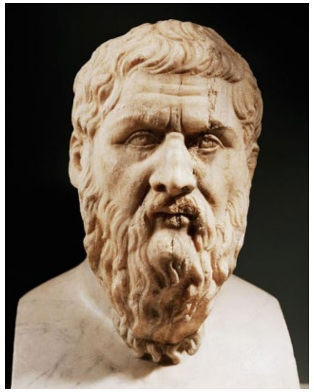
S&DS 101
Intro Statistics: Life Sciences

Overview

Parametric tests for comparing 2 means

The bootstrap

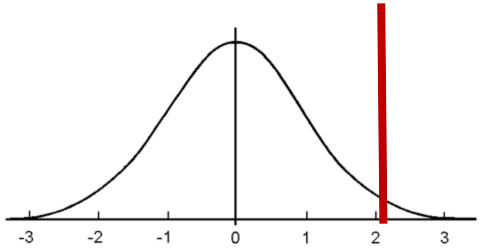
Five steps of hypothesis testing



1. State H_0 and H_A
   - Assume Gorgias (H_0) was right

2. Calculate the actual observed statistic

3. Create a distribution of what statistics would look like if Gorgias is right
   - Create the **null distribution** (that is consistent with H_0)



4. Get the probability we would get a statistic more than the observed statistic from the null distribution
   - p-value

5. Make a judgement
   - Assess whether the results are statistically significant

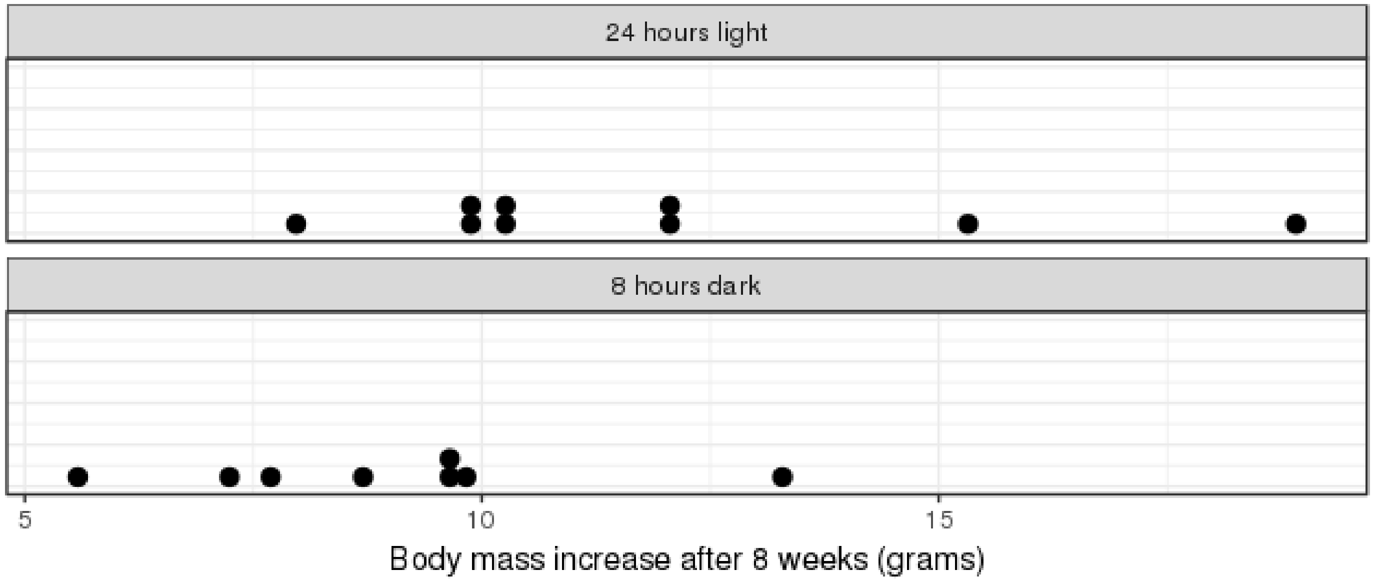
Do mice who eat late at night get fat?

A study by Fonken et al, 2010, wanted to examine whether more weight was gained by mice who could eat late at night

Mice were randomly divided into 2 groups:

- <u>Dark condition</u>: 8 mice were given 8 hours of darkness at night (when they couldn't eat)
- <u>Light condition</u>: 9 were constantly exposed to light for 24 hours (so they could always eat)

What's a good first thing to do when analyzing data?



Hypothesis tests for differences in two group means

1. State the null and alternative hypothesis
   - H_0: $\mu_{Dark} = \mu_{Light}$ or $\mu_{Dark} - \mu_{Light} = 0$
   - H_A: $\mu_{Dark} > \mu_{Light}$ or $\mu_{Dark} - \mu_{Light} > 0$

2. Calculate statistic of interest
   - $\bar{x}_{effect} = \bar{x}_{Dark} - \bar{x}_{Light}$

Do mice who eat late at night get fat?

You can get the data using:

```
>  download_class_data('mice.Rda')

> dark_BM_increase       # length(dark_BM_increase)
> light_BM_increase      # length(light_BM_increase)
```

Can you calculate the observed statistic (step 2)?

```
> obs_stat <- mean(light_BM_increase) - mean(dark_BM_increase)
```

What’s next?

3. Create the null distribution!

Mice in the light

Mice in the dark

Reconstructed sample of all mice

Shuffle data for random assignment consistent with H_0

Shuffled ‘light group’

Shuffled ‘dark group’

One null distribution statistic: $\bar{x}_{Shuff\_Dark} - \bar{x}_{Shuff\_Light}$

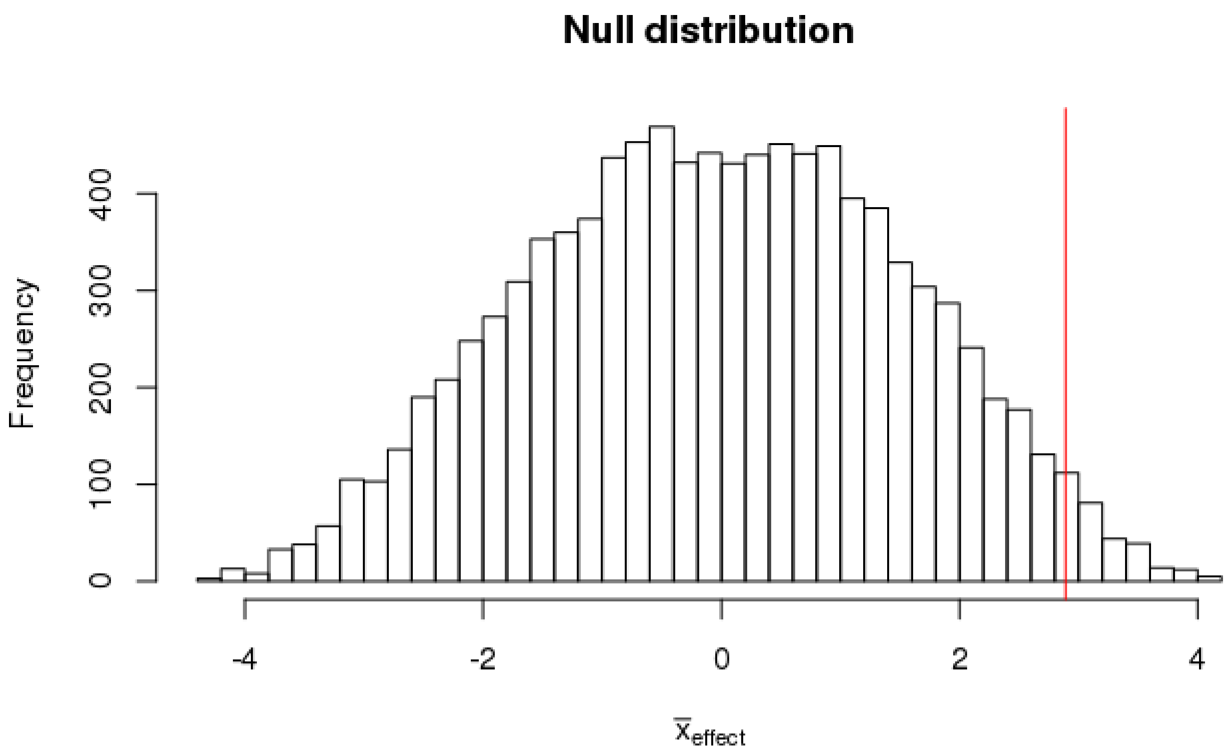
Do mice who eat late at night get fat?

What is the first thing we need to do for creating the null distribution?

```
    combo_data <- c(light_BM_increase, dark_BM_increase)

null_dist <- NULL

for (i in 1:10000) {

    shuff_data <- sample(combo_data)
    shuff_light <- shuff_data[1:9]
    shuff_dark <- shuff_data[10:17]
    null_dist[i] <- mean(shuff_light) - mean(shuff_dark)

}
```

Do mice who eat late at night get fat?

Plot the null distribution: hist(null_dist, nclass = 50)



What do we do next?

Do mice who eat late at night get fat?

Get the p-value

```
p_val <- sum(null_dist >= obs_stat)/10000
```

p-value = 0.02

Two-sided tests

If we did not have a belief about whether the light or the dark group would gain more weight, then we would state our hypotheses as follows:

State the null and alternative hypothesis

H_0: $\mu_{Dark} = \mu_{Light}$ or $\mu_{Dark} - \mu_{Light} = 0$

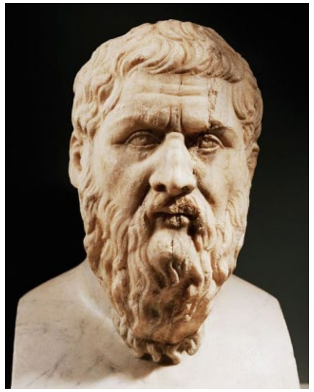
H_A: $\mu_{Dark} \neq \mu_{Light}$ or $\mu_{Dark} - \mu_{Light} \neq 0$

Steps 1-3 of hypothesis testing would still be the same, but for step 4 we would need to calculate the p-values by examining extreme values in both tails

```
p_val <- sum(abs(null_distr) >= abs(obs_stat))/10000
```

Parametric methods for comparing 2 means

Five steps of hypothesis testing

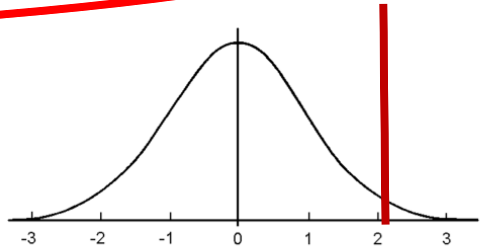


1. State H_0 and H_A
   - Assume Gorgias (H_0) was right

2. Calculate the actual observed statistic

3. Create a distribution of what statistics would look like if Gorgias is right
   - Create the **null distribution** (that is consistent with H_0)

4. Get the probability we would get a statistic more than the observed statistic from the null distribution
   - p-value



5. Make a judgement
   - Assess whether the results are statistically significant

Parametric methods for comparing 2 means

If we are testing the hypothesis comparing two means

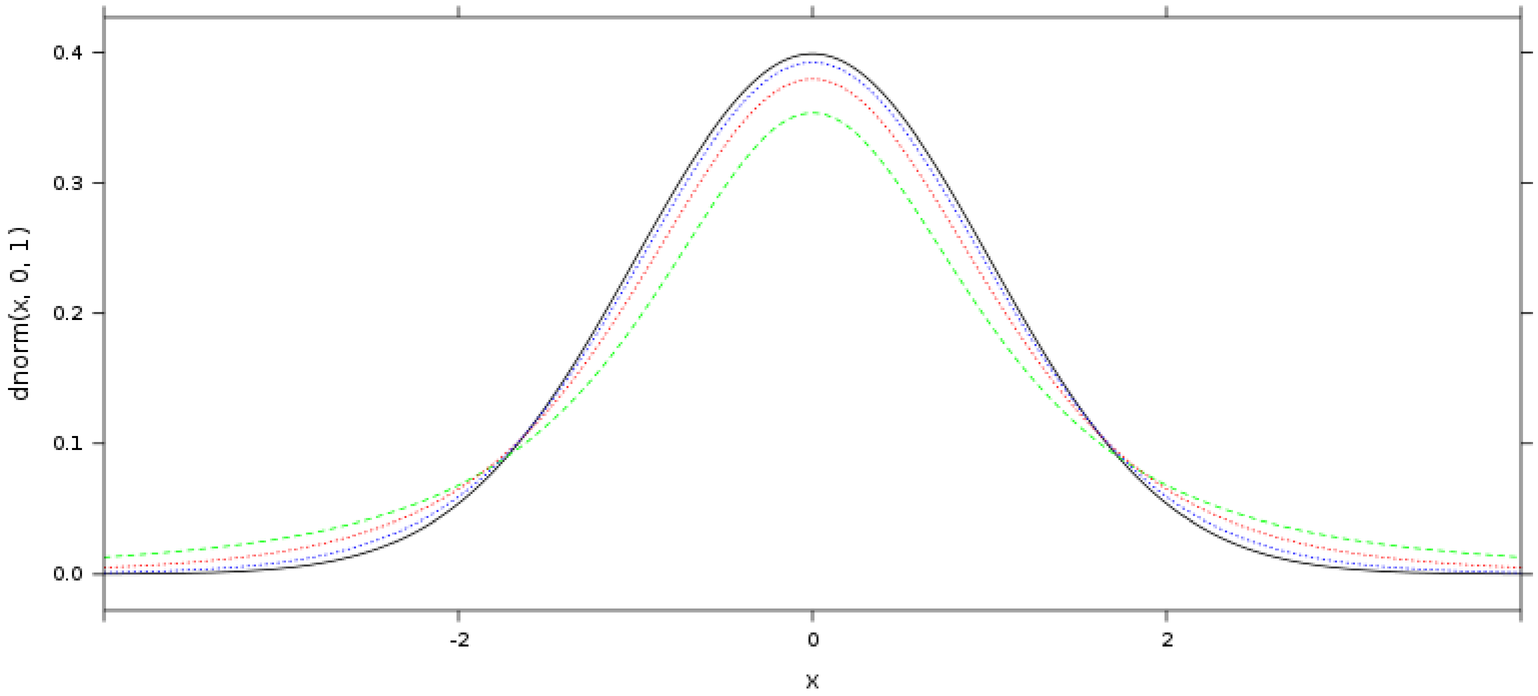
- H_0: $\mu_{Dark} - \mu_{Light} = 0$
- H_A: $\mu_{Dark} - \mu_{Light} > 0$

And we have two samples of size n_1 and n_2

We can calculate a t-statistic $$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

When H_0 is true, we can use a t-distribution, with degrees of freedom equal to the minimum of $n_1 - 1$ and $n_2 - 1$ as a null distribution

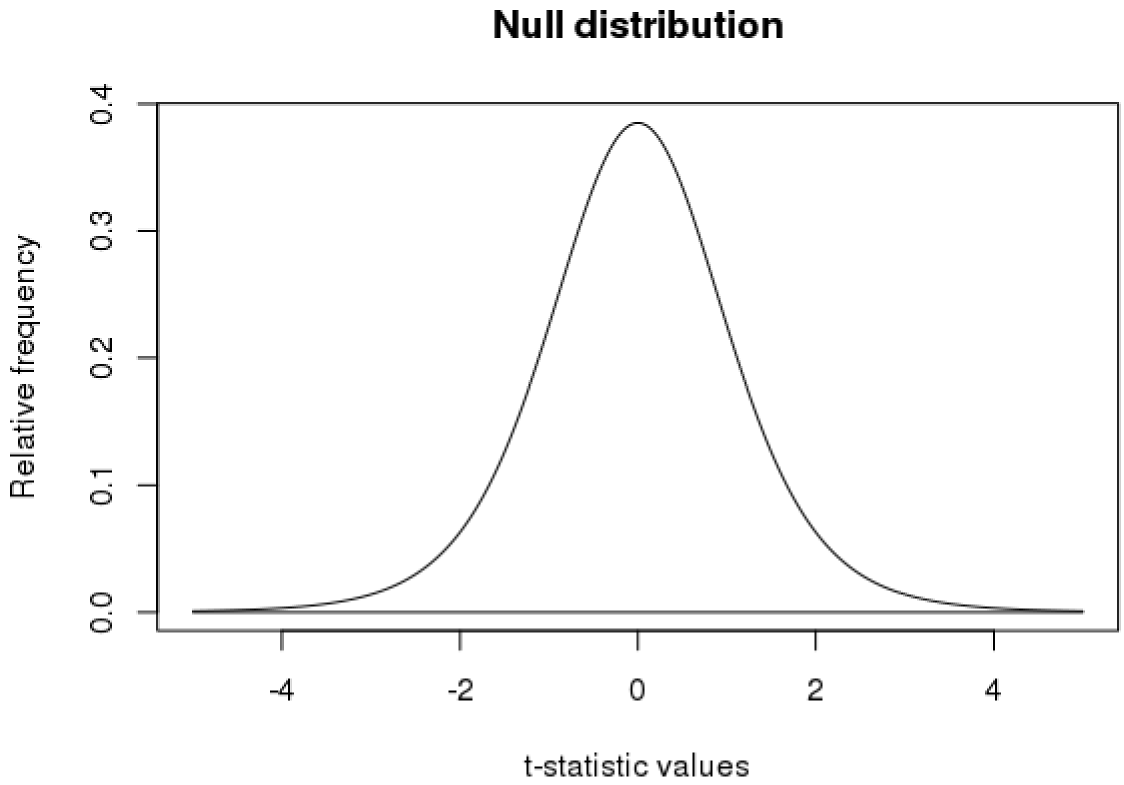
t-distributions



N(0, 1), df = 2, df = 5, df = 15

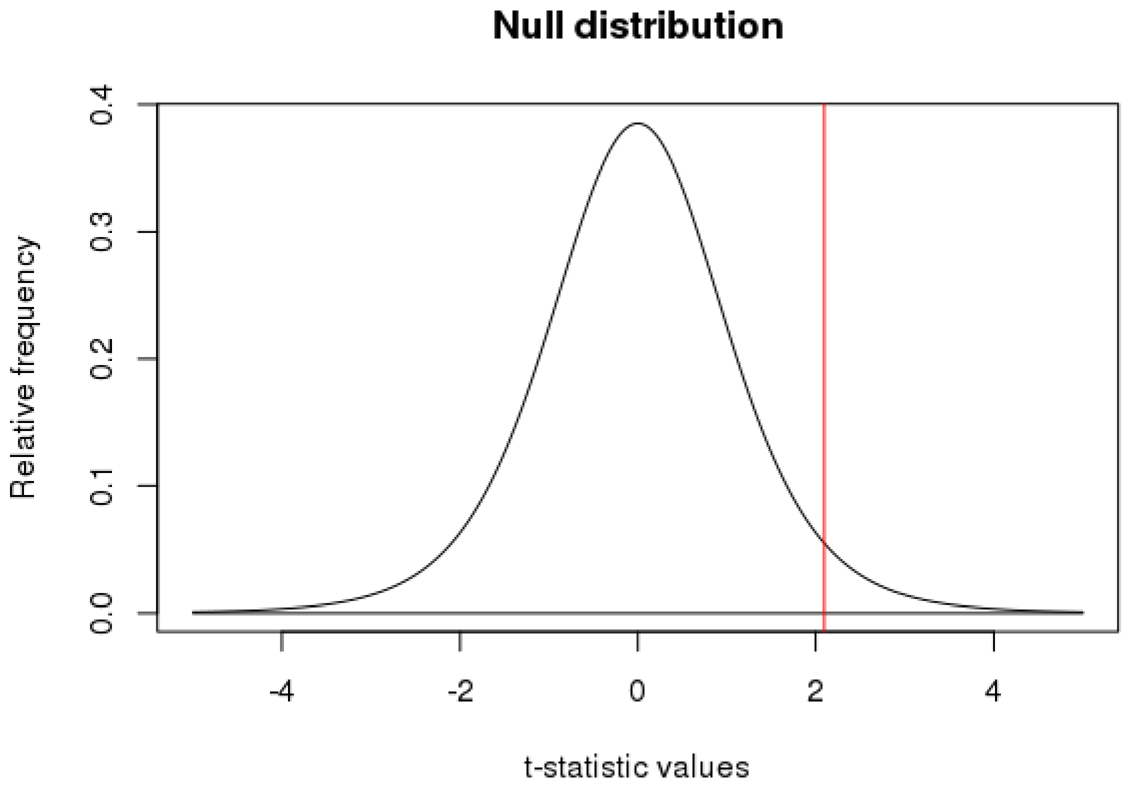
Null distribution

We can get the null density function using: dt(x_vals, df)



p-value

We can get the p-value using: pt(obs_stat, df, lower.tail = FALSE)



Do mice who eat late at night get fat?

Get the p-value

```
p_val <- pt(obs_stat, df,  lower.tail = FALSE)
```

p-value = 0.037

Two-sided tests

If we did not have a belief about whether the light or the dark group would gain more weight, then we would state our hypotheses as follows:

State the null and alternative hypothesis

H_0: $\mu_{Dark} = \mu_{Light}$ or $\mu_{Dark} - \mu_{Light} = 0$

H_A: $\mu_{Dark} \neq \mu_{Light}$ or $\mu_{Dark} - \mu_{Light} \neq 0$

Steps 1-3 of hypothesis testing would still be the same, but for step 4 we would need to calculate the p-values by examining extreme values in both tails

```
p_val <- 2 * pt(abs(obs_stat), df,  lower.tail = FALSE)
```

Let’s try it in R…