_A is true.
- Problem: A decrease in one causes an increase in the other.
- Also: We can never have a Type I or II error equal to zero!

Setting the Type I Error Rate

The solution to our quandary is to **set the Type I Error Rate** to be small, and hope for a small Type II error also. The logic is as follows:

1. Assume the data come from a population with **unknown distribution** but which has mean (μ) and standard deviation (σ).
2. For any sample of size **n** from this population we can compute a sample mean, $\bar{y}$.
3. Sample means from repeated studies having samples of size **n** will have the **sampling distribution** of $\bar{y}$ following a normal distribution with mean μ and a standard deviation of $\sigma/\sqrt{n}$ (the standard error of the mean). This is the Central Limit Theorem.
4. If the Null Hypothesis is true and $\mu = \mu_0$, then we deduce that with probability α we will observe a sample mean greater than $\mu_0 + z_\alpha \sigma/\sqrt{n}$. (For example, for $\alpha = 0.05$, z_α=1.645.)

Setting Rejection Regions for Given Type I Error

Following this same logic, rejection rules for all three possible alternative hypotheses can be derived.

Reject H_o if:

$H_A^1: \quad \mu > \mu_0$	$\dfrac{\bar{y} - \mu_0}{\sigma / \sqrt{n}} > z_\alpha$
$H_A^2: \quad \mu < \mu_0$	$\dfrac{\bar{y} - \mu_0}{\sigma / \sqrt{n}} < -z_\alpha$
$H_A^3: \quad \mu \neq \mu_0$	$\dfrac{\lvert\bar{y} - \mu_0\rvert}{\sigma / \sqrt{n}} > z_{\alpha/2}$

Note: It is just easier to work with a test statistic that is <u>standardized</u>. We only need the standard normal table to find critical values.

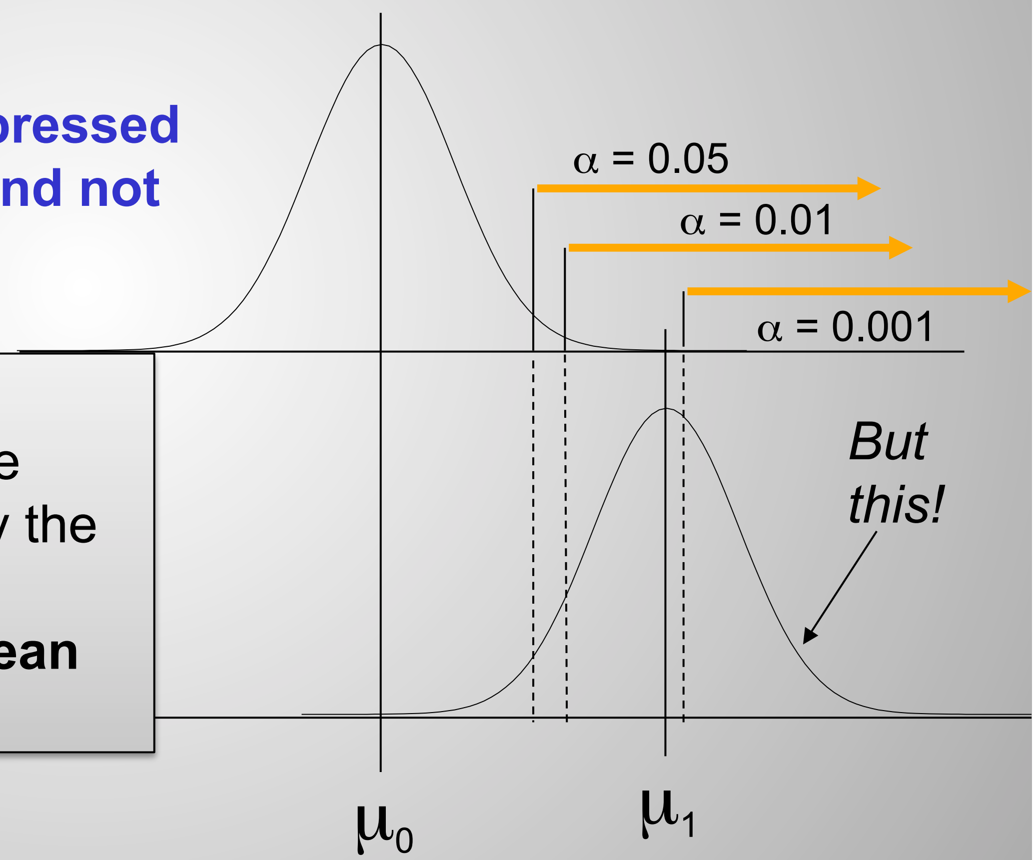
For (1-α)100% of repeated studies, if the true population mean is μ_0 as conjectured by H_0, then the decision concluded from the statistical test will be correct. On the other hand, in α100% of studies, the wrong decision (a Type I error) will be made.

Risk

The value $0 < \alpha < 1$ represents the risk **we** are willing to take of making the wrong decision.

But, what if I don't wish to take any risk?
Why not set $\alpha = 0$?

What if the true situation is expressed by the alternative hypothesis and not the null hypothesis?



Suppose H_A^1 is really the true situation. Then the samples come from the distribution described by the alternative hypothesis and the **sampling distribution of the mean** will have mean μ_1, not μ_0.

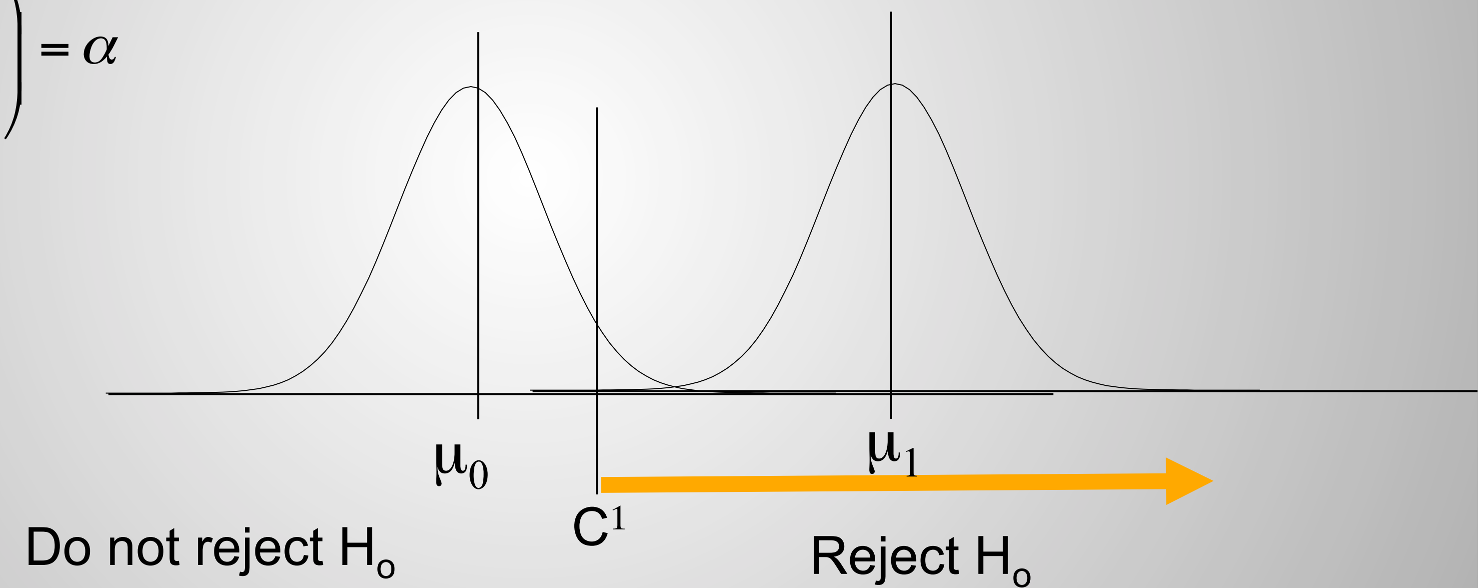
$$\bar{y} \sim N(\mu_1, \sigma/\sqrt{n})$$

The Rejection Rule

(at an α Type I error probability)

$$\text{Rule: Reject } H_0 \text{ if } \frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha$$ or if $$\bar{y} > C^1 = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$\Pr\left(\frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha\right) = \alpha$$

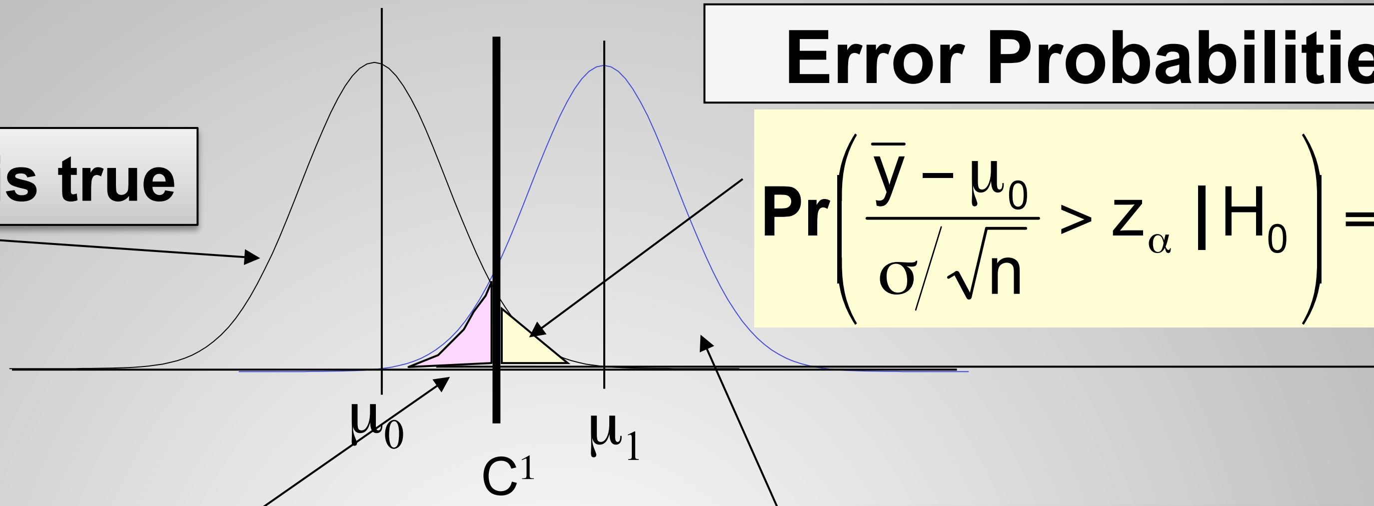


Pretty clear cut when μ_1 much greater than μ_0

Error Probabilities

When H_0 is true

$$\mathbf{Pr}\left(\frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha \mid H_0\right) = \alpha$$



When H_A is true

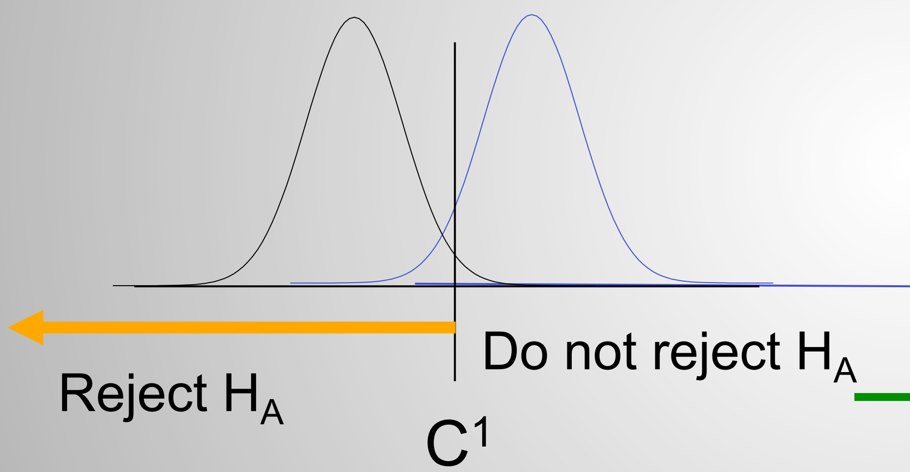
$$\mathbf{Pr}\left(\frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}} < z_\alpha \mid H_A\right) = \beta$$

$$\mathbf{Pr}\left(\frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha \mid H_A\right) = 1-\beta$$

If H_A is the true situation, then any sample whose mean $\bar{y}$ is larger than $\mu_0 + z_\alpha \sigma/\sqrt{n}$ will lead to the correct decision (reject H_0, accept H_A).

If H_A is the true situation

Then any sample such that its sample mean $\bar{y}$ is less than $\mu_0 + z_\alpha \sigma/\sqrt{n}$ will lead to the <u>wrong</u> decision (do not reject H_0, reject H_A).



		True Situation	
		H_0 : True	H_A : True
Decision	Reject H_0	Type I Error α	Correct $1 - \beta$
	Accept H_0	Correct $1 - \alpha$	Type II Error β

Computing Error Probabilities

$$\mathbf{Pr}\left(\frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha \,\middle|\, H_0\right) = \alpha$$

Type I Error Probability

(Reject H_0 when H_0 true)

$$\mathbf{Pr}\left(\frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha \,\middle|\, H_A\right) = 1-\beta$$

Power of the test.

(Reject H_0 when H_A true)

$$\mathbf{Pr}\left(\frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}} < z_\alpha \,\middle|\, H_A\right) = \beta$$

Type II Error Probability

(Reject H_A when H_A true)

Example

Sample Size: **n = 50**

Sample Mean: $\bar{y} = 40.1$

Sample Standard deviation: **s = 5.6**

H_0: $\mu = \mu_0 = 38$

H_A: $\mu > 38$

What risk are we willing to take that we reject H_0 when in fact H_0 is true?

P(Type I Error) = α = .05   Critical Value: $z_{.05}$ = 1.645

Rejection Region

$$\frac{\bar{y} - \mu_0}{\sigma / \sqrt{n}} > z_\alpha$$

$$\frac{40.1 - 38}{5.6/\sqrt{50}} = 2.651 > 1.645$$

Conclusion: Reject H_0

Type II Error.

To compute the Type II error for a hypothesis test, we need to have a **specific alternative hypothesis** stated, rather than a vague alternative

Vague

H_A: $\mu > \mu_0$

H_A: $\mu < \mu_0$

H_A: $\mu \neq \mu_0$

H_A: $\mu \neq 10$

Specific

H_A: $\mu = \mu_1 > \mu_0$

H_A: $\mu = \mu_1 < \mu_0$

H_A: $\mu = 5 < 10$

H_A: $\mu = 20 > 10$

Note:

As the difference between μ_0 and μ_1 gets **larger**, the probability of committing a Type II error (β) **decreases**.

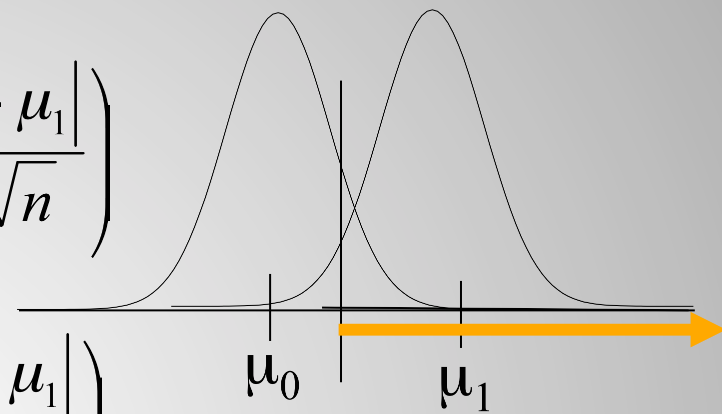
Significant Difference: $\Delta = \mu_0 - \mu_1$

Computing the probability of a Type II Error (β)

H_0: $\mu = \mu_0$

H_A: $\mu = \mu_1 > \mu_0$

$$\beta = \Pr\left(Z < z_{\alpha} - \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}} \right)$$



H_0: $\mu = \mu_0$

H_A: $\mu = \mu_1 < \mu_0$

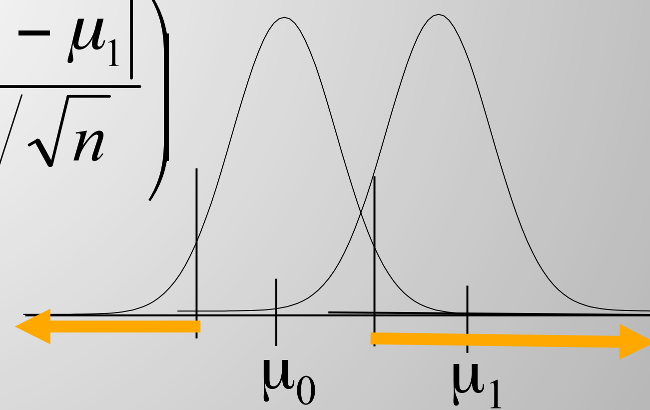
$$\beta = \Pr\left(Z < z_{\alpha} - \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}} \right)$$

H_0: $\mu = \mu_0$

H_a: $\mu \neq \mu_0$

$\mu = \mu_1 \neq \mu_0$

$$\beta = \Pr\left(Z < z_{\alpha/2} - \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}} \right)$$



$\Delta = |\mu_0 - \mu_1|$ = critical difference

1 - β = Power of the test

Example: Power

$n = 50$

$\bar{y} = 40.1 \qquad s = 5.6$

$\mathbf{Pr(}\text{Type I error}\mathbf{)} = \alpha = .05$

$z_\alpha = 1.645$

$H_0 : \mu = \mu_0 = 38$

$H_A : \mu = \mu_1 = 40$

$$\beta = \mathbf{Pr}\left(Z < z_\alpha - \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}} \right) = \mathbf{Pr}\left(Z < 1.645 - \frac{|\Delta|}{\sigma/\sqrt{50}} \right)$$

$$= \mathbf{Pr}\left(Z < 1.645 - \frac{|\Delta|}{\sigma}\sqrt{n} \right)$$

Assuming s is actually equal to σ (usually what is done):

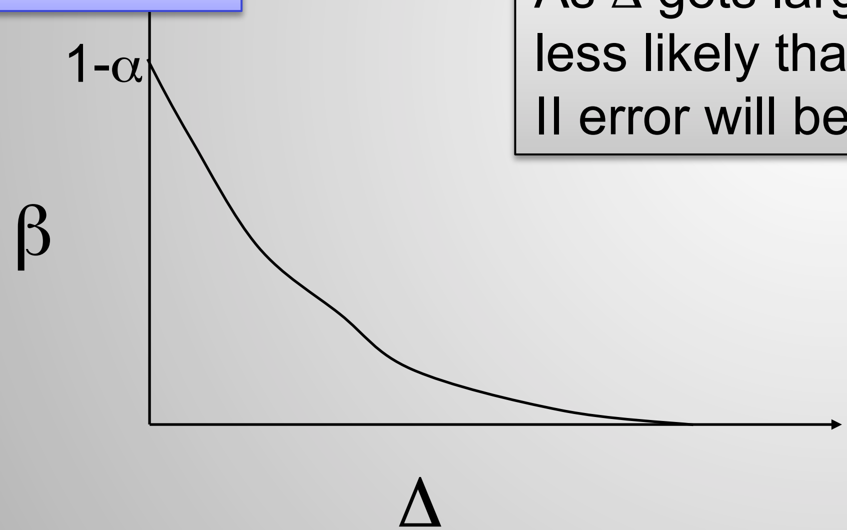
$$\beta = \mathbf{Pr}\left(Z < 1.645 - \frac{2}{5.6}(\sqrt{50}) \right) = \mathbf{Pr(}Z < -0.88\mathbf{)} = 0.1894$$

Power versus Δ (fixed n and σ)

$$\beta = \Pr\left(Z < z_\alpha - \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}} \right) = \Pr\left(Z < z_\alpha - \frac{|\Delta|}{\sigma}\sqrt{n} \right)$$

Type II error

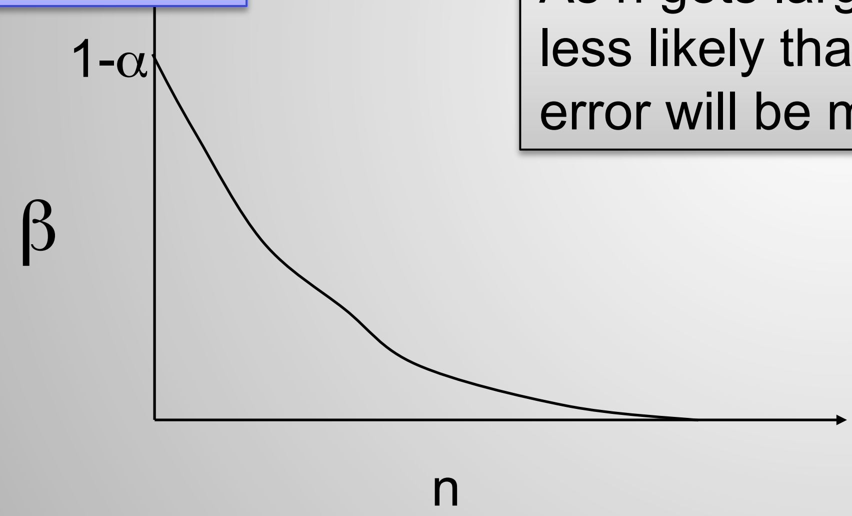
As Δ gets larger, it is less likely that a Type II error will be made.



Power versus n (fixed Δ and σ)

$$\beta = \Pr\left(Z < z_\alpha - \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}} \right) = \Pr\left(Z < z_\alpha - \frac{|\Delta|}{\sigma}\sqrt{n} \right)$$

Type II error



As n gets larger, it is less likely that a Type II error will be made.

What happens for larger σ?

Power vs. Sample Size

$$\beta = \Pr\left(Z < z_\alpha - \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}} \right) = \Pr\left(Z < 1.645 - \frac{|\Delta|}{\sigma}\sqrt{n} \right)$$

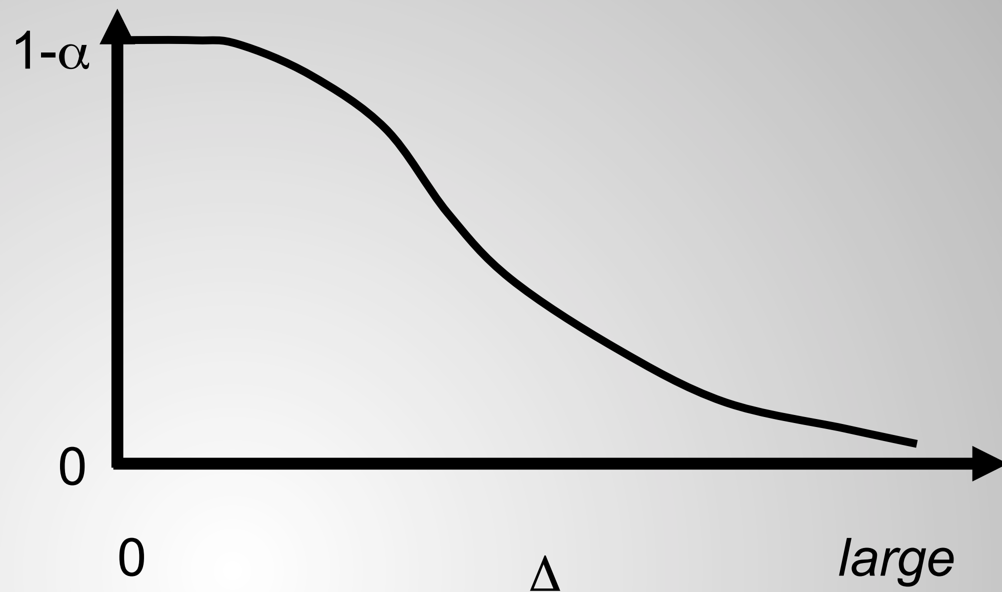
Power = 1-β

Δ	n	β	power
1	50	.755	**.245**
2	50	.286	**.714**
4	50	.001	**.999**

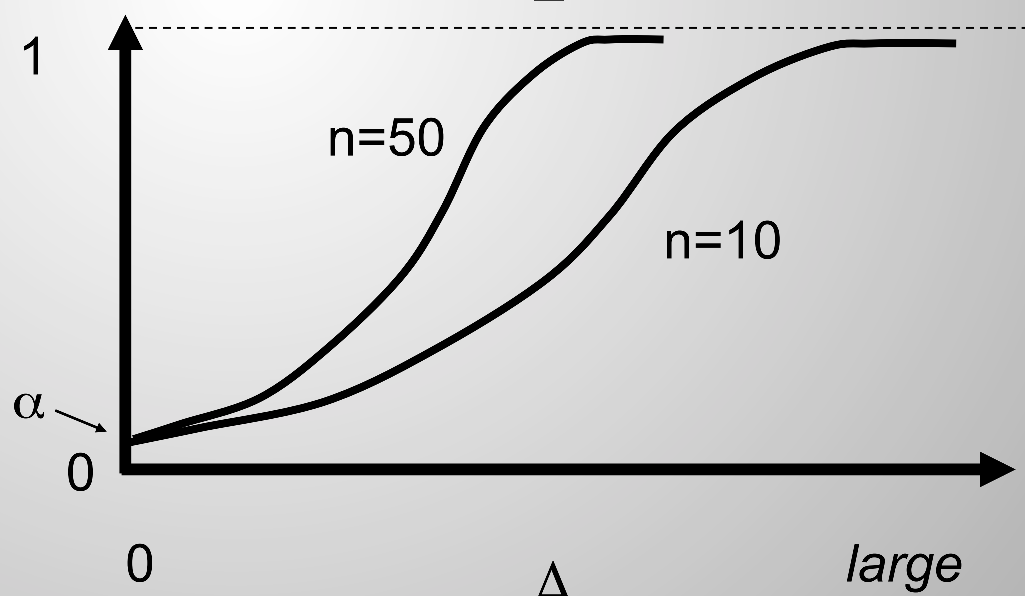
Δ	n	β	power
1	25	.857	**.143**
2	25	.565	**.435**
4	25	.054	**.946**

Power Curve

Pr(Type II Error)
β



Power
1−β



Summary

$$\beta = \Pr\left(Z < z_\alpha - \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}}\right) = \Pr\left(Z < z_\alpha - \frac{|\mu_0 - \mu_1|}{\sigma}\sqrt{n}\right)$$

$$= \Pr\left(Z < z_\alpha - \frac{|\Delta|}{\sigma}\sqrt{n}\right)$$

1) **For fixed σ and n,**

   β decreases (power increases) as Δ increases.

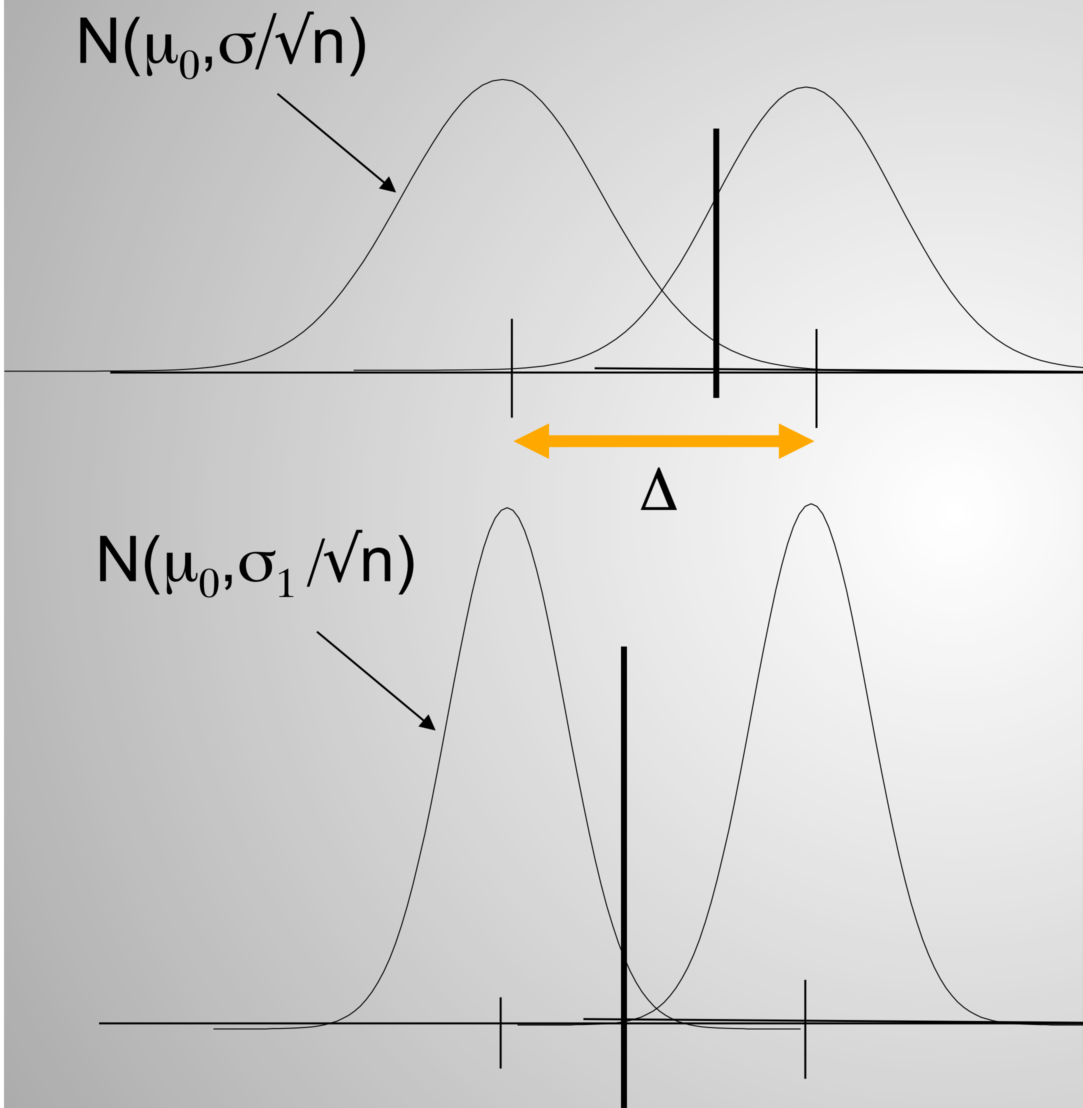
2) **For fixed σ and Δ,**

   β decreases (power increases) as n increases.

3) **For fixed Δ and n,**

   β decreases (power increases) as σ decreases.

Increasing Precision for fixed Δ increases Power



$\sigma_1 < \sigma$

Decreasing σ **decreases** the spread in the sampling dist of $\bar{y}$

Note: z_α changes.

Same thing happens if you increase n.

Same thing happens if Δ is increased.

Sample Size Determination

1) Specify the critical difference, Δ (assume σ is known).

2) Choose **P(Type I error) = α** and **Pr(Type II error) = β** based on traditional and/or personal levels of risk.

3) One-sided tests:

$$n = \frac{\sigma^2}{\Delta^2}(z_\alpha + z_\beta)^2$$

4) Two-sided tests:

$$n = \frac{\sigma^2}{\Delta^2}(z_{\alpha/2} + z_\beta)^2$$

Example: One-sided test, $\sigma = 5.6$, and we wish to show a difference of $\Delta = .5$ as significant (i.e. reject H_0: $\mu = \mu_0$ for H_A: $\mu = \mu^1 = \mu_0 + \Delta$) with $\alpha = .05$ and $\beta = .2$.

$$n = \frac{(5.6)^2}{(.5)^2}(1.65 + 0.84)^2 = 777.7 \sim 778$$