Confidence Intervals & Hypothesis Testing

- Types of error and power

Types of error

- Type I error – significance level of the test
  α=P(reject H_0 | H_0 is true)
- Occurs when we incorrectly reject the null
- We take a sample from the population, calculate the statistics, and make inference about the true population. If we did this repeatedly, we would incorrectly reject the null 5% of the time *that it is true* if α is set to 0.05.

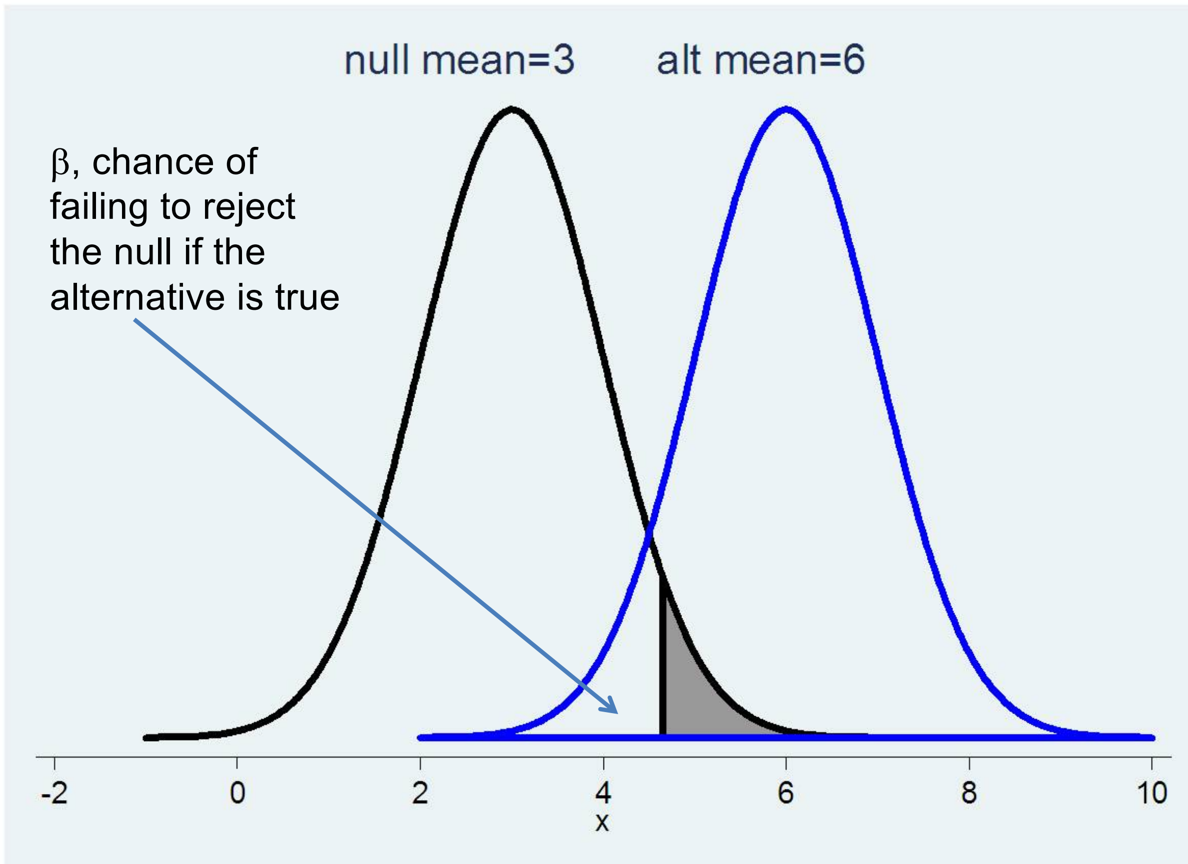
Types of error

- Type II error – β = P(do not reject H_0 | H_0 is false)

This occurs when we incorrectly fail to reject the null, i.e. when we do not reject the null *when the null is false*

	True state	
Decision	**H_0 is true**	**H_0 is false**
Do not reject H_0	Correct	Type II error=β
Reject H_0	Type I error=α	Correct

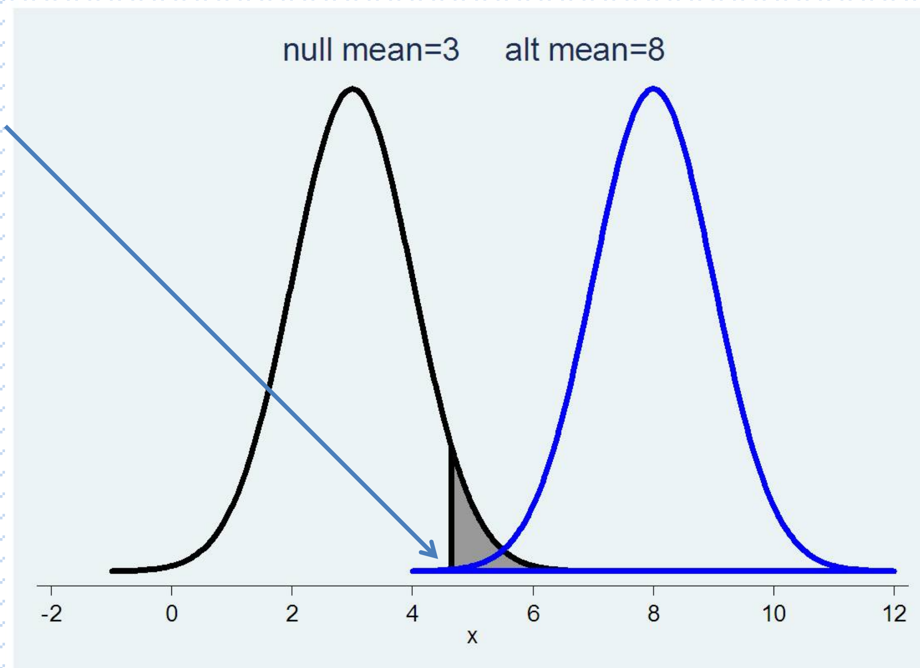
Chance of a type II error



Chance of a type II error

If the alternative is very different from the null, the chance of a Type II error is low

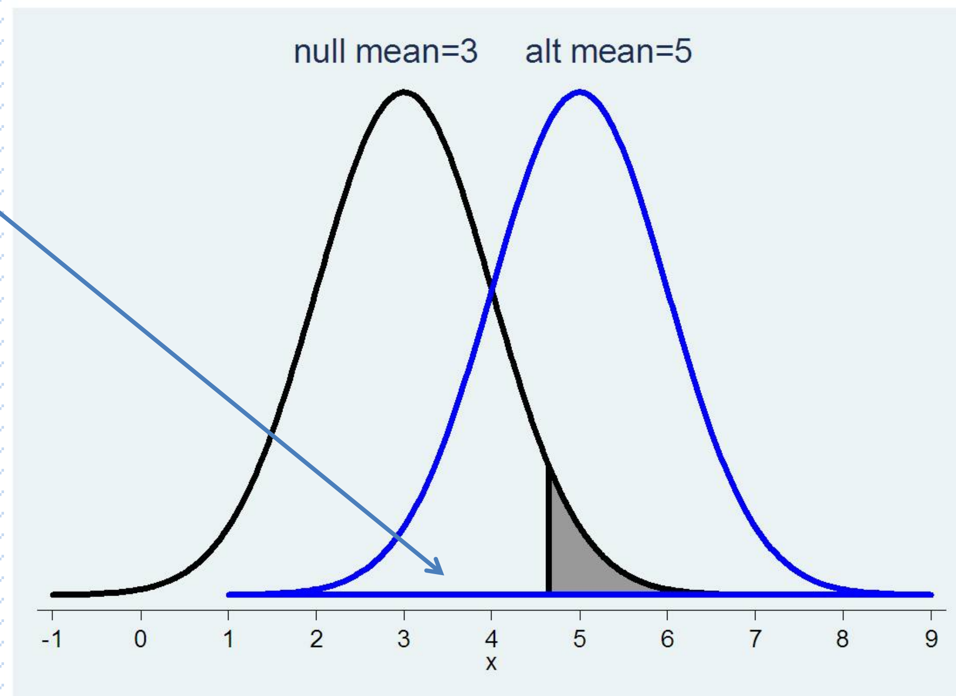
β, chance of failing to reject the null if the alternative is true



Chance of a type II error

If the alternative is not very different from the null, the chance of a Type II error is high

β, chance of failing to reject the null if the alternative is true



Finding β, P(Type II error)

- Example: Mean age of walking
  - H_0: μ< 11.4 months (μ_0) H_A: μ>11.4 months
  - Known SD=2
  - Sample size=9

- We will reject the null if the z_{stat} (assuming σ known) > 1.645
- So we will reject the null if $z_{stat} \geq \dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \geq 1.645$

$$\bar{x} \geq 1.645\sigma/\sqrt{n} + \mu_0$$

- For our example, the null will be rejected if
  $\bar{X}$> 1.645*2/3 + 11.4 > 12.5

Finding β, P(Type II error)

- If the true mean μ is really 14 meaning the null is false, what is the probability that the null will not be rejected? Type II error?
- The null will be rejected if the sample mean is >12.5, not rejected if it is ≤12.5
- What is the probability of getting a sample mean of ≤12.5 if the true mean is 14? This is a z statistic.
- P(Z<(12.5-14)/(2/3))

```
di normal((12.5-14)/(2/3))
.01222447
```

So if the true mean is 14 and the sample size is 9, the probability of incorrectly failing to reject the null is 0.012

Finding β, P(Type II error)

- Note that this depended on this other population mean (e.g. 14)
  - The chance of failing to reject the null will increase if this true population mean is closer to the null value
- What is the probability of failing to reject the null if the true population mean is 13?
  - P(Z<(12.5-13)/(2/3))

    ```
    di normal((12.5-13)/(2/3))
    .22662735
    ```
- So for an alternative mean that is closer to the null value, the chances of failing to reject are higher

Power

- 1-β is called the power of a statistical test
- The power of a test is the probability that you will reject the null hypothesis when you should
- You can construct a power curve by plotting power versus different alternative hypotheses
- You can also construct a power curve by plotting power versus different sample sizes (n's)
  - This curve will allow you to see what gains you might have versus cost of adding participants
  - Power curves are not linear

Power

- The power of a statistical test is lower for alternative values that are closer to the null hypothesis value.
    - When the alternate mean was 14, power =1-.012=.988
    - When the alternate mean was 13, power was 1-.227=.773

- The power of a statistical test can also be increased by increasing n

- For a fixed n, the power of a statistical test can be increased by increasing α, the probability of a Type I error

Power

- The balance between type I error and type II error (or power) should depend on the context of the study

- When setting up a study, most investigators make sure that there will be at least 80% power.

- Sample size formulas allow you to specify the desired α and β levels.

Why is this kind of analysis important?

- The power calculation for a research study can determine the sample size OR
- Your sample size will tell you how much power/level of significance you have for your hypotheses
- How do we build this into our research and statistical plans?