One-way ANOVA

Inference for one-way ANOVA

IPS Chapter 12.1

Objectives (IPS Chapter 12.1)

Inference for one-way ANOVA

- Comparing means
- The two-sample *t* statistic
- An overview of ANOVA
- The ANOVA model
- Testing hypotheses in one-way ANOVA
- The *F*-test
- The ANOVA table

The idea of ANOVA

Reminders: A **factor** is a variable that can take one of several **levels** used to differentiate one group from another.

An experiment has a **one-way,** or **completely randomized, design** if several levels of one factor are being studied and the individuals are randomly assigned to its levels.

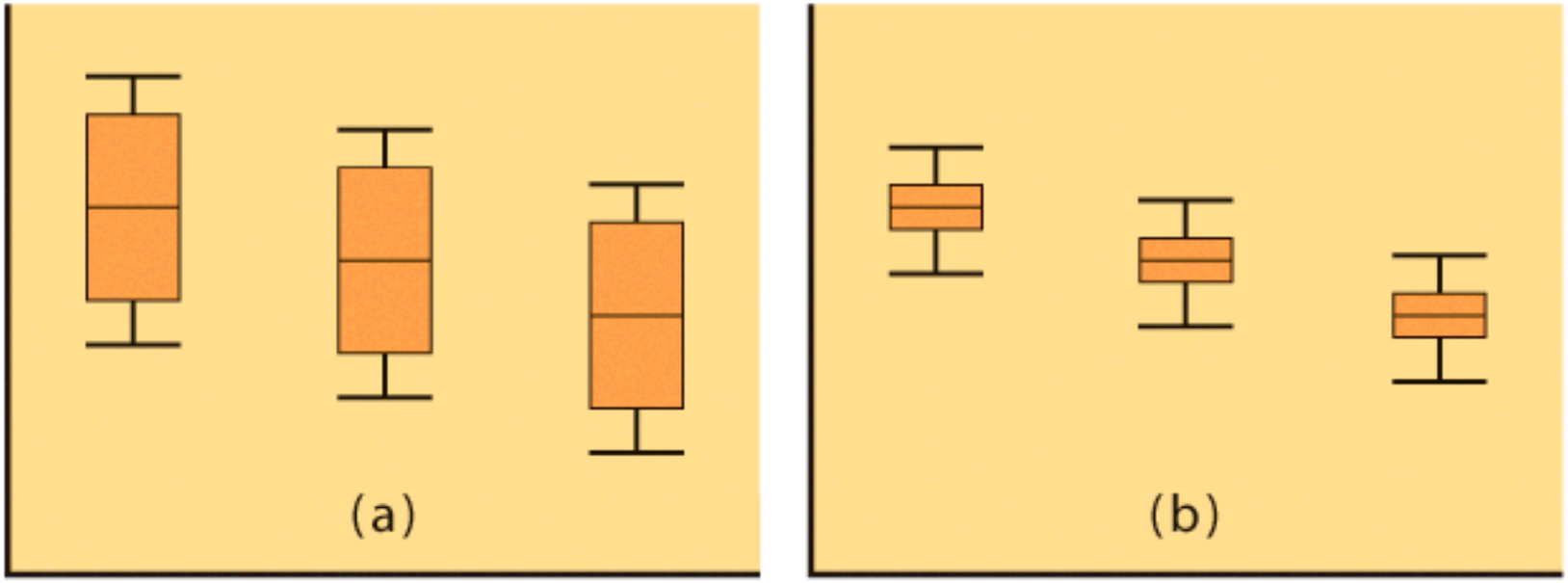
- Example: Four levels of nematode quantity in seedling growth experiment.
  - Two seed species and four levels of nematodes would be a two-way design.

Analysis of variance (ANOVA) is the technique used to determine if there are any differences among the means of the treatment groups.

One-way ANOVA is used for completely randomized, one-way designs.

Comparing means

We want to know if the observed differences in sample means are likely to have occurred by chance just because of random sampling.



This will likely depend on both the difference between the sample means and how much variability there is within each sample.

Reminder: Two-sample *t* statistic

A two sample *t*-test assuming equal variance or an ANOVA comparing only two groups will give you the exact same *p*-value (for a two-sided hypothesis).

H_0: $\mu_1 = \mu_2$ H_a: $\mu_1 \neq \mu_2$	H_0: $\mu_1 = \mu_2$ H_a: $\mu_1 \neq \mu_2$
One-way ANOVA	*t*-test assuming equal variance
F-statistic	*t*-statistic

F = t^2 and both *p*-values are the same.

But the *t*-test is more flexible: You may choose a one-sided alternative instead, or you may want to run a *t*-test assuming unequal variance if you are not sure that your two populations have the same standard deviation σ.

An Overview of ANOVA

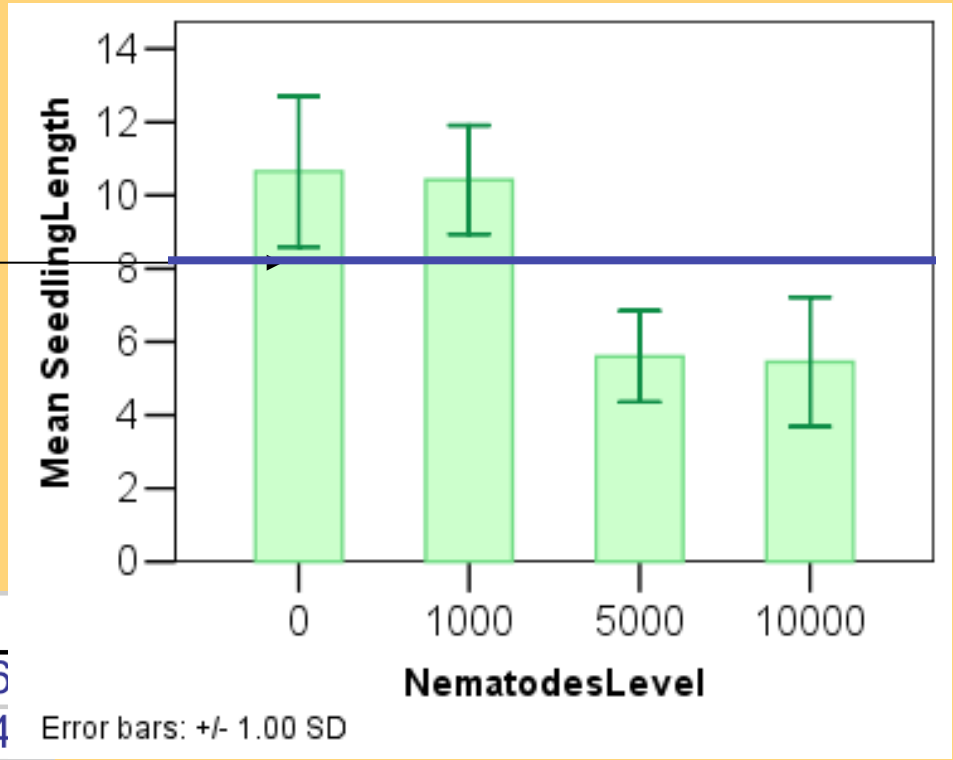
- We **first** examine the multiple populations or multiple treatments to test for overall statistical significance as evidence of any difference among the parameters we want to compare using the **ANOVA F-test**
- **If** that overall test shows statistical significance, then a detailed follow-up analysis is legitimate.
  - If we planned our experiment with specific alternative hypotheses in mind (before gathering the data), we can test them using **contrasts.**
  - If we do not have specific alternatives, we can examine all pair-wise parameter comparisons to define which parameters differ from which, using **multiple comparisons procedures.**

Nematodes and plant growth

Do nematodes affect plant growth? A botanist prepares 16 identical planting pots and adds different numbers of nematodes into the pots. Seedling growth (in mm) is recorded two weeks later.

Hypotheses: All μ_i are the same (H_0)

versus not All μ_i are the same (H_a)



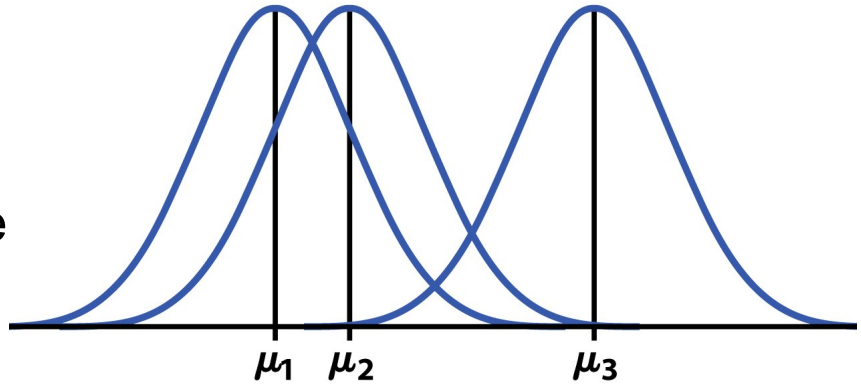
Nematodes	Seedling growth				$\bar{x}_i$
0	10.8	9.1	13.5	9.2	10.6
1,000	11.1	11.1	8.2	11.3	10.4
5,000	5.4	4.6	7.4	5	5.6
10,000	5.8	5.3	3.2	7.5	5.45
overall mean 8.03					

The ANOVA model

Random sampling always produces chance variations. Any "factor effect" would thus show up in our data as the factor-driven differences plus chance variations ("error"):

Data = fit ("group mean") **+ residual** ("error")

The one-way ANOVA model analyzes situations where chance variations are normally distributed $N(0,\sigma)$ so that:



$$x_{ij} = \mu_i + \epsilon_{ij}$$

for $i = 1, \ldots, I$ and $j = 1, \ldots, n_i$. The ϵ_{ij} are assumed to be from an $N(0, \sigma)$ distribution. The **parameters of the model** are the population means $\mu_1, \mu_2, \ldots, \mu_I$ and the common standard deviation σ.

The ANOVA table

Source of variation	Sum of squares SS	DF	Mean square MS	F	P value
Groups	$SSG = \sum n_i(\bar{x}_i - \bar{x})^2$	I -1	SSG/DFG	MSG/MSE	Tail area above F
Error	$SSE = \sum (n_i - 1)s_i^2$	N - I	SSE/DFE		
Total	SST=SSG+SSE $= \sum (x_{ij} - \bar{x})^2$	N – 1			

R^2 = SSG/SST
Coefficient of determination

$\sqrt{}$MSE = s_p
Pooled standard deviation

The sum of squares represents variation in the data: SST = SSG + SSE.

The degrees of freedom likewise reflect the ANOVA model: DFT = DFG + DFE.

Testing hypotheses in one-way ANOVA

We have ***I* independent SRSs,** from I populations or treatments.

The i^{th} population has a **normal distribution** with unknown mean μ_i.

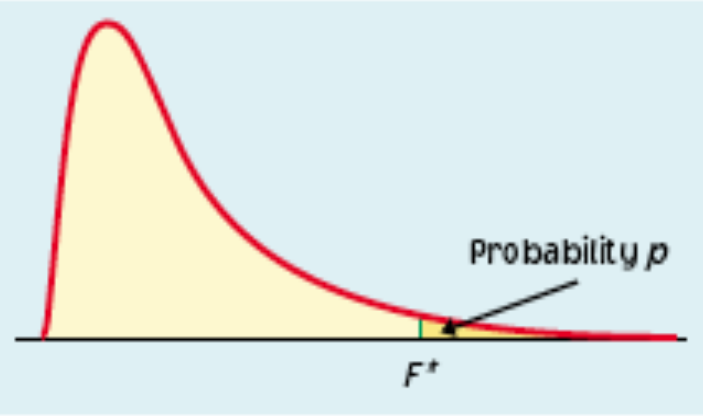
All *I* populations have the **same standard deviation σ**, unknown.

The ANOVA *F* statistic tests:

$$F = \frac{\text{SSG}/(I-1)}{\text{SSE}/(N-I)}$$

H_0: $\mu_1 = \mu_2 = \ldots = \mu_I$

H_a: not all the μ_i are equal.



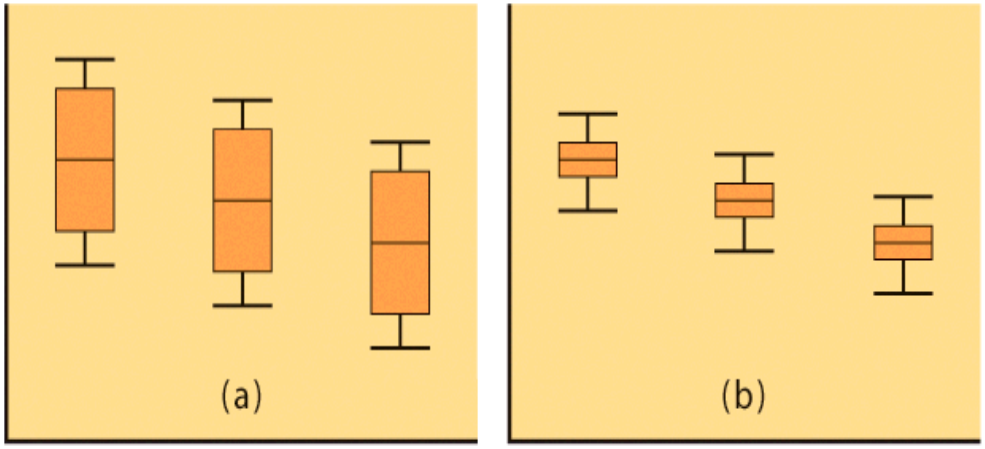
When H_0 is true, *F* has the **F distribution** with I − 1 (*numerator*) and N − I (*denominator*) degrees of freedom.

The ANOVA *F*-test

The **ANOVA F-statistic** compares variation due to specific sources (levels of the factor) with variation among individuals who should be similar (individuals in the same sample).

$$F = \frac{\text{variation among sample means}}{\text{variation among individuals in same sample}}$$

Difference in means small relative to overall variability



Difference in means large relative to overall variability

➔ F tends to be small ➔ F tends to be large

Larger F-values typically yield more significant results. How large depends on the degrees of freedom (I − 1 and N − I).

Checking our assumptions

Each of the *I* populations must be **normally distributed** (histograms or normal quantile plots). But the test is robust to normality deviations for large enough sample sizes, thanks to the central limit theorem.

The ANOVA F-test requires that all populations have the **same standard deviation σ.** Since σ is unknown, this can be hard to check.

Practically: The results of the ANOVA F-test are approximately correct when the largest sample standard deviation is no more than twice as large as the smallest sample standard deviation.

(Equal sample sizes also make ANOVA more robust to deviations from the equal σ rule)

Do nematodes affect plant growth?

	Seedling growth				$\bar{x}_i$	s_i
0 nematode	10.8	9.1	13.5	9.2	10.65	2.053
1000 nematodes	11.1	11.1	8.2	11.3	10.425	1.486
5000 nematodes	5.4	4.6	7.4	5.0	5.6	1.244
10000 nematodes	5.8	5.3	3.2	7.5	5.45	1.771

Conditions required:

- equal variances: checking that largest s_i no more than twice smallest s_i

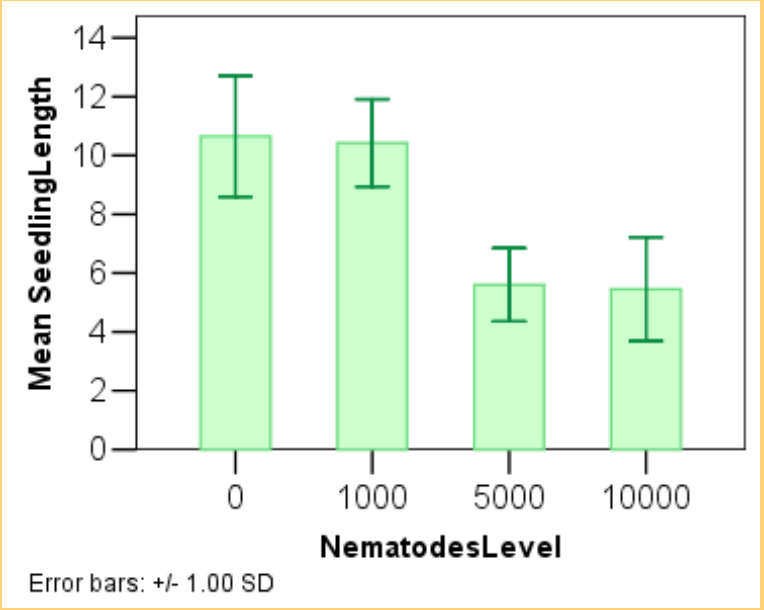
  Largest s_i = 2.053; smallest s_i = 1.244

- Independent SRSs

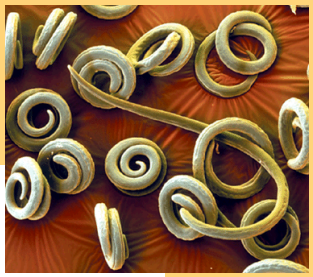
  Four groups obviously independent

- Distributions "roughly" normal

  It is hard to assess normality with only four points per condition. But the pots in each group are identical, and there is no reason to suspect skewed distributions.



Excel output for the one-way ANOVA



Menu/Tools/DataAnalysis/AnovaSingleFactor

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
0 nematode	4	42.6	10.65	4.21667
1000 nematodes	4	41.7	10.425	2.20917
5000 nematodes	4	22.4	5.6	1.54667
10000 nematodes	4	21.8	5.45	3.13667

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	100.647	3	33.549	12.0797	0.00062	3.4902996
Within Groups	33.3275	12	2.77729			
Total	133.974	15				

numerator

denominator

Here, the calculated F-value (12.08) is larger than $F_{critical}$ (3.49) for α=0.05.

Thus, the test is significant at α=5% ➔ Not all mean seedling lengths are the same; the number of nematodes is an influential factor.

SPSS output for the one-way ANOVA

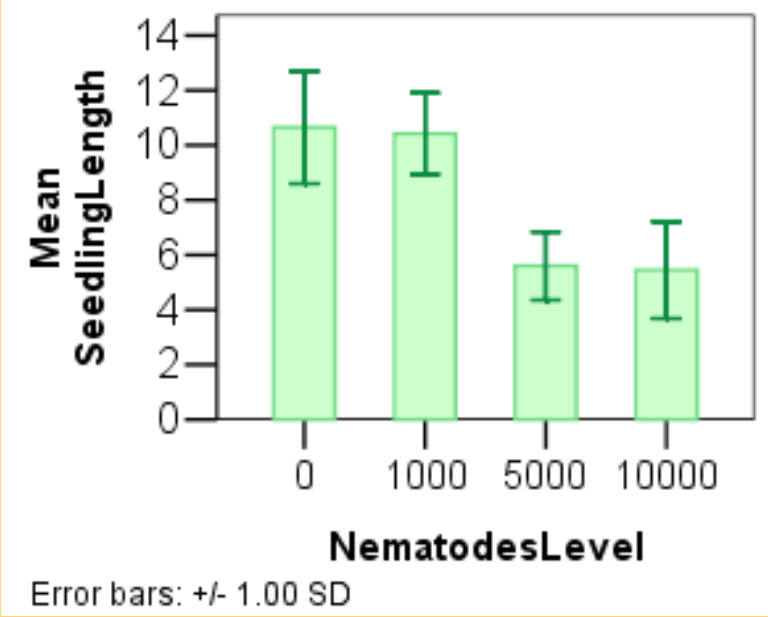
ANOVA

SeedlingLength

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	100.647	3	33.549	12.080	.001
Within Groups	33.328	12	2.777		
Total	133.974	15			

The **ANOVA** found that the amount of nematodes in pots significantly impacts seedling growth.

The **graph** suggests that nematode amounts above 1,000 per pot are detrimental to seedling growth.



Using Table E

The F distribution is asymmetrical and has two distinct degrees of freedom. This was discovered by Fisher, hence the label "F."

Once again, what we do is calculate the value of F for our sample data and then look up the corresponding area under the curve in Table E.

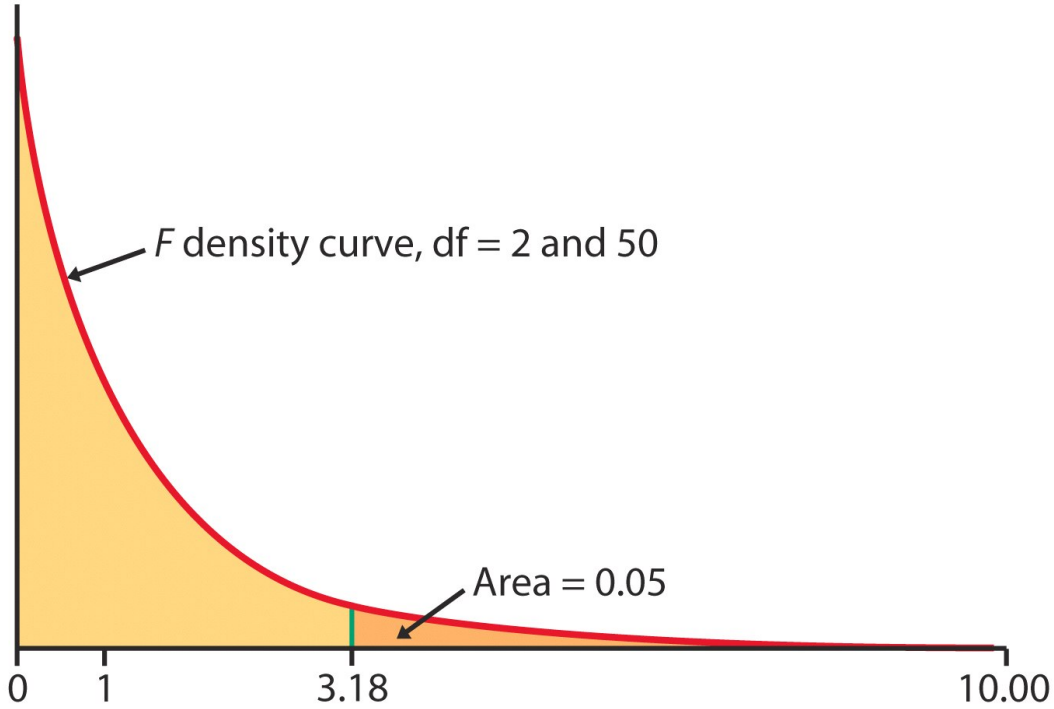


Table E F distribution critical values

$df_{num} = I - 1$

For df: 5,4

Degrees of freedom in the denominator	p	Degrees of freedom in the numerator: 1	2	3	4	5	6	7	8
1	0.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44
	0.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88
	0.025	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66
	0.010	4052.2	4999.5	5403.4	5624.6	5763.6	5859	5928.4	5981.1
	0.001	405284	500000	540379	562500	576405	585937	592873	598144
2	0.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37
	0.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37
	0.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37
	0.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37
	0.001	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37
3	0.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25
	0.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85
	0.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54
	0.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49
	0.001	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62
4	0.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95
	0.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04
	0.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98
	0.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80
	0.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00
5	0.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34
	0.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82
	0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76
	0.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29
	0.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65
6	0.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98
	0.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15
	0.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60
	0.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10
	0.001	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03

p

F

$df_{den} = N - I$

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	101	3	33.5	12.08	0.00062	3.4903
Within Groups	33.3	12	2.78			
Total	134	15				

		Degrees of freedom in the numerator									
	p	1	2	3	4	5	6	7	8	9	10
	0.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42
	0.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
9	0.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96
	0.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26
	0.001	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11	9.89
	0.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
	0.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
10	0.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72
	0.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
	0.001	21.04	14.91	12.55	11.28	10.48	9.93	9.52	9.20	8.96	8.75
	0.100	3.18	2.81	2.61					2.24	2.21	2.19
	0.050	4.75	3.89	3.49					2.85	2.80	2.75
12	0.025	6.55	5.10	4.47					3.51	3.44	3.37
	0.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
	0.001	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48	7.29
	0.10				2.36	2.27	2.21	2.16	2.12	2.09	2.06
	0.05				3.06	2.90	2.79	2.71	2.64	2.59	2.54
15	0.02				3.80	3.58	3.41	3.29	3.20	3.12	3.06
	0.01				4.89	4.56	4.32	4.14	4.00	3.89	3.80
	0.00				8.25	7.57	7.09	6.74	6.47	6.26	6.08
	0.100	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94
	0.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
20	0.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77
	0.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
	0.001	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08

in the denominator

$F_{critical}$ for α 5% is 3.49

F = 12.08 > 10.80
Thus *p* < 0.001

Yogurt preparation and taste

Yogurt can be made using three distinct commercial preparation methods: traditional, ultra filtration, and reverse osmosis.

To study the effect of these methods on taste, an experiment was designed where three batches of yogurt were prepared for each of the three methods. A trained expert tasted each of the nine samples, presented in random order, and judged them on a scale of 1 to 10.

Variables, hypotheses, assumptions, calculations?

ANOVA table

Source of variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between groups	17.3					
Within groups	4.6					
Total	21.9					

TABLE D F distribution critical values

$df_{num} = I - 1$

Degrees of freedom in the denominator	p	Degrees of freedom in the numerator: 1	2	3	4	5	6	7	8
1	0.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44
	0.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88
	0.025	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66
	0.010	4052.2	4999.5	5403.4	5624.6	5763.6	5859	5928.4	5981.1
	0.001	405284	500000	540379	562500	576405	585937	592873	598144
2	0.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37
	0.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37
	0.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37
	0.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37
	0.001	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37
3	0.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25
	0.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85
	0.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54
	0.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49
	0.001	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62
4	0.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95
	0.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04
	0.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98
	0.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80
	0.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00
5	0.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34
	0.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82
	0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76
	0.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29
	0.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65
6	0.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98
	0.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15
	0.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60
	0.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10
	0.001	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03

F

$df_{den} = N - I$

Computation details

$$F = \frac{\text{MSG}}{\text{MSE}} = \frac{\text{SSG}/(I-1)}{\text{SSE}/(N-I)}$$

MSG, the mean square for groups, measures how different the individual means are from the overall mean (weighted average of square distances of sample averages to the overall mean). SSG is the sum of squares for groups.

$$\text{MSG} = \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_I(\bar{x}_I - \bar{x})^2}{I-1}$$

MSE, the mean square for error is the **pooled sample variance s_p^2** and estimates the common variance σ^2 of the I populations (weighted average of the variances from each of the I samples). SSE is the sum of squares for error.

$$\text{MSE} = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + \cdots + (n_I-1)s_I^2}{N-I}$$

One-way ANOVA

Comparing the means

IPS Chapter 12.2

Objectives (IPS Chapter 12.2)

Comparing the means

- Contrasts
- Multiple comparisons
- Power of the one-way ANOVA test

You have calculated a p-value for your ANOVA test. Now what?

If you found a significant result, you still need to determine which treatments were different from which.

- You can gain insight by looking back at your plots (boxplot, mean ± *s*).
- There are several tests of statistical significance designed specifically for multiple tests. You can choose *a priori* **contrasts**, or *a posteriori* **multiple comparisons**.
- You can find the confidence interval for each mean μ_i shown to be significantly different from the others.

- **Contrasts** can be used only when there are clear expectations BEFORE starting an experiment, and these are reflected in the experimental design. Contrasts are **planned comparisons.**
  - Patients are given either drug A, drug B, or a placebo. The placebo is meant to provide a baseline against which the other drugs can be compared.
- **Multiple comparisons** should be used when there are no specific planned comparisons. Those are *a posteriori*, **pair-wise tests** of significance.
  - We compare gas mileage for eight brands of SUVs. We have no prior knowledge to expect one brand to perform differently from the rest. Pair-wise comparisons should be performed here, but only if an ANOVA test on all eight brands reached statistical significance first.

It is NOT appropriate to use a contrast test when suggested comparisons appear only after the data is collected.

Contrasts: planned comparisons

When an experiment is designed to test one or more specific hypotheses that some treatments are different from other treatments, we can use contrasts to test for significant differences between these specific treatments, EVEN if the overall ANOVA F-test is not significant.

- Contrasts are more powerful than multiple comparisons because they are more specific. They are more able to pick up a significant difference.
- You can use a *t*-test on the contrasts or calculate a *t*-confidence interval.
- The results are valid regardless of the results of your multiple sample ANOVA test (you are still testing a valid hypothesis).

A contrast is a combination of population means of the form :

$$\psi = \sum a_i \mu_i$$

Where the coefficients a_i have sum 0.

The corresponding sample contrast is :

$$c = \sum a_i \bar{x}_i$$

The standard error of c is :

$$SE_c = s_p \sqrt{\sum \frac{a_i^2}{n_i}} = \sqrt{MSE \sum \frac{a_i^2}{n_i}}$$

To test the null hypothesis H_0: ψ = 0 use the *t*-statistic:

$$t = c / SE_c$$

With degrees of freedom **DFE** that is associated with s_p. The alternative hypothesis can be one- or two-sided.

A level *C* confidence interval for the contrast ψ is :

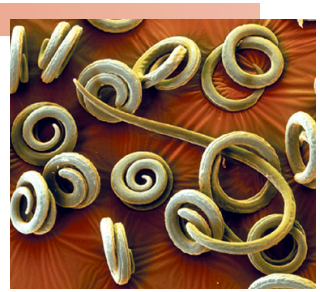
$$c \pm t^* SE_c$$

Where t^* is the critical value defining the middle *C*% of the *t* distribution with **DFE** degrees of freedom.

Contrasts are not always readily available in statistical software packages (when they are, you need to assign the coefficients "a_i"), or may be limited to comparing each sample to a control.

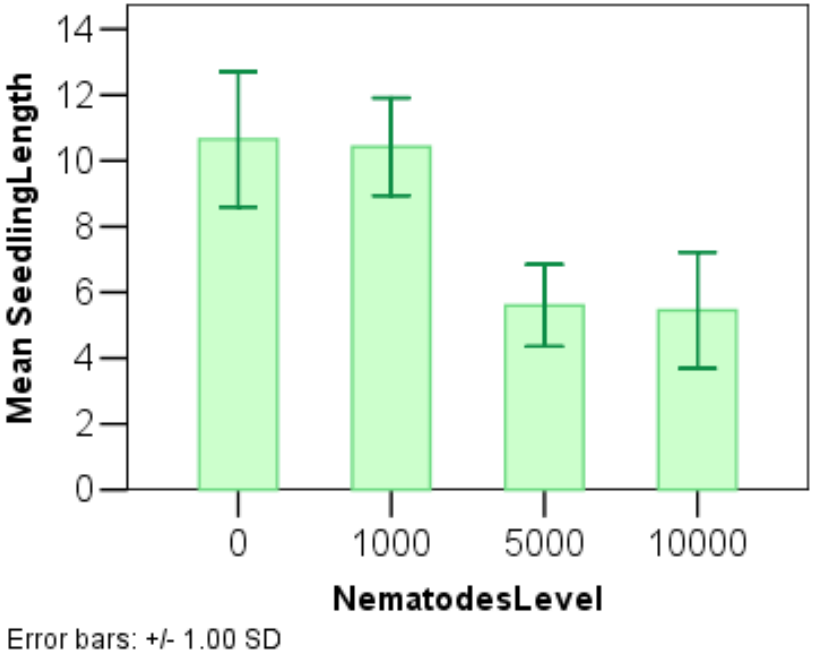
If your software doesn't provide an option for contrasts, you can test your contrast hypothesis with a regular *t*-test using the formulas we just highlighted. Remember to use the pooled variance and degrees of freedom as they reflect the better estimate of the population variance.

Nematodes and plant growth



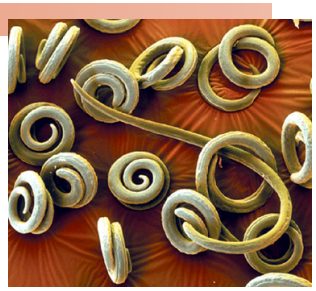
Do nematodes affect plant growth? A botanist prepares 16 identical planting pots and adds different numbers of nematodes into the pots. Seedling growth (in mm) is recorded two weeks later.

Nematodes	Seedling growth				$\bar{x}_i$
0	10.8	9.1	13.5	9.2	10.65
1,000	11.1	11.1	8.2	11.3	10.43
5,000	5.4	4.6	7.4	5	5.6
10,000	5.8	5.3	3.2	7.5	5.45
overall mean 8.03					



One group contains no nematodes at all. If the botanist planned this group as a baseline/control, then a contrast of all the nematode groups against the control would be valid.

Nematodes: planned comparison



Contrast of all the nematode groups against the control:

Combined contrast hypotheses:

H_0: $\mu_1 = 1/3\ (\mu_2 + \mu_3 + \mu_4)$

H_a: $\mu_1 > 1/3\ (\mu_2 + \mu_3 + \mu_4)$

	$\bar{x}_i$	s_i
G1: 0 nematode	10.65	2.053
G2: 1,000 nematodes	10.425	1.486
G3: 5,000 nematodes	5.6	1.244
G4: 1,0000 nematodes	5.45	1.771

Contrast coefficients: (+1 −1/3 −1/3 −1/3) or (+3 −1 −1 −1)

$$c = \sum a_i \bar{x}_i = 3*10.65 - 10.425 - 5.6 - 5.45 = 10.475$$

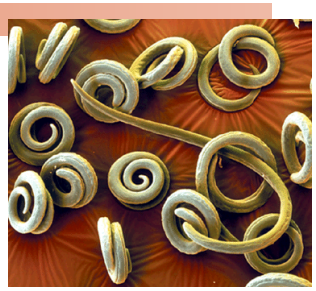
$$SE_c = s_p \sqrt{\sum \frac{a_i^2}{n_i}} = \sqrt{2.78} * \sqrt{\left(\frac{3^2}{4} + 3*\frac{(-1)^2}{4}\right)} \approx 2.9$$

$$t = c/SE_c = 10.5/2.9 \approx 3.6 \qquad \text{df} : N\text{-}I = 12$$

In Excel: TDIST(3.6,12,1) = *tdist(t, df, tails)* ≈ 0.002 (*p*-value).

Nematodes result in significantly shorter seedlings (alpha 1%).

ANOVA vs. contrasts in SPSS



- **ANOVA:** H_0: all μ_i are equal vs. H_a: not all μ_i are equal

ANOVA

SeedlingLength

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	100.647	3	33.549	12.080	.001
Within Groups	33.328	12	2.777		
Total	133.974	15			

➔ not all μ_i are equal

- **Planned comparison:**

H_0: $\mu_1 = 1/3\ (\mu_2 + \mu_3 + \mu_4)$ vs.

H_a: $\mu_1 > 1/3\ (\mu_2 + \mu_3 + \mu_4)$ → *one tailed*

Contrast coefficients: (+3 −1 −1 −1)

Contrast Coefficients

	NematodesLevel			
Contrast	0	1000	5000	10000
1	-3	1	1	1

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
SeedlingLength	Assume equal variances	1	-10.4750	2.88650	-3.629	12	.003
	Does not assume equal	1	-10.4750	3.34823	-3.129	4.139	.034

Nematodes result in significantly shorter seedlings (alpha 1%).

Multiple comparisons

Multiple comparison tests are variants on the two-sample *t*-test.

- They use the pooled standard deviation $s_p = \sqrt{\text{MSE}}$,
- the pooled degrees of freedom **DFE,**
- and they compensate for the multiple comparisons.

We compute the *t*-statistic for all pairs of means:

$$t_{ij} = \frac{\bar{x}_i - \bar{x}_j}{s_p\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}$$

A given test is significant (μ_i and μ_j significantly different), when

$$|t_{ij}| \geq t^{**} \text{ (df = DFE).}$$

The value of t^{**} depends on which procedure you choose to use.

The Bonferroni procedure

The **Bonferroni procedure** performs a number of pairwise comparisons with *t*-tests and then multiplies each *p*-value by the number of comparisons made. This ensures that the probability of making *one or more* false rejections among all comparisons made is no greater than the chosen significance level α.

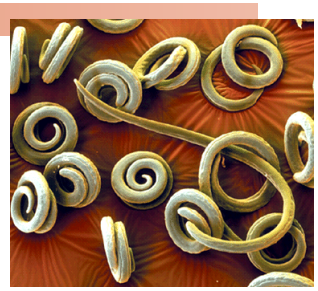
As a consequence, the higher the number of pair-wise comparisons you make, the more difficult it will be to show statistical significance for each test. But the chance of committing a type I error also increases with the number of tests made. The Bonferroni procedure lowers the working significance level of each test to compensate for the increased chance of type I errors among all tests performed.

Simultaneous confidence intervals

We can also calculate simultaneous level C **confidence intervals for all pair-wise differences ($\mu_i - \mu_j$)** between population means:

$$CI : (\bar{x}_i - \bar{x}_j) \pm t^{**} s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

- s_p is the pooled variance, MSE.
- t^{**} is the t critical with degrees of freedom DFE = $N - I$, adjusted for multiple, simultaneous comparisons (e.g., Bonferroni procedure).



SYSTAT

File contains variables: GROWTH NEMATODES$
Categorical values encountered during processing are:
NEMATODES$ (four levels): 10K, 1K, 5K, none

Dep Var: GROWTH N: 16 Multiple R: 0.867 Squared multiple R: 0.751

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
NEMATODES$	100.647	3	33.549	12.080	0.001
Error	33.328	12	2.777		

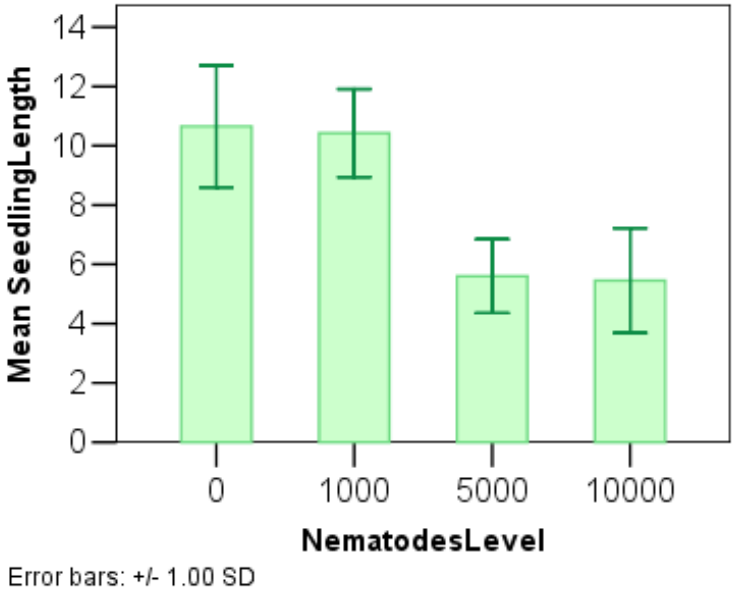
Post Hoc test of GROWTH **Using model MSE of 2.777 with 12 df.**

Matrix of pairwise mean differences:

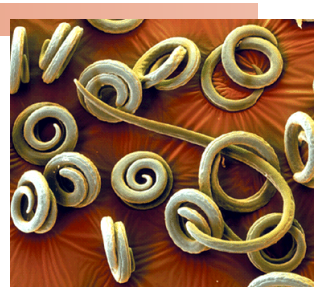
	1	2	3	4
1	0.000			
2	4.975	0.000		
3	0.150	−4.825	0.000	
4	5.200	0.225	5.050	0.000

Bonferroni Adjustment
Matrix of pairwise comparison probabilities:

	1	2	3	4
1	1.000			
2	0.007	1.000		
3	1.000	0.009	1.000	
4	0.005	1.000	0.006	1.000



SigmaStat—One-Way Analysis of Variance



Normality Test: Passed (P > 0.050)
Equal Variance Test: Passed (P = 0.807)

Group Name	N	Missing	Mean	Std dev	SEM
None	4	0	10.650	2.053	1.027
1K	4	0	10.425	1.486	0.743
5K	4	0	5.600	1.244	0.622
10K	4	0	5.450	1.771	0.886

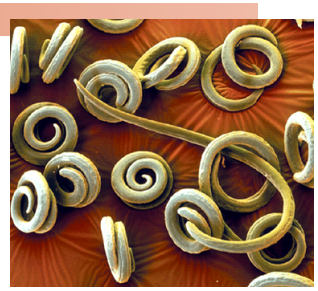
Source of variation	DF	SS	MS	F	P
Between groups	3	100.647	33.549	12.080	<0.001
Residual	12	33.328	2.777		
Total	15	133.974			

Power of performed test with alpha = 0.050: 0.992

All Pairwise Multiple Comparison Procedures (Bonferroni *t*-test):
Comparisons for factor: Nematodes

Comparison	Diff of means	*t*	P	P<0.050
None vs. 10K	5.200	4.413	0.005	Yes
None vs. 5K	5.050	4.285	0.006	Yes
None vs. 1K	0.225	0.191	1.000	No
1K vs. 10K	4.975	4.222	0.007	Yes
1K vs. 5K	4.825	4.095	0.009	Yes
5K vs. 10K	0.150	0.127	1.000	No

SPSS



ANOVA

SeedlingLength

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	100.647	3	33.549	12.080	.001
Within Groups	33.328	12	2.777		
Total	133.974	15			

Multiple Comparisons

Dependent Variable: SeedlingLength

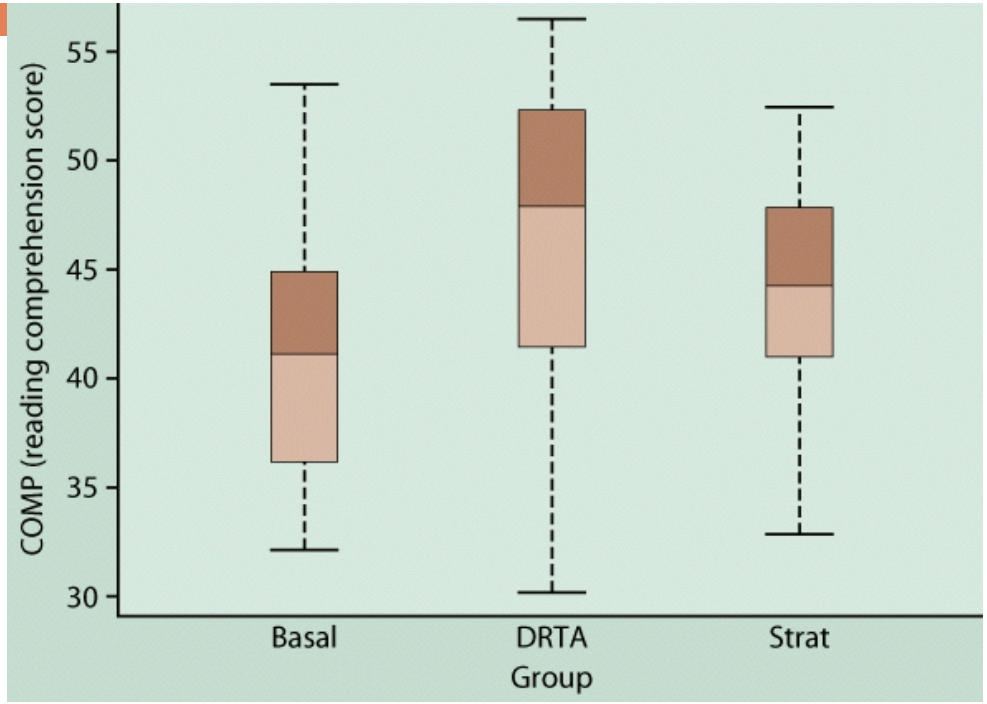
Bonferroni

(I) NematodesLevel	(J) NematodesLevel	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
0	1000	.22500	1.17841	1.000	-3.4901	3.9401
	5000	5.05000*	1.17841	.006	1.3349	8.7651
	10000	5.20000*	1.17841	.005	1.4849	8.9151
1000	0	-.22500	1.17841	1.000	-3.9401	3.4901
	5000	4.82500*	1.17841	.009	1.1099	8.5401
	10000	4.97500*	1.17841	.007	1.2599	8.6901
5000	0	-5.05000*	1.17841	.006	-8.7651	-1.3349
	1000	-4.82500*	1.17841	.009	-8.5401	-1.1099
	10000	.15000	1.17841	1.000	-3.5651	3.8651
10000	0	-5.20000*	1.17841	.005	-8.9151	-1.4849
	1000	-4.97500*	1.17841	.007	-8.6901	-1.2599
	5000	-.15000	1.17841	1.000	-3.8651	3.5651

*. The mean difference is significant at the .05 level.

Teaching methods

A study compares the reading comprehension ("COMP," a test score) of children randomly assigned to one of three teaching methods: basal, DRTA, and strategies.

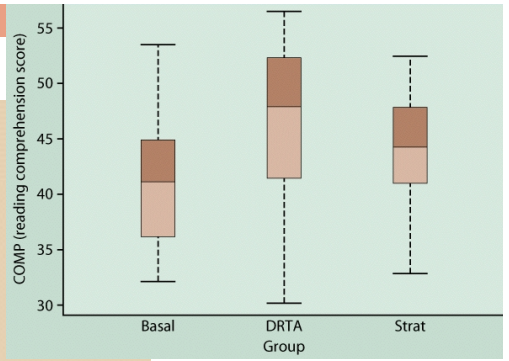


We test: H_0: $\mu_{Basal} = \mu_{DRTA} = \mu_{Strat}$ vs. H_a: H_0 not true

Dependent Variable: COMP

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	357.30303	178.65152	4.48	0.0152
Error	63	2511.68182	39.86797		
Corrected Total	65	2868.98485			

The ANOVA test is significant (α 5%): we have found evidence that the three methods do not all yield the same population mean reading comprehension score.



Bonferroni (Dunn) T tests for variable: COMP

Alpha= 0.05 Confidence= 0.95 df= 63 MSE= 39.86797

Critical Value of T= 2.45958

Minimum Significant Difference= 4.6825

Comparisons significant at the 0.05 level are indicated by '***'.

GROUP Comparison	Simultaneous Lower Confidence Limit	Difference Between Means	Simultaneous Upper Confidence Limit	
DRTA - Strat	-2.228	2.455	7.137	
DRTA - Basal	0.999	5.682	10.364	***
Basal - Strat	-7.910	-3.227	1.455	

What do you conclude?

The three methods do not yield the same results: We found evidence of a significant difference between DRTA and basal methods (DRTA gave better results on average), but the data gathered does not support the claim of a difference between the other methods (DRTA vs. strategies or basal vs. strategies).

Power

The power, or sensitivity, of a one-way ANOVA is the probability that the test will be able to detect a difference among the groups (i.e. reach statistical significance) when there really is a difference.

Estimate the power of your test while designing your experiment to select sample sizes appropriate to detect an amount of difference between means that you deem important.

- Too small a sample is a waste of experiment, but too large a sample is also a waste of resources.
- A power of at least 80% is often suggested.

Power computations

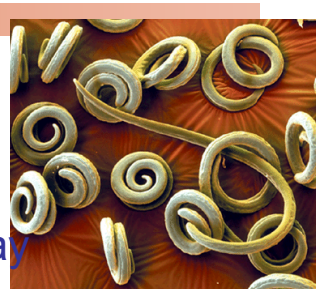
ANOVA power is affected by

- The significance level α
- The sample sizes and number of groups being compared
- The differences between group means μ_i
- The guessed population standard deviation

You need to decide what alternative H_a you would consider important, detect statistically for the means $\mu_{i,}$ and to guess the common standard deviation σ (from similar studies or preliminary work).

The power computations then require calculating a **non-centrality paramenter λ,** which follows the F distribution with DFG and DFE degrees of freedom to arrive at the power of the test.

Systat: Power analysis



If we anticipated a gradual decrease of seedling length for increasing amounts of nematodes in the pots and would consider gradual changes of 1 mm on average to be important enough to be reported in a scientific journal…

→

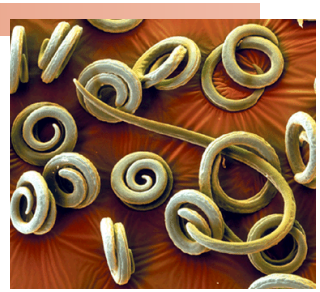
Alpha =	0.05
Model =	Oneway
Number of groups =	4
Within cell S.D. =	1.5 guessed
Mean(01) =	7.0
Mean(02) =	8.0
Mean(03) =	9.0
Mean(04) =	10.0
Effect Size =	0.745
Noncentrality parameter =	2.222 * sample size

…then we would reach a power of 80% or more when using six pots or more for each condition.
(Four pots per condition would only bring a power of 55%.)

→

SAMPLE SIZE (per cell)	POWER
3	0.373
4	0.551
5	0.695
6	0.802
7	0.876
8	0.925

Systat: Power analysis



If we anticipated that only large amounts of nematodes in the pots would lead to substantially shorter seedling lengths and would consider step changes of 4 mm on average an important effect…

→

Alpha = 0.05
Power = 0.80
Model = One-way
Number of groups = 4
Within cell S.D. = 1.7 guessed
Mean(01) = 6.0
Mean(02) = 6.0
Mean(03) = 10.0
Mean(04) = 10.0
Effect size = 1.176
Noncentrality parameter = 5.536 * sample size

SAMPLE SIZE (per cell)	POWER
2	0.394
3	0.766
4	0.931

Total sample size = 16

…then we would reach a power of 80% or more when using four pots or more in each group (three pots per condition might be close enough though).

→