

Fruit Bursts

I need **4** volunteers...

How many choices does the **4th** person
have in their choice of Fruit Burst?



Fruit Bursts

I need **4** volunteers...

How many choices does the **4th** person have?... Only (**4** - **1**) = **3** of you have the freedom to choose your Fruit Burst!

Constraint: there are only **4** Fruit Bursts and each person “has” to eat a unique one

Degrees of freedom



What is a degree of freedom?

The degree of freedom is

- the number of variables that are **free to vary** without breaking a constraint,
- the number of **independent** observations or measurements that can be made in order to calculate some statistic.

Example

Given data sample $x = \{1, 5, 6\}$

How many elements of x are **free to vary** (independently)?

Example (1-SAMPLE T TEST)

Given data sample $x = \{1, 5, 6\}$

Constraint: we would like to estimate the mean of the population.

Due to maths... **mean(x) * (number of observations in x) = sum of observations**

So sum of observations (must) equal $4 * 3 = 12$

How many elements of x are **free to vary** (independently)?

```
> t.test(x = c(1, 5, 6), mu = 0)
```

One Sample t-test

```
data:  c(1, 5, 6)
```

```
t = 2.6186, df = 2, p-value = 0.1201
```

```
alternative hypothesis: true mean is not equal to 0
```

```
95 percent confidence interval:
```

```
-2.572411 10.572411
```

```
sample estimates:
```

```
mean of x
```

```
4
```

Example (2 SAMPLE T TEST)

Given data sample $x = \{1, 5, 6\}$ and $k = \{7, 9, 11\}$

Constraint: we would like to estimate the differences between means.

Due to maths... **mean(.) * (number of observations in .) = sum of observations**

So, sum of x (must) equal $4 * 3 = 12$ & sum of k (must) equal $9 * 3 = 27$

How many elements of x and k combined are **free to vary** (independently)?


```
> t.test(x = c(1, 5, 6), y = c(7, 9, 11), var.equal = TRUE)
```

Two Sample t-test

data: c(1, 5, 6) and c(7, 9, 11)

t = -2.6112, df = 4, p-value = 0.05935

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-10.3164876 0.3164876

sample estimates:

mean of x mean of y

4

9

Example (PAIRED T TEST)

Given data sample $x = \{1, 5, 6, 7, 9, 11\}$ measurements taken from **three** different individuals **before** and **after** some event

Constraint: we would like to estimate the differences between/change in means.

Due to maths... **mean(.) * (number of observations in .) = sum of observations**

How many elements of x are **free to vary (independently)**?

```
> t.test(x = c(1, 5, 6), y = c(7, 9, 11), var.equal = TRUE, paired = TRUE)
```

Paired t-test

data: c(1, 5, 6) and c(7, 9, 11)

t = -8.6603, df = 2, p-value = 0.01307

alternative hypothesis: true mean difference is not equal to 0

95 percent confidence interval:

-7.484138 -2.515862

sample estimates:

mean difference

-5

Example (ANOVA)

Given data sample $x = \{1, 5, 6\}$, $k = \{7, 9, 11\}$ and $h = \{1, 10, 22\}$

Constraint: we would like to estimate the differences between means.

Due to maths... **mean(.) * (number of observations in .) = sum of observations**

So, sum of x (must) equal $4 * 3 = 12$ & sum of k (must) equal $9 * 3 = 27$ & sum of h (must) equal $11 * 3 = 33$

How many elements of x , k , and h combined are **free to vary** (independently)?

```

> data <- data.frame(y = c(1, 5, 6, 7, 9, 11, 1, 10, 22), group = rep(c("x", "k", "h"), each = 3))
> aov(y ~ group, data = data) |> summary()
      Df Sum Sq Mean Sq F value Pr(>F)
group    2     78   39.00   0.959  0.435
Residuals  6    244   40.67
> lm(y ~ group, data = data) |> summary()

```

Call:

```
lm(formula = y ~ group, data = data)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|-----|----|--------|----|-----|
| -10 | -2 | 0 | 2 | 11 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 11.000 | 3.682 | 2.988 | 0.0244 * |
| groupk | -2.000 | 5.207 | -0.384 | 0.7141 |
| groupx | -7.000 | 5.207 | -1.344 | 0.2274 |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.377 on 6 degrees of freedom

Multiple R-squared: 0.2422, Adjusted R-squared: -0.01035

F-statistic: 0.959 on 2 and 6 DF, p-value: 0.4351

Example (REGRESSION)

```
lm(logAUC ~ Disease*Organ, data = data)
```

| Disease | Organ | Animal | Sample | logAUC |
|----------|---------|--------|--------|--------|
| Healthy | innerLV | 1 | 1 | 9.40 |
| Healthy | outerLV | 2 | 2 | 8.83 |
| Healthy | innerLV | 3 | 1 | 10.33 |
| Healthy | outerLV | 4 | 2 | 10.49 |
| Healthy | innerLV | 5 | 1 | 9.74 |
| Healthy | outerLV | 6 | 2 | 10.98 |
| Diabetic | innerLV | 7 | 1 | 7.92 |
| Diabetic | outerLV | 8 | 2 | 9.37 |
| Diabetic | innerLV | 9 | 1 | 8.69 |
| Diabetic | outerLV | 10 | 2 | 11.31 |
| Diabetic | innerLV | 11 | 1 | 7.01 |
| Diabetic | outerLV | 12 | 2 | 9.29 |

```
> data <- readr::read_csv("https://raw.githubusercontent.com/STATS-UOA/databunker/master/data/factorial_expt.csv")
```

```
Rows: 12 Columns: 5
```

```
0s— Column specification
```

```
Delimiter: ","
```

```
chr (2): Disease, Organ
```

```
dbl (3): Animal, Sample, logAUC
```

```
i Use `spec()` to retrieve the full column specification for this data.
```

```
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
> aov(logAUC ~ Disease*Organ, data = data) |> summary()
```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---------------|----|--------|---------|---------|----------|
| Disease | 1 | 3.183 | 3.183 | 3.630 | 0.0932 . |
| Organ | 1 | 4.296 | 4.296 | 4.900 | 0.0578 . |
| Disease:Organ | 1 | 2.539 | 2.539 | 2.896 | 0.1272 |
| Residuals | 8 | 7.014 | 0.877 | | |

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> lm(logAUC ~ Disease*Organ, data = data) |> summary()
```

```
Call:
```

```
lm(formula = logAUC ~ Disease * Organ, data = data)
```

```
Residuals:
```

| | Min | 1Q | Median | 3Q | Max |
|--|----------|----------|----------|---------|---------|
| | -1.27000 | -0.64000 | -0.01833 | 0.58417 | 1.32000 |

```
Coefficients:
```

| | Estimate | Std. Error | t value | Pr(> t) |
|-----------------------------|----------|------------|---------|--------------|
| (Intercept) | 7.8733 | 0.5406 | 14.565 | 4.84e-07 *** |
| DiseaseHealthy | 1.9500 | 0.7645 | 2.551 | 0.0341 * |
| OrganouterLV | 2.1167 | 0.7645 | 2.769 | 0.0243 * |
| DiseaseHealthy:OrganouterLV | -1.8400 | 1.0812 | -1.702 | 0.1272 |

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.9363 on 8 degrees of freedom
```

```
Multiple R-squared:  0.5882,    Adjusted R-squared:  0.4338
```

```
F-statistic: 3.809 on 3 and 8 DF,  p-value: 0.05788
```

Example (REGRESSION)

```
lm(logAUC ~ Disease*Organ + Animal, data =  
data)
```

| Disease | Organ | Animal | Sample | logAUC |
|----------|---------|--------|--------|--------|
| Healthy | innerLV | 1 | 1 | 9.40 |
| Healthy | outerLV | 1 | 2 | 8.83 |
| Healthy | innerLV | 2 | 1 | 10.33 |
| Healthy | outerLV | 2 | 2 | 10.49 |
| Healthy | innerLV | 3 | 1 | 9.74 |
| Healthy | outerLV | 3 | 2 | 10.98 |
| Diabetic | innerLV | 4 | 1 | 7.92 |
| Diabetic | outerLV | 4 | 2 | 9.37 |
| Diabetic | innerLV | 5 | 1 | 8.69 |
| Diabetic | outerLV | 5 | 2 | 11.31 |
| Diabetic | innerLV | 6 | 1 | 7.01 |
| Diabetic | outerLV | 6 | 2 | 9.29 |


```

> data <- readr::read_csv("https://raw.githubusercontent.com/STATS-UOA/databunker/master/data/split_plot.csv")
Rows: 12 Columns: 5
0s— Column specification
Delimiter: ","
chr (2): Disease, Organ
dbl (3): Animal, Sample, logAUC

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
> aov(logAUC ~ Disease*Organ + Animal, data = data) |> summary()
          Df Sum Sq Mean Sq F value Pr(>F)
Disease    1  3.183    3.183   3.309 0.1117
Organ       1  4.296    4.296   4.467 0.0724 .
Animal      1  0.281    0.281   0.292 0.6054
Disease:Organ 1  2.539    2.539   2.640 0.1482
Residuals   7  6.732    0.962
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> lm(logAUC ~ Disease*Organ + Animal, data = data) |> summary()

Call:
lm(formula = logAUC ~ Disease * Organ + Animal, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-1.08250 -0.54625 -0.00083  0.55312  1.32000

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    6.9358     1.8238   3.803  0.00669 **
DiseaseHealthy  2.5125     1.3127   1.914  0.09718 .
OrganouterLV    2.1167     0.8007   2.643  0.03326 *
Animal          0.1875     0.3467   0.541  0.60544
DiseaseHealthy:OrganouterLV -1.8400     1.1324  -1.625  0.14822
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9807 on 7 degrees of freedom
Multiple R-squared:  0.6047,    Adjusted R-squared:  0.3788
F-statistic: 2.677 on 4 and 7 DF,  p-value: 0.121

```