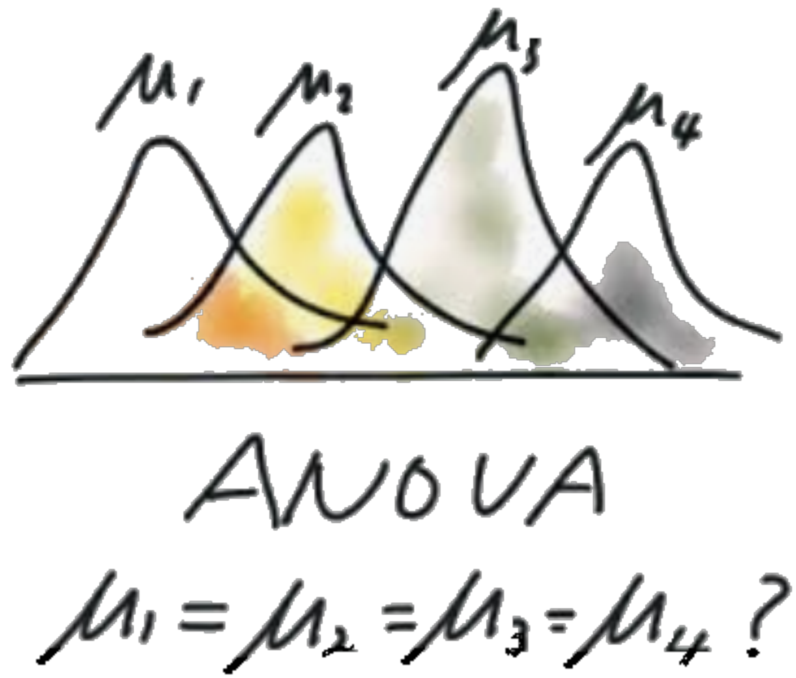
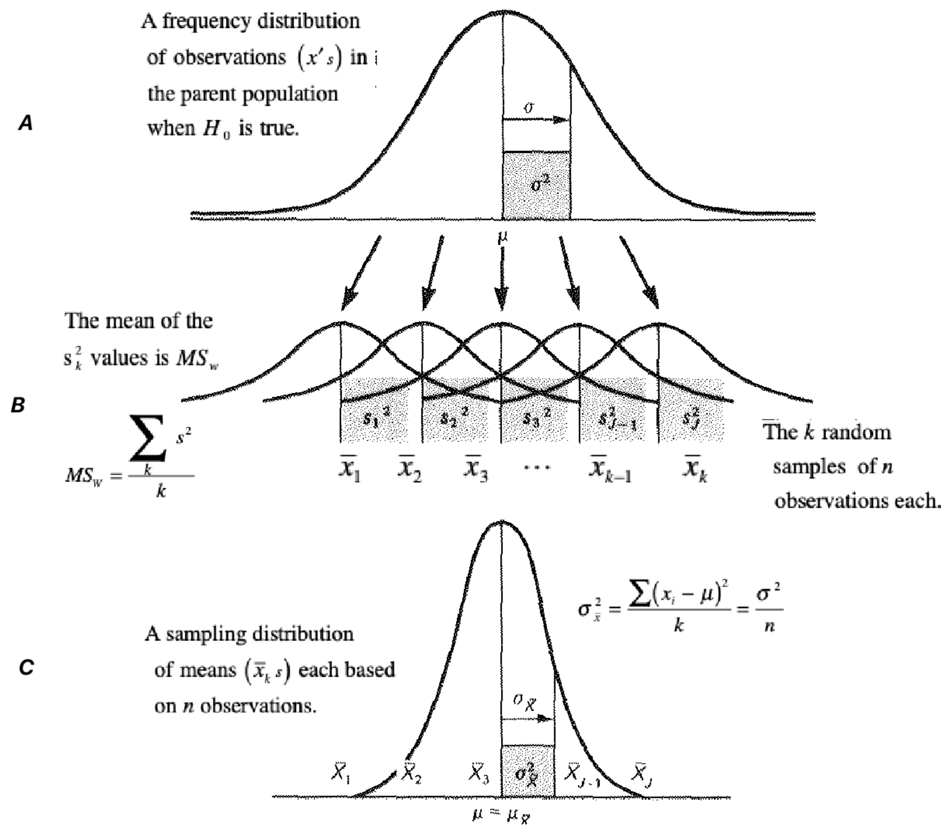
Analysis of Variance



Analysis of Variance



1. Since $s_{\bar{x}}^2 = \frac{s^2}{n}$, $ns_{\bar{x}}^2 = s^2$. Hence, $E\left(ns_{\bar{x}}^2\right) = \sigma^2$;

$ns_{\bar{x}}^2$ is called the mean square between or MS_b

2. $\frac{\sum_{k=1}^{n} s_k^2}{n} = \frac{\left(s_1^2 + s_2^2 + \cdots + s_n^2\right)}{n}, s_w^2;\ E\left(s_w^2\right) = \sigma^2;$

s_w^2 is called the mean square within or MS_w

3. Therefore $F = \frac{MS_b}{MS_w}$ should be "about" 1

when the null hypothesis is true.

Analysis of Variance

Penicillin is produced by the *Penicillium* fungus, which is grown in a broth whose sugar content must be carefully controlled. Several samples of broth were taken on three successive days, and the amount of dissolved sugars, in milligrams per milliliter, was measured on each sample. The results were as follows. Can we conclude that the mean sugar concentration differs among the three days?

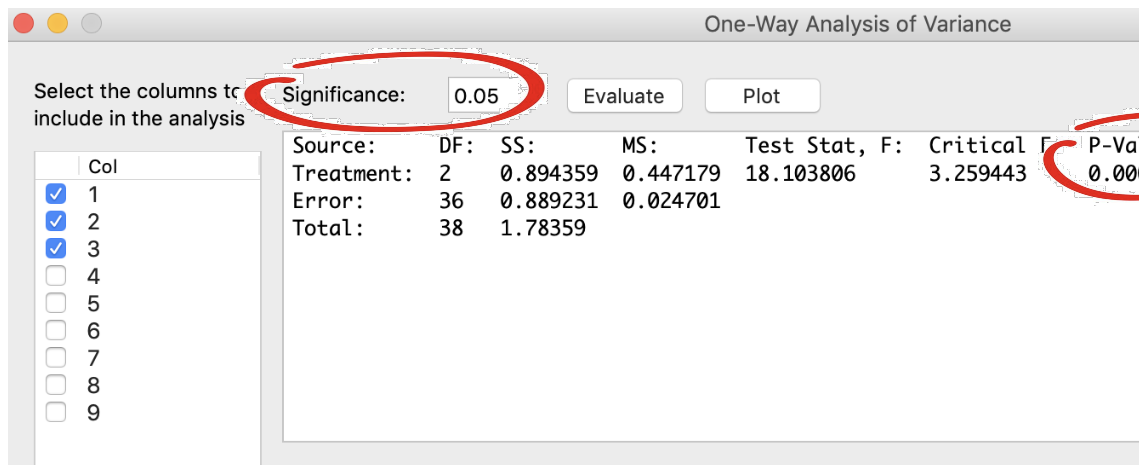
Day 1	Day 2	Day 3
4.8	5.4	5.5
5.1	5.0	5.1
5.1	5.0	5.3
4.8	5.1	5.5
5.2	5.2	5.3
4.9	5.1	5.5
5.3	5.3	5.1
4.9	5.2	5.6
5.0	5.2	5.3
4.8	5.1	5.2
4.8	5.4	5.5
5.1	5.2	5.3
5.0	5.4	5.4

- Normality – that each sample is taken from a normally distributed population.
- Sample independence – that each sample has been drawn independently of the other samples.
- Variance equality – that the variance of data in the different groups should be the same.

Analysis of Variance

Penicillin is produced by the *Penicillium* fungus, which is grown in a broth whose sugar content must be carefully controlled. Several samples of broth were taken on three successive days, and the amount of dissolved sugars, in milligrams per milliliter, was measured on each sample. The results were as follows. Can we conclude that the mean sugar concentration differs among the three days?

Day 1	Day 2	Day 3
4.8	5.4	5.4
5.1	5.2	5.2
5.1	4.8	4.8
4.8	5.1	5.1
5.2	5.2	5.2
4.9	5.1	5.1
5.3	5.3	5.3
4.9	5.2	5.2
5.0	4.9	4.9
4.8	5.1	5.1



Reject Ho!

Analysis of Variance

Welch's ANOVA compares two means to see if they are equal. It is an alternative to the Classic ANOVA (Fisher) and can be used even **if your data violates the assumption of homogeneity of variances.**

You should run Welch's test in all cases where you have normally distributed data that violates the assumption of homogeneity of variance. ANOVA (and the non-parametric alternative Kruskal-Wallis) are very unstable for these situations, producing Type I error rates that are:

Conservative for large sample sizes and
Inflated for small sample size.

Welch One Way ANOVA

$$F = \frac{\frac{1}{k-1}\sum_{j=1}^{k} w_j(\bar{x}_j - \bar{x}')^2}{1 + \frac{2(k-2)}{k^2-1}\sum_{j=1}^{k}\left(\frac{1}{n_j-1}\right)\left(1-\frac{w_j}{w}\right)^2}$$

where

$$w_j = \frac{n_j}{s_j^2} \qquad w = \sum_{j=1}^{k} w_j \qquad \bar{x}' = \frac{\sum_{j=1}^{k} w_j \bar{x}_j}{w}$$

then

$$F \sim F(k-1, df)$$

where

$$df = \frac{k^2-1}{3\sum_{j=1}^{k}\left(\frac{1}{n_j-1}\right)\left(1-\frac{w_j}{w}\right)^2}$$

Analysis of Variance

Welch One Way ANOVA

	Day 1	Day 2	Day 3	
	4.8	5.4	5.5	
	5.1	5.0	5.1	
	5.1	5.0	5.3	
	4.8	5.1	5.5	
	5.2	5.2	5.3	
	4.9	5.1	5.5	
	5.3	5.3	5.1	
	4.9	5.2	5.6	
	5.0	5.2	5.3	
	4.8	5.1	5.2	
	4.8	5.4	5.5	
	5.1	5.2	5.3	
	5.0	5.4	5.4	
α	0.05			
count	13	13	13	
mean	5.0	5.2	5.4	
variance	0.028	0.020	0.026	
w	463.014	650.000	499.507	1612.521
mean'	2307.945	3380.000	2674.286	5.186
a	18.743	0.131	14.104	16.489
b	0.042	0.030	0.040	0.112
k	3.000			
F (TS)	16.041		**CV:**	3.422
df1	2.000			
df2	23.865		**Reject Ho!**	
p-value	0.000			

Analysis of Variance

1	2	3
50	44	16
39	31	60
42	50	24
45	22	19
38	30	31
44	27	37
40	32	44
49	25	55
42	40	
41		

One-Way Analysis of Variance

Select the columns to include in the analysis

Col: 1 ☑, 2 ☑, 3 ☑, 4 ☐, 5 ☐, 6 ☐, 7 ☐, 8 ☐, 9 ☐

Significance: 0.05 | Evaluate | Plot

Source:	DF:	SS:	MS:	Test Stat, F:	Critical F:	P-Value:
Treatment:	2	474.277778	237.138889	2.109681	3.402829	0.143208
Error:	24	2697.722222	112.405093			
Total:	26	3172.00				

DO NOT Reject Ho!

Welch One Way ANOVA

	New	Old	Control	
	50	44	16	
	39	31	60	
	42	50	24	
	45	22	19	
	38	30	31	
	44	27	37	
	40	32	44	
	49	25	55	
	42	40		
	41			
α	0.05			
count	10	9	8	
mean	43.000	33.444	35.750	
variance	16.222	86.528	265.643	
w	0.616	0.104	0.030	0.75
mean'	26.507	3.479	1.077	41.38
a	1.608	6.558	0.956	4.56
b	0.004	0.093	0.132	0.22
k	3.000			
F (TS)	4.315		**CV:**	3.98
df1	2.000			
df2	11.700		**Reject Ho!**	
p-value	0.041			

Two-way Analysis of Variance

TABLE : Crash Test Force on Femur with Two Factors: Femur Side and Vehicle Size Category

	Small	Midsize	Large	SUV
Left Femur	1.6 1.4 0.5 0.2 0.4	0.4 0.7 1.1 0.7 0.5	0.6 1.8 0.3 1.3 1.1	0.4 0.4 0.6 0.2 0.2
Right Femur	2.8 1.0 0.3 0.3 0.2	0.6 0.8 1.3 0.5 1.1	1.5 1.7 0.2 0.6 0.9	0.7 0.7 3.0 0.2 0.2

Two-way analysis of variance involves *two* factors, such as vehicle size (small, midsize, large SUV) and femur side (left, right) as shown in the table. The two-way analysis of variance procedure requires that we test for (1) an interaction effect between the two factors; (2) an effect from the row factor; (3) an effect from the column factor.

- **There does not appear to be an interaction effect.**
- **The car crash force measurements do not appear to be affected by whether the femur is in the left leg or right leg.**
- **The femur crash force measurements do not appear to be affected by the size of the vehicle.**