Classes of statistical designs

Dependent Variable	Independent Variable	
	Continuous	Categorical
Continuous	Regression	t-tests and **ANOVA**
Categorical	Logistic Regression	Tabular

COMPARING THE MEANS OF THREE OR MORE SAMPLES or GROUPS (often called *treatments* in experiments)

THE ANALYSIS OF VARIANCE (ANOVA):

One of the most important and used tools in statistics

THE ANALYSIS OF VARIANCE (ANOVA) for comparing multiple sample means (groups or treatments)

The problem about "The knees who say night"

By Whitlock and Schluter (2009)

OR

"Bright light behind the knees is just bright light behind the knees"

http://www.genomenewsnetwork.org/articles/08_02/bright_knees.shtml

Resetting the human Circadian rhythm

Extraocular Circadian Phototransduction in Humans

Scott S. Campbell* and Patricia J. Murphy

Physiological and behavioral rhythms are governed by an endogenous circadian clock. The response of the human circadian clock to extraocular light exposure was monitored by measurement of body temperature and melatonin concentrations throughout the circadian cycle before and after light pulses presented to the popliteal region (behind the knee). A systematic relation was found between the timing of the light pulse and the magnitude and direction of phase shifts, resulting in the generation of a phase response curve. These findings challenge the belief that mammals are incapable of extraretinal circadian phototransduction and have implications for the development of more effective treatments for sleep and circadian rhythm disorders.

SCIENCE • VOL. 279 • 16 JANUARY 1998

Data challenged as subjects were exposed to light while knees being illuminated

Our core body temperature fluctuates around 37°C, but fluctuates by about 1°C or so throughout the night.

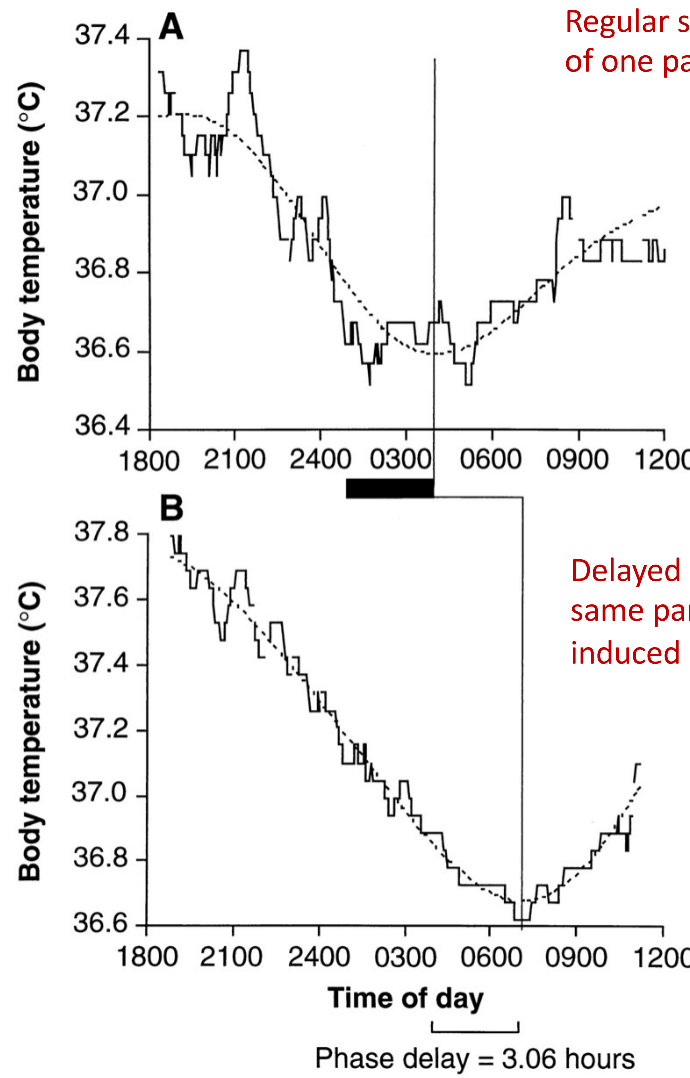
The drop in temperature starts about two hours before you go to sleep, coinciding with the release of the sleep hormone melatonin.

Extraocular Circadian Phototransduction in Humans

Scott S. Campbell* and Patricia J. Murphy

Example of a delay in circadian phase in response to a 3-hour bright light presentation to the popliteal region. Light was presented on one occasion between 0100 and 0400 on night 2 in the laboratory (black bar) while the participant (a 29-year-old male) remained awake and seated in a dimly lit room (ambient illumination <20 lux).

The circadian phase was determined by fitting a complex cosine curve (dotted line

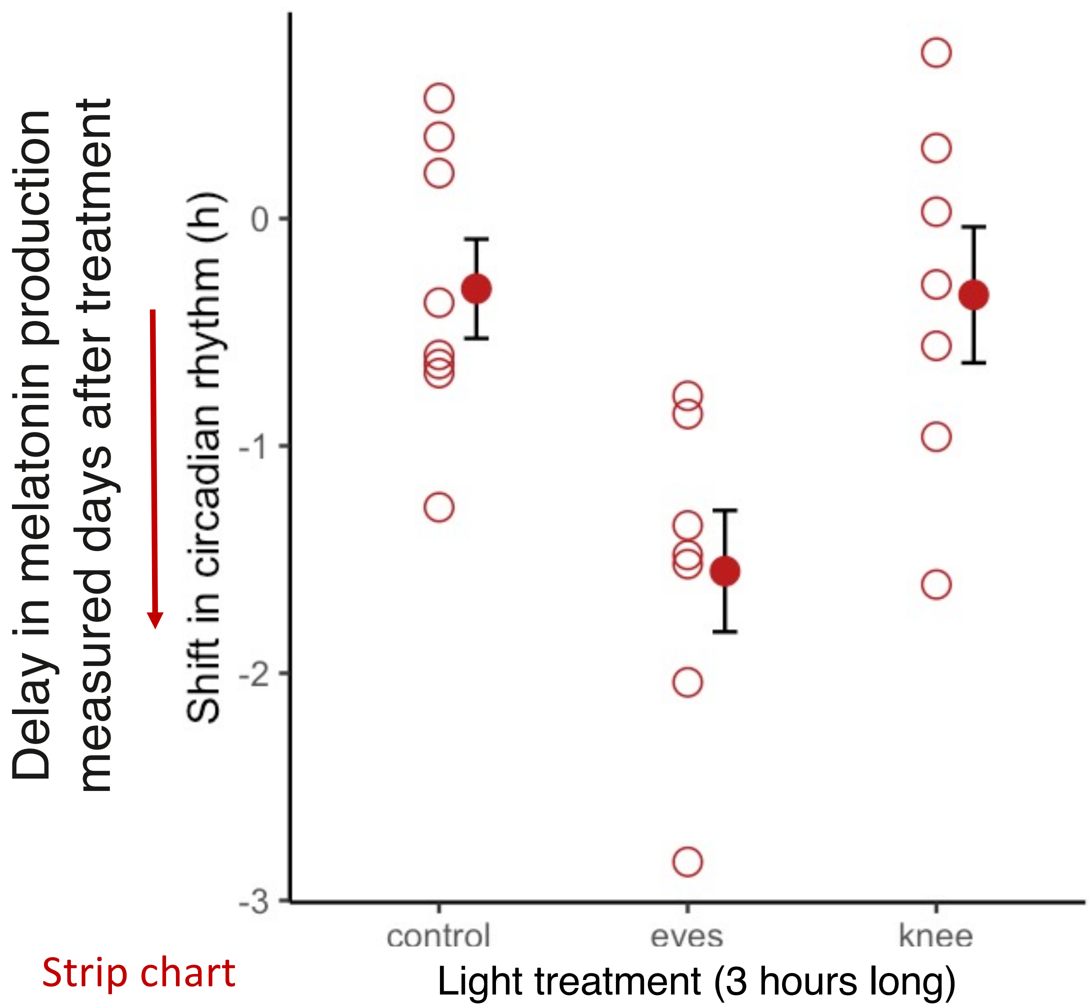


Regular sleep of one participant

Delayed sleep of the same participant induced by light.

The resulting phase delay was 3.06 hours

THE ANALYSIS OF VARIANCE (ANOVA)
for comparing multiple sample means (groups or treatments)



Strip chart

Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company

PHYSIOLOGY SCIENCE VOL 297 26 JULY 2002

Absence of Circadian Phase Resetting in Response to Bright Light Behind the Knees

Kenneth P. Wright Jr.* and Charles A. Czeisler

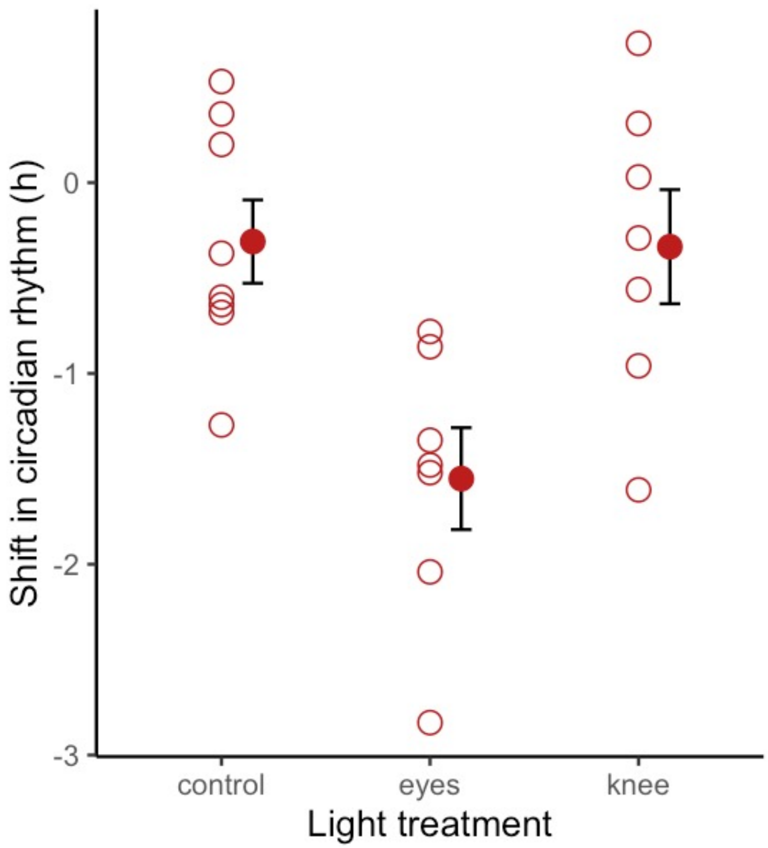
New study challenged the original study (Wright & Czeiler 2002): subjects were exposed to light to knees only.

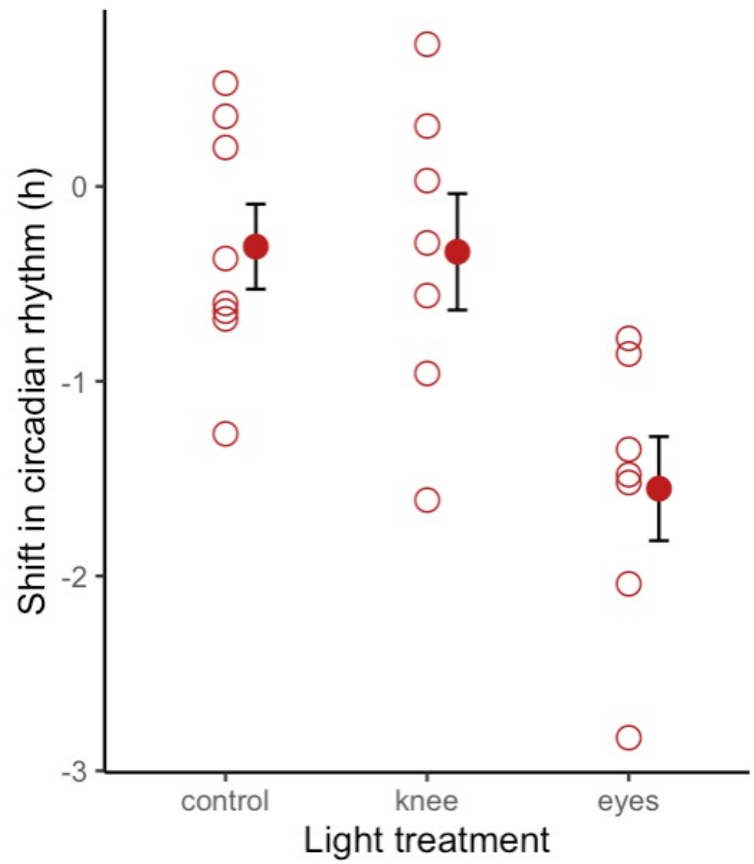
22 people randomly assigned to one of the three light treatments.

Do these means come from the same statistical population, i.e., do these samples only differ from each other due to sampling variation from the same statistical population?

THE ANALYSIS OF VARIANCE (ANOVA)
for comparing multiple sample means (groups or treatments)

Which order of treatments work best?

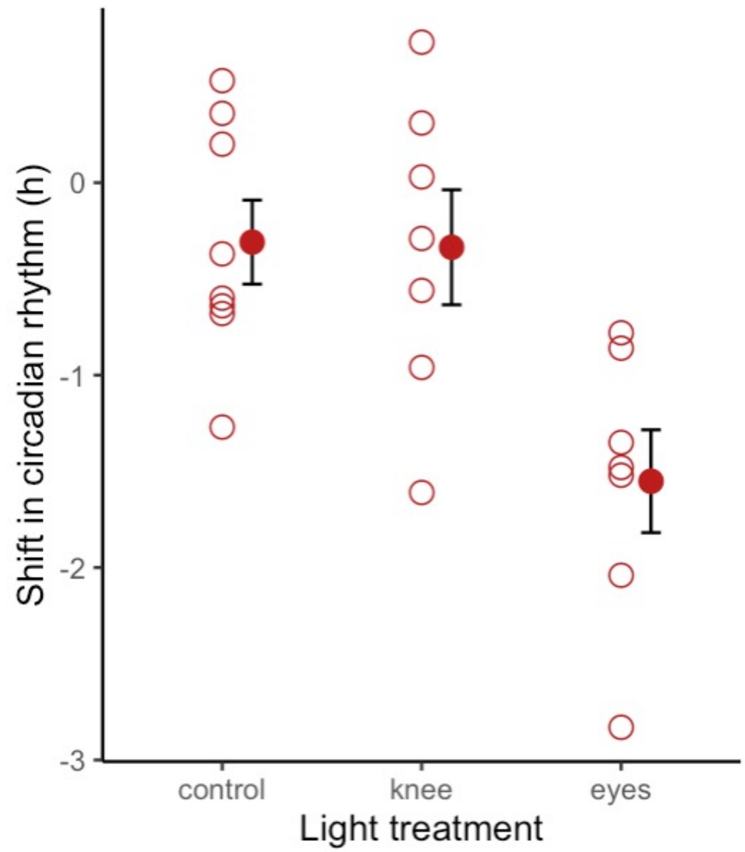




THE ANALYSIS OF VARIANCE (ANOVA)
for comparing multiple sample means (groups or treatments)

H_0: The samples come from statistical populations with the same mean, i.e., $\mu_{control} = \mu_{knee} = \mu_{eyes}$.

H_A: At least two samples come from different statistical populations with different means.



THE ANALYSIS OF VARIANCE (ANOVA)
for comparing multiple sample means (groups or treatments)

H_0: The samples come from statistical populations with the same mean, i.e., $\mu_{control} = \mu_{knee} = \mu_{eyes}$.

H_A: At least two samples come from different statistical populations with different means.

Which is to say:

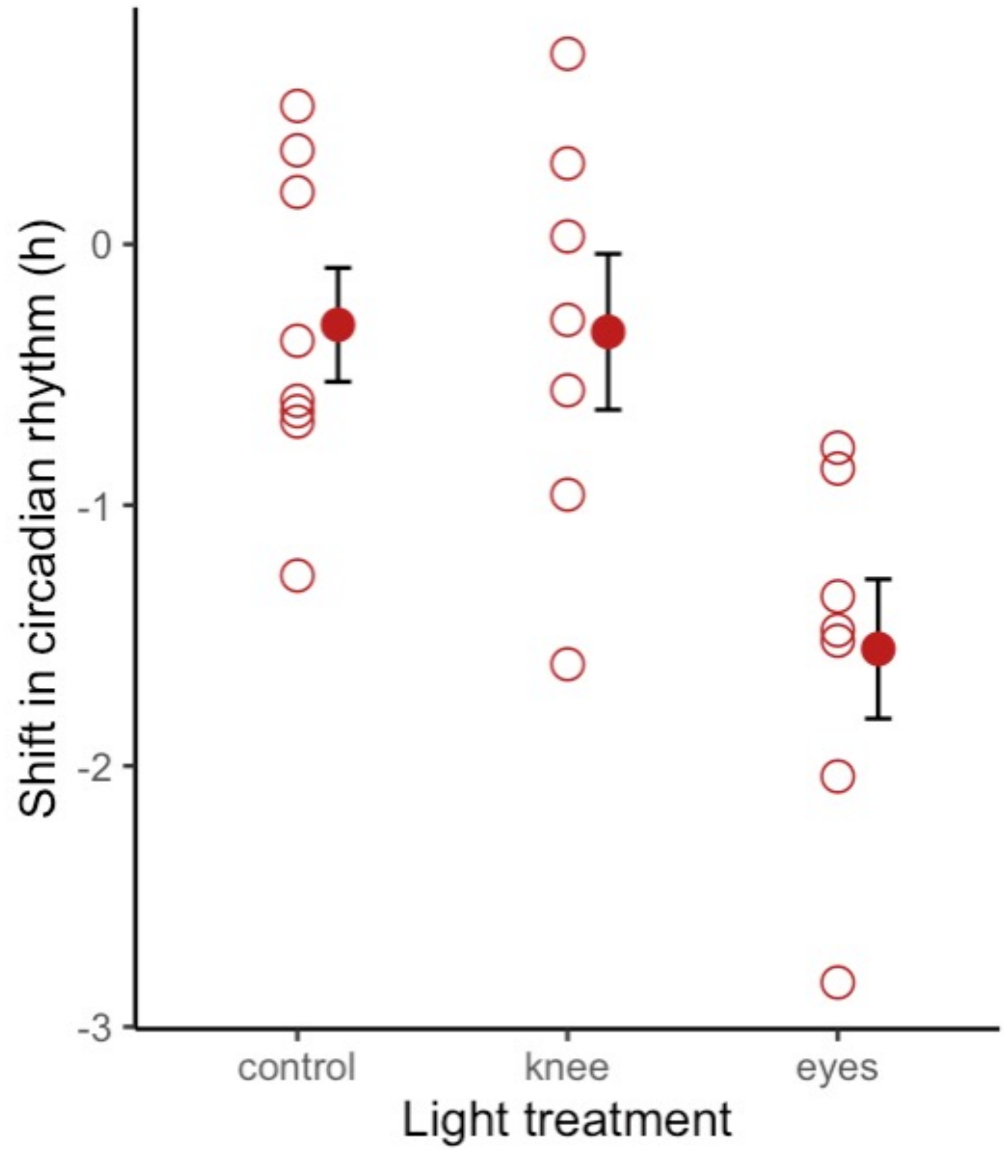
H_0: Differences in means among groups are due to **sampling error from the same population**.
H_A: Differences in means among groups are NOT due to **sampling error from the same population**.

Remember: ***Sampling error*** is due to sampling variation, i.e., samples that come from the same statistical population may differ in their means just due to chance alone.

THE ANALYSIS OF VARIANCE (ANOVA)
for comparing multiple sample means (groups or treatments)

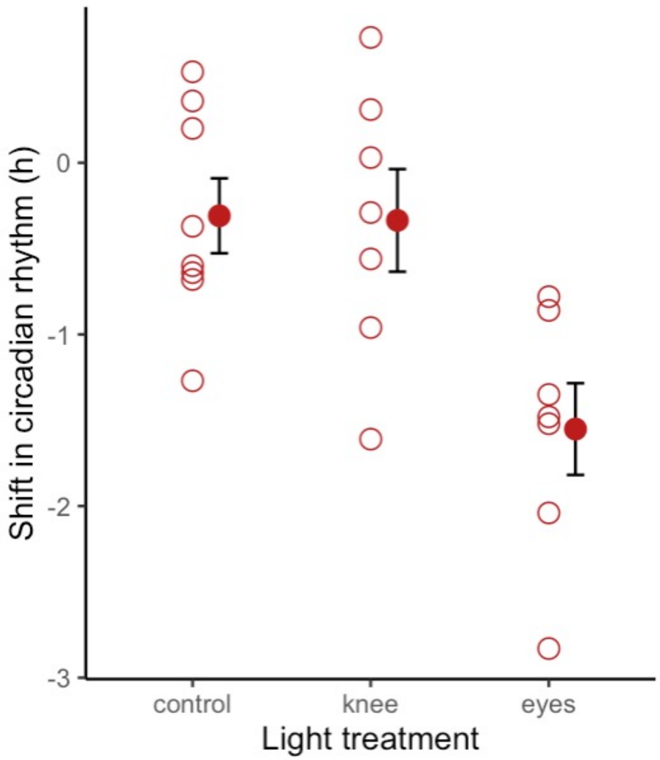
An ANOVA always involves one ***continuous*** variable (e.g., shift in circadian rhythm) and **one** ***categorical*** (e.g., treatment or factor) variable.

The categorical variable is divided into groups (also called treatments or levels).



THE ANALYSIS OF VARIANCE (ANOVA)
for comparing multiple sample means (groups or treatments)

H_0: The samples come from statistical populations with the same mean, i.e., $\mu_{control} = \mu_{knee} = \mu_{eyes}$.
H_A: At least two samples come from different statistical populations with different means.



We are studying one single factor (light), we use a ***one-way ANOVA***.

If two factors were involved (say light and time of experimentations: two-way ANOVA

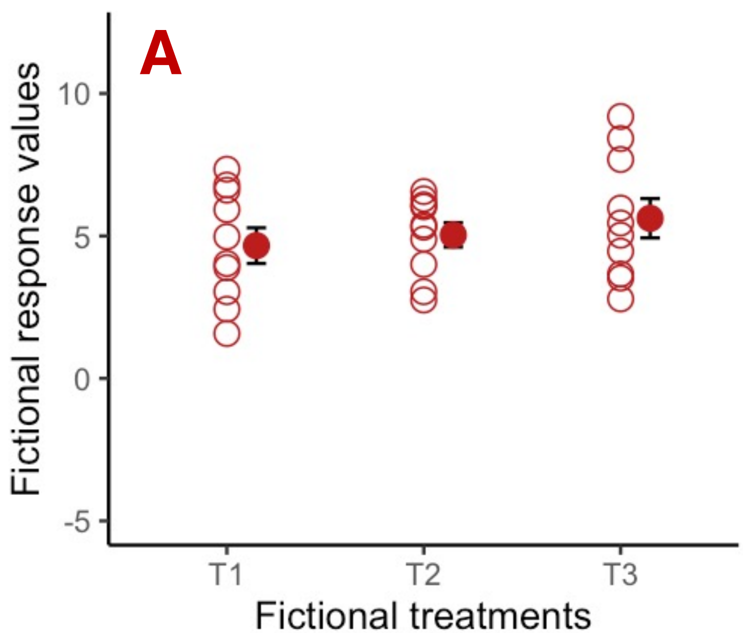
We need a test statistic that is sensitive to mean variation across multiple groups (or treatments):
The F statistic does that by considering the ratio of two variances (variance components):

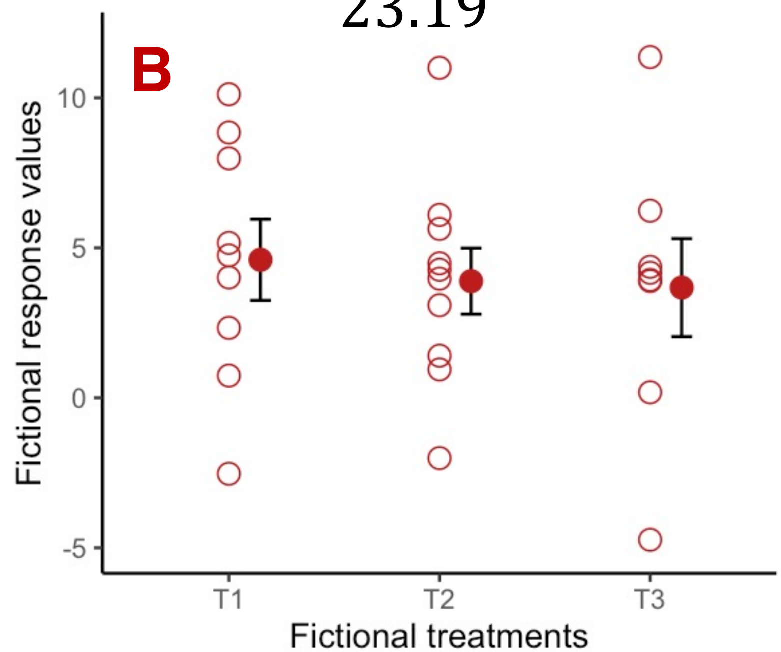
$$F = \frac{\text{variance among group means (due to “treatment”)}}{\text{variance within groups (caller error or residual variation not accounted by the differences in mean among groups)}}$$

Means among groups don't vary much in both data **A** and **B**,
but residual variation (within groups) is smaller in **A** than **B**.

$$F_A = \frac{2.34}{3.51} = 0.67$$

$$F_B = \frac{11.63}{23.19} = 0.50$$



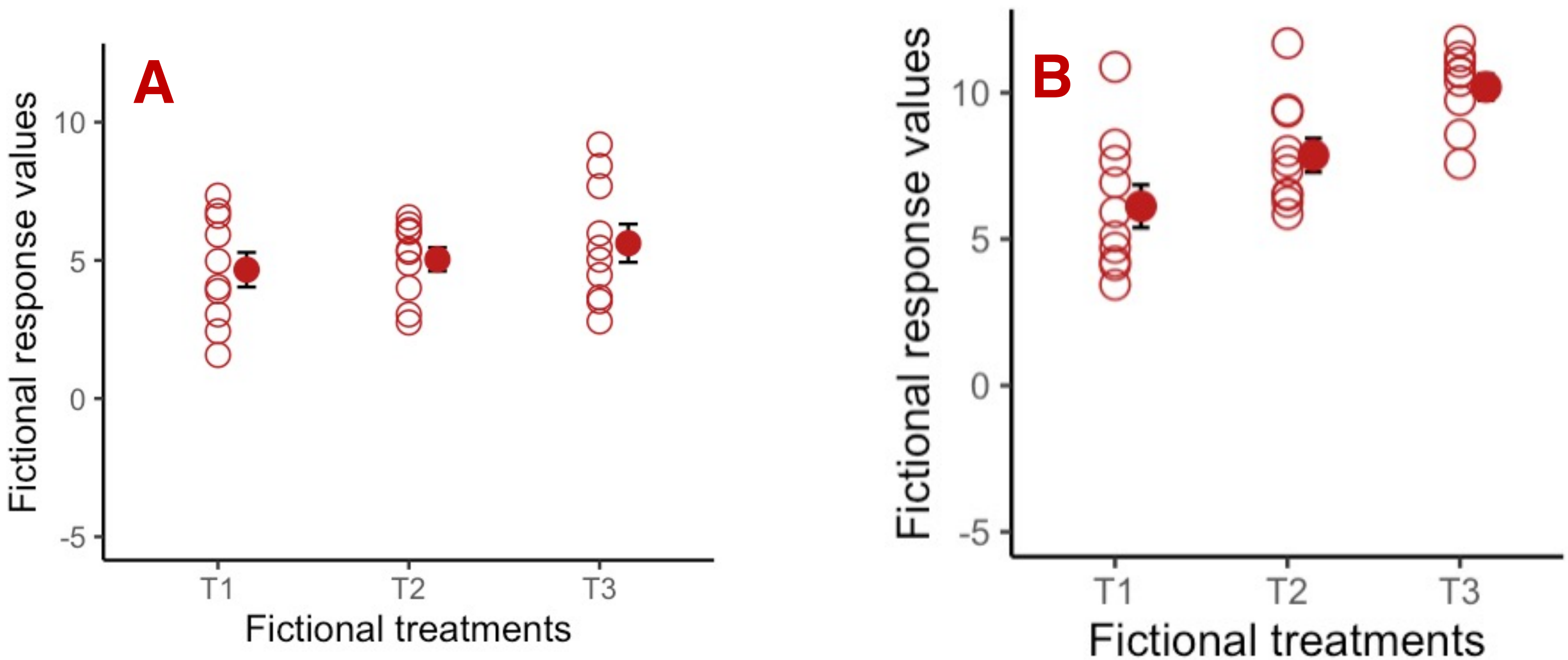


We need a test statistic that is sensitive to mean variation across multiple groups (or treatments):
The F statistic does that by considering the ratio of two variances (variance components):

Means among groups don't vary in **A** but vary in **B**;
residuals variation is similar in **A** than **B**.

$$F_A = \frac{2.34}{3.51} = 0.67$$

$$F_B = \frac{47.41}{3.64} = 13.03$$

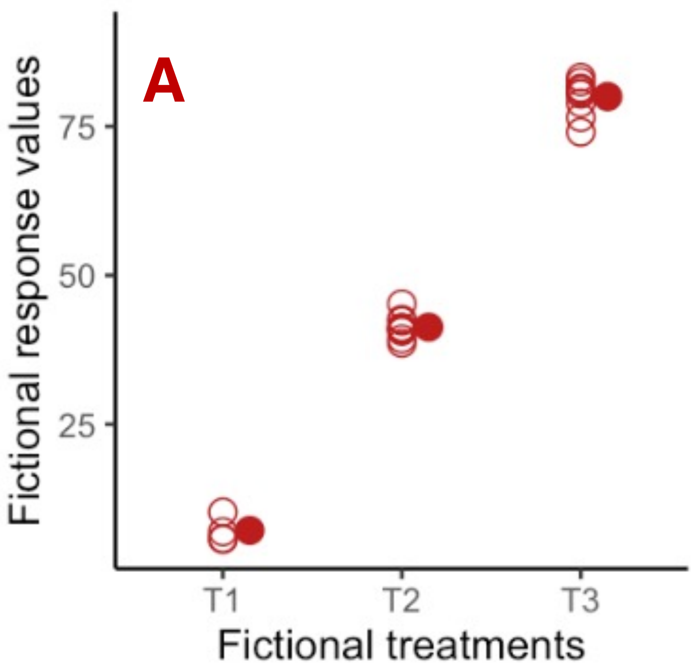


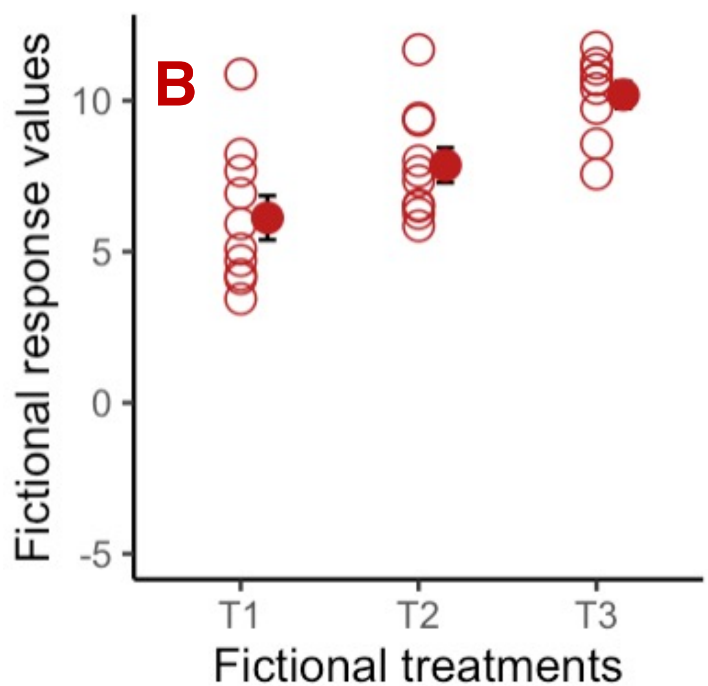
We need a test statistic that is sensitive to mean variation across multiple groups (or treatments):
The F statistic does that by considering the ratio of two variances (variance components):

Means among groups are much bigger in **A** than **B**;
residuals variation is similar in **A** than **B**. Notice the differences in their Y-scales (the mean differences among groups is huge in **A**).

$$F_A = \frac{14078.0}{5.71} = 2456.90$$

$$F_B = \frac{47.41}{3.64} = 13.03$$



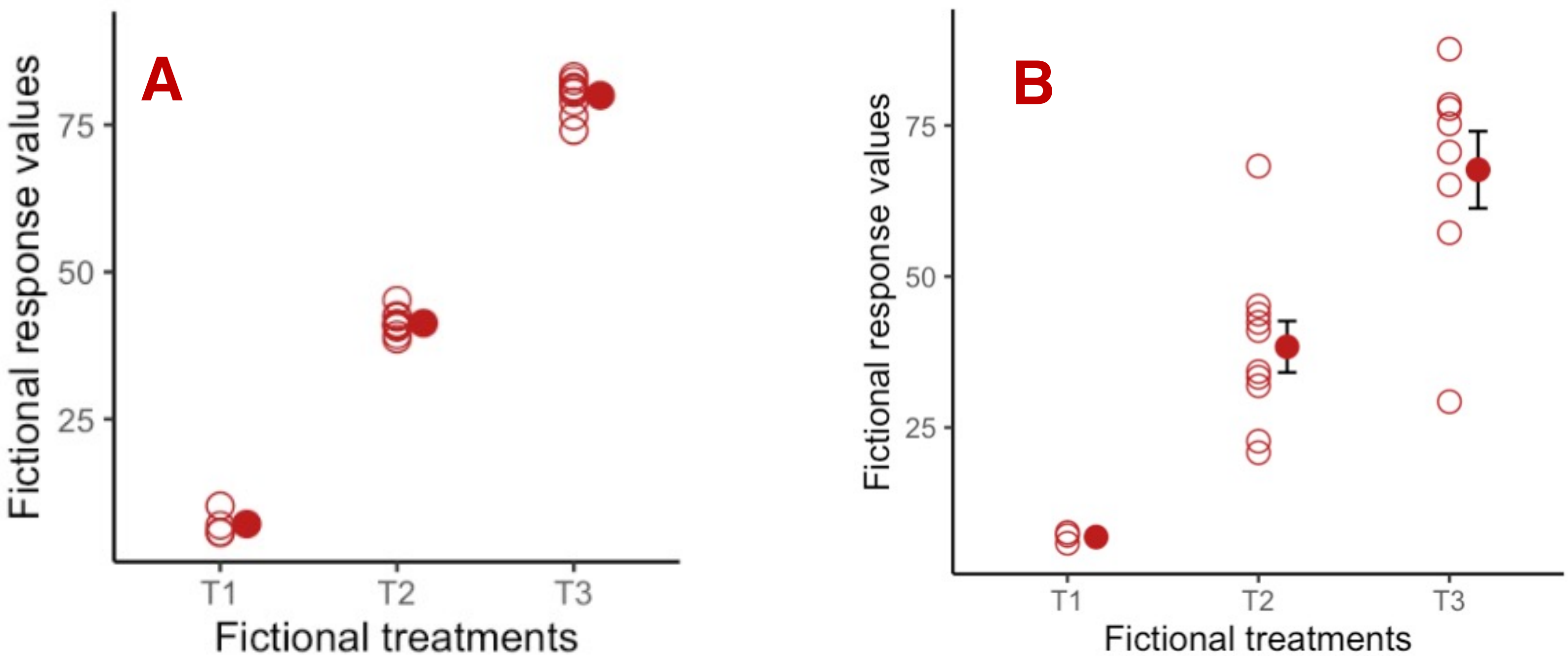


HETEROSCEDASTICITY reduces the F-ratio ability to differentiate among differences in means among groups

Means among groups are somewhat similar in **A** than **B**;
A is homoscedastic **B** heteroscedastic

$$F_A = \frac{14078.0}{5.71} = 2456.90$$

$$F_B = \frac{12275.0}{217.9} = 56.34$$



We need a test statistic that is sensitive to mean variation across multiple groups (or treatments): The F statistic does that by considering the ratio of two variances (variance components):

Let's talk ANOVA "jargon"

$$F = \frac{\text{variance among group means (due to "treatment")}}{\text{variance within groups (caller error or residual variation not explained by the mean within groups)}}$$

You can interpret ANOVA without knowing how it works, but you are less likely to use ANOVA inappropriately if you have some idea of how it works (*Motulsky*)

We need a test statistic that is sensitive to mean variation across multiple groups (or treatments): The F statistic does that by considering the ratio of two variances (variance components):

Let’s talk ANOVA “jargon”

$$F = \frac{\text{variance among group means (due to “treatment”)}}{\text{variance within groups (caller error or residual variation not explained by the mean within groups)}}$$

$$F = \frac{\text{Group Mean Square}}{\text{Error Mean Square}} = \frac{MS_{groups}}{MS_{error}}$$

The F statistic measures the variance among groups but accounting for the variance within groups

The F statistic in the ANOVA context is so important that is more than worth knowing how it works!

Group Mean Square

MS_{groups}

(b=between or among)

Mean of each group

Total mean!

$$F = \frac{s_b^2}{s_w^2} = \frac{\dfrac{\sum_{i=1}^{g} n_i(\bar{X}_i - \bar{\bar{X}})^2}{g-1}}{}$$

Degrees of freedom of MS_{groups}

MS_{errors}

(w=within groups)

Error Mean Square

The F statistic measures the variance among groups but accounting for the variance within groups

The F statistic in the ANOVA context is so important that is more than worth knowing how it works!

$$F = \frac{s_b^2}{s_w^2} = \frac{\dfrac{\sum_{i=1}^{g} n_i(\bar{X}_i - \bar{\bar{X}})^2}{g-1}}{\dfrac{\sum_{i=1}^{g}(n_i - 1)s_i^2}{\sum_{i=1}^{g}(n_i - 1)}}$$

$$\sum_{i=1}^{g}(n_i - 1) = (N-g)$$

Group Mean Square

MS_{groups}

(b=between or among)

Mean of each group

Total mean!

Degrees of freedom of MS_{groups}

Variance of each group

Big "N"; sum of all sample sizes across groups

Number of groups

Degrees of freedom of MS_{groups}

Sample size of each group

MS_{errors}

(w=within groups)

Error Mean Square

A small example: worth doing it "by hand"!

Let's suppose two groups for simplicity!

group 1

1	2	3	4	5

$\overline{X}_1 = 3.0$

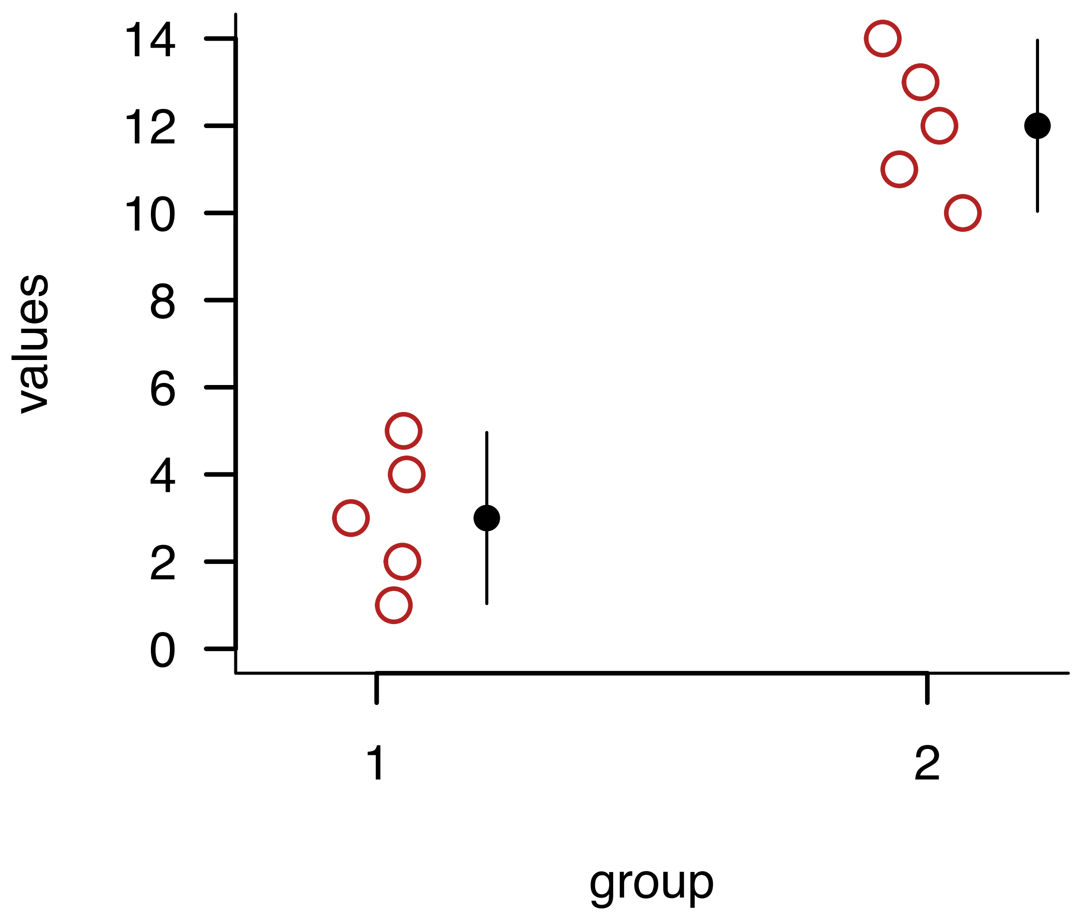
$s_1^2 = 2.5$

group 2

10	11	12	13	14

$\overline{X}_2 = 12.0$

$s_2^2 = 2.5$



g_1: 1 2 3 4 5

g_2: 10 11 12 13 14

$\bar{X}_1 = 3.0 \quad \bar{X}_2 = 12.0$

$s_1^2 = 2.5 \quad s_2^2 = 2.5$

$\bar{\bar{X}}$ = (1+2+3+4+5+10+11+12+13+14)/10 = **7.5**

MS_{groups} = variance among group means (due to "treatment")

= (5x(**3.0** - 7.5)^2 + 5x(**12.0** - 7.5)^2)/(2-1) =

202.5/(2-1) = **202.5**

df(MS_{groups}) = g - 1

$$F = \frac{202.5}{???} = ???$$

Mean of each group

Total mean!

$$F = \frac{s_b^2}{s_w^2} = \frac{\dfrac{\sum_{i=1}^{g} n_i(\bar{X}_i - \bar{\bar{X}})^2}{g-1}}{\dfrac{\sum_{i=1}^{g}(n_i - 1)s_i^2}{\sum_{i=1}^{g}(n_i - 1)}}$$

MS_{groups}

Variance of each group

Big "N"; sum of all sample sizes across groups

$\sum_{i=1}^{g}(n_i - 1)$ → = (N-g)

g_1: 1 2 3 4 5

g_2: 10 11 12 13 14

Mean of each group — Total mean!

$$F = \frac{s_b^2}{s_w^2} = \frac{\dfrac{\sum_{i=1}^{g} n_i(\bar{X}_i - \bar{\bar{X}})^2}{g-1}}{\dfrac{\sum_{i=1}^{g}(n_i - 1)s_i^2}{\sum_{i=1}^{g}(n_i - 1)}}$$

Variance of each group

Big "N"; sum of all sample sizes across groups

$\sum_{i=1}^{g}(n_i - 1)$ → = (N-g)

MS_{error}

$\bar{X}_1 = 3.0$ $\quad \bar{X}_2 = 12.0$

$s_1^2 = 2.5$ $\quad s_2^2 = 2.5$

MS_{error} = variance within groups (residuals)

$MSE_1 = (1-3.0)^2+(2-3.0)^2+(3-3.0)^2+(4-3.0)^2+(5-3.0)^2 =$ **10**

$MSE_2 = (10-12.0)^2+(11-12.0)^2+(12-12.0)^2+(13-12.0)^2+(14-12.0)^2 =$ **10**

$MS_{error} = (MSE_1 + MSE_2)/(N-g) = (10+10) / (10-2) = 20/8 =$ **2.5**

$df(MS_{error}) = N-g = 10 - 2 = 8$

$\bar{\bar{X}}$ = (1+2+3+4+5+10+11+12+13+14)/10 = **7.5**

MS_{groups}=

= $(5x(\mathbf{3.0} - 7.5)^2 + 5x(\mathbf{12.0} - 7.5)^2)/(2-1)$ =

202.5/(2-1) = **202.5**

$df(MS_{groups})$ = g - 1 = 2-1

$$F = \frac{\mathbf{202.5}}{\mathbf{2.5}} = 81$$

MS_{error} = variance within groups (residuals)

$MSE_1 = (1-3.0)^2+(2-3.0)^2 +(3-3.0)^2 +(4-3.0)^2 +(5-3.0)^2 = \mathbf{10}$

$MSE_2 = (10-12.0)^2 +(11-12.0)^2 +(12-12.0)^2 +(13-12.0)^2 +(14-12.0)^2 = \mathbf{10}$

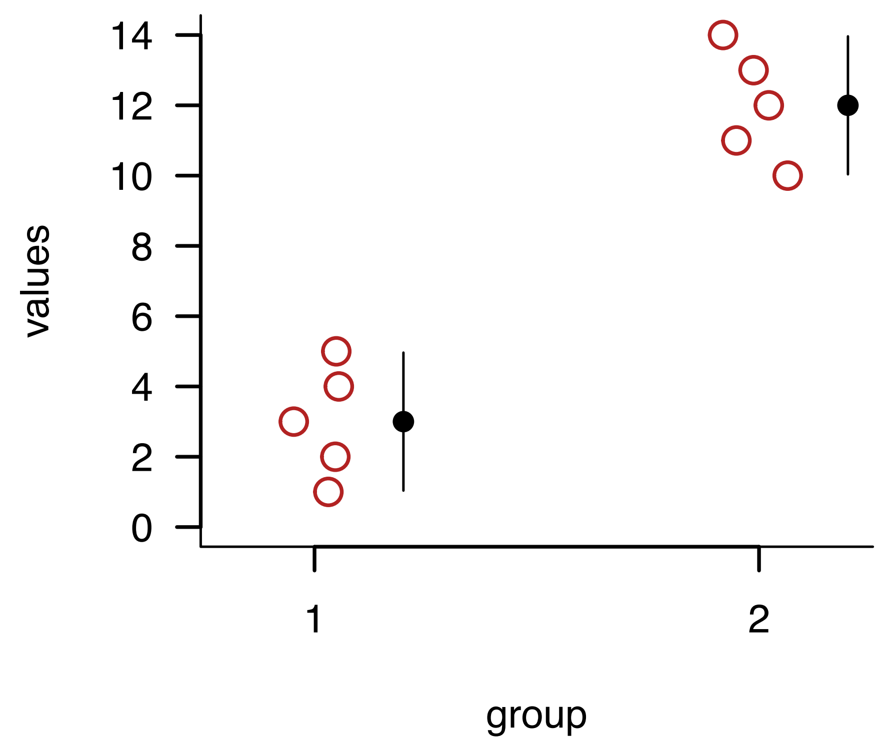
$MS_{error} = (MSE_1 + MSE_2)/(N-g)=(10+10) / (10-2) = 20/8 = \mathbf{2.5}$

$df(MS_{error})$ = N-g = 10 – 2 = 8

Let’s take a power break – 1 minute

ANOVA in R - step 1: organizing data in a csv file

E	F
values	group
1	1
2	1
3	1
4	1
5	1
10	2
11	2
12	2
13	2
14	2



ANOVA in R

Function to
run **A**n **A**nalysis of **V**ariance (aov)

Vector of observations
(1,2,3,4,5,10,11,12,13,14)

Factor identifying group of
the observation
(1,1,1,1,1,2,2,2,2,2)

```
> aov(data.points~groups)
Call:
   aov(formula = data.points ~ groups)

Terms:
                groups Residuals
Sum of Squares   202.5      20.0
Deg. of Freedom      1         8

Residual standard error: 1.581139
Estimated effects may be unbalanced
```

$$F = \frac{202.5}{2.5} = 81$$

ANOVA for two groups is equivalent to the two-sample t-test

```
> aov(data.points~groups)
Call:
   aov(formula = data.points ~ groups)

Terms:
                groups Residuals
Sum of Squares   202.5      20.0
Deg. of Freedom      1         8

Residual standard error: 1.581139
Estimated effects may be unbalanced
```

$$F = \frac{\mathbf{202.5}}{\mathbf{2.5}} = 81$$

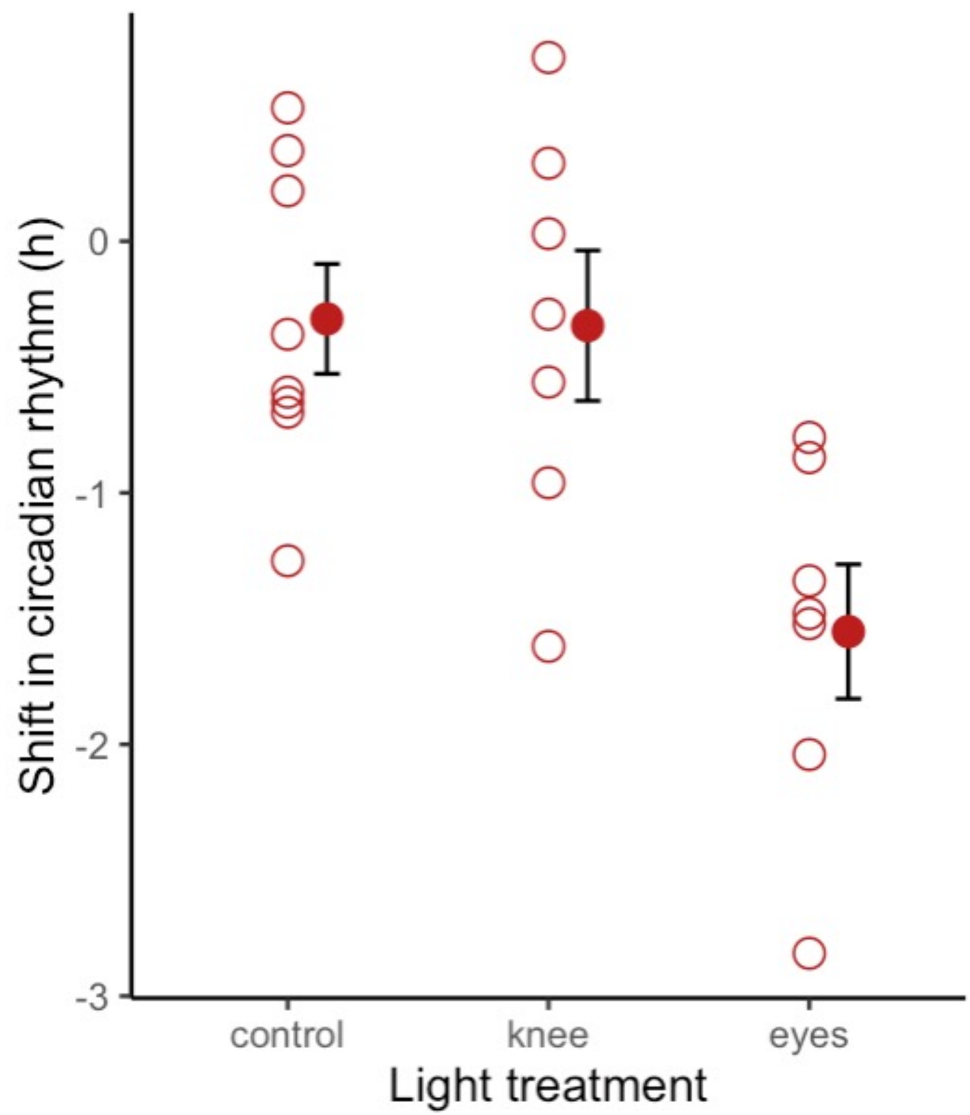
$$t = \sqrt{F} = \sqrt{\frac{\mathbf{202.5}}{\mathbf{2.5}}} = 9$$

```
t.test(data.points~groups,var.equal = TRUE)

     Two Sample t-test

data:  data.points by groups
t = -9, df = 8, p-value = 1.853e-05
```

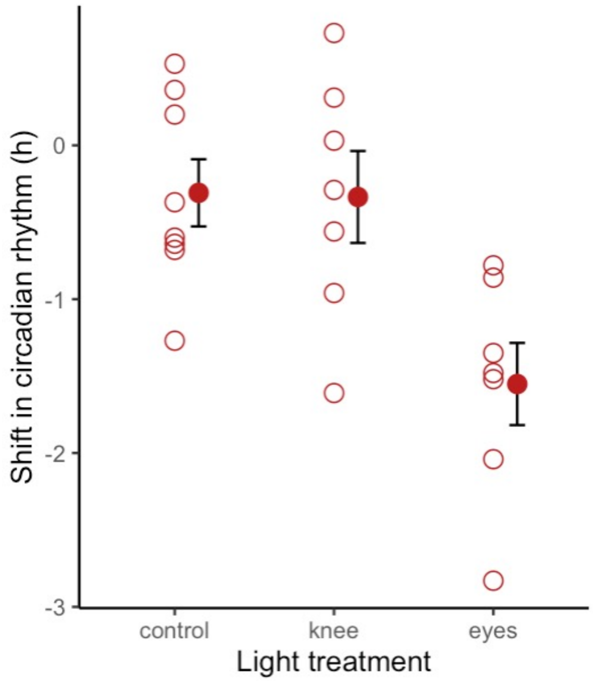
LET's go back to the "The knees who say night"

data in a csv file

A	B
treatment	shift
control	0.53
control	0.36
control	0.2
control	-0.37
control	-0.6
control	-0.64
control	-0.68
control	-1.27
knee	0.73
knee	0.31
knee	0.03
knee	-0.29
knee	-0.56
knee	-0.96
knee	-1.61
eyes	-0.78
eyes	-0.86
eyes	-1.35
eyes	-1.48
eyes	-1.52
eyes	-2.04
eyes	-2.83



“The knees who say night”



Statistical Conclusion?

H_0: The samples come from the same population.

H_A: At least two samples come from different populations.

```
summary(aov(shift ~ treatment, data=circadian))
            Df Sum Sq Mean Sq F value  Pr(>F)
treatment    2  7.224   3.612   7.289 0.00447 **
Residuals   19  9.415   0.496
---
```

“The knees who say night”

```
summary(aov(shift ~ treatment, data=circadian))
            Df Sum Sq Mean Sq F value  Pr(>F)
treatment    2  7.224   3.612   7.289 0.00447 **
Residuals   19  9.415   0.496
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

ANOVA Table – reporting quality

Source of variation	Sum of squares	df	Mean square	F	P
Between	7.224	2	3.612	7.289	0.00447
Within	9.415	19	0.496		

Remembering the role of degrees of freedom

Source of variation	Sum of squares	df	Mean square	F	P
Between	7.224	2	3.612	7.289	0.00447
Within	9.415	19	0.496		

Remember that the calculations of sum of squares involve subtractions from means so that they would be biased if not divided by adjustments (degrees of freedom) to produce mean square deviations.

"The knees who say night"

ANOVA Table

Source of variation	Sum of squares	df	Mean square	F	P
Between	7.224	2	3.612	7.289	0.00447
Within	9.415	19	0.496		

H_0: The samples come from the same population.

H_A: At least two samples come from different populations.

Reject H_0

How does the ANOVA significance test work?

How was the F distribution built?

The statistical “machinery”:

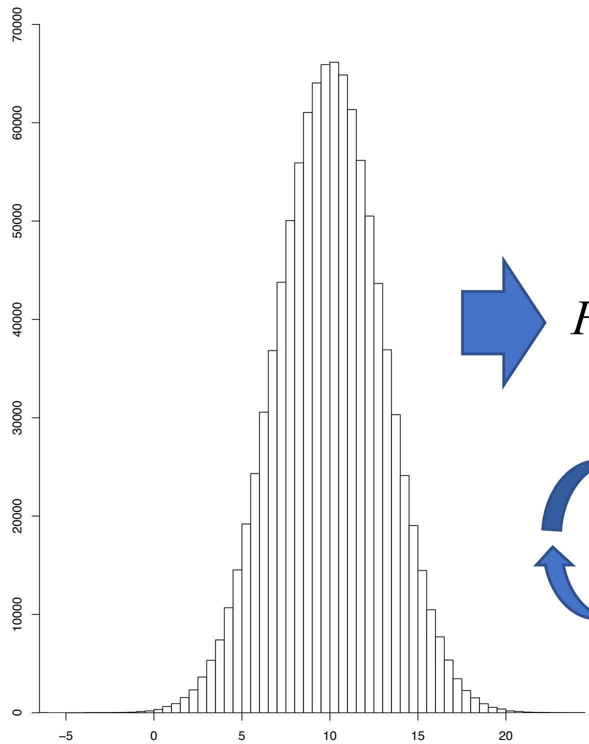
1) Assume that **H_0** is true (i.e., samples come from the same population; i.e., population having the same mean and same variance).

2) Sample from the population the appropriate number of groups (samples) respecting the sample size of each group.

3) Repeat step 2 a large (or infinite) number of times and each time calculate the F statistic.

The F (sampling) distribution assuming that H_0 is true

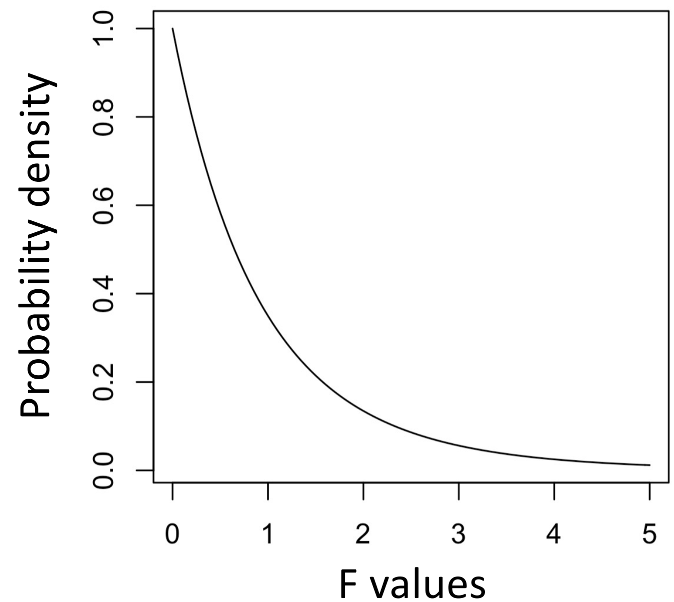
H_0: Differences in means among groups are due to **sampling error from the same population**.

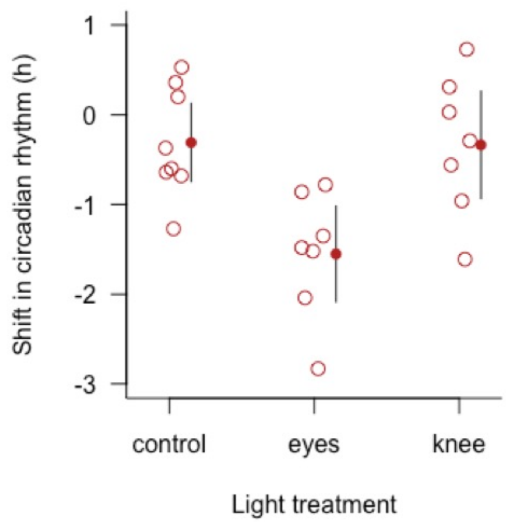


Sample from the same (normally distributed) population (i.e., ***H_0 is true***), respecting the original number of groups and their sample sizes.

$$F = \frac{s_b^2}{s_w^2} = \frac{\dfrac{\sum_{i=1}^{g} n_i(\overline{X}_i - \overline{\overline{X}})^2}{g-1}}{\dfrac{\sum_{i=1}^{g}(n_i - 1)s_i^2}{\sum_{i=1}^{g}(n_i - 1)}}$$

(8,7,7) observations

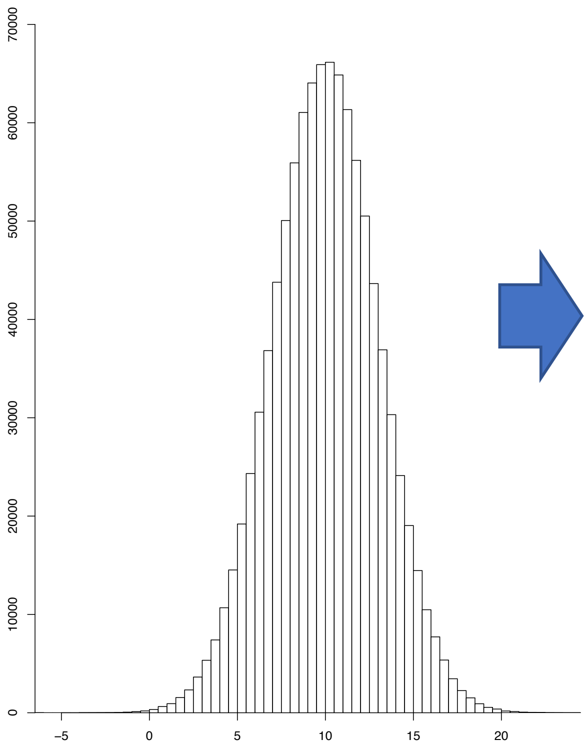




Control: 8 observations
Eyes: 7 observations
Knee: 7 observations

The F (sampling) distribution assuming that H_0 is true

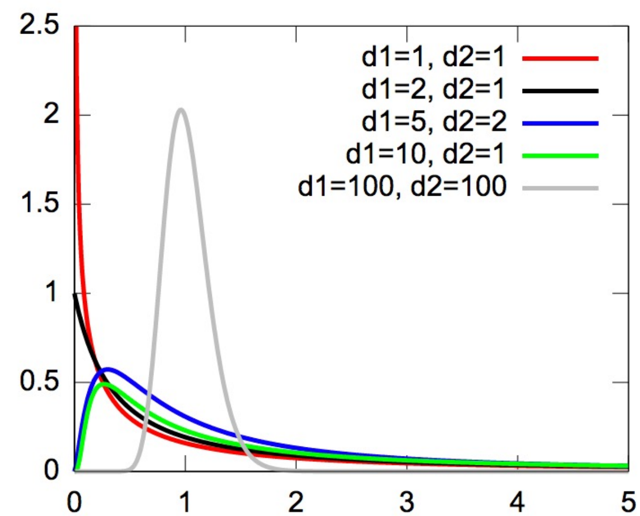
H_0: Differences in means among groups are due to **sampling error from the same population**.



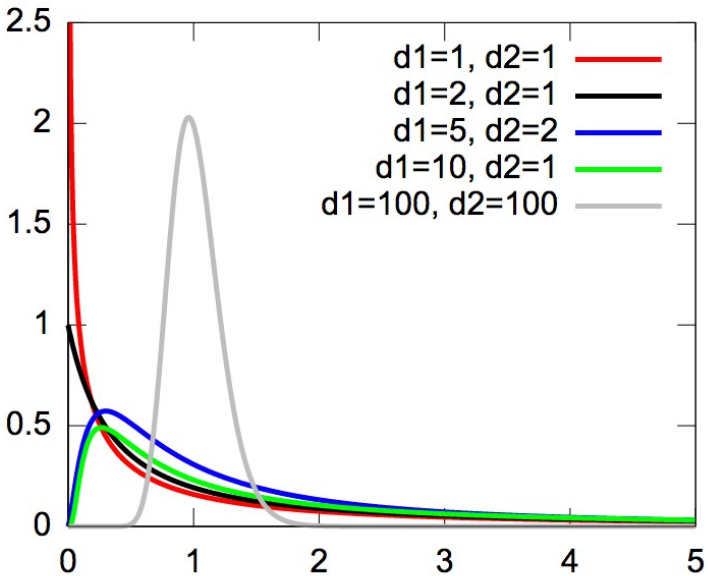
Sample from the same (normally distributed) population (i.e., ***H_0 is true***), respecting the original number of groups and their sample sizes.

$$F = \frac{s_b^2}{s_w^2} = \frac{\dfrac{\sum_{i=1}^{g} n_i(\overline{X}_i - \overline{\overline{X}})^2}{g-1}}{\dfrac{\sum_{i=1}^{g}(n_i - 1)s_i^2}{\sum_{i=1}^{g}(n_i - 1)}}$$

Different number of groups and different number of observations per group generate different shapes for the F distribution.



The F distribution assuming that H_0 is true (i.e., the sampling distribution of the test statistic F when H_0 is true).



Mean of each group

Total mean!

$$F = \frac{s_b^2}{s_w^2} = \frac{\dfrac{\sum_{i=1}^{g} n_i(\overline{X}_i - \overline{\overline{X}})^2}{g-1}}{\dfrac{\sum_{i=1}^{g}(n_i - 1)s_i^2}{\sum_{i=1}^{g}(n_i - 1)}}$$

$\sum_{i=1}^{g}(n_i - 1)$ → = (N-g)

Variance of each group

Big "N"; sum of all sample sizes across groups

df_1

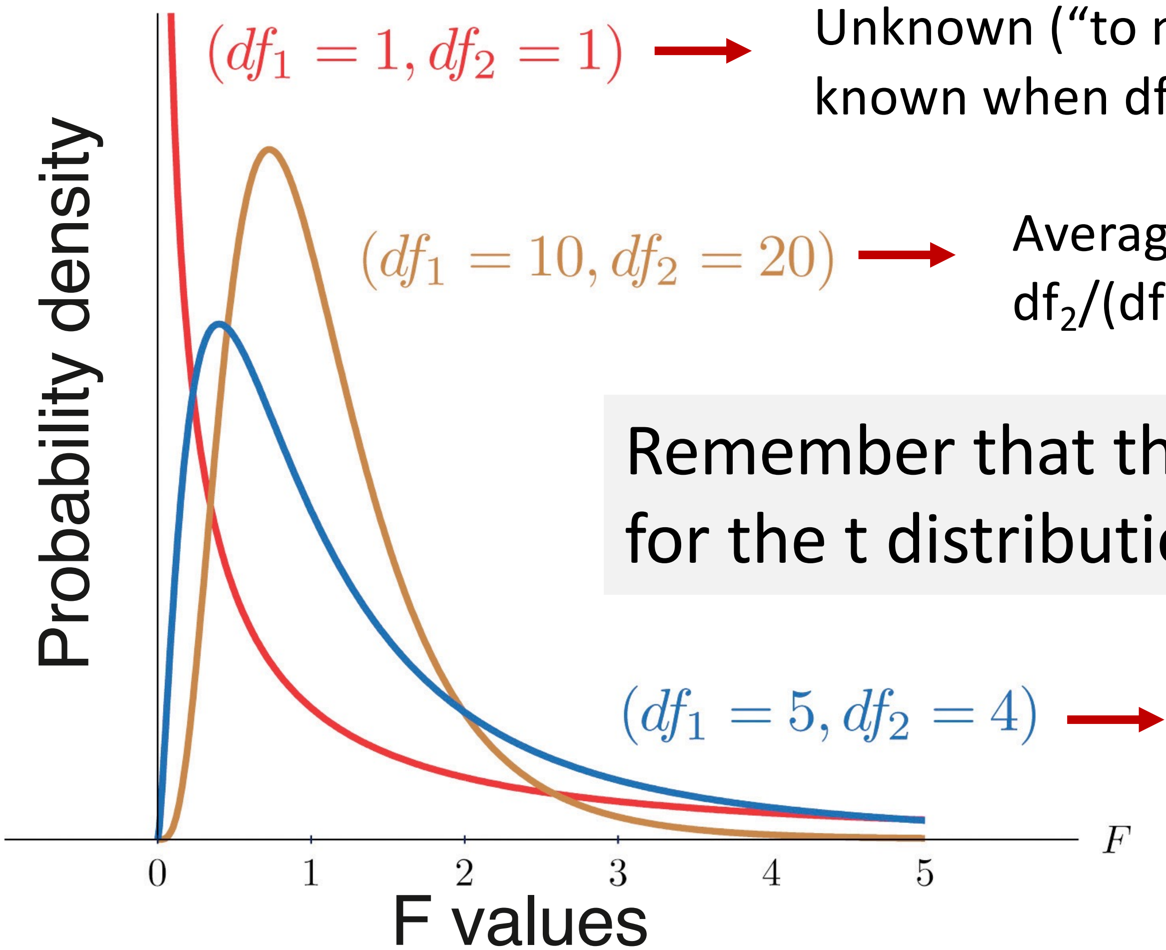
df_2

The numerator degrees of freedom is based on the number of groups (g-1) and the denominator degrees of freedom depends on the total number of observations (N-g)

The F (sampling) distribution assuming that H_0 is true

Remember: The average of the t-distribution under $\mathbf{H_0}$ is always zero;

However, the average of the F-distribution expected under $\mathbf{H_0}$ depends on its degrees of freedom.



Unknown ("to me"); known when $df_2 > 2$

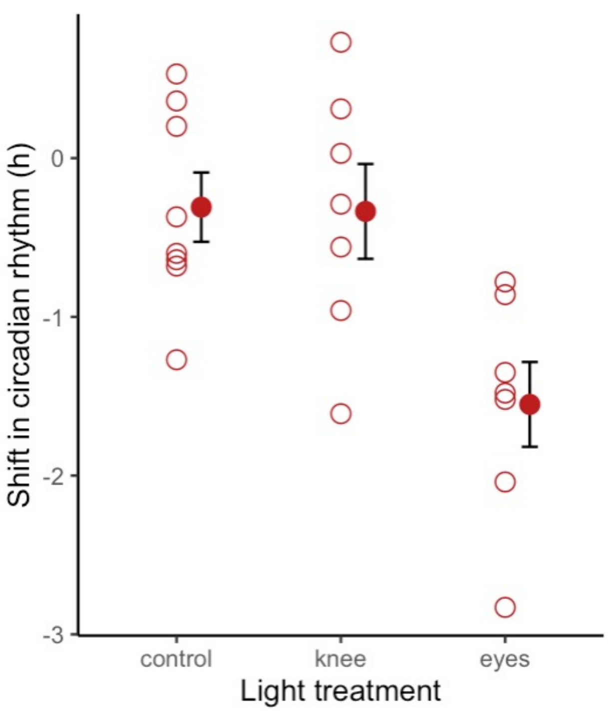
Average F value is $df_2/(df_2-2) = 20/18 = 1.11$

Remember that the average mean value for the t distribution is zero

Average F value is $df_2/(df_2-2) = 4/2 = 2.00$

The expected mean value for the F distribution under the null hypothesis as a way to also express the null and alternative hypothesis

H_0: The samples come from statistical populations with the same mean, i.e., $\mu_{control} = \mu_{knee} = \mu_{eyes}$.
H_A: At least two samples come from different statistical populations with different means.



Which is to say:

H_0: $F = df_2/(df_2 - 2)$
H_A: $F \neq df_2/(df_2 - 2)$

```
summary(aov(shift ~ treatment, data=circadian))
            Df Sum Sq Mean Sq F value  Pr(>F)
treatment    2  7.224   3.612   7.289 0.00447 **
Residuals   19  9.415   0.496
```

Degrees of freedom

Observed F-value (observed test statistic)

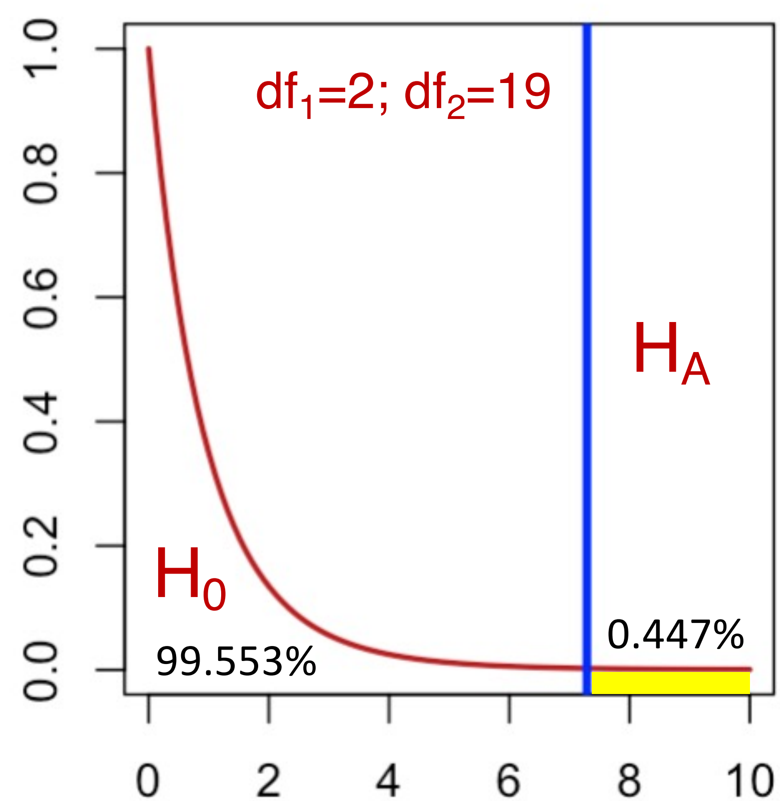
P-value

ANOVA is a one-sided (one-tail) statistical test by design; even though the null hypothesis was set as a two-tailed test.

H_0: The samples come from statistical populations with the same mean, i.e., $\mu_{control} = \mu_{knee} = \mu_{eyes}$.

H_A: At least two samples come from different statistical populations with different means.

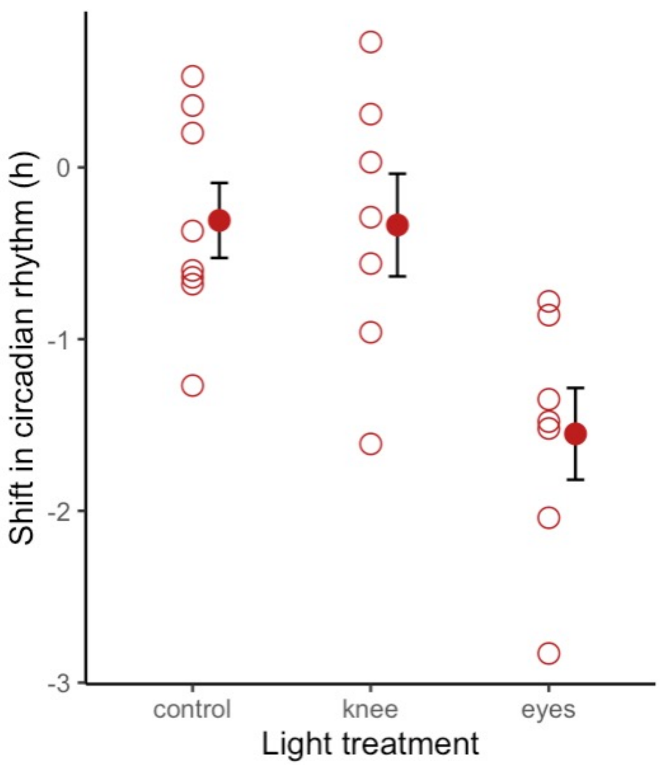
The probability of rejection of **H_0** (P-value) is estimated as the number of F-values in the null distribution equal or greater than the observed F-value (i.e., one tailed-test).



THE ANALYSIS OF VARIANCE (ANOVA)
for comparing multiple sample means (groups or treatments)

H_0: The samples come from statistical populations with the same mean, i.e., $\mu_{control} = \mu_{knee} = \mu_{eyes}$.

H_A: At least two samples come from different statistical populations with different means.



Research conclusion: Light treatment influences shifts in circadian rhythm.

ANOVA

Assumptions are the same as for the independent two sample t-test:

- Each of the observations is a random sample from its population (whether they are the same or different populations).

- The variable (e.g., shift in circadian rhythm) is normally distributed in each (treatment) population. More on that in another lecture.

- The variances are equal among all populations from which the treatments were sampled (otherwise the F values change in ways that may not measure difference among means). More on that in another lecture.

"The knees who say night"

H_0: $\mu_{control} = \mu_{knee} = \mu_{eyes}$
H_A: at least one population mean (μ) is different from another population mean or other population means.

Conclusion?
Significant, but how?

How do we know which group means differ from one another?

Why not simply not contrast all pairs of means using a two-sample mean t-test?
Control vs. knee; control vs. eyes; knee vs. eyes?

More later in the course!