

Power Analysis (in R)

Josh Sumner

Outline

- **Scientific Process**
- **Definitions**
- **Statistical Power and Significance**
- **Statistics and Power Analysis Workflow**
 - Formulaic vs Simulated
- **Example Scenarios (5)**

The Scientific Process

- 1) Theorize question or problem
- 2) Develop Hypotheses
- 3) Design Experiment
- 4) Record observations
- 5) Analyze observations

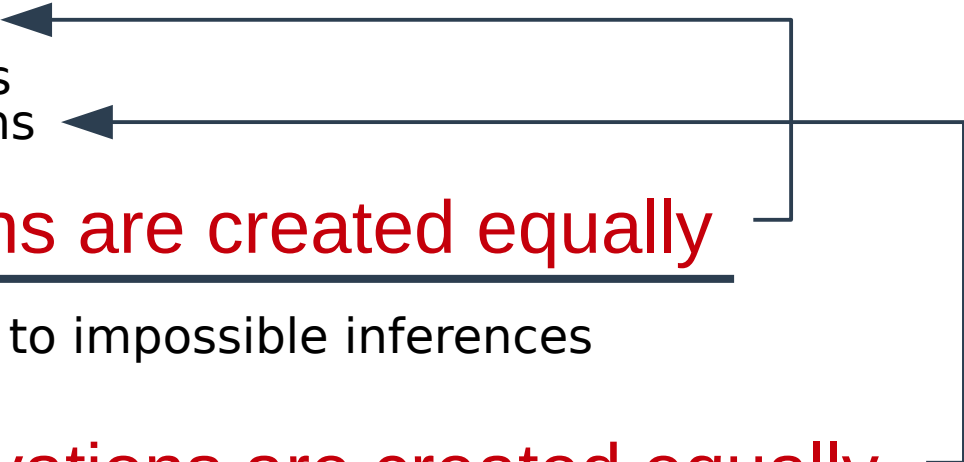
The Scientific Process

- 1) Theorize question or problem
- 2) Develop Hypotheses
- 3) Design Experiment
- 4) Record observations
- 5) Analyze observations

- Not all designs are created equally

- poor design leads to impossible inferences

The Scientific Process

- 1) Theorize question or problem
 - 2) Develop Hypotheses
 - 3) Design Experiment
 - 4) Record observations
 - 5) Analyze observations
- 
- ```
graph TD; 4[4) Record observations] --> 3[3) Design Experiment]; 5[5) Analyze observations] --> 3;
```

- Not all designs are created equally

- poor design leads to impossible inferences

- Not all observations are created equally

- poor statistical methods leads to false-positives and false-negatives
- **continuous > ordinal >= nominal**

# The Scientific Process

- 1) Theorize question or problem
- 2) Develop Hypotheses
- 3) Design Experiment
- 4) Record observations
- 5) Analyze observations

## - Not all designs are created equally

- poor design leads to impossible inferences

## - Not all observations are created equally

- poor statistical methods leads to false-positives and false-negatives
- **continuous > ordinal >= nominal**

## - Not all methods are created equally

- many methods can only test a handful of hypotheses and use p-values

# The Scientific Process

- 1) Theorize question or problem
- 2) Develop Hypotheses
- 3) Design Experiment
- 4) Record observations
- 5) Analyze observations

## - Before starting you should consider

---

What outcomes are possible?

Are the questions you're interested in answerable with the design?

What format will your data be in?

What test will you use and what hypotheses are possible in it?

Do you have the replication required?

# The Scientific Process

- 1) Theorize question or problem
- 2) Develop Hypotheses
- 3) Design Experiment
- 4) Record observations
- 5) Analyze observations

The **Stats in RCR** Workshop gets into These steps some And talks about how To know when you're Off of the flow chart.

The **Troubleshooting In R** workshop gets into some coding problems in this step.

## - Before starting you should consider

What outcomes are possible?

Are the questions you're interested in answerable with the design?

What format will your data be in?

What test will you use and what hypotheses are possible in it?

Do you have the replication required?

Focus of the **Stats In R** and **Stats in pcvr** Workshops, but will be part of today as well



# The Scientific Process

- 1) Theorize question or problem
- 2) Develop Hypotheses
- 3) Design Experiment
- 4) Record observations
- 5) Analyze observations

## - Before starting you should consider

What outcomes are possible?

Are the questions you're interested in answerable with the design?

What format will your data be in?

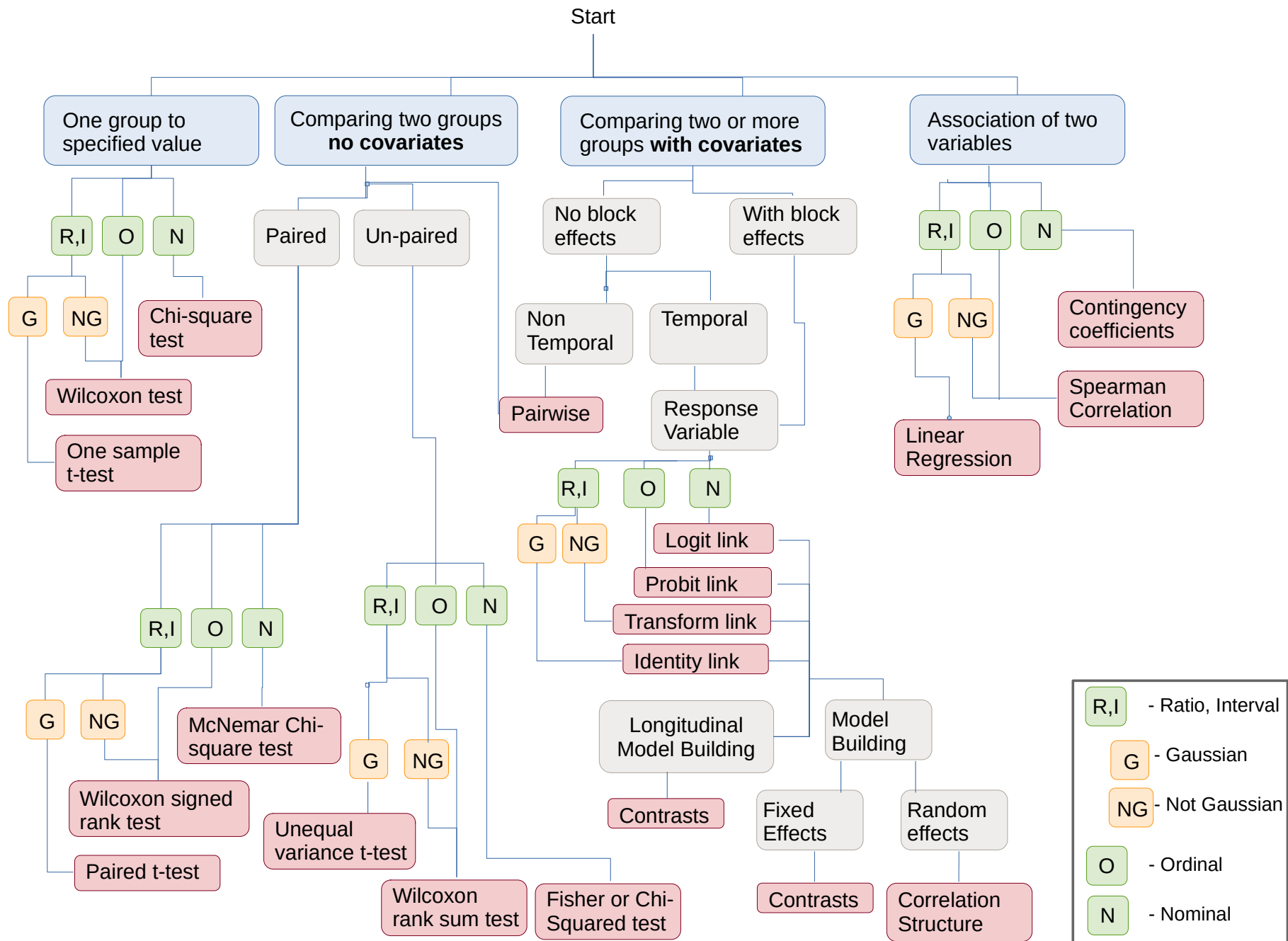
What test will you use and what hypotheses are possible in it?

Do you have the replication required?

Power analysis is useful in method development but here we'll focus on experimental design

# Outline

- Scientific Process
- **Definitions**
- **Statistical Power and Significance**
  - Formulaic vs Simulated
- **Statistics and Power Analysis Workflow**
- **Example Scenarios**



\* If the outcome and main predictor are both two-level factors, Breslow-Day and Cochren-Mantel-Hanzel tests are better

R,I

- **Ratio, Interval**

- \* Continuous scale measurements
- \* Ex: 1.3, 3.45, 2.98, ...
  - \* Plant height, CFUs, q-PCR, Watersoaking area, etc

O

- **Ordinal**

- \* Ordered discrete scale measurements
- \* Ex: 19, 13, 18, ...
  - \* Severity scores, hull vertices, number of leaves, etc

N

- **Nominal**

- \* Non-ordered discrete scale measurements
- \* Ex: Red, Yellow, Cyan, ...
  - \* Diseased vs not-diseased, dead vs alive, punnett squares, etc

R,I

- **Ratio, Interval**
  - Continuous scale measurements
  - Ex: 1.3, 3.45, 2.98, ...
    - Plant height, CFUs, q-PCR, Watersoaking area, etc

R,I

- **Ratio, Interval**

- Continuous scale measurements
- Ex: 1.3, 3.45, 2.98, ...
  - Plant height, CFUs, q-PCR, Watersoaking area, etc

NG

- **Not Gaussian distributed**

- Usually means non-parametric test
  - If it follows a different distribution,  
likelihood ratio test or other methods using  
that distribution

G

- **Gaussian distributed**

- Parametric testing

R,I

- **Ratio, Interval**

- Continuous scale measurements
- Ex: 1.3, 3.45, 2.98, ...
  - Plant height, CFUs, q-PCR, Watersoaking area, etc

NG

- **Not Gaussian distributed**

- Usually means non-parametric test
- If it follows a different distribution,  
likelihood ratio test or other methods using  
that distribution

G

- **Gaussian distributed**

- Parametric testing

**- Is data gaussian?**

- 1) Visualize data
- 2) shapiro.test()
- 3) ks.test()

```
> shapiro.test(dat$Values[dat$Group == "Gaussian"])
```

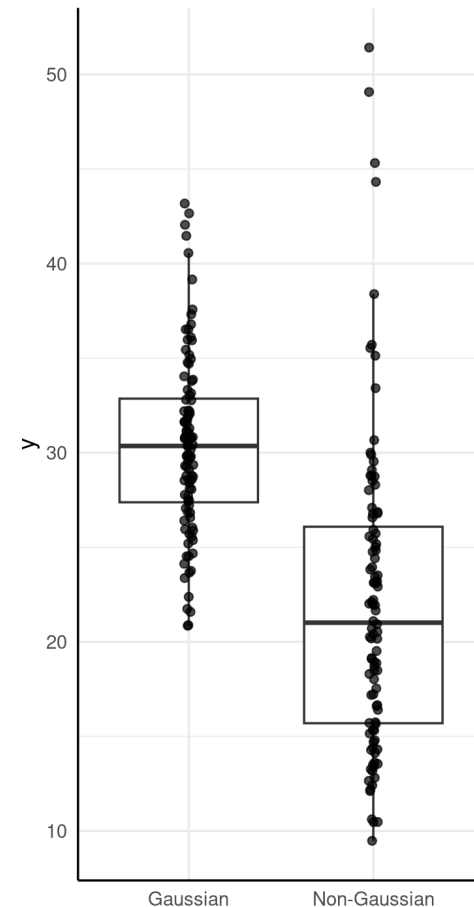
Shapiro-Wilk normality test

```
data: dat$Values[dat$Group == "Gaussian"]
W = 0.98484, p-value = 0.3094
```

```
> shapiro.test(dat$Values[dat$Group == "Not Gaussian"])
```

Shapiro-Wilk normality test

```
data: dat$Values[dat$Group == "Not Gaussian"]
W = 0.86899, p-value = 6.384e-08
```



R,I

- **Ratio, Interval**

- Continuous scale measurements
- Ex: 1.3, 3.45, 2.98, ...
  - Plant height, CFUs, q-PCR, Watersoaking area, etc

NG

- **Not Gaussian distributed**

- Usually means non-parametric test
- If it follows a different distribution,  
likelihood ratio test or other methods using  
that distribution

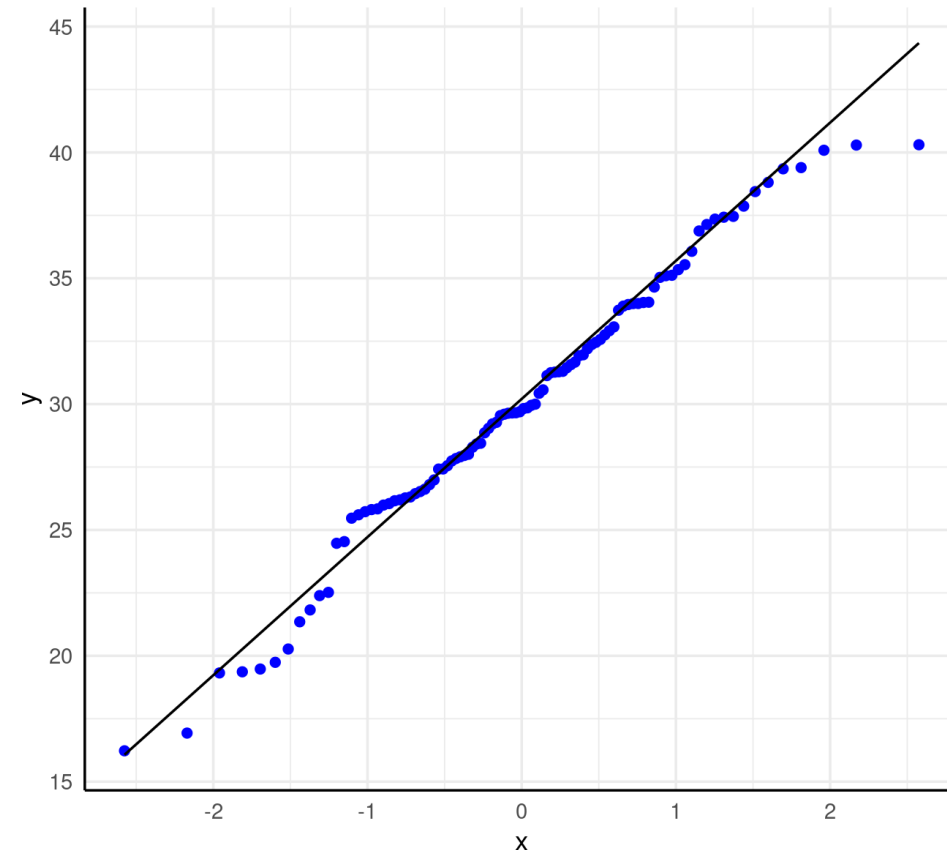
G

- **Gaussian distributed**
- Parametric testing

**- Is data gaussian?**

- 1) Visualize data
  - QQ plots
- 2) shapiro.test()
- 3) ks.test()

Gaussian Sample QQ plot





R,I

- Ratio, Interval

- Continuous scale measurements
- Ex: 1.3, 3.45, 2.98, ...
  - Plant height, CFUs, q-PCR, Watersoaking area, etc

NG

- Not Gaussian distributed

- Usually means non-parametric test
- If it follows a different distribution,  
likelihood ratio test or other methods using  
that distribution

G

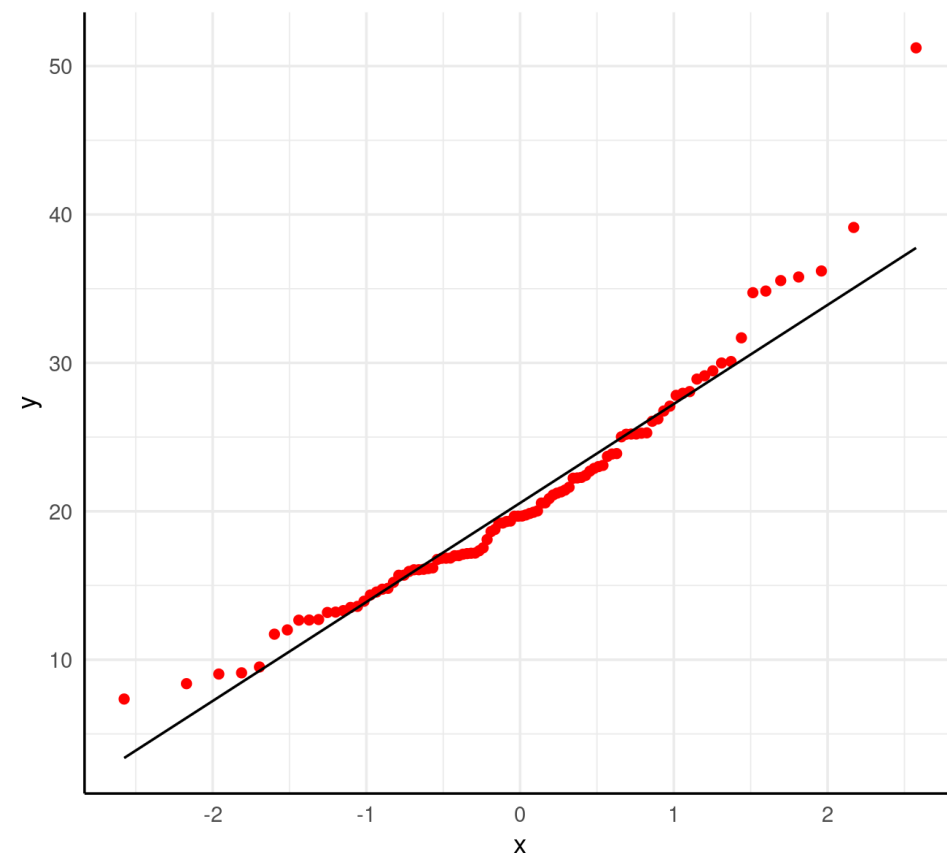
- Gaussian distributed

- Parametric testing

- Is data gaussian?

- 1) Visualize data
  - QQ plots
- 2) shapiro.test()
- 3) ks.test()

Non-Gaussian Sample QQ plot



R,I

- Ratio, Interval

- Continuous scale measurements
- Ex: 1.3, 3.45, 2.98, ...
  - Plant height, CFUs, q-PCR, Watersoaking area, etc

NG

- Not Gaussian distributed

- Usually means non-parametric test
- If it follows a different distribution,  
likelihood ratio test or other methods using  
that distribution

G

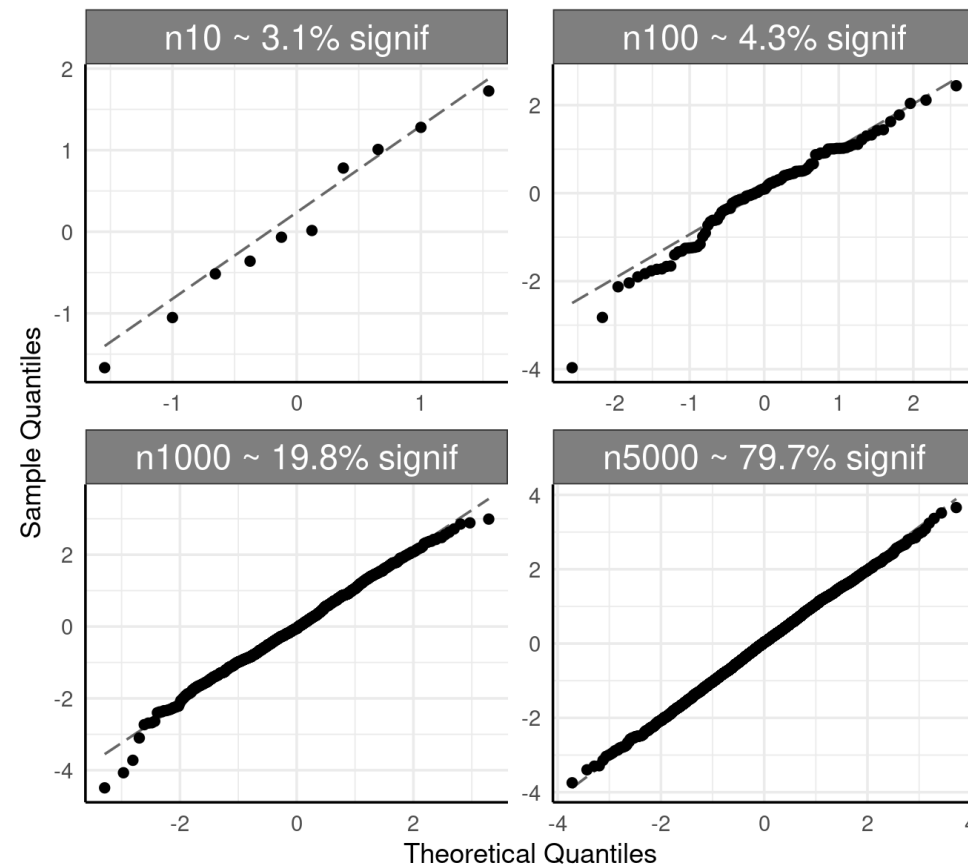
- Gaussian distributed
- Parametric testing

- Is data gaussian?

- 1) Visualize data
  - QQ plots
- 2) shapiro.test()
- 3) ks.test()

Broadly,  
visualizations are  
more useful than  
tests of normality.

Percentage of Significant Shapiro Tests



R,I

- **Ratio, Interval**

- Continuous scale measurements
- Ex: 1.3, 3.45, 2.98, ...
  - Plant height, CFUs, q-PCR, Watersoaking area, etc

NG

- **Not Gaussian distributed**

- Usually means non-parametric test
- If it follows a different distribution,  
likelihood ratio test or other methods using  
that distribution

G

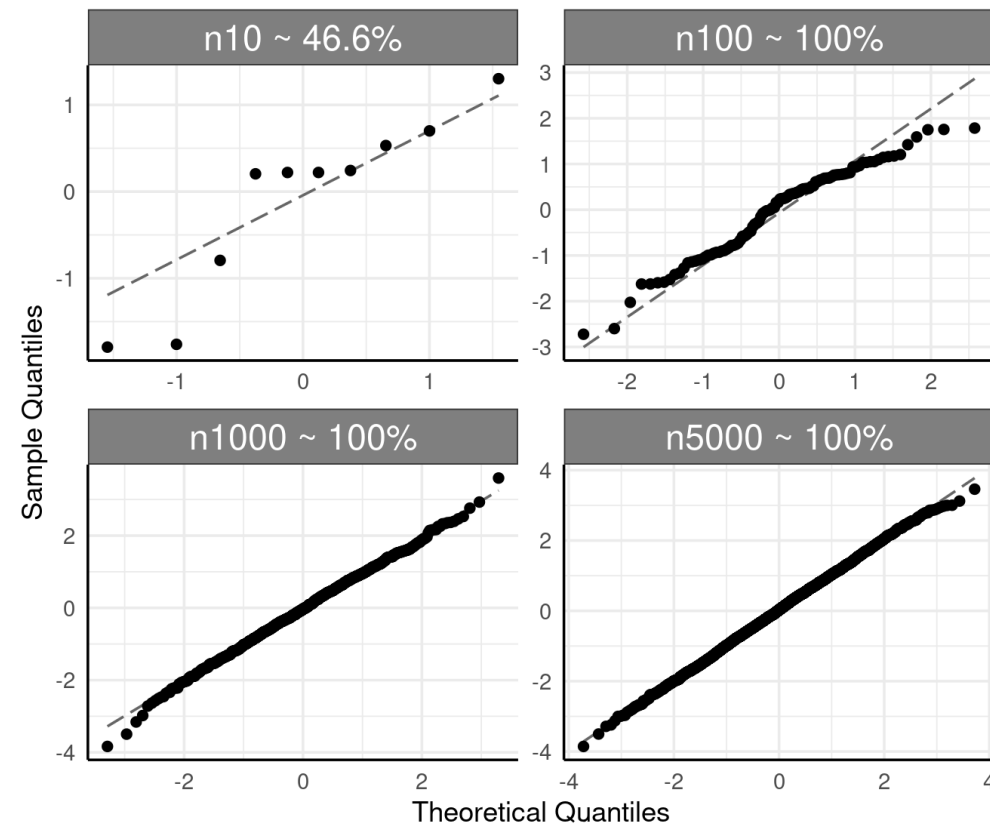
- **Gaussian distributed**
  - Parametric testing

- **Is data gaussian?**

- 1) Visualize data
  - QQ plots
- 2) shapiro.test()
- 3) ks.test()

Broadly,  
visualizations are  
more useful than  
tests of normality.

Percentage of Significant KS Tests



R,I

- Ratio, Interval

- Continuous scale measurements
- Ex: 1.3, 3.45, 2.98, ...
  - Plant height, CFUs, q-PCR, Watersoaking area, etc

NG

- Not Gaussian distributed

- Usually means non-parametric test
- If it follows a different distribution,  
likelihood ratio test or other methods using  
that distribution

G

- Gaussian distributed

- Parametric testing

O

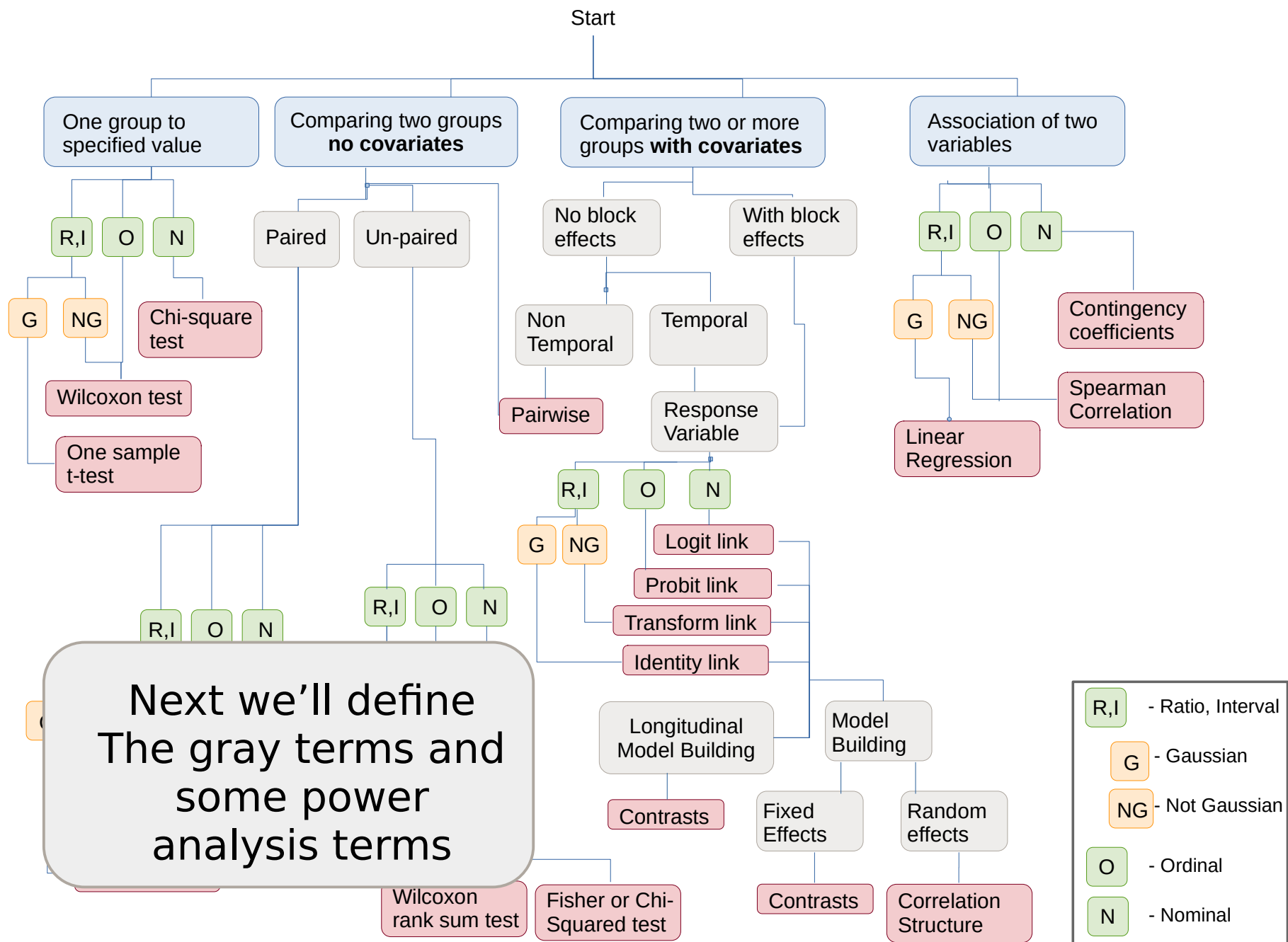
- Ordinal

- Ordered discrete scale measurements
- Ex: 19, 13, 18, ...
  - Severity scores, hull vertices, number of leaves, etc

N

- Nominal

- Non-ordered discrete scale measurements
- Ex: Red, Yellow, Cyan, ...
  - Diseased vs not-diseased, dead vs alive, punnett squares, etc



\* If the outcome and main predictor are both two-level factors, Breslow-Day and Cochren-Mantel-Hanzel tests are better

## Type 1 Error

- **A false positive**
  - Detecting an effect when none truly exists.

## Type 2 Error

- **A false negative**
  - Failing to detect an effect that does truly exist.

Type 1  
Error

- **A false positive**
  - Detecting an effect when none truly exists.

Type 2  
Error

- **A false negative**
  - Failing to detect an effect that does truly exist.

Alpha

- **Probability of a false Positive**
  - This is generally set to 0.05 just based on convention

Power

- **Probability of a false negative**
  - Generally we aim for >80% power

Type 1  
Error

- **A false positive**
  - Detecting an effect when none truly exists.

Type 2  
Error

- **A false negative**
  - Failing to detect an effect that does truly exist.

Alpha

- **Probability of a false Positive**
  - This is generally set to 0.05 just based on convention

Power

- **Probability of a false negative**
  - Generally we aim for >80% power

Statistical  
Significance

- **Conditional probability of the data given no effect**
  - Having a P-value below alpha, used for saying that the Null hypothesis is unlikely.

Effect Size

- **The magnitude of an effect**
  - Stronger effects are easier to detect.



Type 1  
Error

- **A false positive**
  - Detecting an effect when none truly exists.

Type 2  
Error

- **A false negative**
  - Failing to detect an effect that does truly exist.

Alpha

- **Probability of a false Positive**
  - This is generally set to 0.05 just based on convention

Power

- **Probability of a false negative**
  - Generally we aim for >80% power

Statistical  
Significance

- **Conditional probability of the data given no effect**
  - Having a P-value below alpha, used for saying that the Null hypothesis is unlikely.

Effect Size

- **The magnitude of an effect**
  - Stronger effects are easier to detect.

Null  
Hypothesis

- **The hypothesis you aim to reject with your experiment**
  - This is almost always that there is no effect.

Alternative  
Hypothesis

- **The hypothesis you propose in place of the Null**
  - This is the hypothesis you would state in your introduction

# Outline

- Scientific Process
- Definitions
- **Statistical Power and Significance**
  - Formulaic vs Simulated
- **Statistics and Power Analysis Workflow**
- **Example Scenarios**

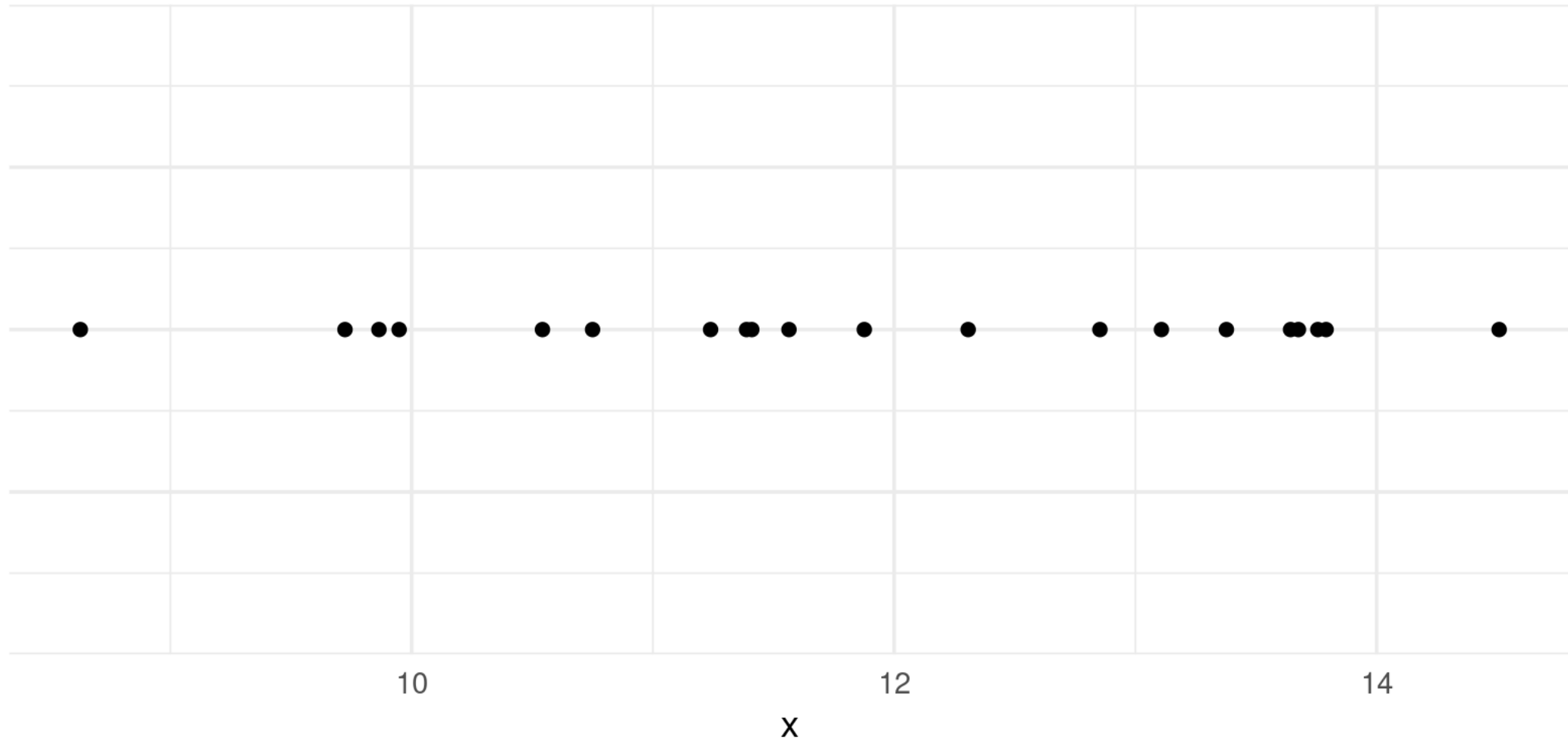
That was a lot of definitions. If you have questions let's go over them, then let's take a break.

# Statistical Significance

- What goes into a statistically significant result?

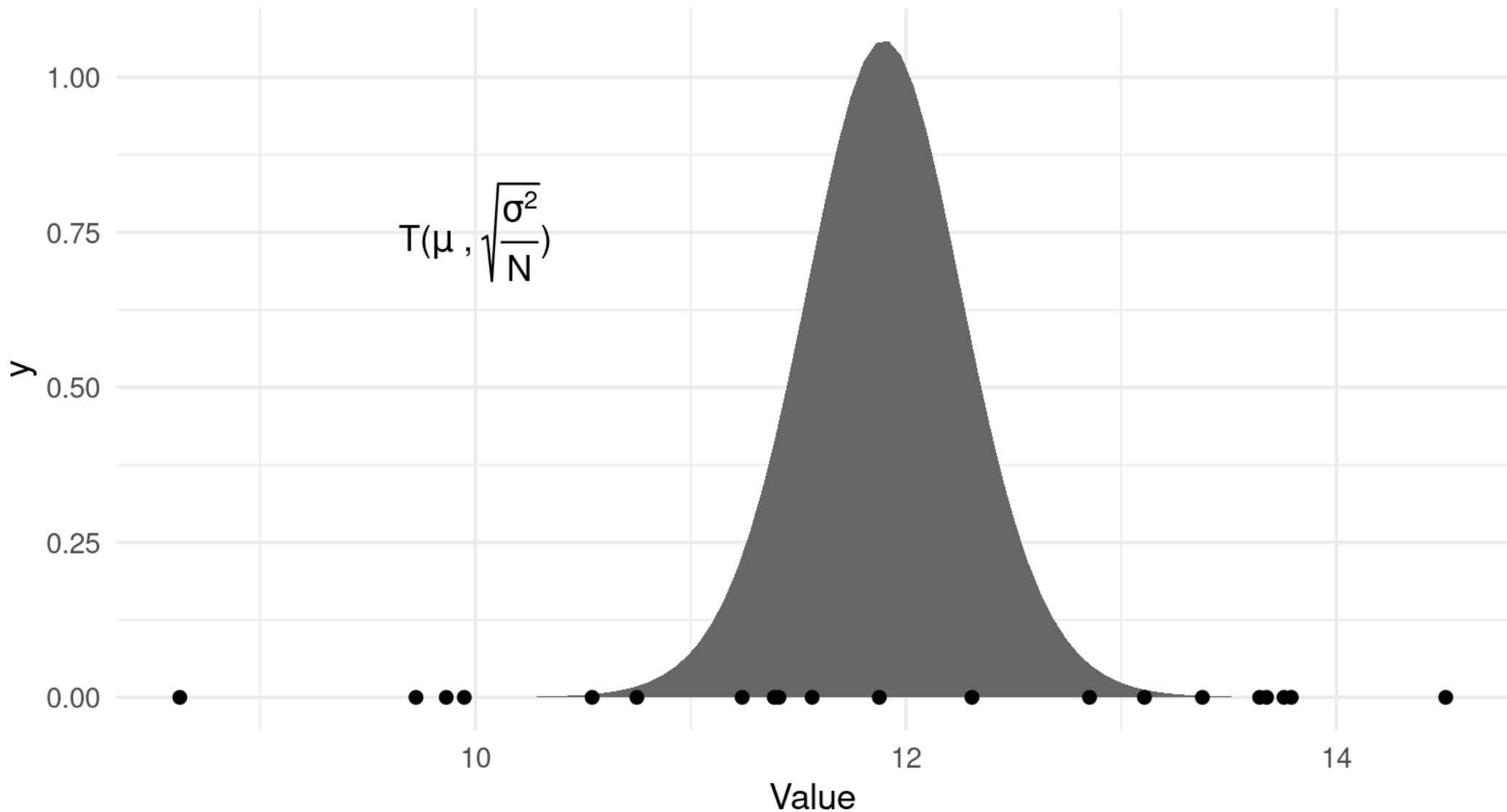
# Statistical Significance

We collect one sample of data



# Statistical Significance

We parameterize a T distribution based on this data



# Statistical Significance

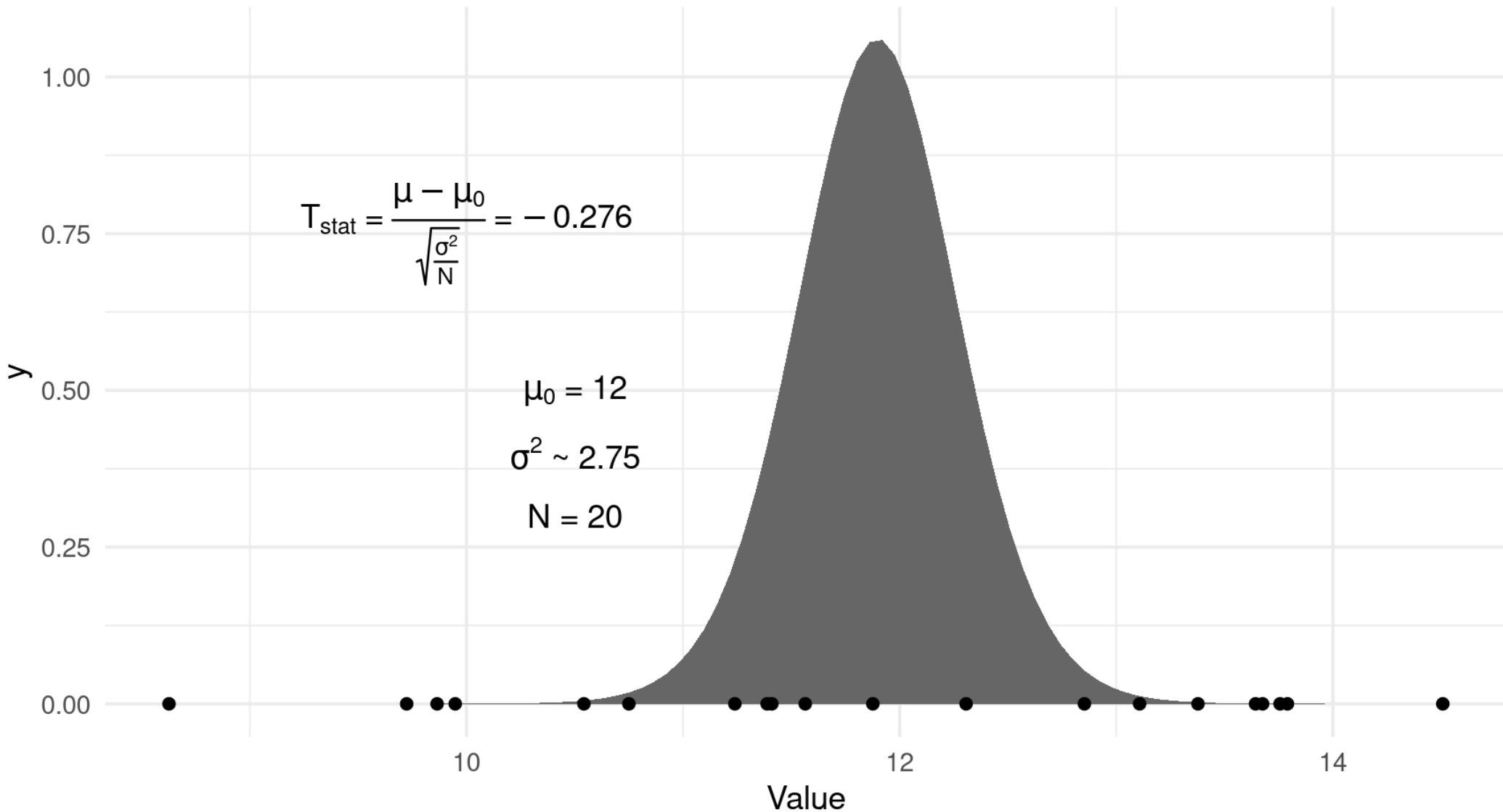
We find our test statistic

$$T_{\text{stat}} = \frac{\mu - \mu_0}{\sqrt{\frac{\sigma^2}{N}}} = -0.276$$

$$\mu_0 = 12$$

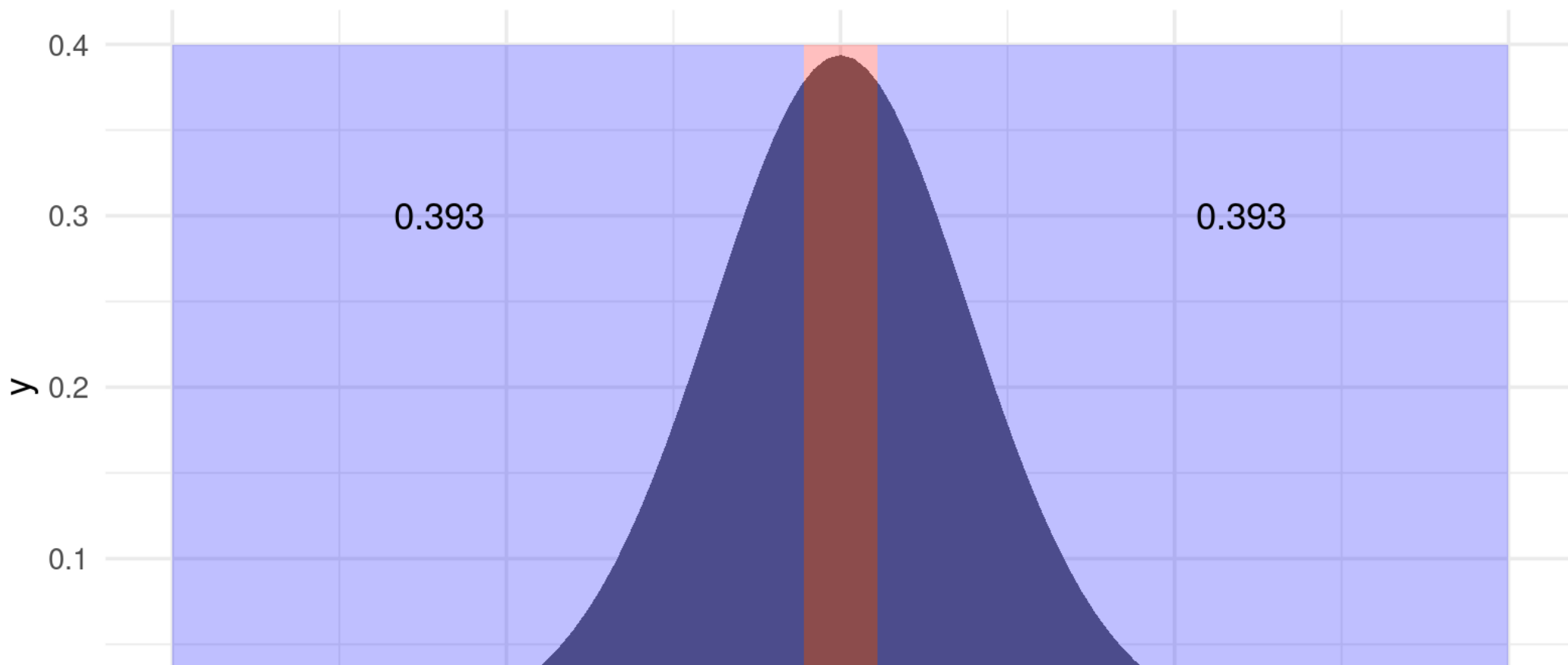
$$\sigma^2 \sim 2.75$$

$$N = 20$$



# Statistical Significance

And place the T stat on a standard T distribution



```
> pt(-abs(tstat), nu1, lower.tail = TRUE) + pt(abs(tstat), nu1, lower.tail = FALSE)
[1] 0.7853765
> t.test(s1$x, mu= nullMean)$p.value
[1] 0.7853765
```

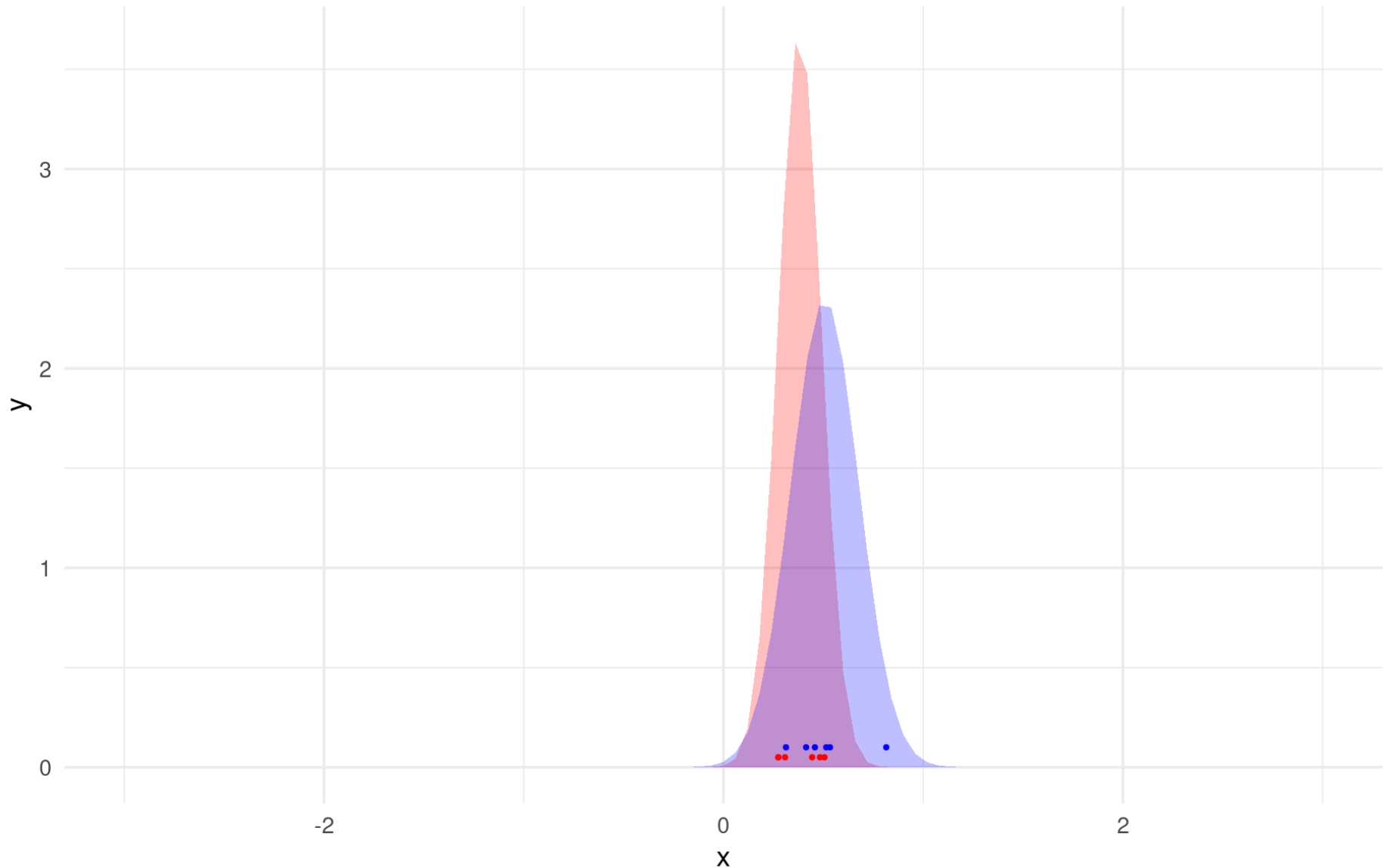
# Statistical Significance

- P-values are a product of the data and assumptions.
- Posterior Probabilities are a product of data, priors, and assumptions.
- Next we'll look at how power is a function of effect size, alpha, and replication.



# Visualizing Power

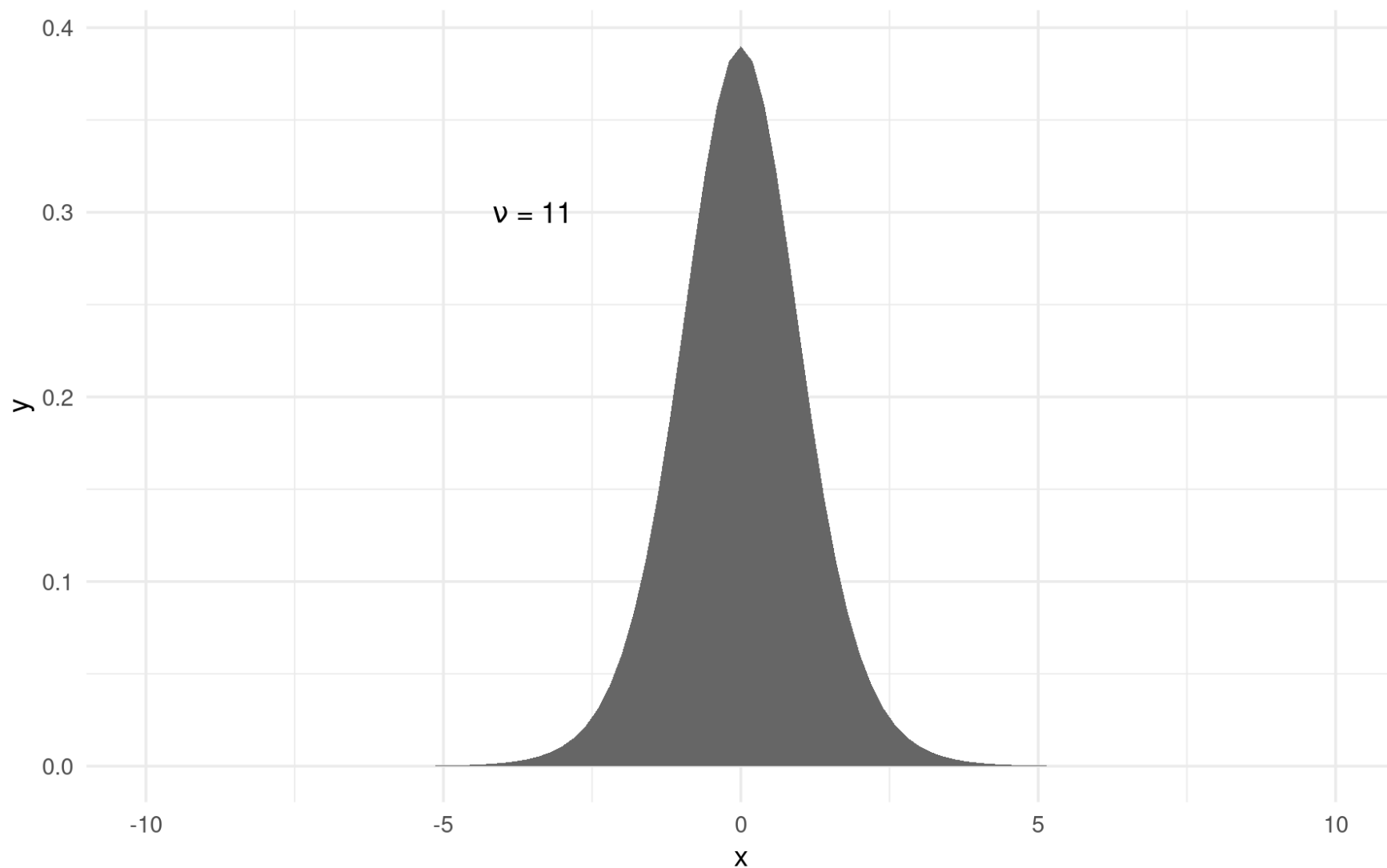
Say we want to compare the means of Gaussian data



# Visualizing Power

We will use a T distribution to compare means

Degrees of Freedom comes from our data



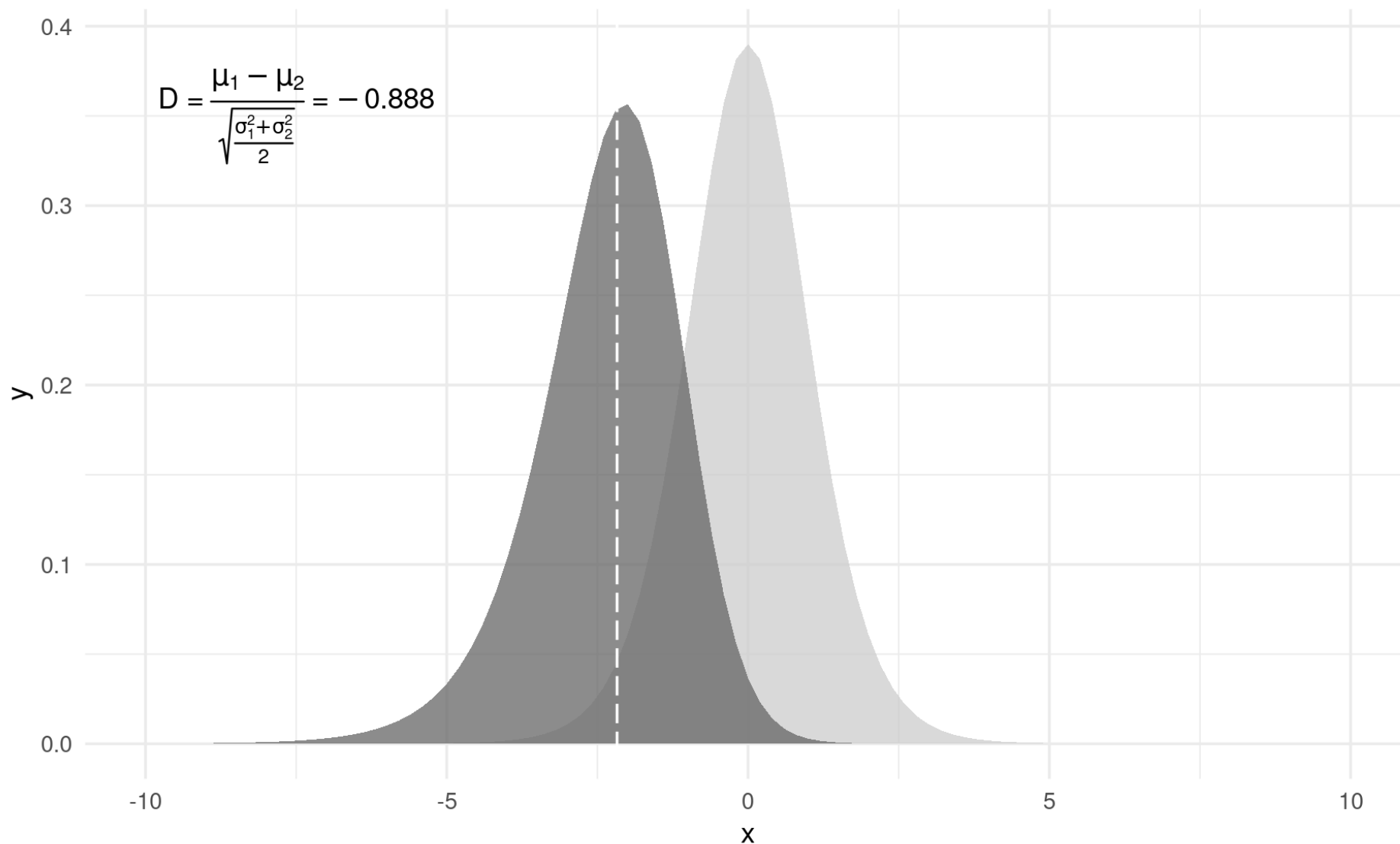
# Non-Centrality in Power Analysis

- Frequentist tests generally have a distribution for the Null hypothesis, but we're interested in the probability under the alternative hypothesis.
- To evaluate that we'll use the Non-Central T Distribution.
  - Non-centrality here changes the distribution from being when the Null is true to when the Null is false (ie, when there is an effect).

# Visualizing Power

We use the Non-Central T Distribution based on our effect size

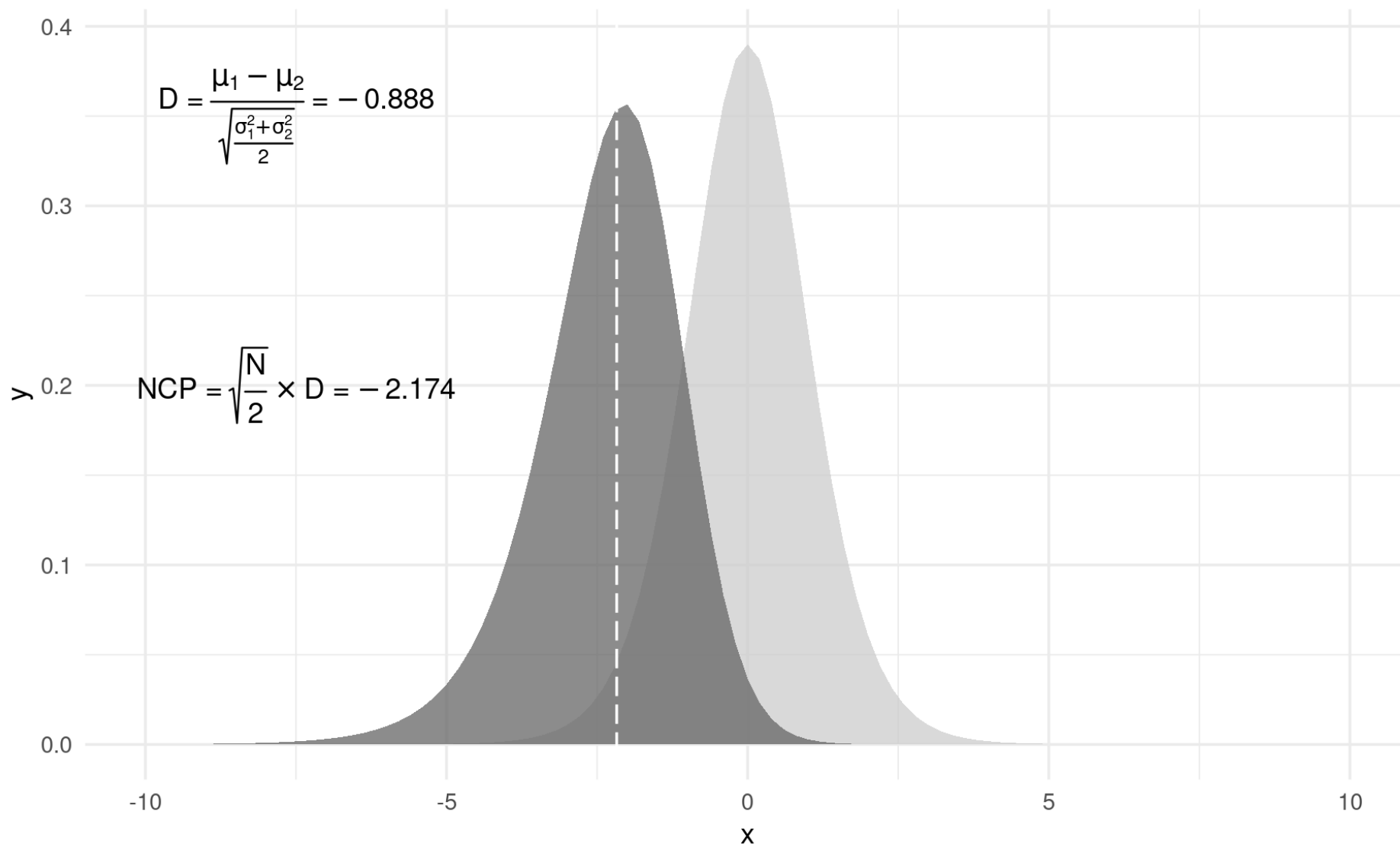
This is the distribution when the Null Hypothesis is False



# Visualizing Power

We use the Non-Central T Distribution based on our effect size

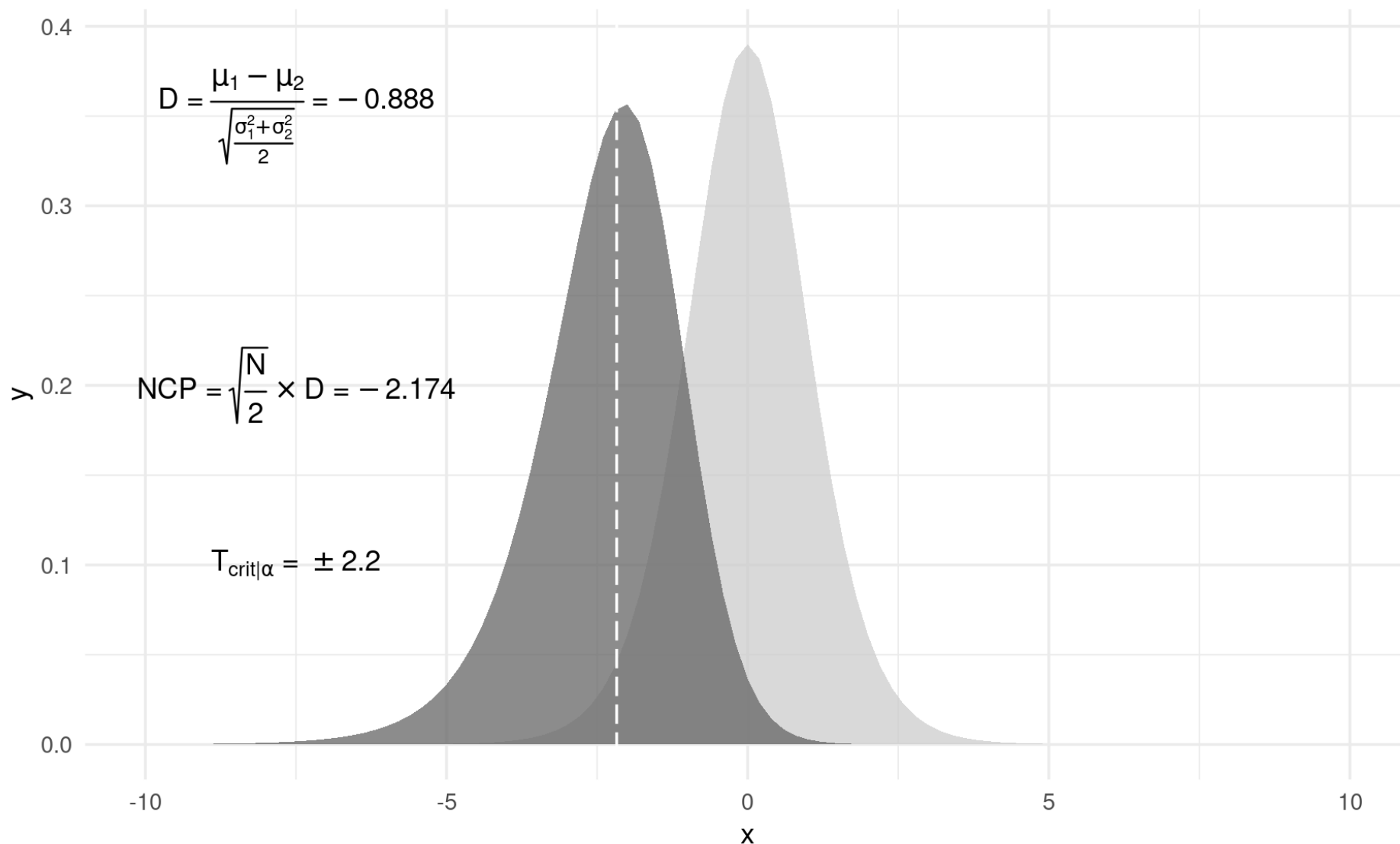
This is the distribution when the Null Hypothesis is False



# Visualizing Power

We use the Non-Central T Distribution based on our effect size

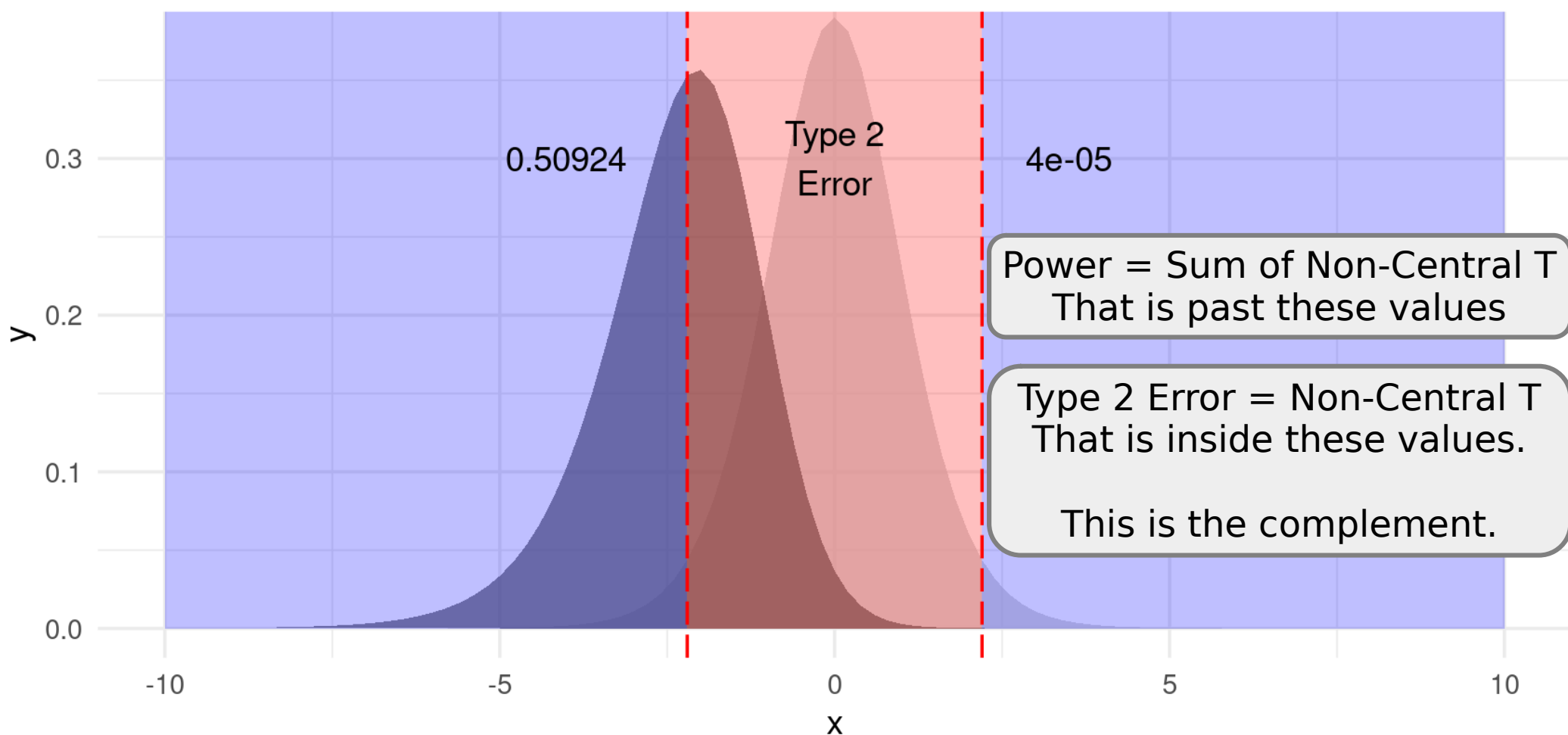
This is the distribution when the Null Hypothesis is False



# Visualizing Power

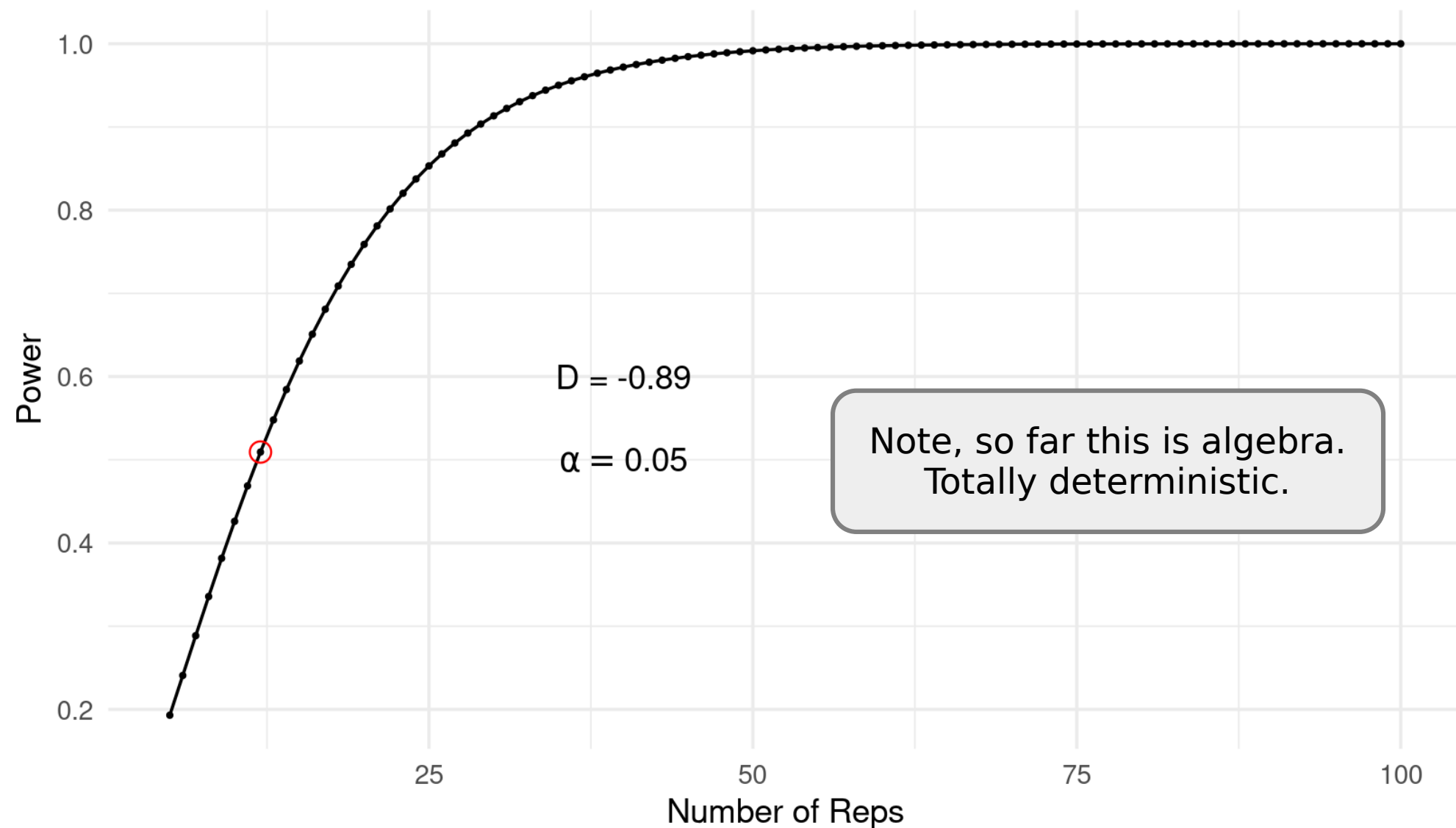
Power  $\sim 0.50927$

We divide the space based on the Centralized (Null) T Distribution's critical values



# Visualizing Power

Keeping  $D$  and  $\alpha$  but changing  $N$  changes Power





# Formulaic Power Analysis

- Given assumptions this is just an equation.
- Formulaic approaches use the steps we just went over to calculate some missing value out of:
  - Effect size
  - Power
  - Alpha
  - Number of Reps

# Formulaic Power Analysis

- Given assumptions this is just an equation.
- Formulaic approaches use the steps we just went over to calculate some missing value out of:
  - Effect size
  - Power
  - Alpha
  - Number of Reps

## Assumptions:

- T test assumptions
- Equal Variance (given 2 samples)
- Formula components are Fixed

# Formulaic Power Analysis

- Given assumptions this is just an equation.
- There are several more formulaic options beyond T tests.
- R packages `pwr`, `PowerUpR`, and `WebPower` expand on the options in `stats::`
- In general these will have more assumptions than their respective tests.

```
?stats::power.anova.test
```

```
?stats::power.t.test
```

```
?stats::power.prop.test
```

```
?pwr::pwr.2p.test
```

```
?pwr::pwr.2p2n.test
```

```
?pwr::pwr.t2n.test
```

```
?pwr::pwr.chisq.test
```

```
?pwr::pwr.f2.test
```

```
?pwr::pwr.r.test
```

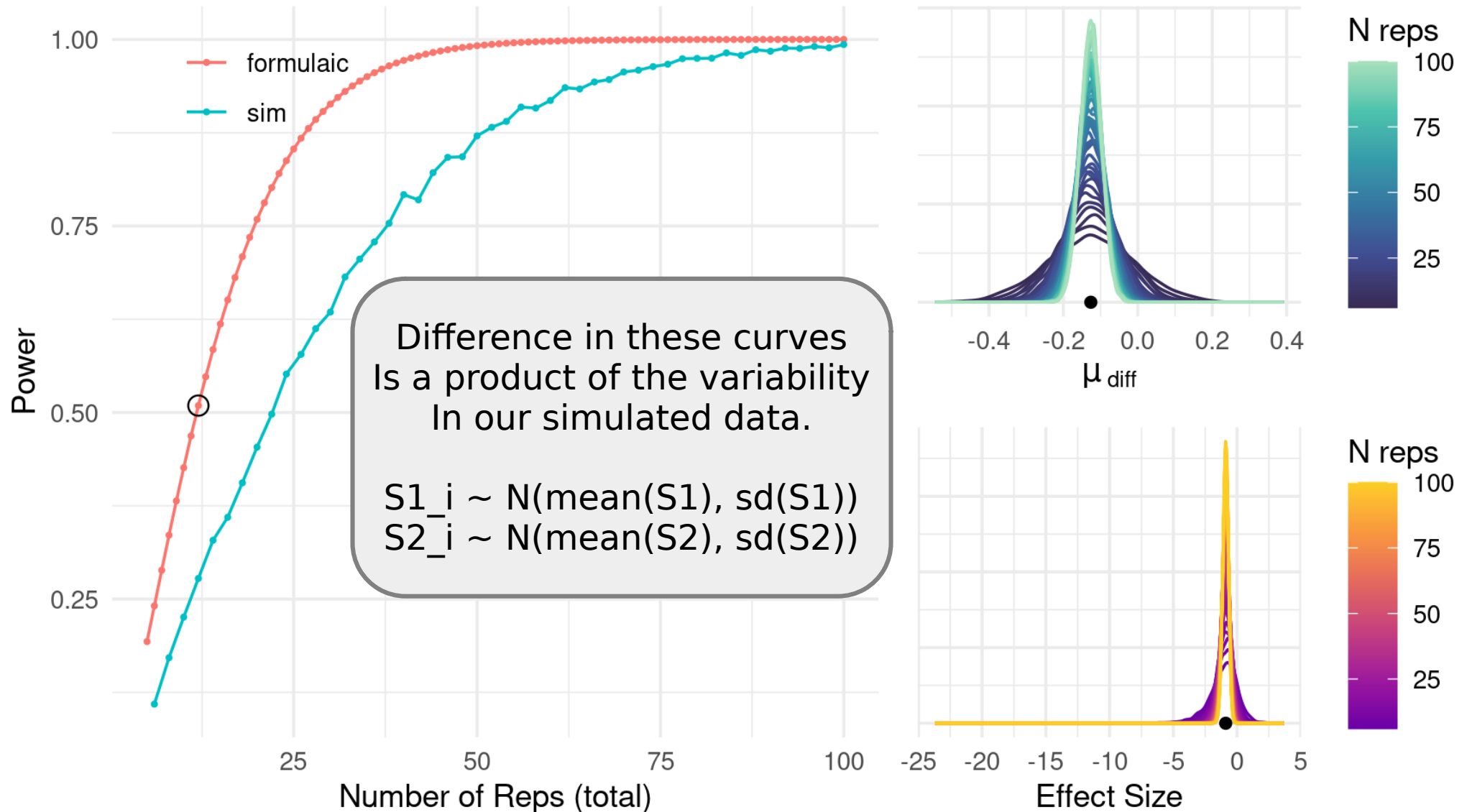
# Simulation Based Power Analysis

- To relax assumptions/test power in more complex settings we'll often turn to simulation based power analysis.

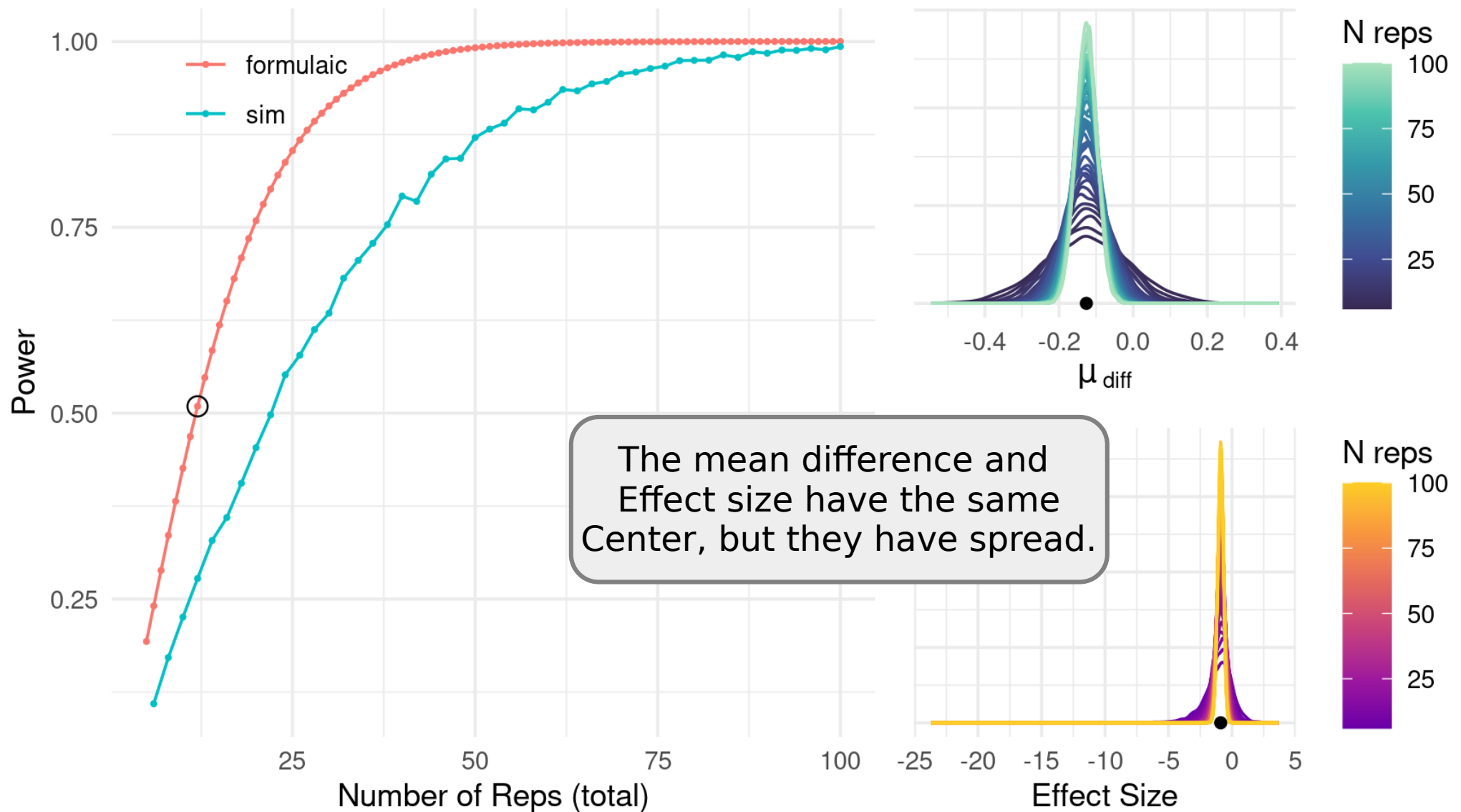
# Simulation Based Power Analysis

- To relax assumptions/test power in more complex settings we'll often turn to simulation based power analysis.
- Generally this will be more conservative than formulaic power analysis because simulated data is more variable.
  - Like models all simulations are wrong, but some are useful.
  - The simulation is often more similar to how newly collected data will behave, but this all depends on how you simulate data.

# Simulation Based Power Analysis



# Simulation Based Power Analysis



# Outline

- Scientific Process
- Definitions
- **Statistical Power and Significance**
  - Formulaic vs Simulated
- **Statistics and Power Analysis Workflow**
- **Example Scenarios**

Questions about Power  
or Significance?

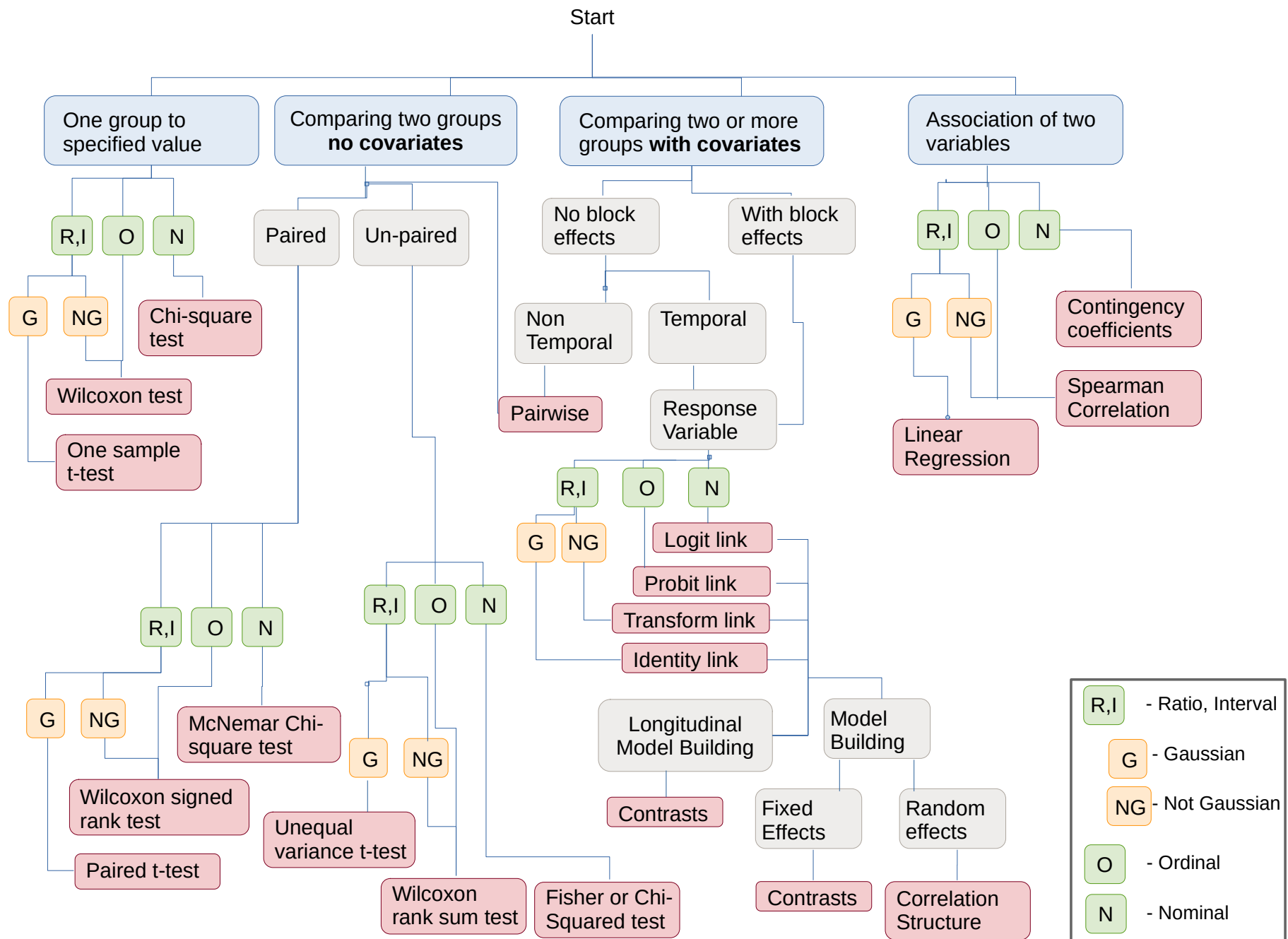


# Outline

- Scientific Process
- Definitions
- Statistical Power and Significance
  - Formulaic vs Simulated
- **Statistics and Power Analysis Workflow**
- **Example Scenarios**

That was some stats

Let's take a break



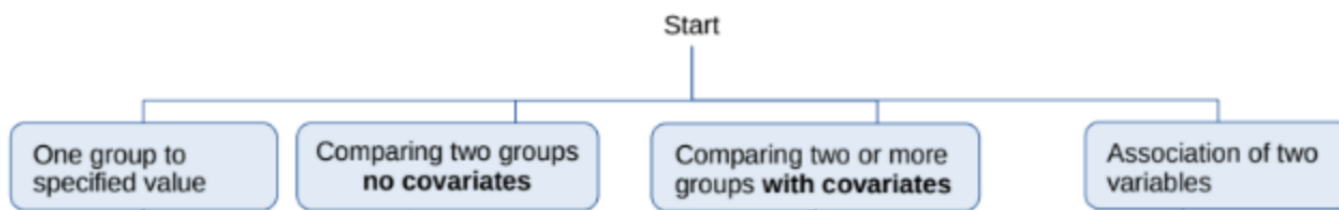
\* If the outcome and main predictor are both two-level factors, Breslow-Day and Cochren-Mantel-Hanzel tests are better

# Scenarios

- Find the test you need given the data
- Find the relevant information for power analysis
- Decide on formulaic vs simulation based

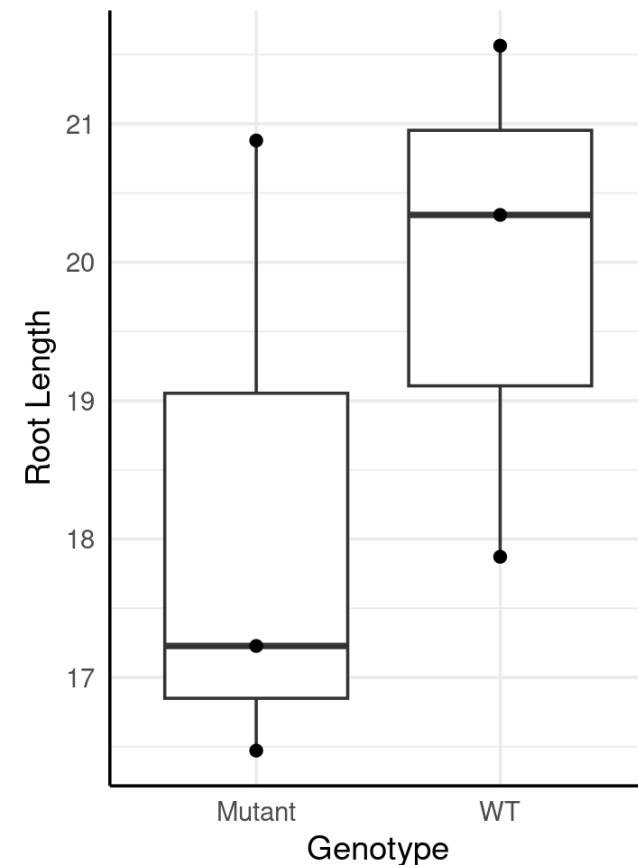
# Outline

- Scientific Process
- Definitions
- Power and Significance
  - Formulaic vs Simulated
- Statistics and Power Analysis Workflow
- **Example Scenarios**



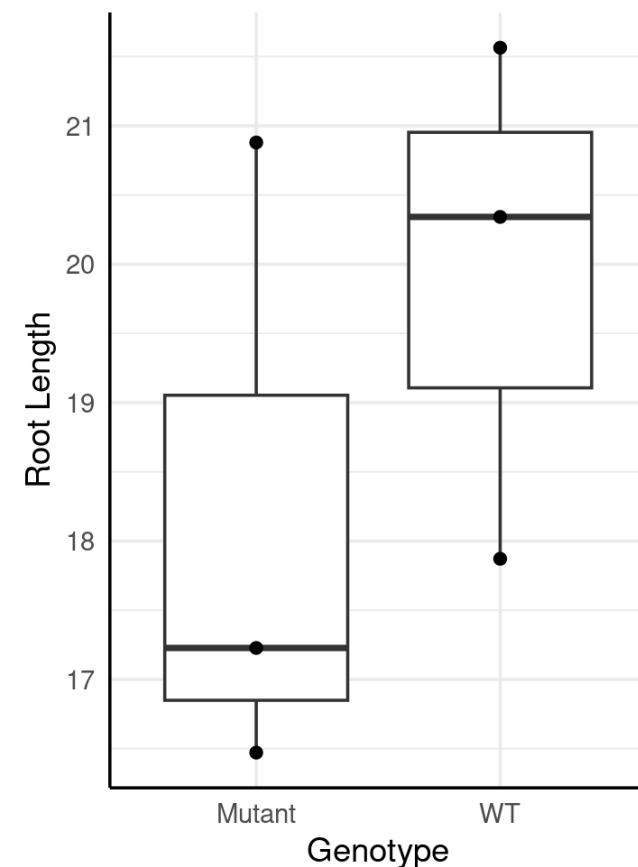
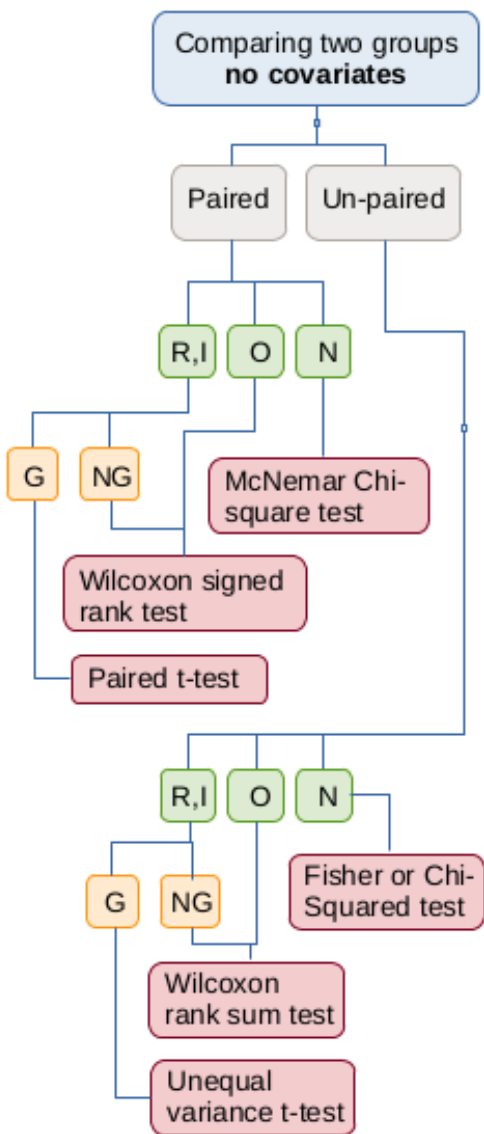
# Scenario 1

- You want to compare root length in WT vs a mutant. You have 3 preliminary reps from each genotype. What test will you use and how many replicates will you need?



# Scenario 1

- You want to compare root length in WT vs a mutant. You have 3 preliminary reps from each genotype. What test will you use and how many replicates will you need?

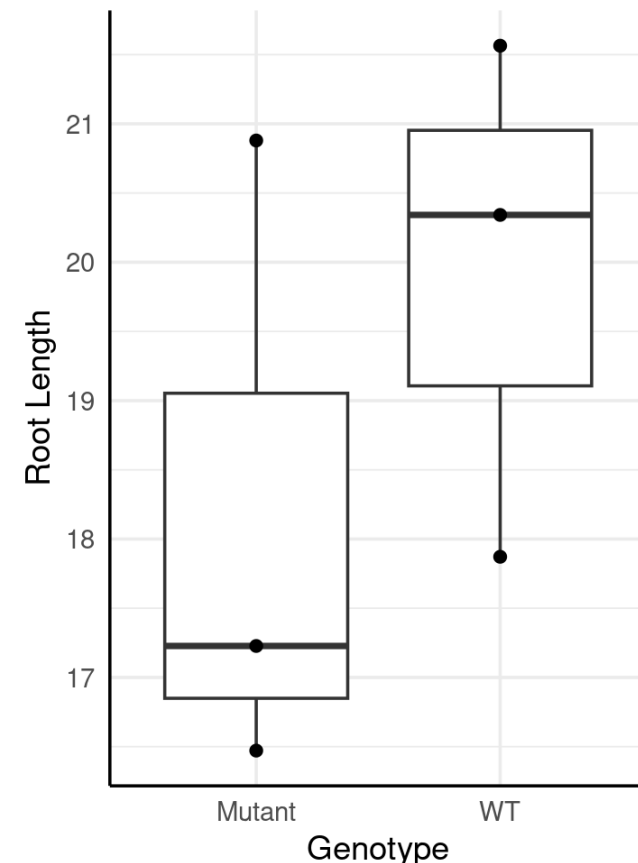


# Scenario 1

- You want to compare root length in WT vs a mutant. You have 3 preliminary reps from each genotype. What test will you use and how many replicates will you need?

## Statistical Test

We have unpaired continuous data. With 3 reps it is hard to say but we'll expect gaussian data so we'll use Welch's T Test.



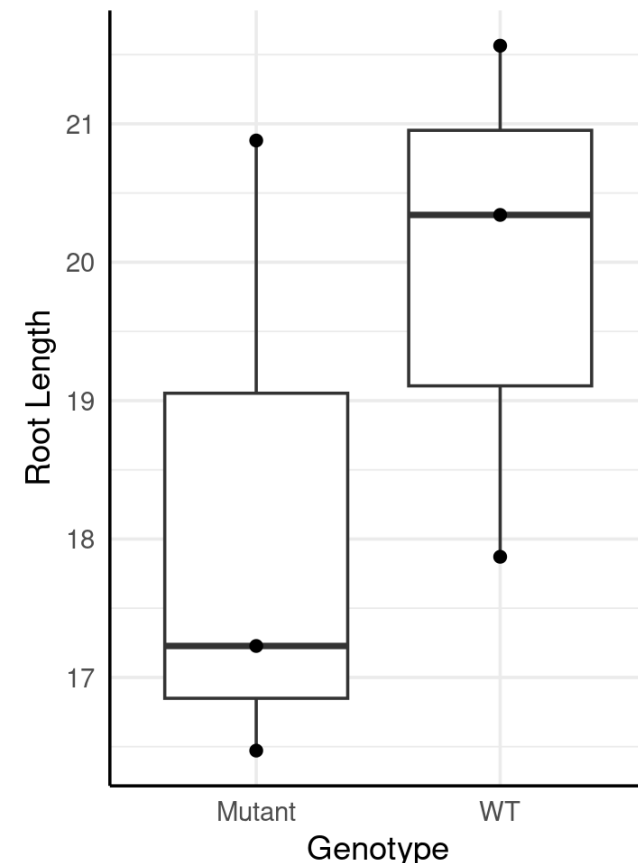
# Scenario 1

- You want to compare root length in WT vs a mutant. You have 3 preliminary reps from each genotype. What test will you use and how many replicates will you need?

## Power Analysis

To calculate reps we need:

- Effect size =
- SD =
- Power = 0.8
- Alpha = 0.05





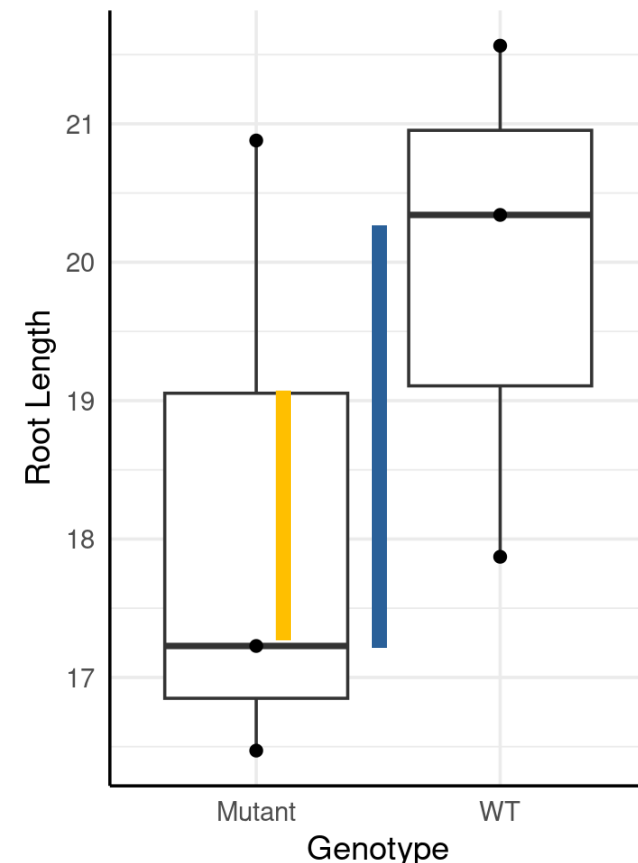
# Scenario 1

- You want to compare root length in WT vs a mutant. You have 3 preliminary reps from each genotype. What test will you use and how many replicates will you need?

## Power Analysis

To calculate reps we need:

- Effect size = 3
- SD = 2
- Power = 0.8
- Alpha = 0.05



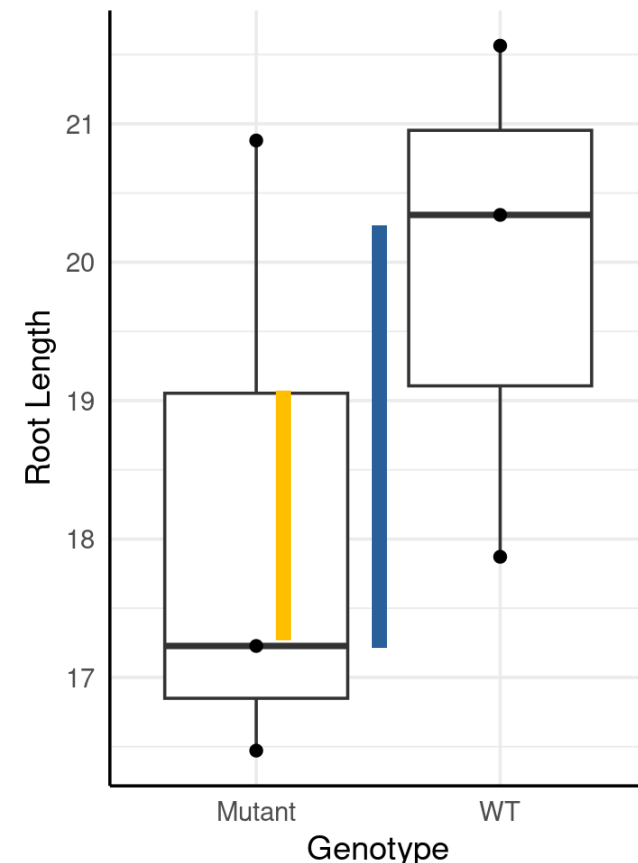
# Scenario 1

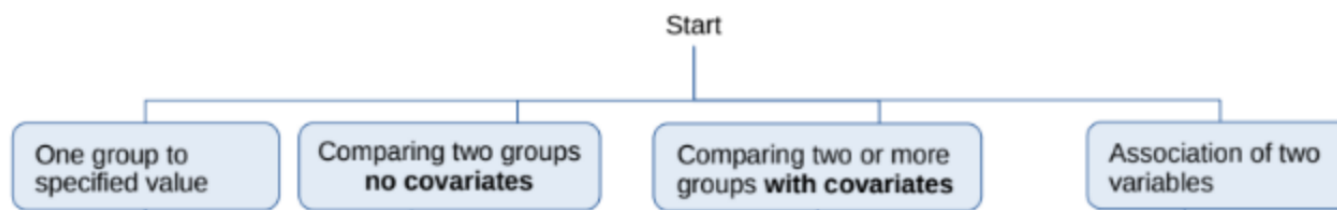
- You want to compare root length in WT vs a mutant. You have 3 preliminary reps from each genotype. What test will you use and how many replicates will you need?

## Power Analysis

Based on a formulaic power analysis we need 9 reps per genotype.

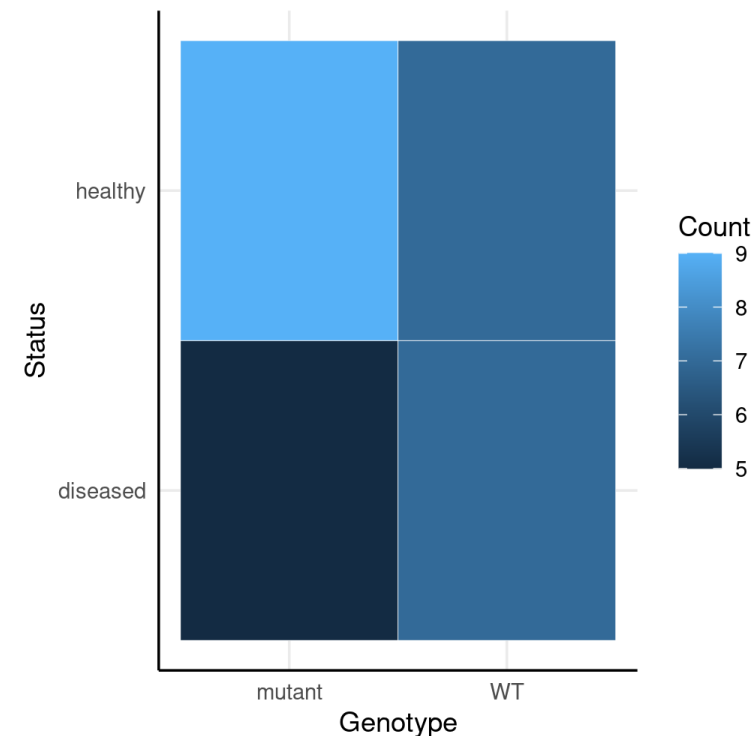
```
> res <- power.t.test(n = NULL, delta = 3,
+ sd = 2, sig.level = 0.05,
+ power = 0.8, type = "two.sample",
+ alternative = "two.sided")
> ceiling(res$n)
[1] 9
```





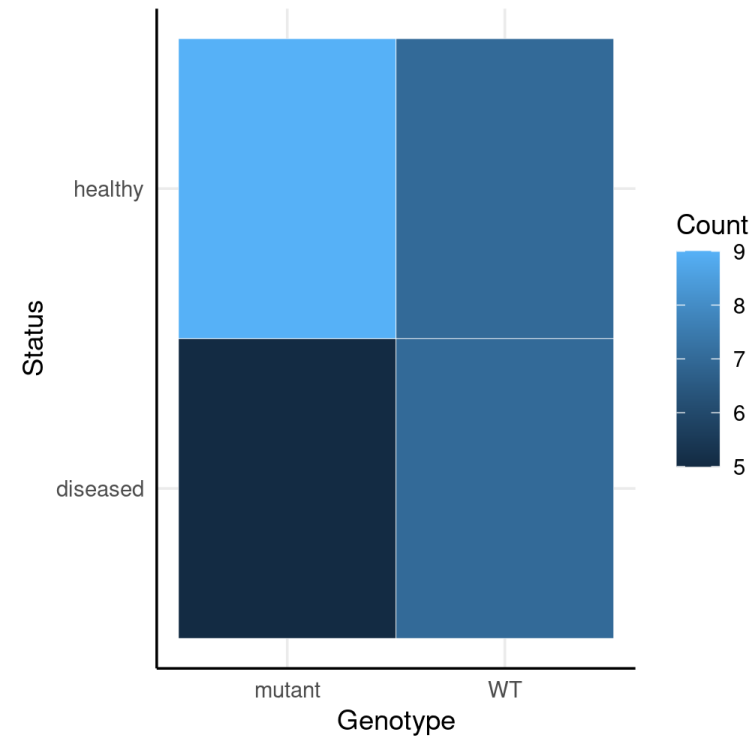
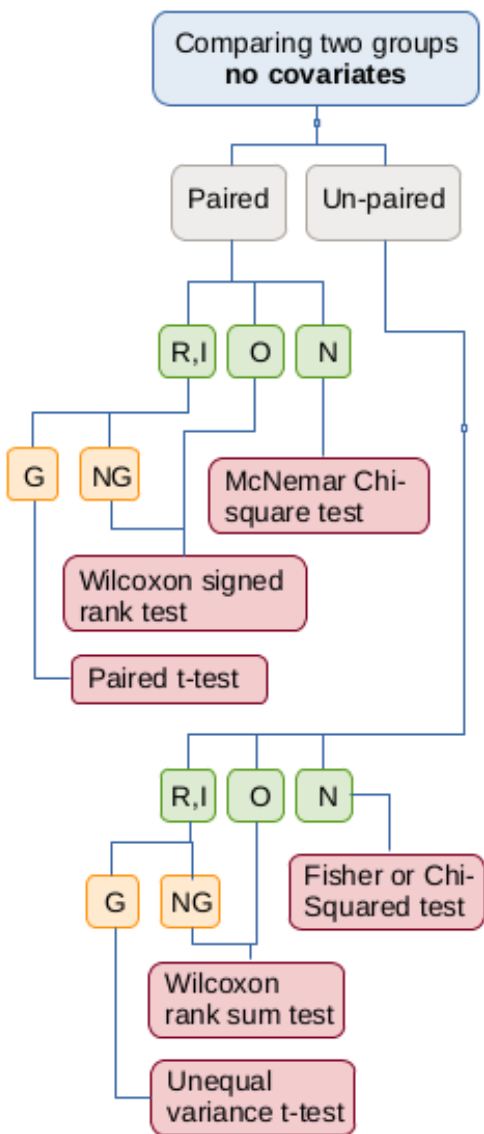
# Scenario 2

- You have categorized plants from two genotypes as diseased or healthy and you want to understand if there is an interaction between genotype and disease.



# Scenario 2

- You have categorized plants from two genotypes as diseased or healthy and you want to understand if there is an interaction between genotype and disease.

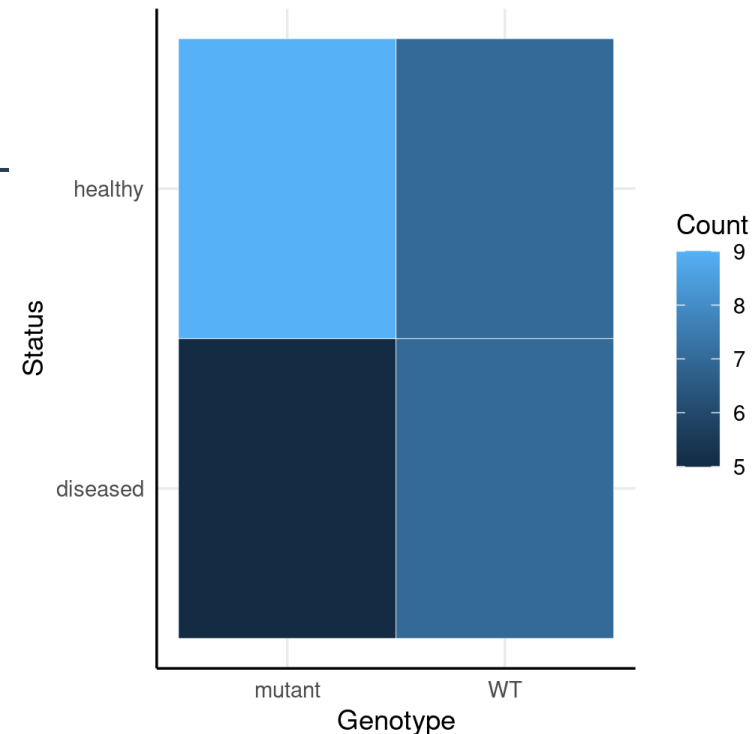


# Scenario 2

- You have categorized plants from two genotypes as diseased or healthy and you want to understand if there is an interaction between genotype and disease.

## Statistical Test

We have unpaired nominal data so we will use a Chi squared test



# Scenario 2

- You have categorized plants from two genotypes as diseased or healthy and you want to understand if there is an interaction between genotype and disease.

## Power Analysis

What power do we have in this experiment and how many reps would we need for 80% power?

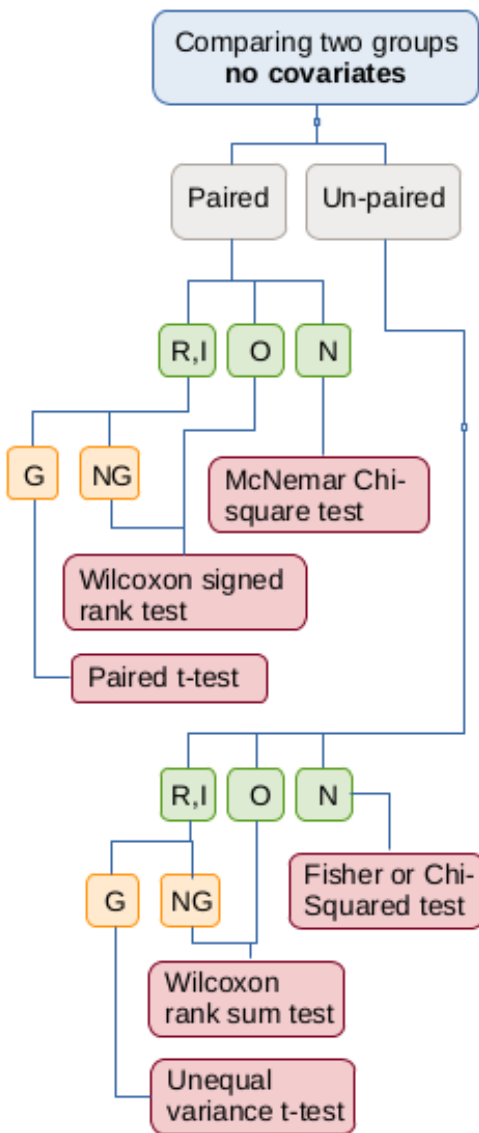
```
> chisq.test(df$genotype, df$status)
```

Pearson's Chi-squared test with Yates' continuity correction

data: df\$genotype and df\$status  
X-squared = 0.14583, df = 1, p-value = 0.7025

```
> table(df$genotype, df$status)
```

|        | diseased | healthy |
|--------|----------|---------|
| mutant | 5        | 9       |
| WT     | 7        | 7       |



# Scenario 2

- You have categorized plants from two genotypes as diseased or healthy and you want to understand if there is an interaction between genotype and disease.

## Power Analysis

To calculate power we need:

- Effect size =
- N =
- DF =
- Alpha = 0.05

```
> table(df$genotype, df$status)
```

|        | diseased | healthy |
|--------|----------|---------|
| mutant | 5        | 9       |
| WT     | 7        | 7       |

# Scenario 2

- You have categorized plants from two genotypes as diseased or healthy and you want to understand if there is an interaction between genotype and disease.

## Power Analysis

To calculate power we need:

- Effect size =
- $N = 5 + 9 + 7 + 7 = 28$
- $DF = (2-1)*(2-1) = 1$
- Alpha = 0.05

```
> table(df$genotype, df$status)
```

|        | diseased | healthy |
|--------|----------|---------|
| mutant | 5        | 9       |
| WT     | 7        | 7       |



# Scenario 2

- You have categorized plants from two genotypes as diseased or healthy and you want to understand if there is an interaction between genotype and disease.

## Power Analysis

To calculate power we need:

- Effect size = 0.144 ( $\chi^2$  stat)
- $N = 5+9+7+7 = 28$
- $DF = (2-1)*(2-1)=1$
- Alpha = 0.05

```
> p <- table(df$genotype, df$status)
> p
```

|        | diseased | healthy |
|--------|----------|---------|
| mutant | 5        | 9       |
| WT     | 7        | 7       |

```
> p <- p / sum(p)
> p
```

|        | diseased  | healthy   |
|--------|-----------|-----------|
| mutant | 0.1785714 | 0.3214286 |
| WT     | 0.2500000 | 0.2500000 |

```
> p_i <- rowSums(p)
> p_j <- colSums(p)
> P0 <- p_i %*% t(p_j)
> sqrt(sum((p - P0)^2 / P0))
[1] 0.1443376
```

# Scenario 2

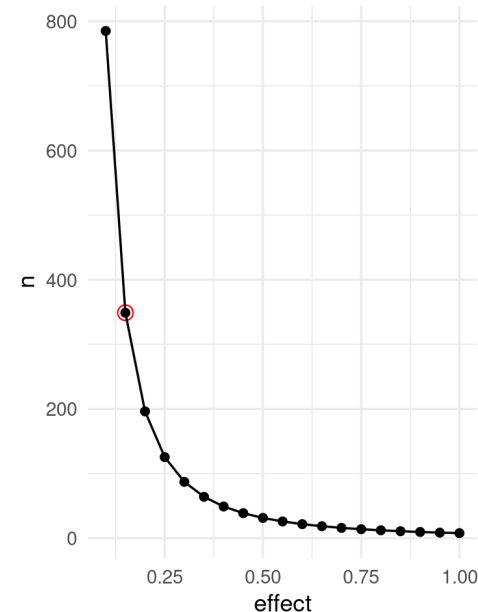
- You have categorized plants from two genotypes as diseased or healthy and you want to understand if there is an interaction between genotype and disease.

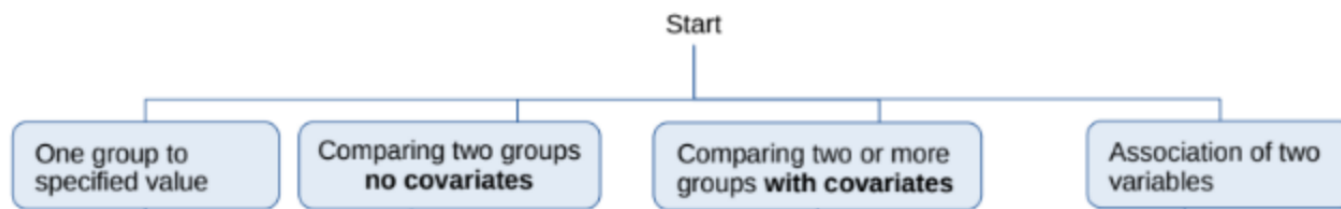
## Power Analysis

Due to a small effect size we would need a very large sample to reach 80% power.

In this data we only have ~12% power.

```
> pwr.chisq.test(w= ES.w2(p),
+ N = 28, df=1, sig.level=0.05)$power
[1] 0.1190365
> pwr.chisq.test(w= ES.w2(p),
+ df=1, power=0.80, sig.level=0.05)$N
[1] 376.7453
```

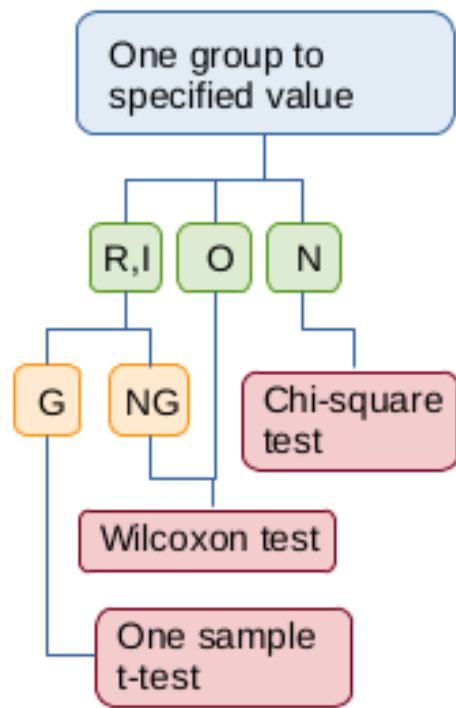




# Scenario 3

- You have identified a gene that you expect increases plant height by about 15%. Based on previous literature you expect WT to grow to ~30cm ( $\pm 3$ cm) in height and you want to know how many replicates you'll need of your mutant, assuming you have to use a less powerful non-parametric test.

# Scenario 3

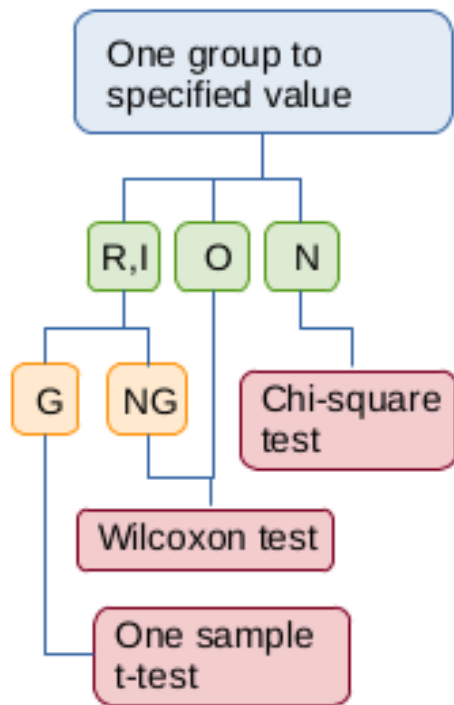


- You have identified a gene that you expect increases plant height by about 15%. Based on previous literature you expect WT to grow to ~30cm ( $\pm 3$ cm) in height and you want to know how many replicates you'll need of your mutant, assuming you have to use a less powerful non-parametric test.

## Why a Non-parametric?

Here we expect 1 gene to have a strong effect on a phenotype, that could push use from a Normal to a Log-Normal

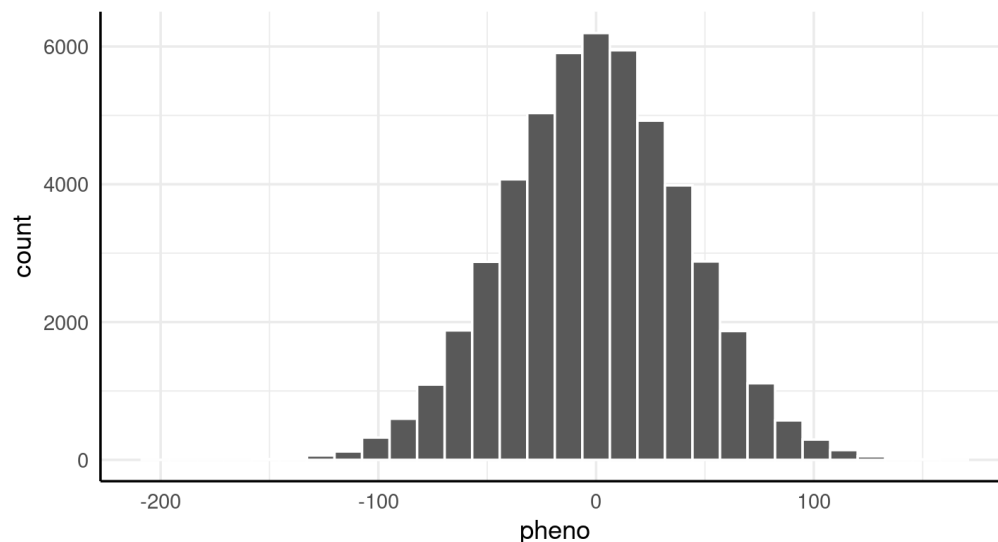
# Scenario 3



- You have identified a gene that you expect increases plant height by about 15%. Based on previous literature you expect WT to grow to ~30cm ( $\pm 3$ cm) in height and you want to know how many replicates you'll need of your mutant, assuming you have to use a less powerful non-parametric test.

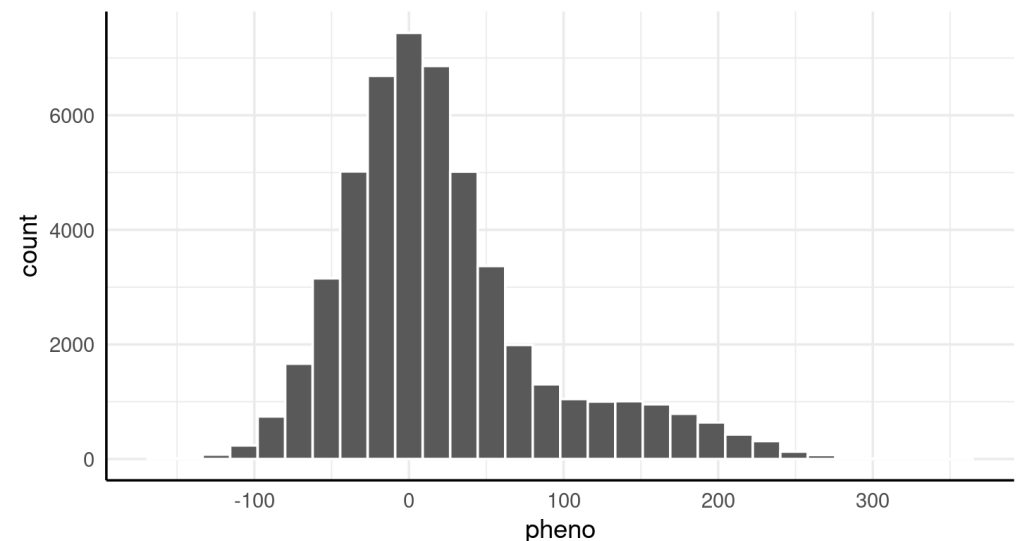
Approx. Normal

5000 Genes with  $U(-1, 1)$  Influence

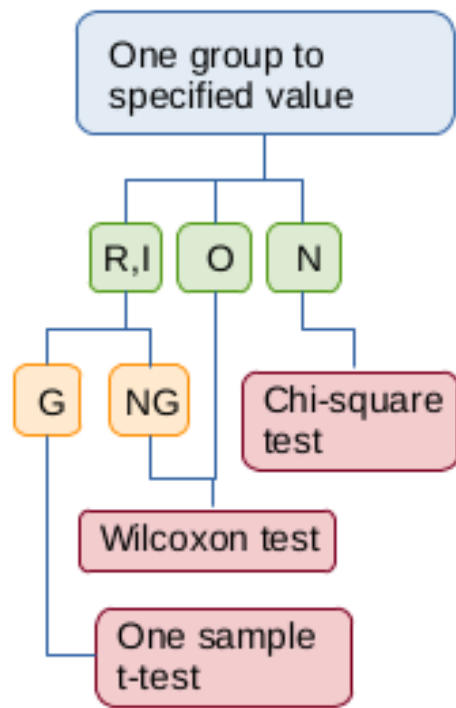


Approx. Log Normal

4999 Genes with  $U(-1, 1)$  Influence, one gene that is sometimes  $U(90, 200)$



# Scenario 3

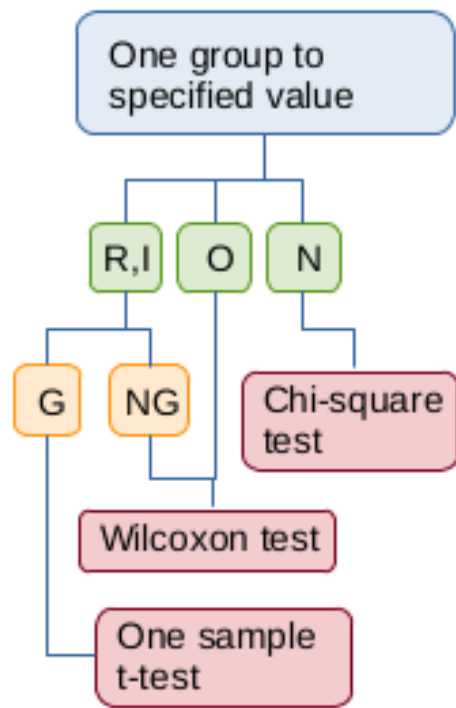


- You have identified a gene that you expect increases plant height by about 15%. Based on previous literature you expect WT to grow to ~30cm ( $\pm 3$ cm) in height and you want to know how many replicates you'll need of your mutant, assuming you have to use a less powerful non-parametric test.

## Statistical Test

We are comparing one group to a specified value and using a non-parametric test, which takes us to the Wilcoxon test.

# Scenario 3



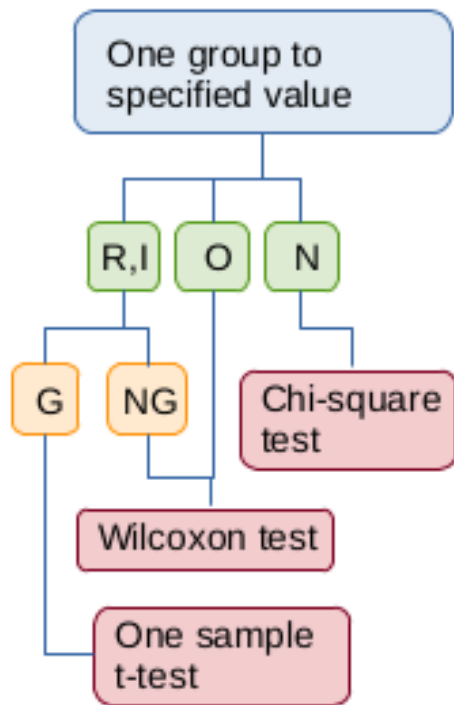
- You have identified a gene that you expect increases plant height by about 15%. Based on previous literature you expect WT to grow to ~30cm ( $\pm 3$ cm) in height and you want to know how many replicates you'll need of your mutant, assuming you have to use a less powerful non-parametric test.

## Power Analysis

We have to simulate this one.

- $\mu =$
- $\sigma =$
- Distribution =  $T_4(\mu, \sigma)$

# Scenario 3



- You have identified a gene that you expect increases plant height by about 15%. Based on previous literature you expect WT to grow to ~30cm ( $\pm 3$ cm) in height and you want to know how many replicates you'll need of your mutant, assuming you have to use a less powerful non-parametric test.

## Power Analysis

We have to simulate this one.

- $\mu = 30 \times 1.15$
- $\sigma = 3$
- Distribution =  $T_4(\mu, \sigma)$

```
sig.level = 0.05
repRange <- seq(5,25,2)
iter <- 5000
simdf<-do.call(rbind, lapply(repRange, function(n){
 iter <- do.call(rbind, lapply(1:iter, function(i){
 s1_i <- extraDistr::rlst(n, 4, 34.5, 3)
 true_mu <- mean(s1_i)
 bool <- wilcox.test(s1_i, mu=30)$p.value < sig.level
 data.frame(bool = bool, mu = true_mu)
 }))
 data.frame(power=mean(iter$bool), n=n, type="sim")
}))
```



# Scenario 3

```
sig.level = 0.05
repRange <- seq(5,25,2)
iter <- 5000
```

We specify some hyperparameters  
To control the simulation.

```
simdf<-do.call(rbind, lapply(repRange, function(n){
 iter <- do.call(rbind, lapply(1:iter, function(i){
 s1_i <- extraDistr::rlst(n, 4, 34.5, 3)
 true_mu <- mean(s1_i)
 bool <- wilcox.test(s1_i, mu=30)$p.value < sig.level
 data.frame(bool = bool, mu = true_mu)
 })))
data.frame(power=mean(iter$bool), n=n, type="sim")
}))
```

# Scenario 3

```
sig.level = 0.05
repRange <- seq(5,25,2)
iter <- 5000
simdf<-do.call(rbind, lapply(repRange, function(n){
 iter <- do.call(rbind, lapply(1:iter, function(i){
 s1_i <- extraDistr::rlst(n, 4, 34.5, 3)
 true_mu <- mean(s1_i)
 bool <- wilcox.test(s1_i, mu=30)$p.value < sig.level
 data.frame(bool = bool, mu = true_mu)
 })))
data.frame(power=mean(iter$bool), n=n, type="sim")
})))
```

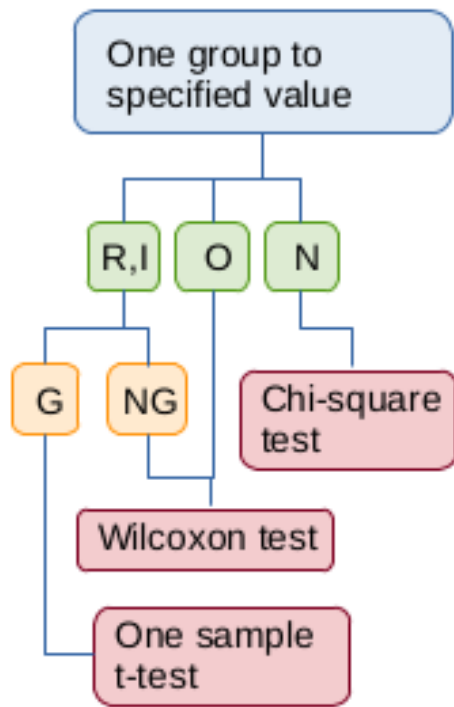
5000 times per each number of reps  
We generate a sample of data and  
Record the mean.

# Scenario 3

```
sig.level = 0.05
repRange <- seq(5,25,2)
iter <- 5000
simdf<-do.call(rbind, lapply(repRange, function(n){
 iter <- do.call(rbind, lapply(1:iter, function(i){
 s1_i <- extraDistr::rlst(n, 4, 34.5, 3)
 true_mu <- mean(s1_i)
 bool <- wilcox.test(s1_i, mu=30)$p.value < sig.level
 data.frame(bool = bool, mu = true_mu)
 }))
}))
data.frame(power=mean(iter$bool), n=n, type="sim")
}))
```

We run our test and record the result.  
Here I returned extra information for  
The first few tests.

# Scenario 3

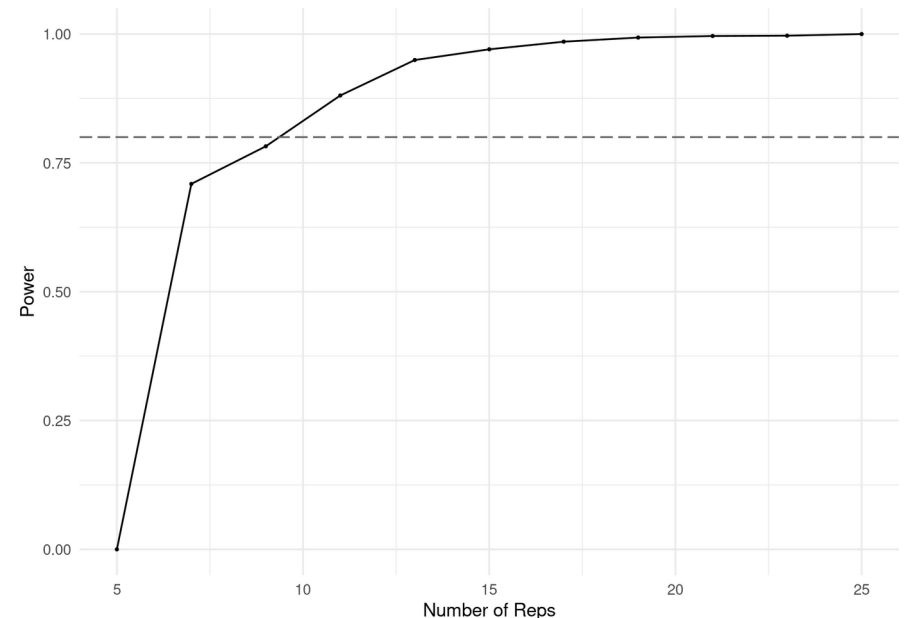


- You have identified a gene that you expect increases plant height by about 15%. Based on previous literature you expect WT to grow to ~30cm ( $\pm 3$ cm) in height and you want to know how many replicates you'll need of your mutant, assuming you have to use a less powerful non-parametric test.

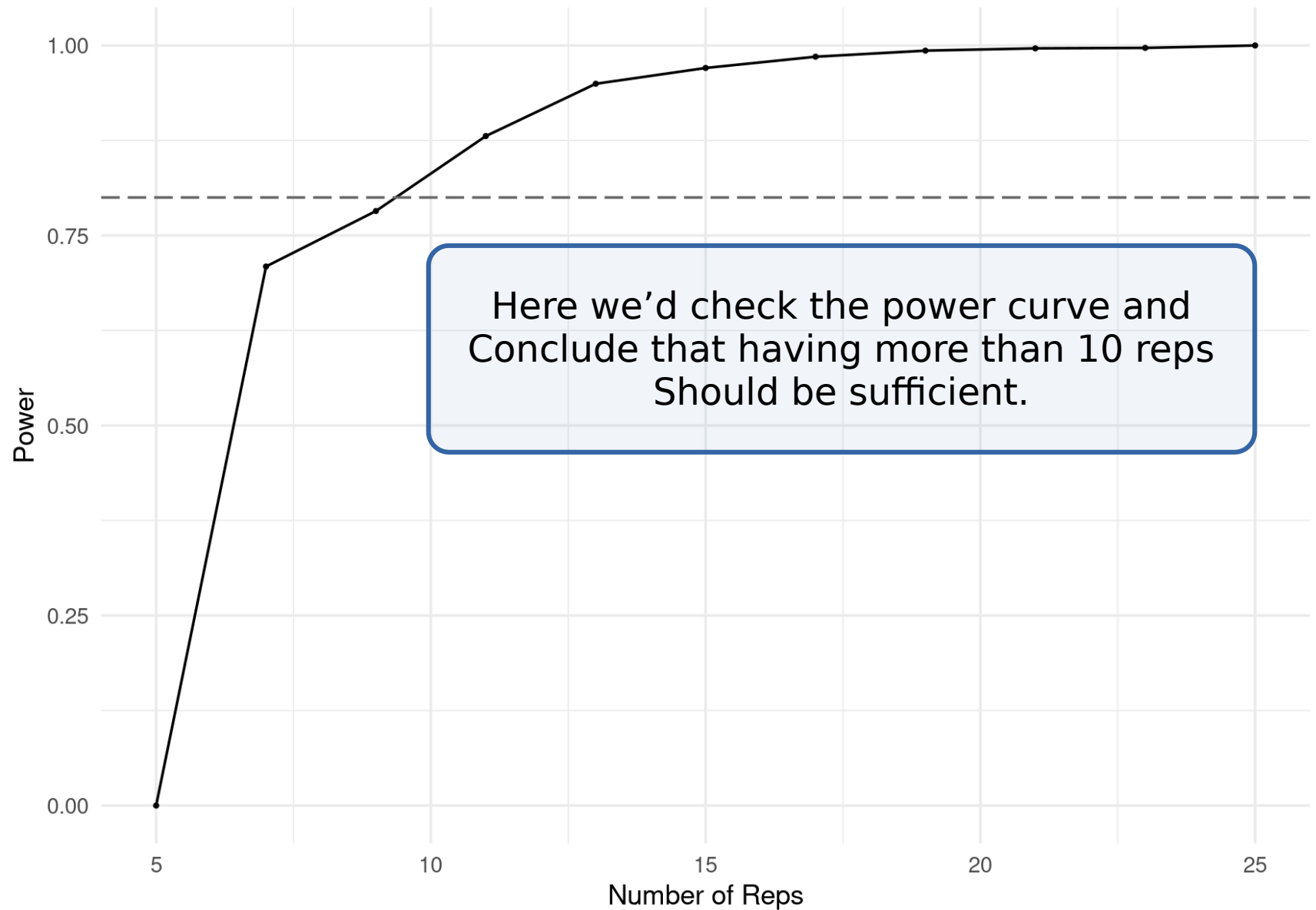
## Power Analysis

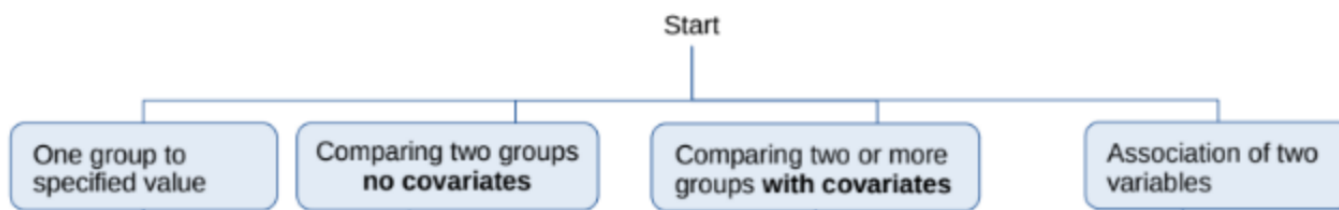
We have to simulate this one.

- $\mu = 30 \times 1.15$
- $\sigma = 3$
- Distribution =  $T_4(\mu, \sigma)$



# Scenario 3



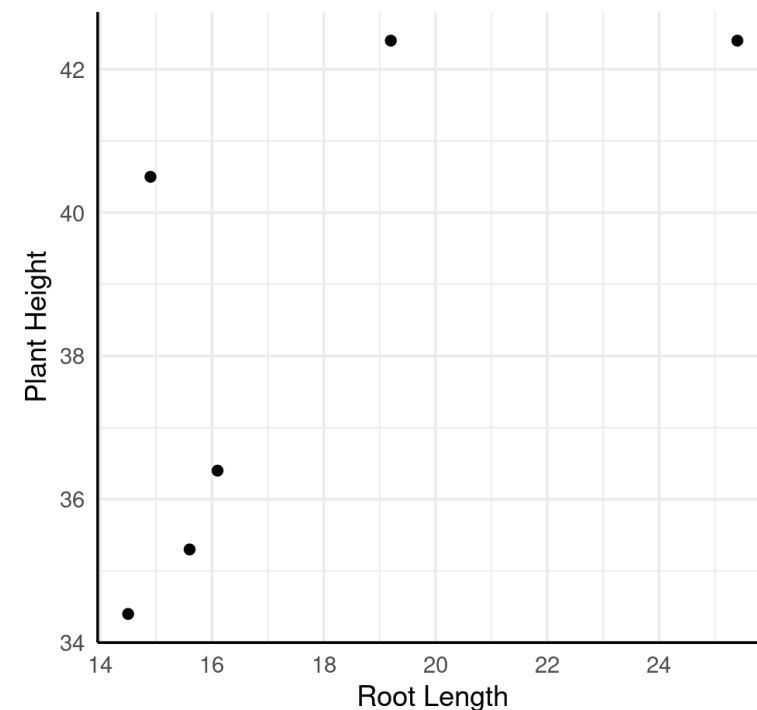


# Scenario 4

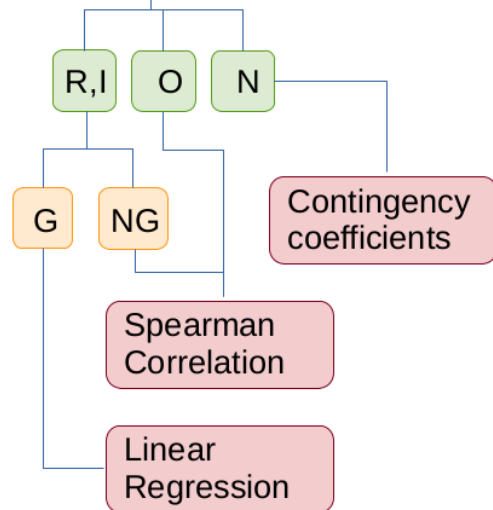
- You noticed an interesting correlation between root length and panicle weight as part of a subgroup analysis. You don't have a statistically significant correlation and want to know how many reps you'd need in a new experiment to see one.

Correlation Test P-value: 0.11

Observed Correlation: 0.71



Association of two variables



# Scenario 4

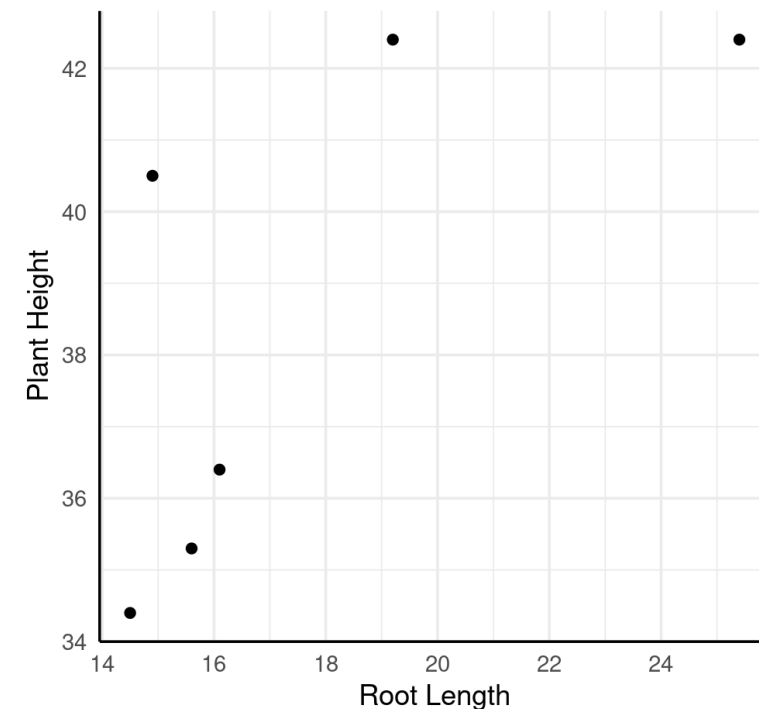
You noticed an interesting correlation between root length and panicle weight as part of a subgroup analysis. You don't have a statistically significant correlation and want to know how many reps you'd need in a new experiment to see one.

## Statistical Test

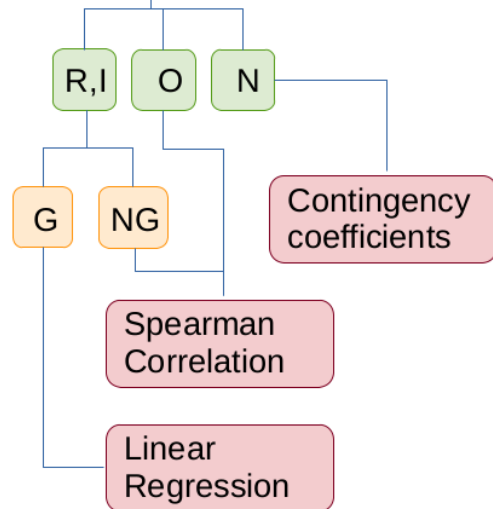
We are explicitly interested in a correlation test here.

Correlation Test P-value: 0.11

Observed Correlation: 0.71



## Association of two variables



# Scenario 4

You noticed an interesting correlation between root length and panicle weight as part of a subgroup analysis. You don't have a statistically significant correlation and want to know how many reps you'd need in a new experiment to see one.

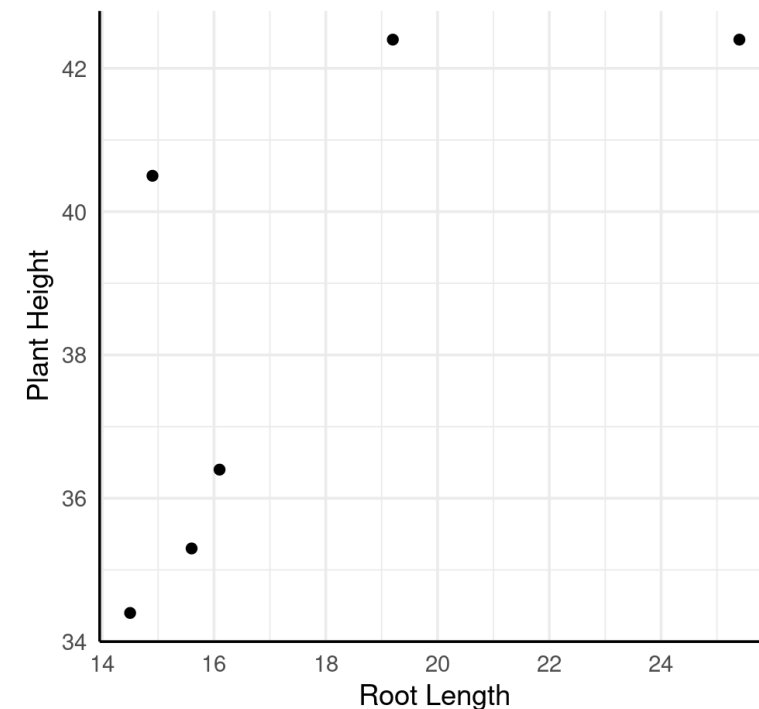
Correlation Test P-value: 0.11

Observed Correlation: 0.71

## Power Analysis

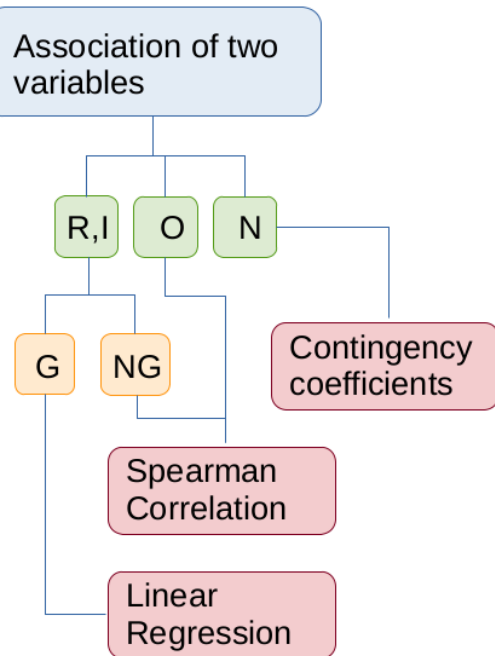
To calculate reps we need:

- Power
- Alpha
- Effect size (correlation)





# Scenario 4



You noticed an interesting correlation between root length and panicle weight as part of a subgroup analysis. You don't have a statistically significant correlation and want to know how many reps you'd need in a new experiment to see one.

## Power Analysis

To calculate reps we need:

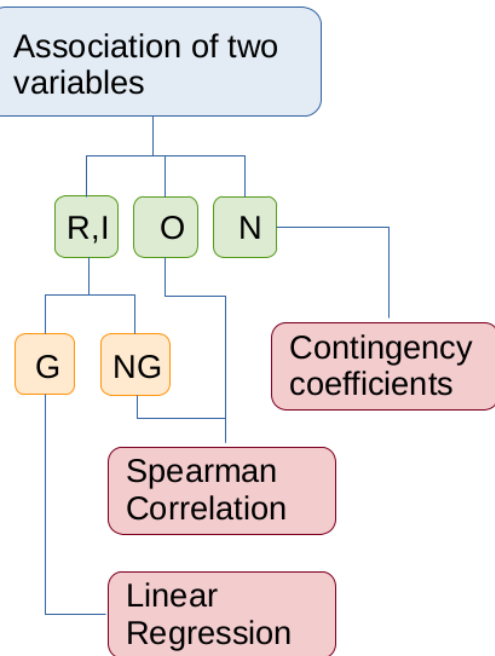
```
> pwr.r.test(n = NULL, r = 0.708, sig.level = 0.05, power = 0.8,
+ alternative = "two.sided")
```

approximate correlation power calculation (arctangh transformation)

```
n = 12.45543
r = 0.708
sig.level = 0.05
power = 0.8
alternative = two.sided
```

- Power
- Alpha
- Effect size (correlation)

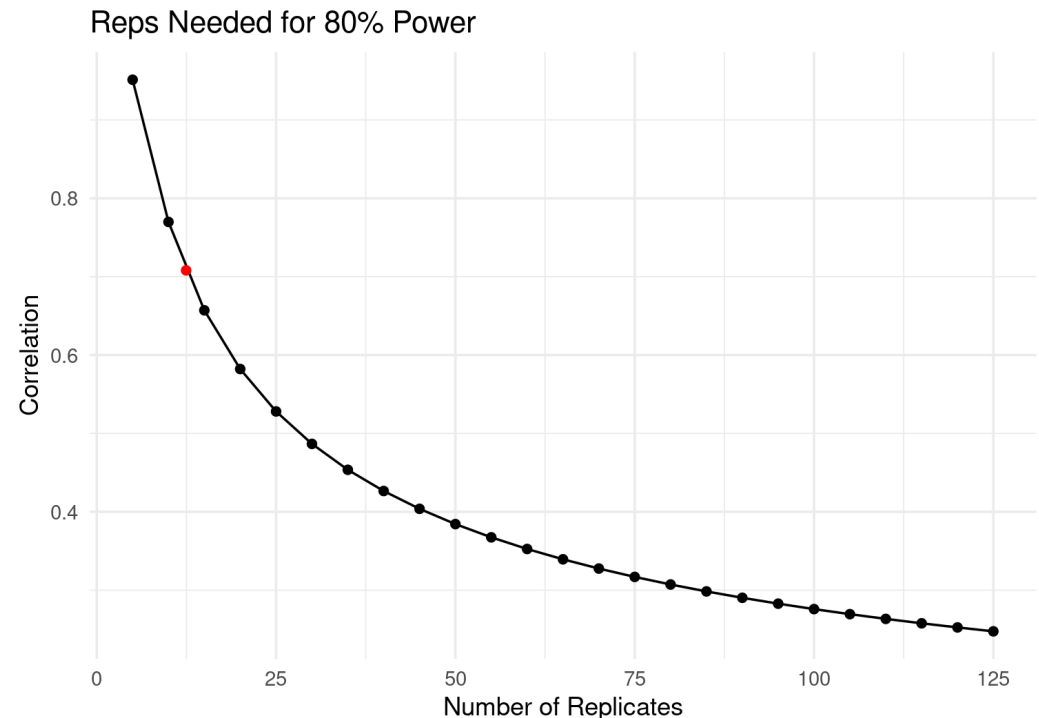
# Scenario 4

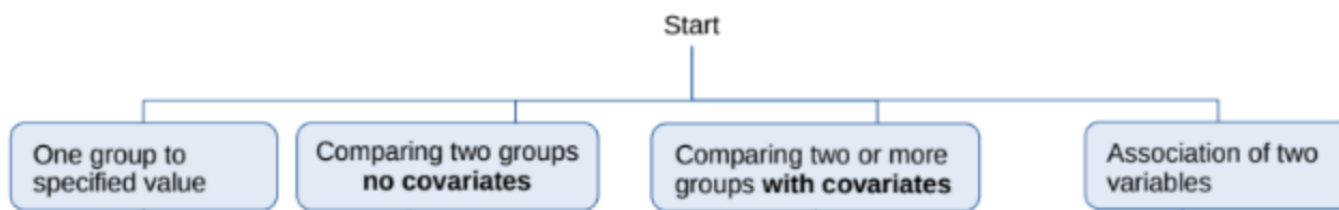


You noticed an interesting correlation between root length and panicle weight as part of a subgroup analysis. You don't have a statistically significant correlation and want to know how many reps you'd need in a new experiment to see one.

Here we show the effect size on Y as though it were a function of the number of reps needed to reach 80% power.

This is a little different, remember any aspect can be the outcome.

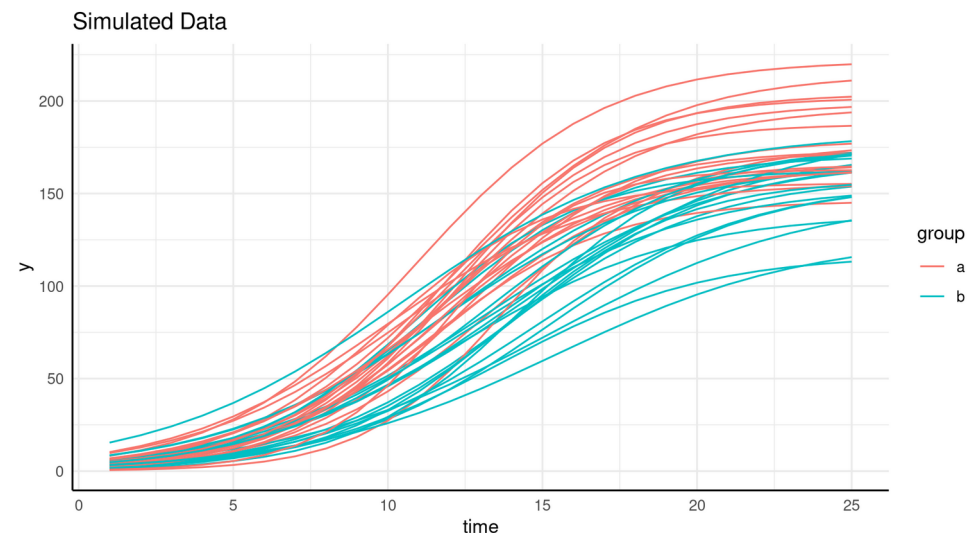




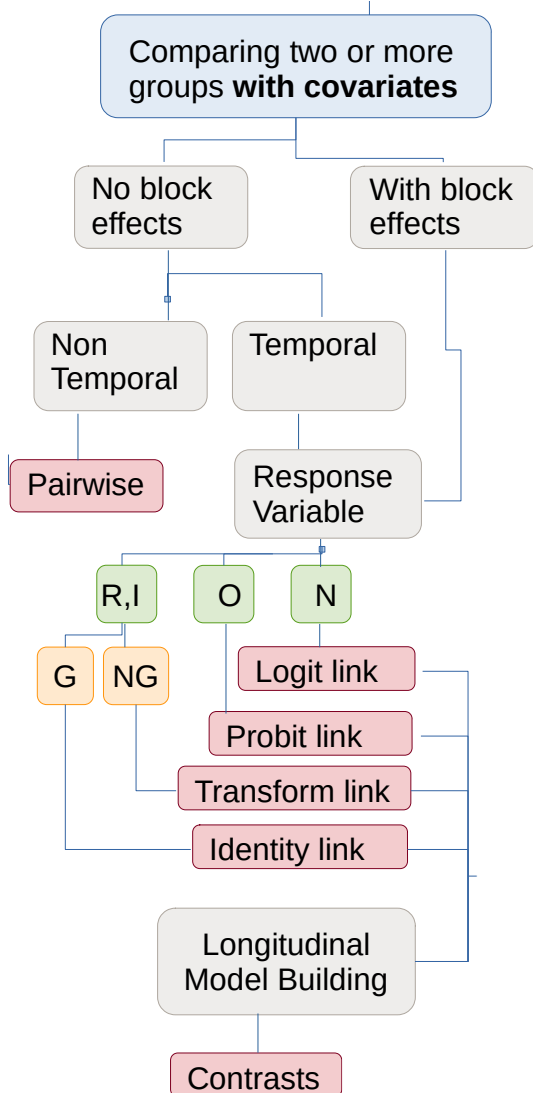
# Scenario 5

- You want to test the difference in two genotypes growth rates under early heat stress over 25 days of imaging. You expect a difference of  $\sim 10\text{cm}^2$  in size at the end of the experiment and relatively minor differences in growth rate and inflection.

How many reps do you need assuming a logistic growth model?

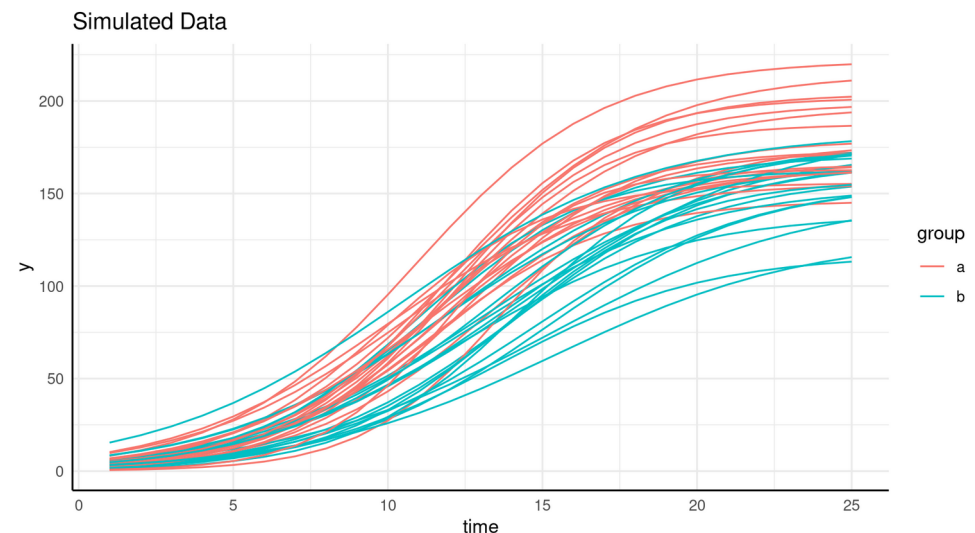


# Scenario 5

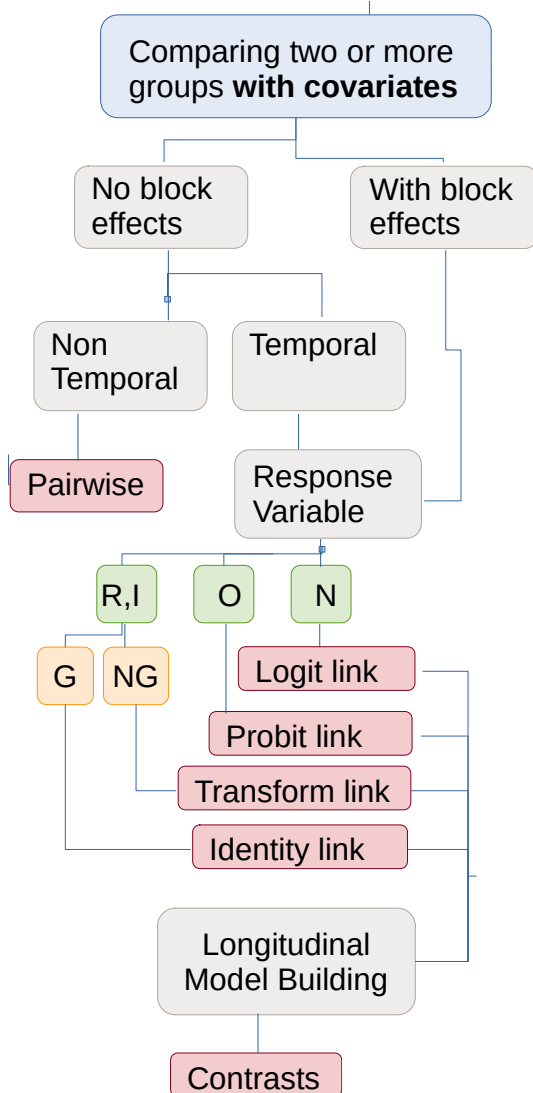


- You want to test the difference in two genotypes growth rates under early heat stress over 25 days of imaging. You expect a difference of  $\sim 10\text{cm}^2$  in size at the end of the experiment and relatively minor differences in growth rate and inflection.

How many reps do you need assuming a logistic growth model?



# Scenario 5

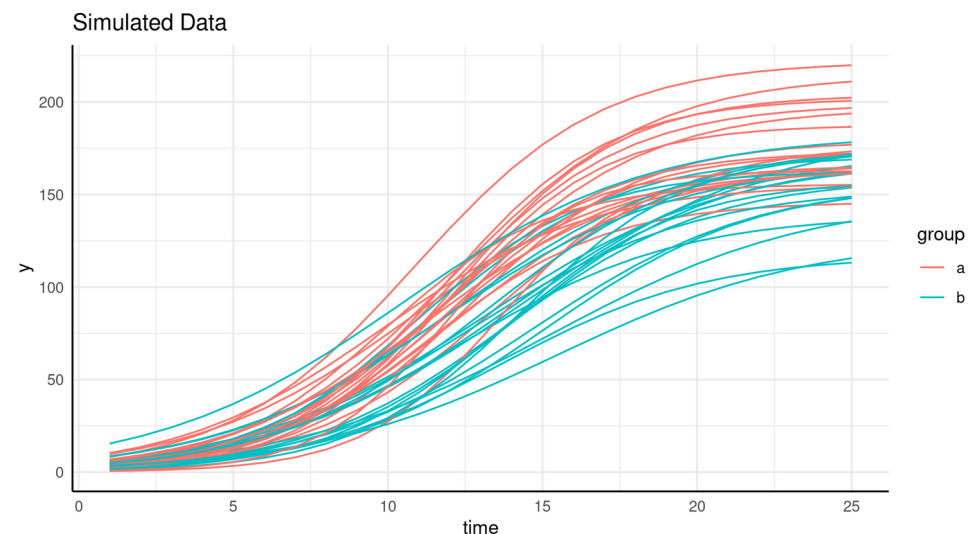


- You want to test the difference in two genotypes growth rates under early heat stress over 25 days of imaging. You expect a difference of  $\sim 10\text{cm}^2$  in size at the end of the experiment and relatively minor differences in growth rate and inflection.

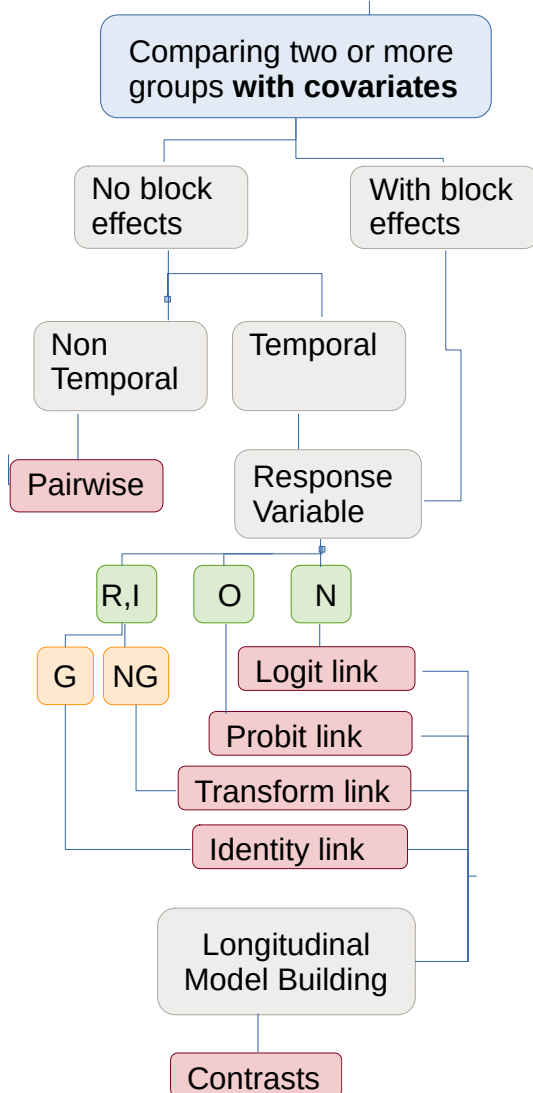
How many reps do you need assuming a logistic growth model?

## Statistical Test

We have temporal data and will need to build a longitudinal model.



# Scenario 5

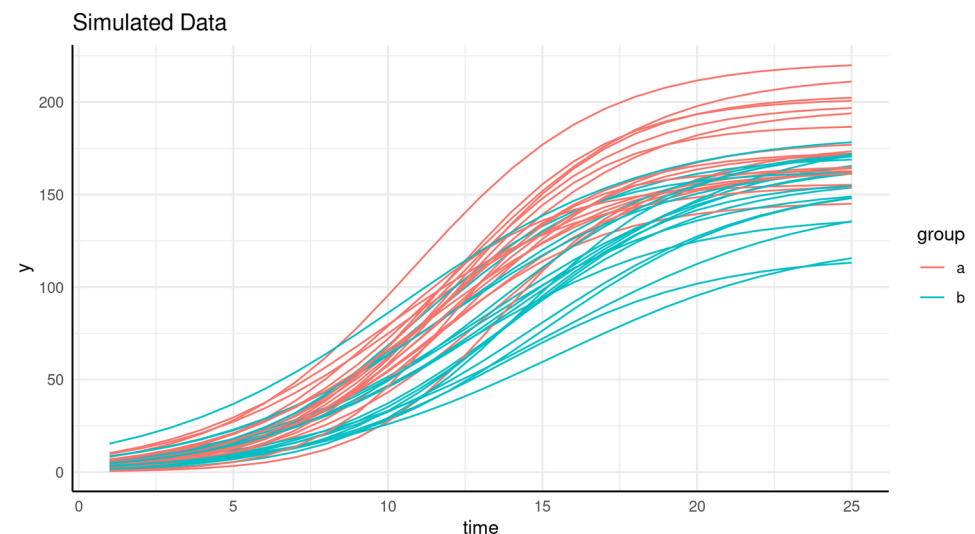


- You want to test the difference in two genotypes growth rates under early heat stress over 25 days of imaging. You expect a difference of  $\sim 10\text{cm}^2$  in size at the end of the experiment and relatively minor differences in growth rate and inflection.

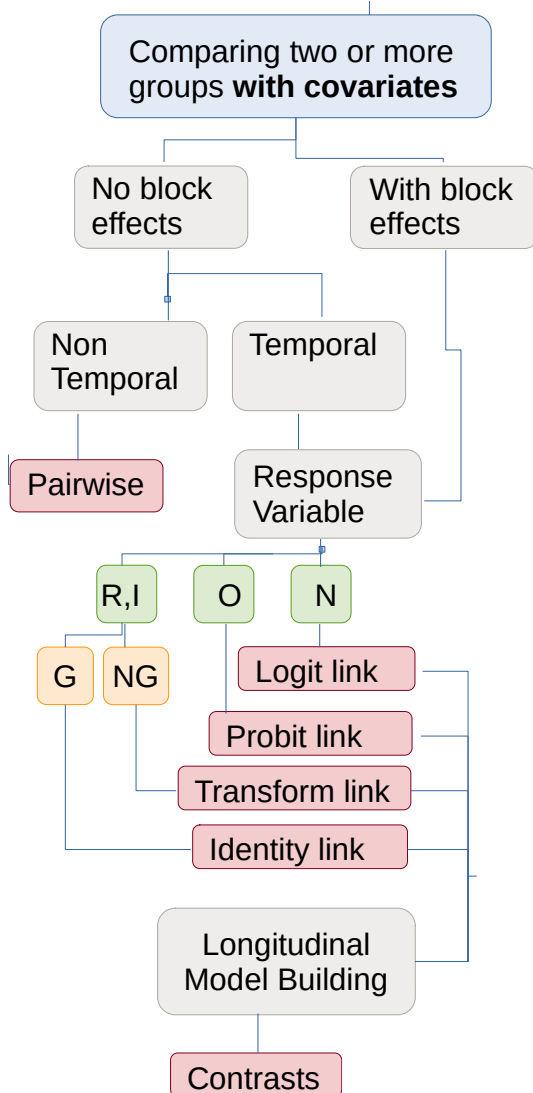
How many reps do you need assuming a logistic growth model?

## Power Analysis

This is a case where we have to use a simulation.



# Scenario 5



- You want to test the difference in two genotypes growth rates under early heat stress over 25 days of imaging. You expect a difference of  $\sim 10\text{cm}^2$  in size at the end of the experiment and relatively minor differences in growth rate and inflection.

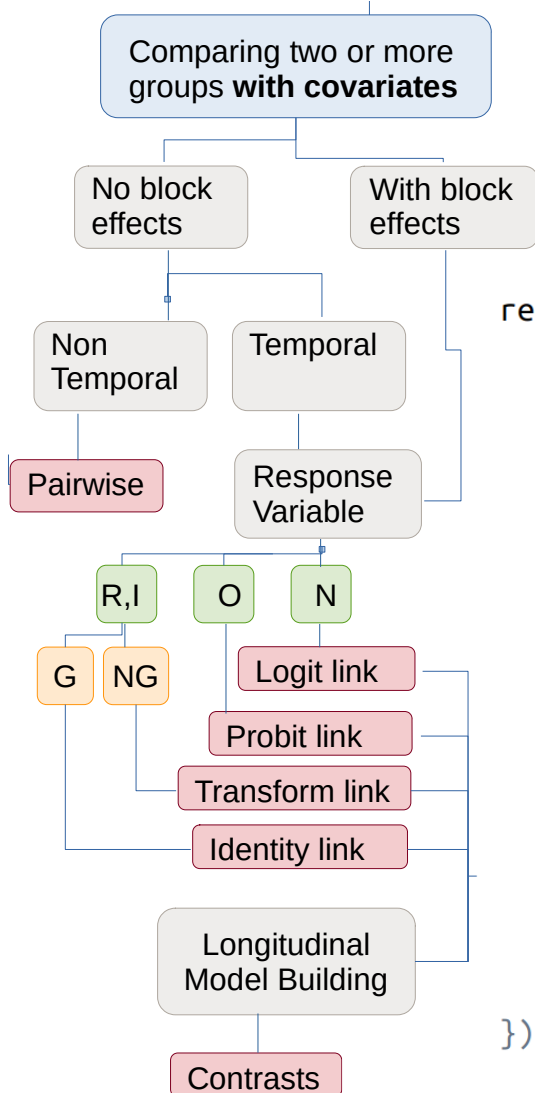
How many reps do you need assuming a logistic growth model?

```
> library(pcvr)
>
> repRange <- seq(5,35,5)
> A1 <- 180
> B1 <- 12
> C1 <- 3
> A2 <- 170
> B2 <- 13
> C2 <- 3.5
> sig.level <- 0.05
> iter = 1000
```

## Power Analysis

This is a case where we have to use a simulation.

# Scenario 5



```
res <- do.call(rbind, lapply(repRange, function(n){
 inner <- parallel::mclapply(1:iter, function(i){
 df <- growthSim("logistic", n=n, t=25,
 params = list("A"=c(A1, A2), "B"=c(B1, B2), "C"=c(C1, C2)))
 ss <- suppressMessages(growthSS("logistic", y ~ time|id/group, df = df, type="nls"))
 m <- fitGrowth(ss)
 t <- testGrowth(ss, m, test = list("A2 - A1", "B2 - B1", "C2 - C1"))
 asymp_test <- t[1,"p-value"]<sig.level
 infl_test <- t[2,"p-value"]<sig.level
 rate_test <- t[3,"p-value"]<sig.level
 return(list("A" = asymp_test, "B" = infl_test, "C"=rate_test))
 }, mc.cores = 10)
 asymp <- unlist(lapply(inner, function(o) o$A))
 infl <- unlist(lapply(inner, function(o) o$B))
 rate <- unlist(lapply(inner, function(o) o$C))
 data.frame(n = n, A = mean(asymp), B = mean(infl), C = mean(rate))
}))
```

## Power Analysis

This is a case where we have to use a simulation.



# For each N in our rep range, iterate

```
res <- do.call(rbind, lapply(repRange, function(n){
 inner <- parallel::mclapply(1:iter, function(i){
 df <- growthSim("logistic", n=n, t=25,
 params = list("A"=c(A1, A2), "B"=c(B1, B2), "C"=c(C1, C2)))
 ss <- suppressMessages(growthSS("logistic", y ~ time|id/group, df = df, type="nls"))
 m <- fitGrowth(ss)
 t <- testGrowth(ss, m, test = list("A2 - A1", "B2 - B1", "C2 - C1"))
 asymp_test <- t[1,"p-value"]<sig.level
 infl_test <- t[2,"p-value"]<sig.level
 rate_test <- t[3,"p-value"]<sig.level
 return(list("A" = asymp_test, "B" = infl_test, "C"=rate_test))
 }, mc.cores = 10)
 asymp <- unlist(lapply(inner, function(o) o$A))
 infl <- unlist(lapply(inner, function(o) o$B))
 rate <- unlist(lapply(inner, function(o) o$C))
 data.frame(n = n, A = mean(asymp), B = mean(infl), C = mean(rate))
}))
```

# Simulate a dataset

```
res <- do.call(rbind, lapply(repRange, function(n){
 inner <- parallel::mclapply(1:iter, function(i){
 df <- growthSim("logistic", n=n, t=25,
 params = list("A"=c(A1, A2), "B"=c(B1, B2), "C"=c(C1, C2)))
 ss <- suppressMessages(growthSS("logistic", y ~ time|id/group, df = df, type="nls"))
 m <- fitGrowth(ss)
 t <- testGrowth(ss, m, test = list("A2 - A1", "B2 - B1", "C2 - C1"))
 asymp_test <- t[1,"p-value"]<sig.level
 infl_test <- t[2,"p-value"]<sig.level
 rate_test <- t[3,"p-value"]<sig.level
 return(list("A" = asymp_test, "B" = infl_test, "C"=rate_test))
 }, mc.cores = 10)
 asymp <- unlist(lapply(inner, function(o) o$A))
 infl <- unlist(lapply(inner, function(o) o$B))
 rate <- unlist(lapply(inner, function(o) o$C))
 data.frame(n = n, A = mean(asymp), B = mean(infl), C = mean(rate))
}))
```

Note that these datasets have  
Some sampling variation, set to  
1/10th of the mean for each  
Parameter.

# Fit a model

```
res <- do.call(rbind, lapply(repRange, function(n){
 inner <- parallel::mclapply(1:iter, function(i){
 df <- growthSim("logistic", n=n, t=25,
 params = list("A"=c(A1, A2), "B"=c(B1, B2), "C"=c(C1, C2)))
 ss <- suppressMessages(growthSS("logistic", y ~ time|id/group, df = df, type="nls"))
 m <- fitGrowth(ss)
 t <- testGrowth(ss, m, test = list("A2 - A1", "B2 - B1", "C2 - C1"))
 asymp_test <- t[1,"p-value"]<sig.level
 infl_test <- t[2,"p-value"]<sig.level
 rate_test <- t[3,"p-value"]<sig.level
 return(list("A" = asymp_test, "B" = infl_test, "C"=rate_test))
 }, mc.cores = 10)
 asymp <- unlist(lapply(inner, function(o) o$A))
 infl <- unlist(lapply(inner, function(o) o$B))
 rate <- unlist(lapply(inner, function(o) o$C))
 data.frame(n = n, A = mean(asymp), B = mean(infl), C = mean(rate))
}))
```

# Record test results

---

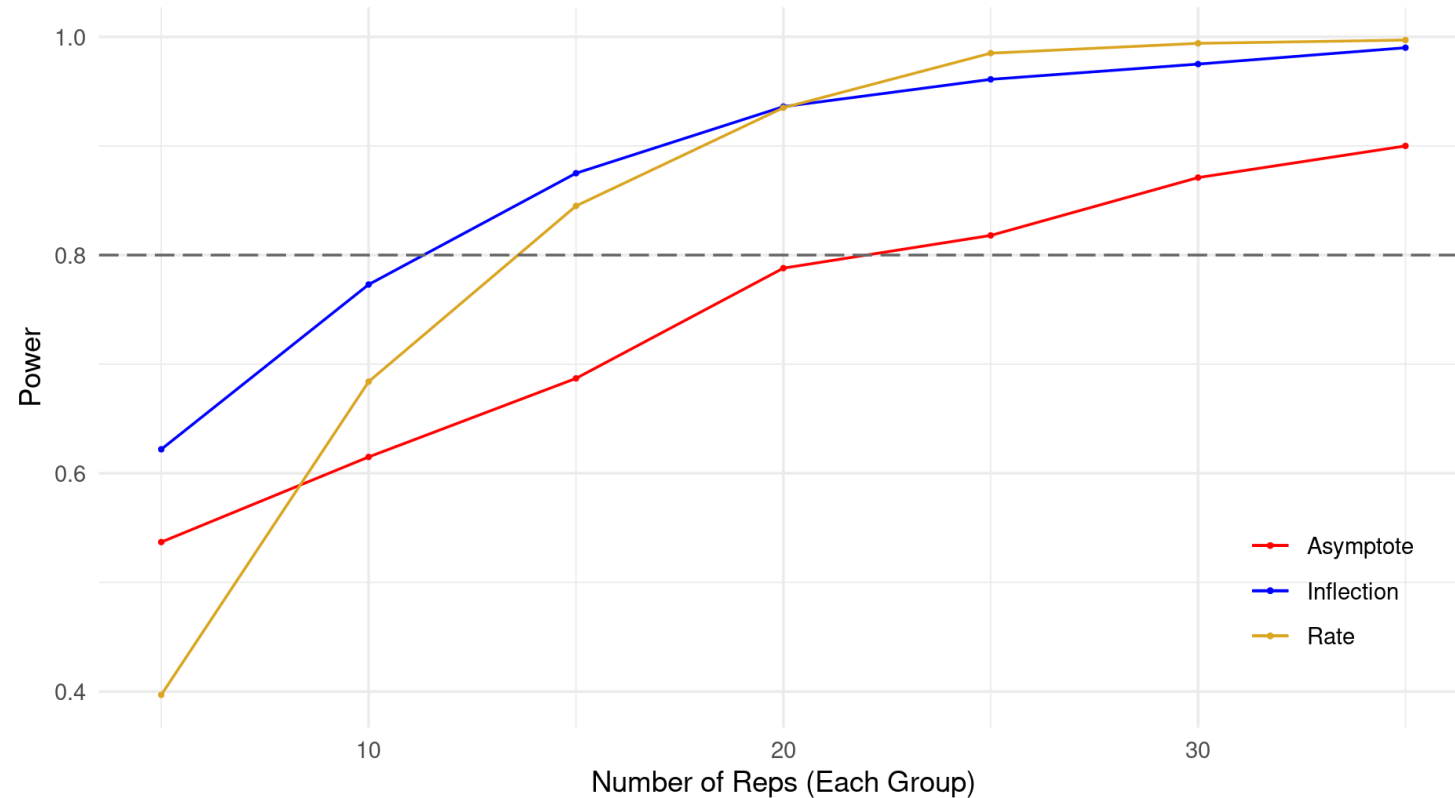
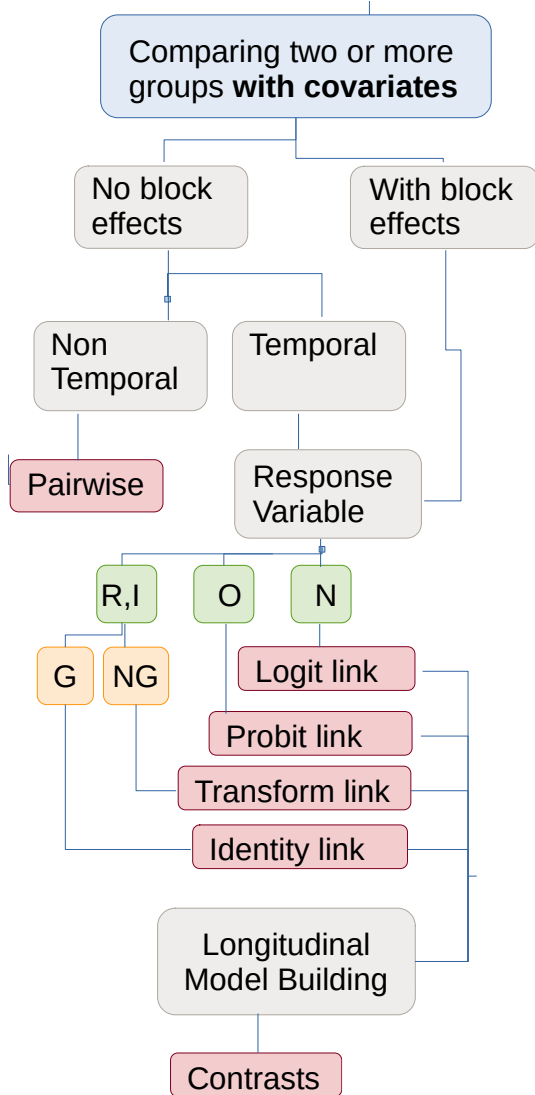
```
res <- do.call(rbind, lapply(repRange, function(n){
 inner <- parallel::mclapply(1:iter, function(i){
 df <- growthSim("logistic", n=n, t=25,
 params = list("A"=c(A1, A2), "B"=c(B1, B2), "C"=c(C1, C2)))
 ss <- suppressMessages(growthSS("logistic", y ~ time|id/group, df = df, type="nls"))
 m <- fitGrowth(ss)
 t <- testGrowth(ss, m, test = list("A2 - A1", "B2 - B1", "C2 - C1"))
 asymp_test <- t[1,"p-value"]<sig.level
 infl_test <- t[2,"p-value"]<sig.level
 rate_test <- t[3,"p-value"]<sig.level
 return(list("A" = asymp_test, "B" = infl_test, "C"=rate_test))
 }, mc.cores = 10)
 asymp <- unlist(lapply(inner, function(o) o$A))
 infl <- unlist(lapply(inner, function(o) o$B))
 rate <- unlist(lapply(inner, function(o) o$C))
 data.frame(n = n, A = mean(asymp), B = mean(infl), C = mean(rate))
}))
```

# Summarize results by N reps

---

```
res <- do.call(rbind, lapply(repRange, function(n){
 inner <- parallel::mclapply(1:iter, function(i){
 df <- growthSim("logistic", n=n, t=25,
 params = list("A"=c(A1, A2), "B"=c(B1, B2), "C"=c(C1, C2)))
 ss <- suppressMessages(growthSS("logistic", y ~ time|id/group, df = df, type="nls"))
 m <- fitGrowth(ss)
 t <- testGrowth(ss, m, test = list("A2 - A1", "B2 - B1", "C2 - C1"))
 asymp_test <- t[1,"p-value"]<sig.level
 infl_test <- t[2,"p-value"]<sig.level
 rate_test <- t[3,"p-value"]<sig.level
 return(list("A" = asymp_test, "B" = infl_test, "C"=rate_test))
 }, mc.cores = 10)
 asymp <- unlist(lapply(inner, function(o) o$A))
 infl <- unlist(lapply(inner, function(o) o$B))
 rate <- unlist(lapply(inner, function(o) o$C))
 data.frame(n = n, A = mean(asymp), B = mean(infl), C = mean(rate))
}))
```

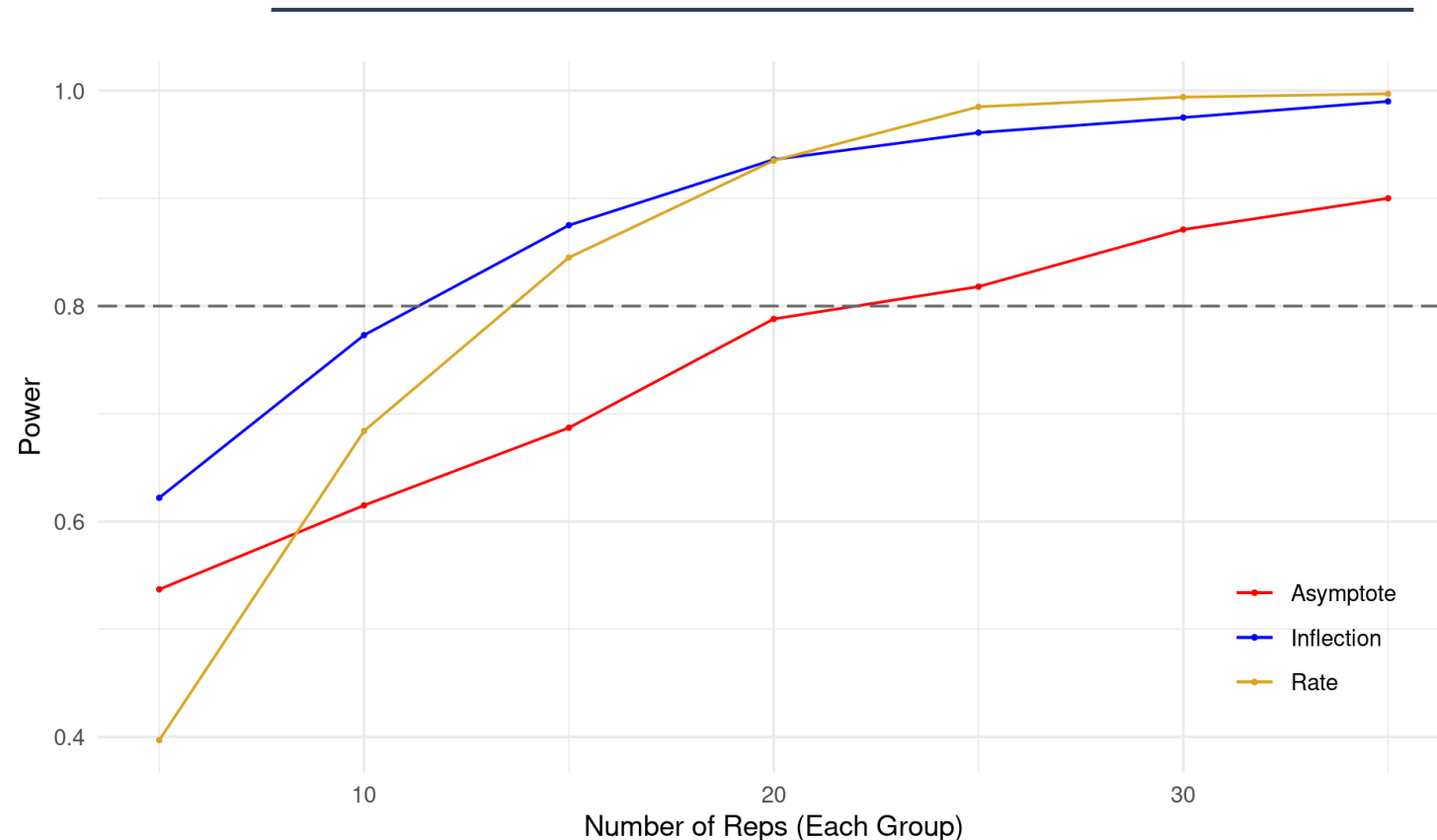
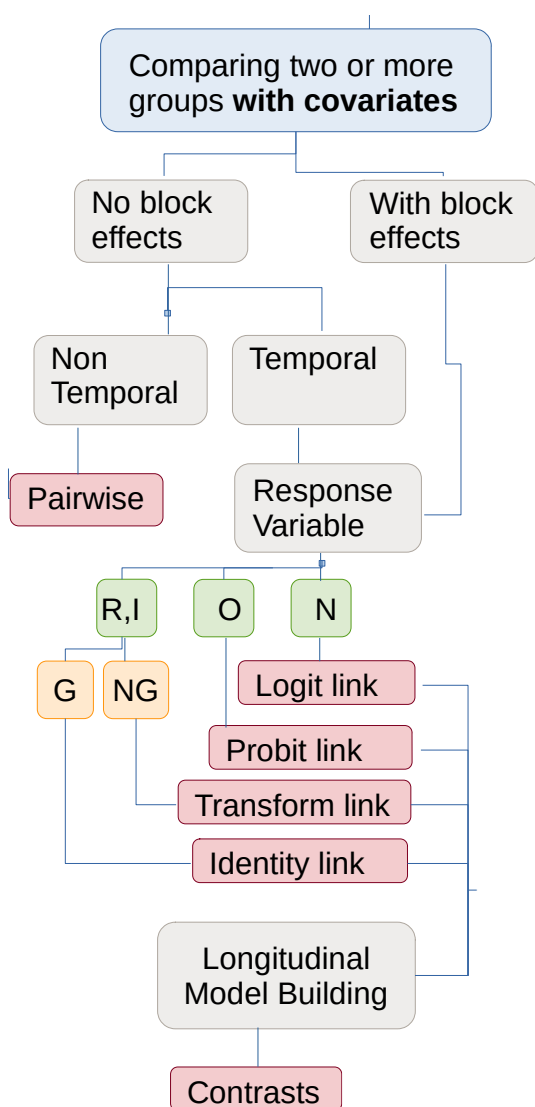
# Scenario 5



## Power Analysis

This is a case where we have to use a simulation.

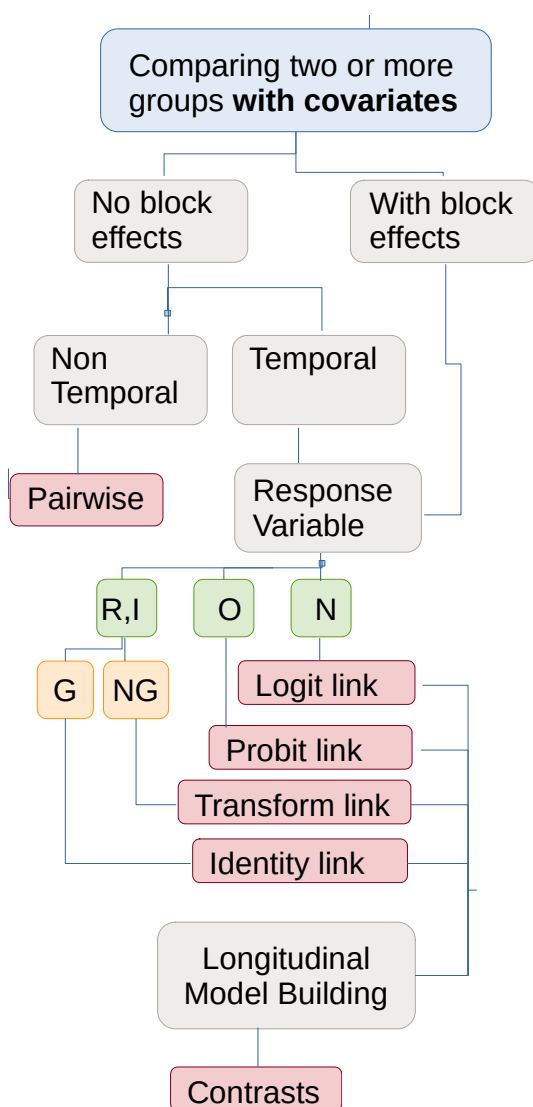
# Scenario 5



## Power Analysis

But here we input fixed parameters, what if we don't have that information?

# Scenario 5

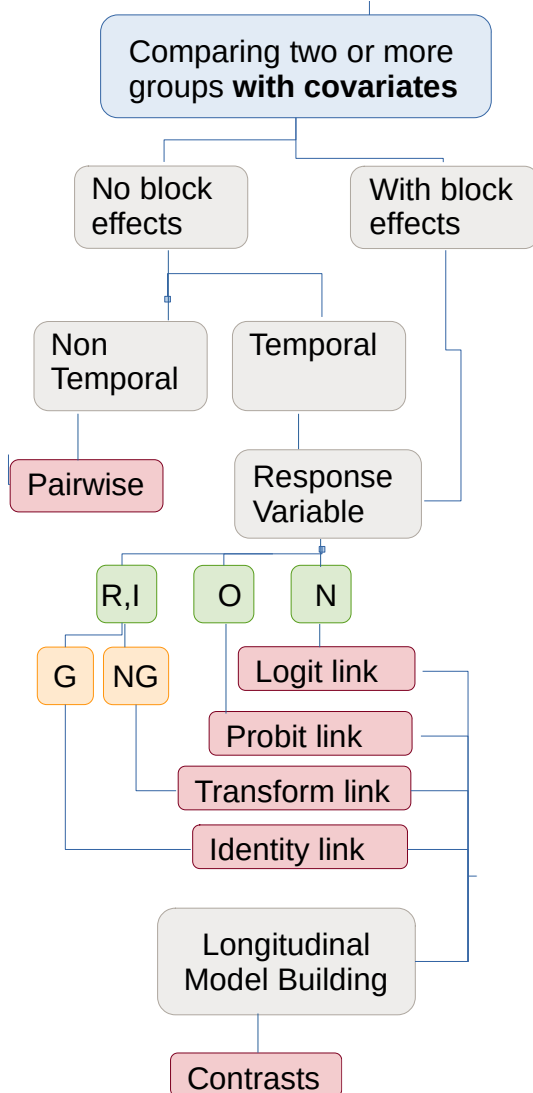


- Instead of  $A_1 = 180$  we now will say  $A_1 \sim N(180, 5)$ 
  - $A_2 \sim N(160, 5)$
  - $B_1 \sim N(12, 1)$
  - $B_2 \sim N(13, 1)$
  - $C_1 \sim N(3, 0.25)$
  - $C_2 \sim N(3.5, 0.25)$

## Power Analysis



# Scenario 5



```
res2 <- do.call(rbind, lapply(repRange, function(n){
 inner <- parallel::mclapply(1:iter, function(i){
 A1 <- rnorm(1, 180, 5)
 A2 <- rnorm(1, 160, 5)
 B1 <- rnorm(1, 12, 1)
 B2 <- rnorm(1, 13, 1)
 C1 <- rnorm(1, 3, 0.25)
 C2 <- rnorm(1, 3.5, 0.25)
 df <- growthSim("logistic", n=n, t=25,
 params = list("A"=c(A1, A2), "B"=c(B1, B2), "C"=c(C1, C2)))
 ss <- suppressMessages(growthSS("logistic", y ~ time|id/group, df = df, type="nls"))
 m <- fitGrowth(ss)
 t <- testGrowth(ss, m, test = list("A2 - A1", "B2 - B1", "C2 - C1"))
 asymp_test <- t[1,"p-value"]<sig.level
 infl_test <- t[2,"p-value"]<sig.level
 rate_test <- t[3,"p-value"]<sig.level
 return(list("A" = asymp_test, "B" = infl_test, "C"=rate_test))
 }, mc.cores = 10)
 asymp <- unlist(lapply(inner, function(o) o$A))
 infl <- unlist(lapply(inner, function(o) o$B))
 rate <- unlist(lapply(inner, function(o) o$C))
 data.frame(n = n, A = mean(asymp), B = mean(infl), C = mean(rate))
}))
```

## Power Analysis

# Scenario 5

```
res2 <- do.call(rbind, lapply(repRange, function(n){
 inner <- parallel::mclapply(1:iter, function(i){
 A1 <- rnorm(1, 180, 5)
 A2 <- rnorm(1, 160, 5)
 B1 <- rnorm(1, 12, 1)
 B2 <- rnorm(1, 13, 1)
 C1 <- rnorm(1, 3, 0.25)
 C2 <- rnorm(1, 3.5, 0.25)
 df <- growthSim("logistic", n=n, t=25,
 params = list("A"=c(A1, A2), "B"=c(B1, B2), "C"=c(C1, C2)))
 ss <- suppressMessages(growthSS("logistic", y ~ time|id/group, df = df, type="nls"))
 m <- fitGrowth(ss)
 t <- testGrowth(ss, m, test = list("A2 - A1", "B2 - B1", "C2 - C1"))
 asymp_test <- t[1,"p-value"]<sig.level
 infl_test <- t[2,"p-value"]<sig.level
 rate_test <- t[3,"p-value"]<sig.level
 return(list("A" = asymp_test, "B" = infl_test, "C"=rate_test))
 }, mc.cores = 10)
 asymp <- unlist(lapply(inner, function(o) o$A))
 infl <- unlist(lapply(inner, function(o) o$B))
 rate <- unlist(lapply(inner, function(o) o$C))
 data.frame(n = n, A = mean(asymp), B = mean(infl), C = mean(rate))
}))
```

This is the only change we made.

But where do we get these distributions?

In practice this is similar to developing priors for a Bayesian model. It relies on your expertise and understanding some.

These datasets now have population Variation AND sampling variation

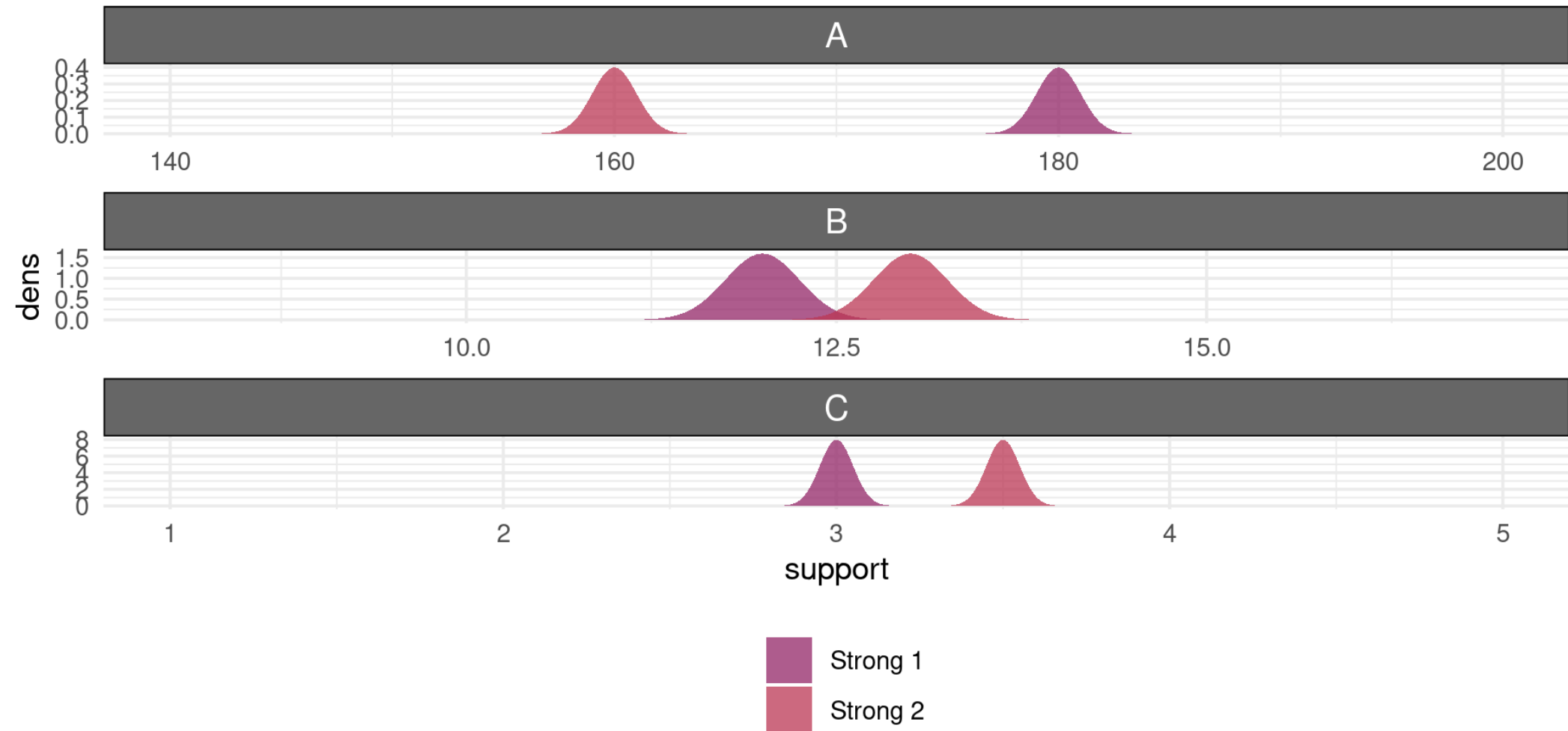
# Simulation Parameters and Priors

- There is more discussion about priors in the [Stats in pcvr](#) workshop.
- “Hard Headed” priors are generally suggested in the Bayesian literature.
  - To make this more concrete we’ll anthropomorphize some distributions.

# Strong Priors

- **Negative:** This guy has OPINIONS and they are not going to change based on your paltry “evidence”
- **Positive:** This guy is far from gullible, he will not exaggerate and aggrandize bad information.

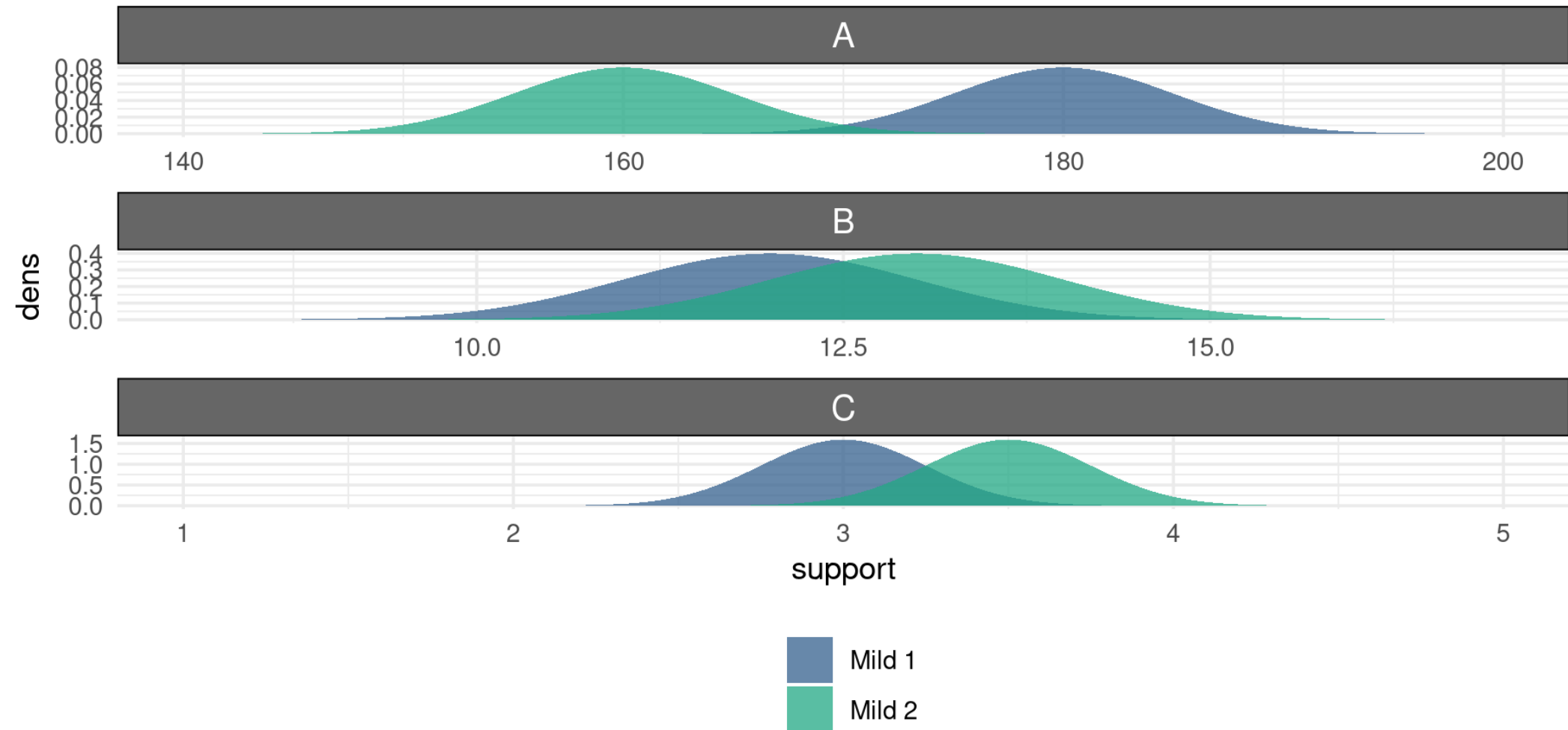
# Strong Priors



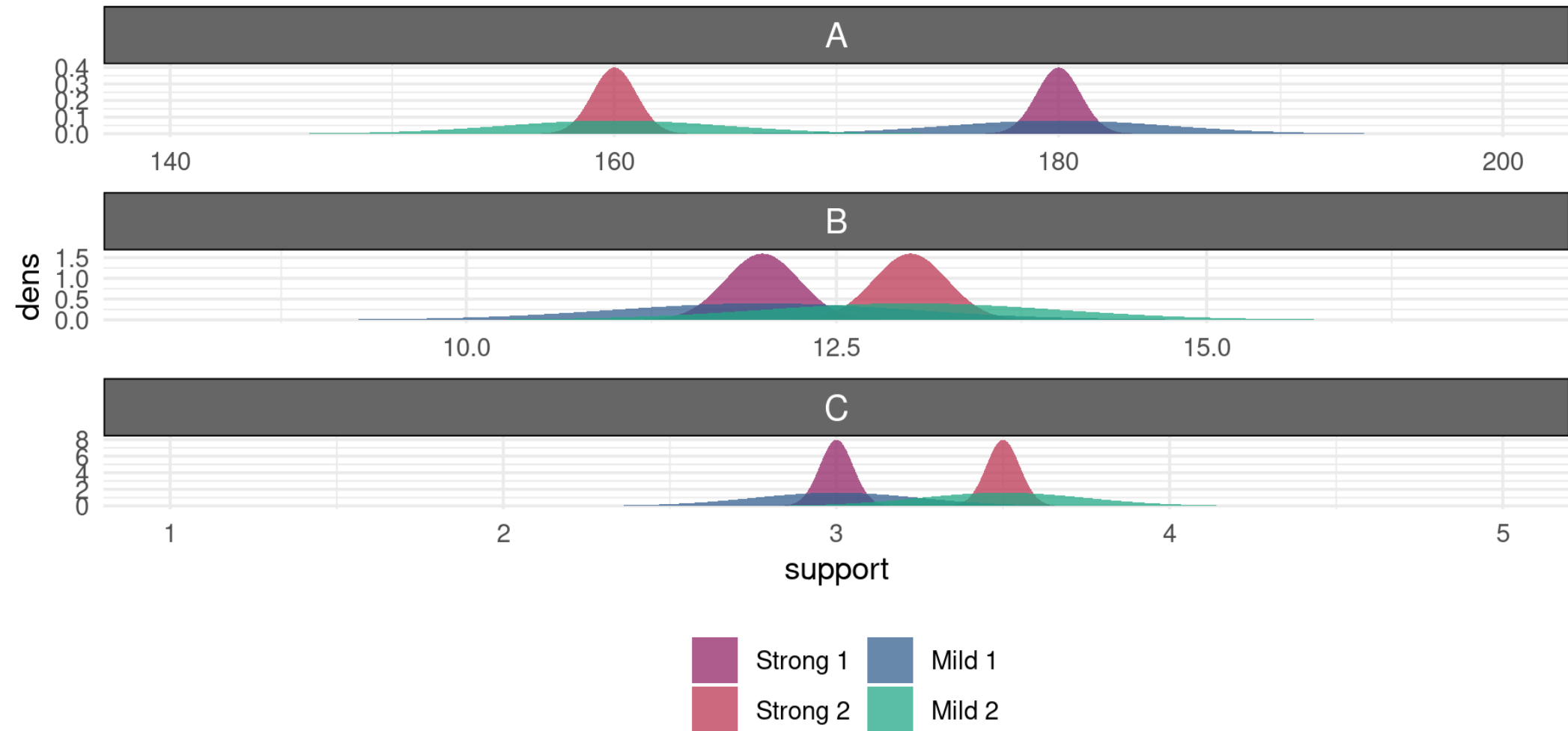
# Mild (weak) Priors

- **Negative:** This guy is not a domain expert, he is not very sure about what to expect. If you don't have much (data) to contribute then your conclusions will be limited.
- **Positive:** This guy knows he is not a domain expert, he is able to contribute to a conversation and collaborate without talking over the evidence you present.

# Mild (weak) Priors



# Mild (weak) Priors

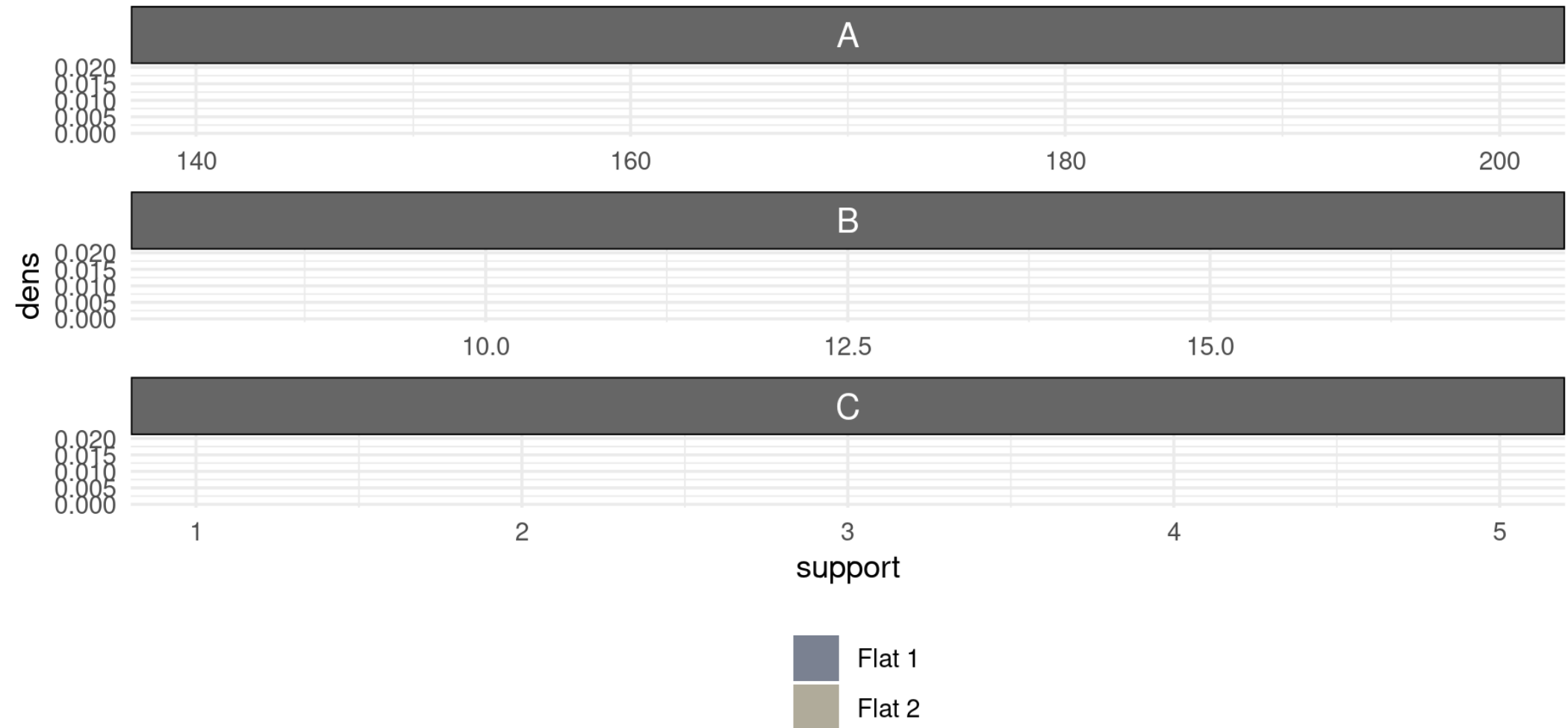




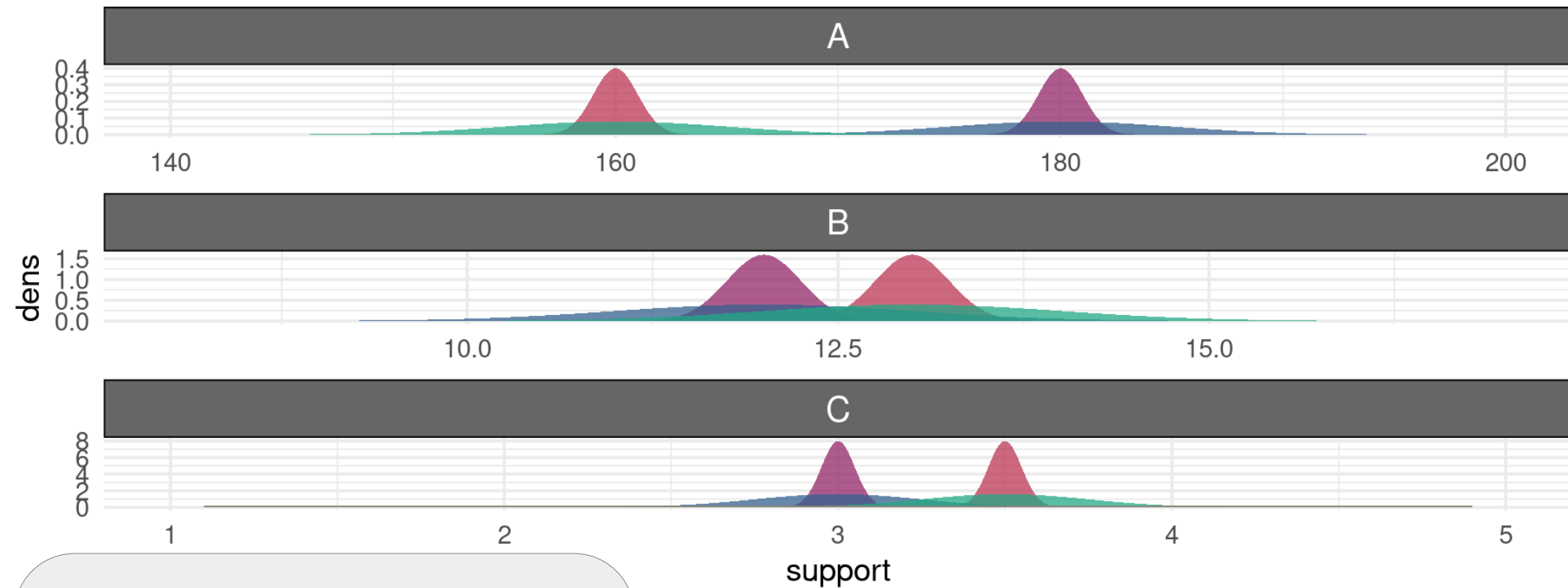
# Flat Priors

- **Negative:** This guy doesn't understand the world at all. He contributes nothing to the conversation and was only invited to round out the numbers.
- **Positive:** “Unbiased” in the eyes of many people, but those people are confusing “unbiased” with “ignorant”.

# Flat Priors



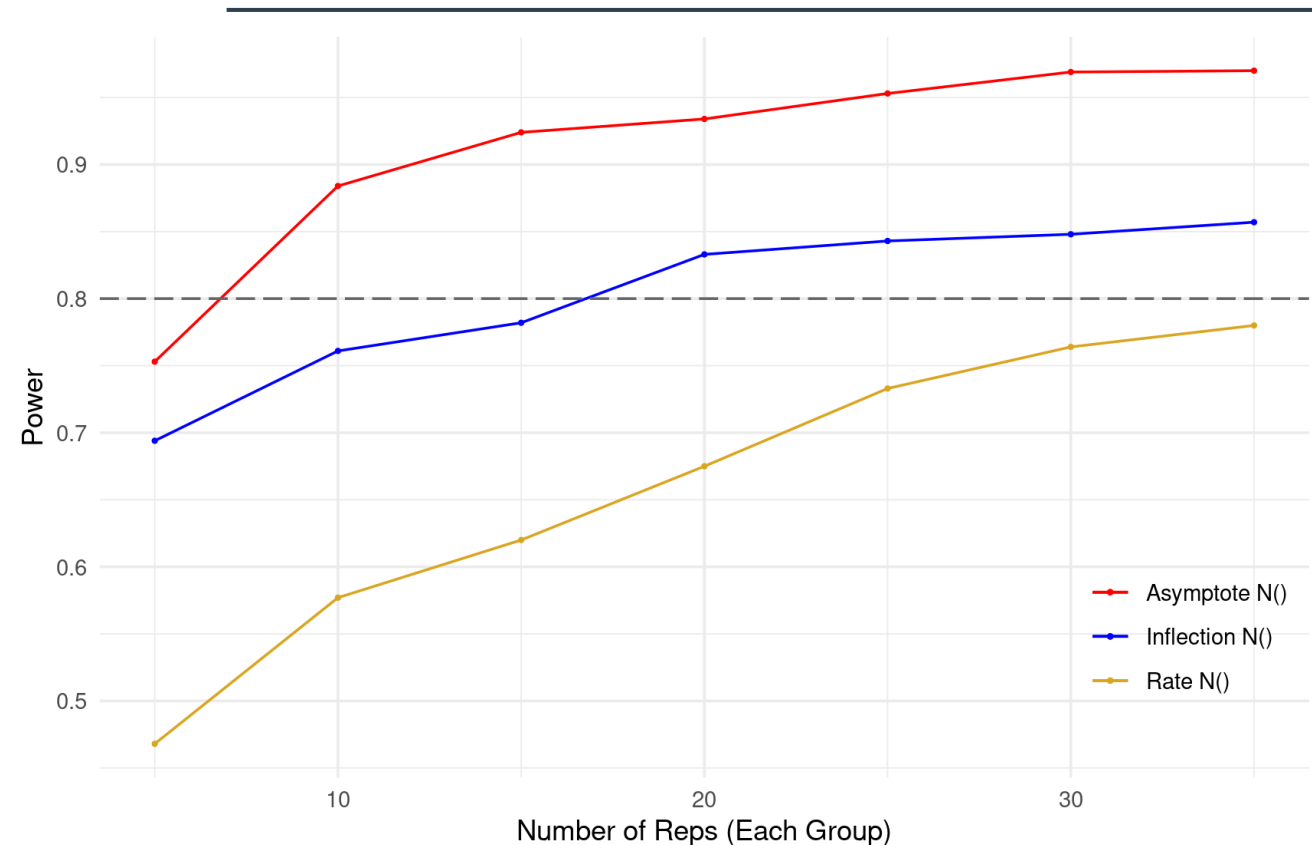
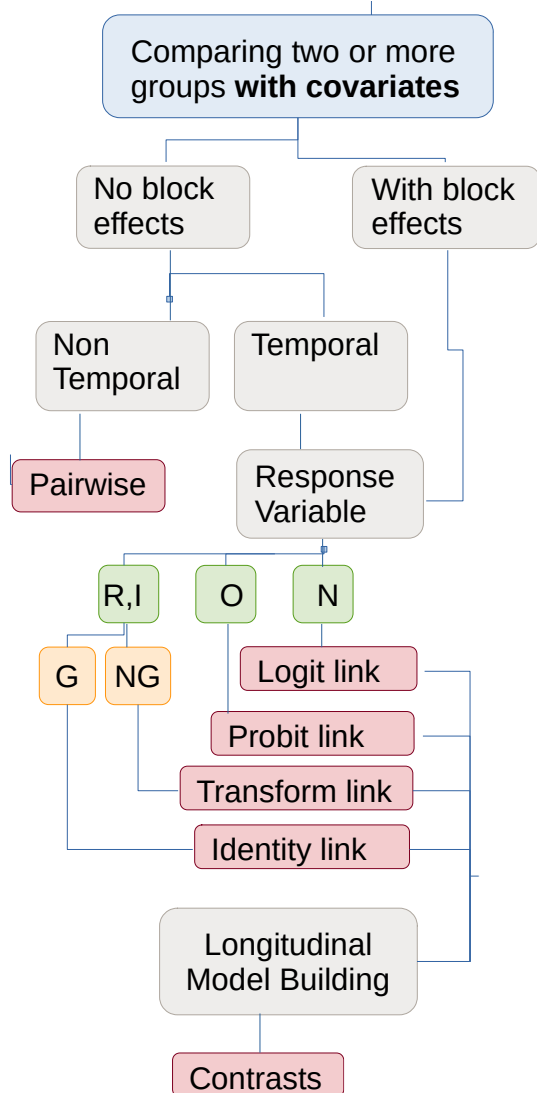
# Priors vs Simulations



The most important difference  
Here is that priors are updated  
With your experiment's data.

In the simulation, the distribution  
is all you have, so having pilot  
Data becomes very important.

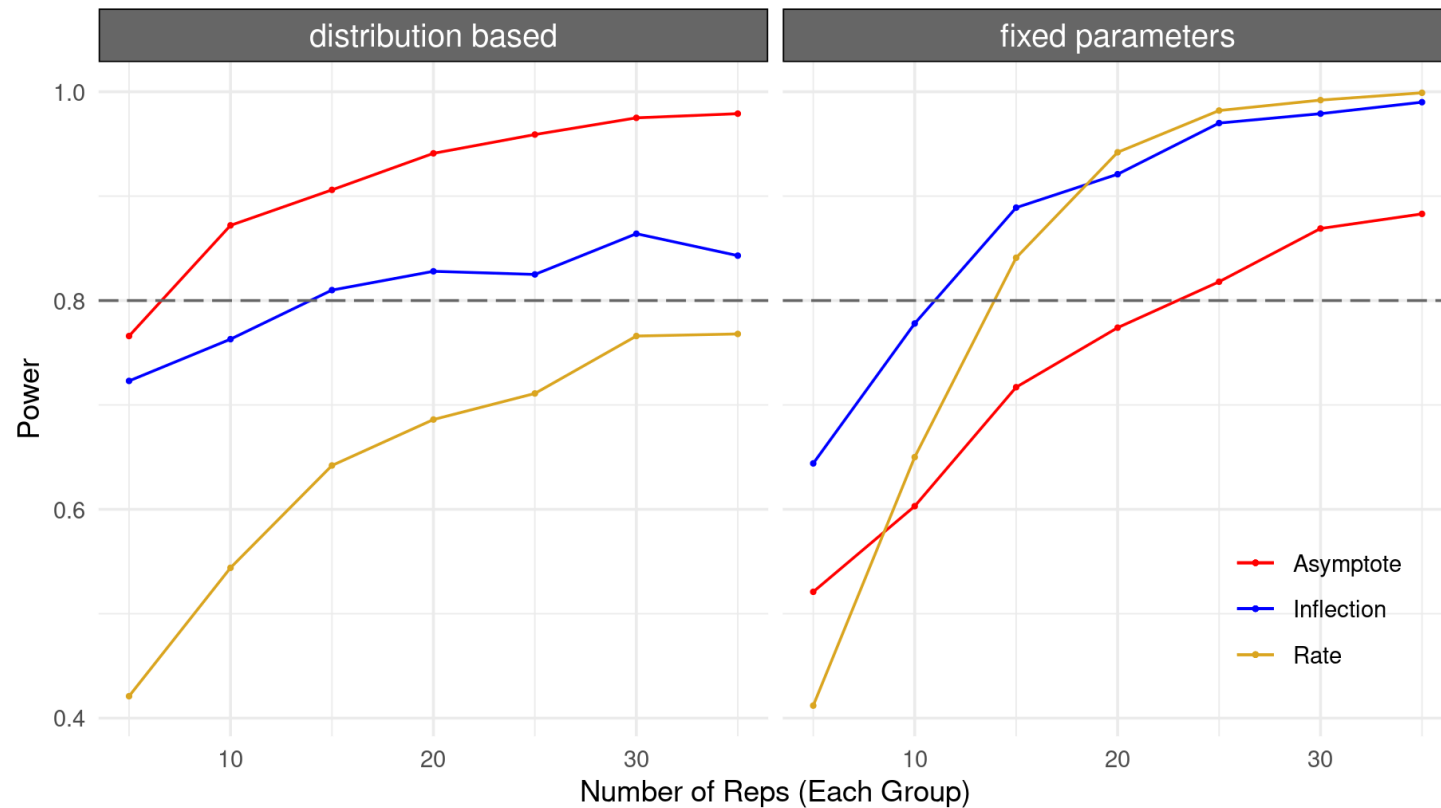
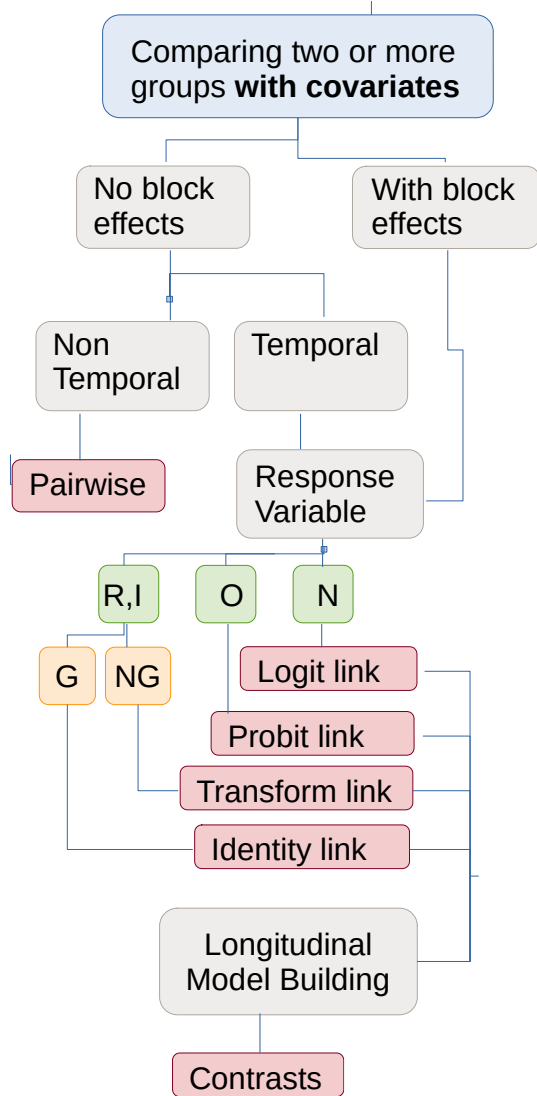
# Scenario 5



## Power Analysis

- Now we have very different power curves

# Scenario 5

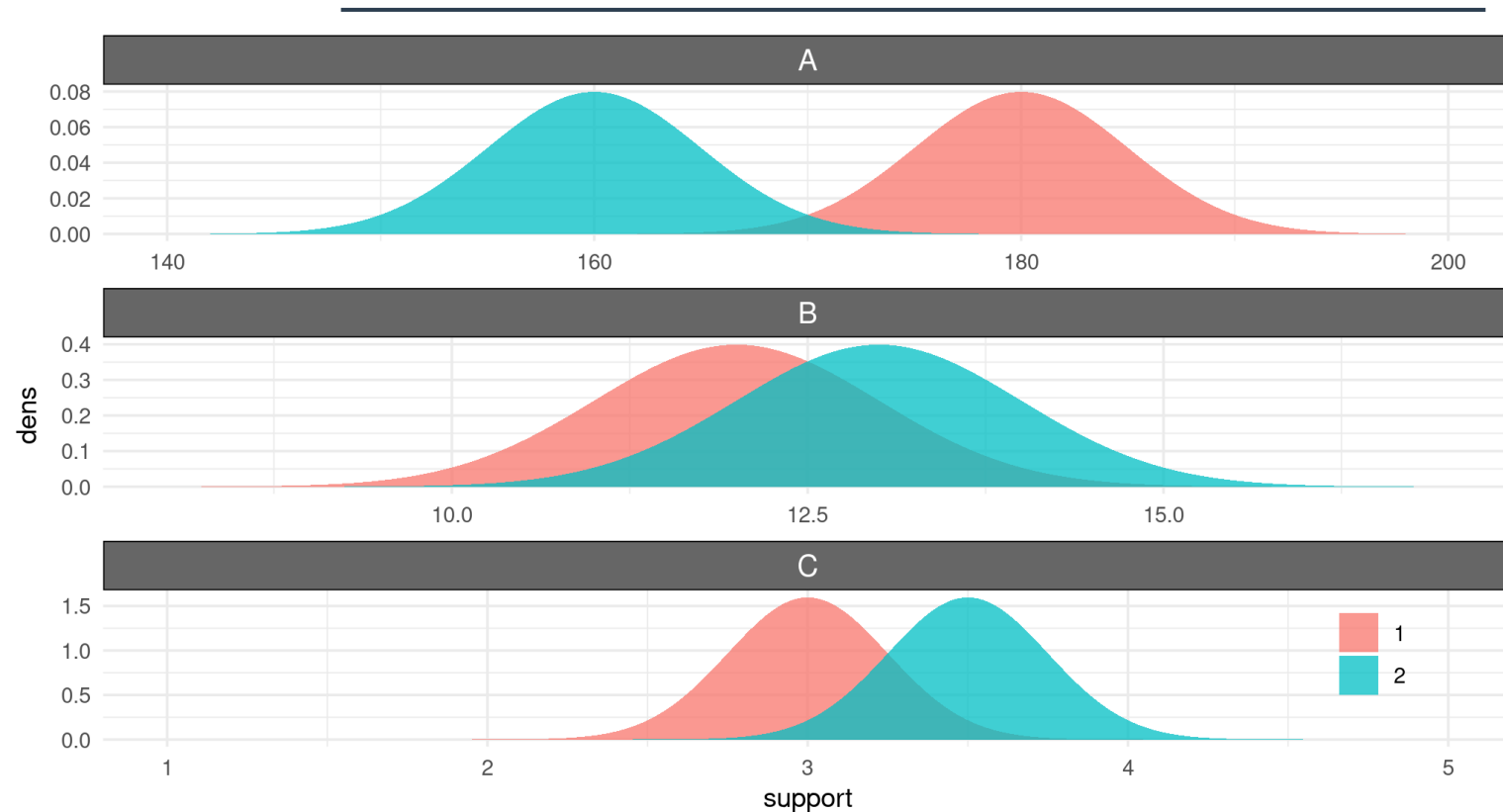
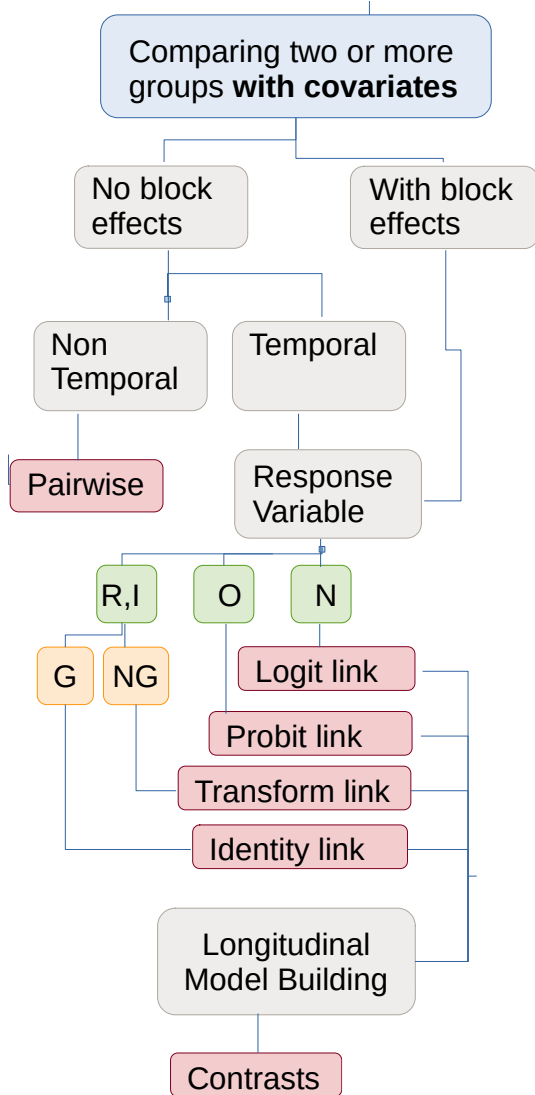


## Power Analysis

- Now we have very different power curves

- These new curves can be much more realistic, *depending on your assumptions.*

# Scenario 5



## Your Assumptions

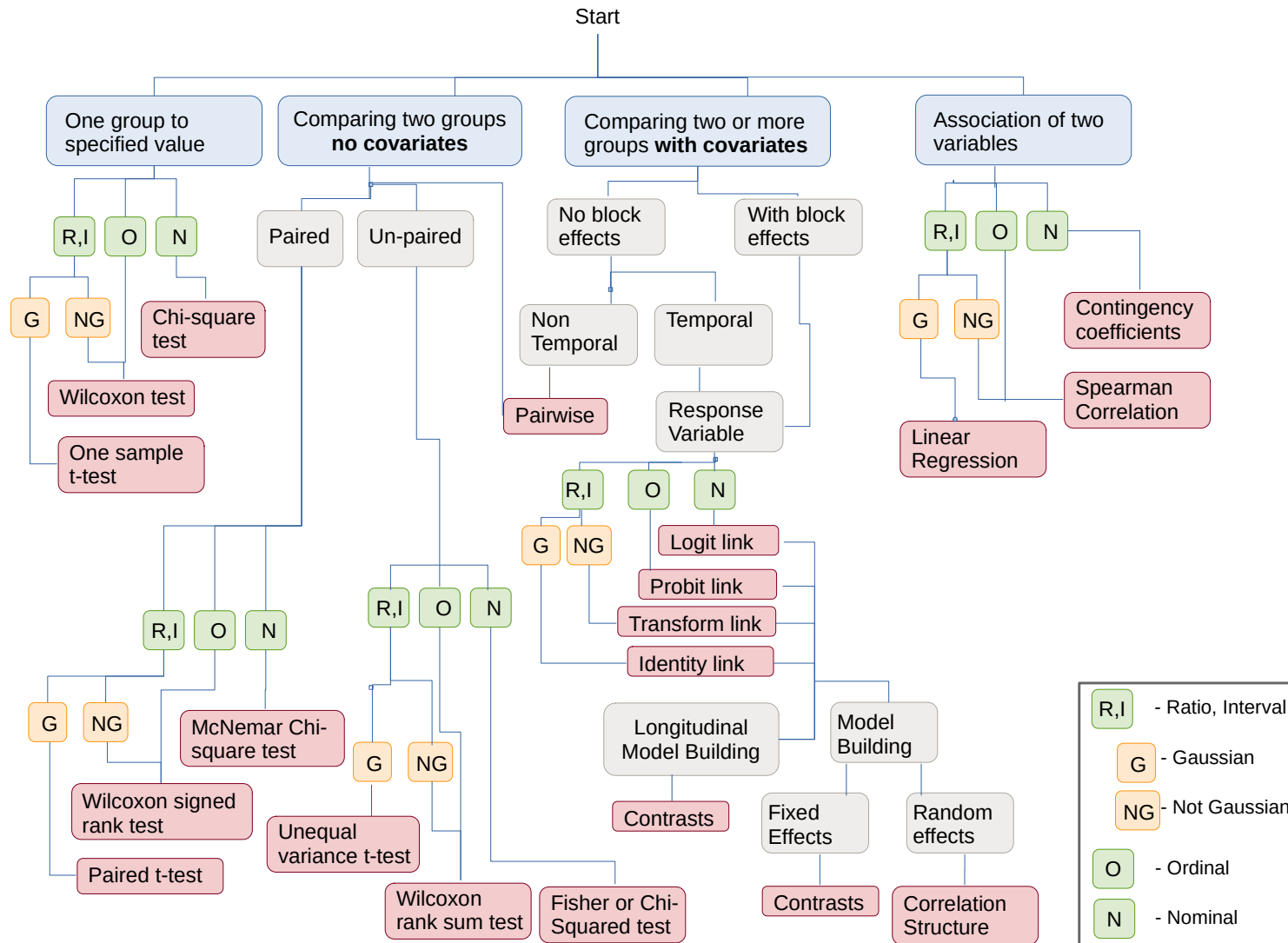
- Think about what you build into the simulation

- Here it's very clear why our asymptote test is higher power, there is less overlap in our simulated parameters.

# Simulations need guidance

- Setting parameters for simulations requires some expectation of what you could see.
  - This can be a similar process conceptually to setting priors for Bayesian analyses.
- Randomness is introduced by simulating data per each iteration or sampling a total dataset per simulation.
  - Your results will be as generalizable as your simulated data.
- If you are aiming to reach a precise power cutoff then this involves iterating over sample sizes.

# End of Scenarios



\* If the outcome and main predictor are both two-level factors, Breslow-Day and Cochren-Mantel-Hanzel tests are better



# Conclusion

- Are there other questions about this content?

# Conclusion

- Results are a function of statistical power
  - Statistical power is a function of experimental design, methods, and replication.
  - Methods are a function of time
  - Replication is a function of money
- You will have to optimize the utility function of **conclusions ~ time + money** for your own research. There is no formulaic approach to that problem.

# Conclusion

- Power is a function of experimental design through effect sizes, replication, and assumptions.
- Planning experiments while keeping statistical power in mind will help you when it comes time to analyze data.
- Both formulaic and simulation based power analyses have their place and can be useful.