Lecture 13

One-Way ANOVA

Ch 12

ANOVA=Analysis of Variance

- Recall:
  - A categorical variable = factor.
  - Its values =levels
- ANOVA in general studies the effect of categorical variables on a quantitative variable (response)
- One-Way = only one factor with several levels

- This is similar with testing if two population means are equal (except that we have more than two populations.

Examples:

Example 1: Numbers of days for healing a standard wound (in an animal) for several treatments.

Example 2: Wages of different ethnic groups in a company.

Example 3: Lifetimes of different brands of tires.

- If comparing means of two groups, ANOVA is *equivalent* to a 2-sample (two-sided) pooled t-test
- ANOVA allows for 3 or more groups.

Inference for One Way Anova

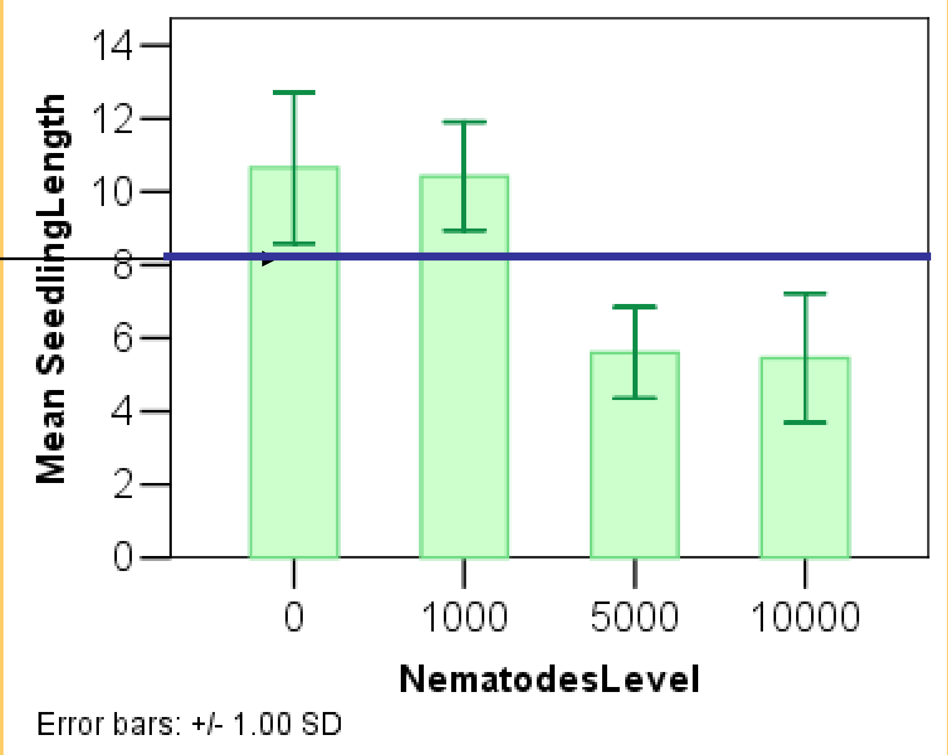
- We first examine the multiple populations or multiple treatments to test for overall statistical significance as evidence of any difference among the parameters we want to compare. ➔ANOVA F-test
- If that overall test showed statistical significance, then a detailed follow-up analysis is legitimate.
  - 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)



Error bars: +/- 1.00 SD

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

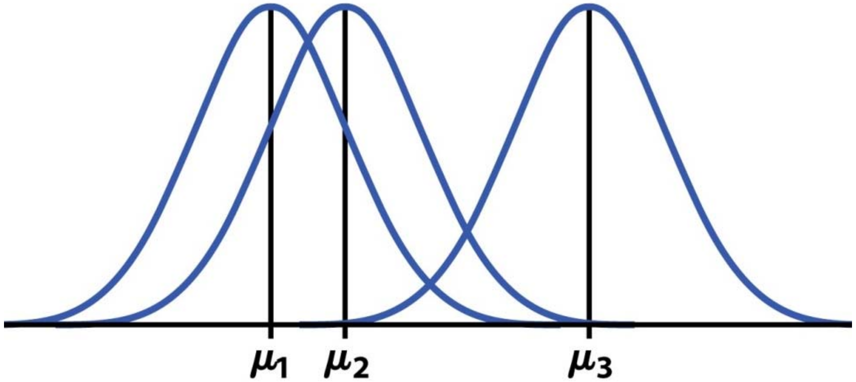
overall mean 8.03

The Model

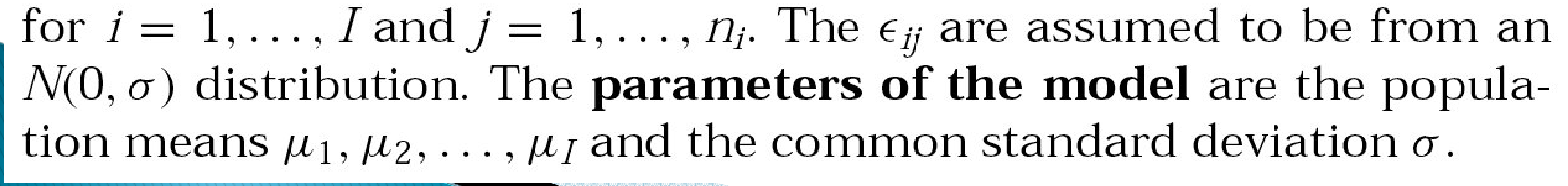
Random sampling always produces chance variations. Any "factor effect" would thus show up in our data as the factor-driven differences plus chance variations ("error"):

Data = **fit** ("factor/groups") + **residual** ("error")

The one-way ANOVA model analyses 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 F-test

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

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

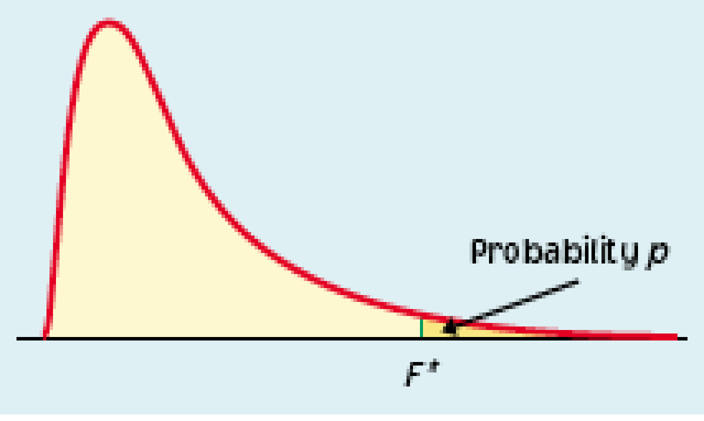
All *I* populations have the **same standard deviation *σ***, unknown.

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.

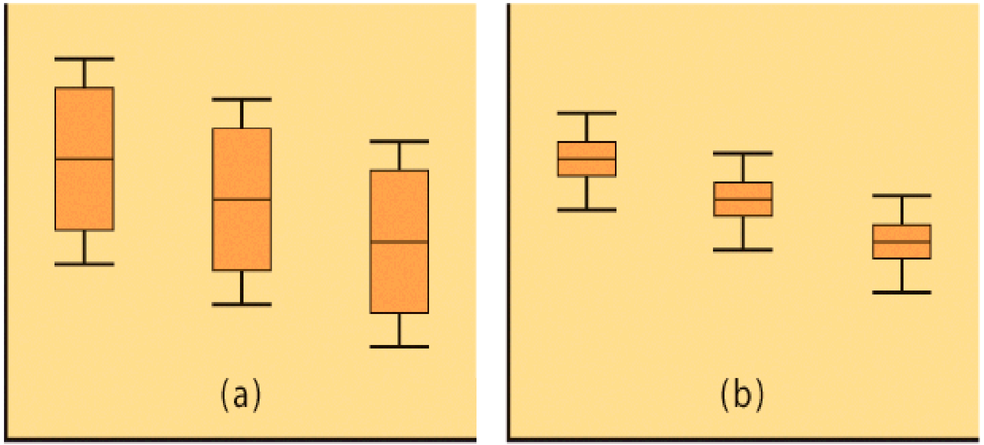


When H_0 is true, *F* has the **F distribution** with I − 1 (*numerator*) and N − I (*denominator*) degrees of freedom.

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

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

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

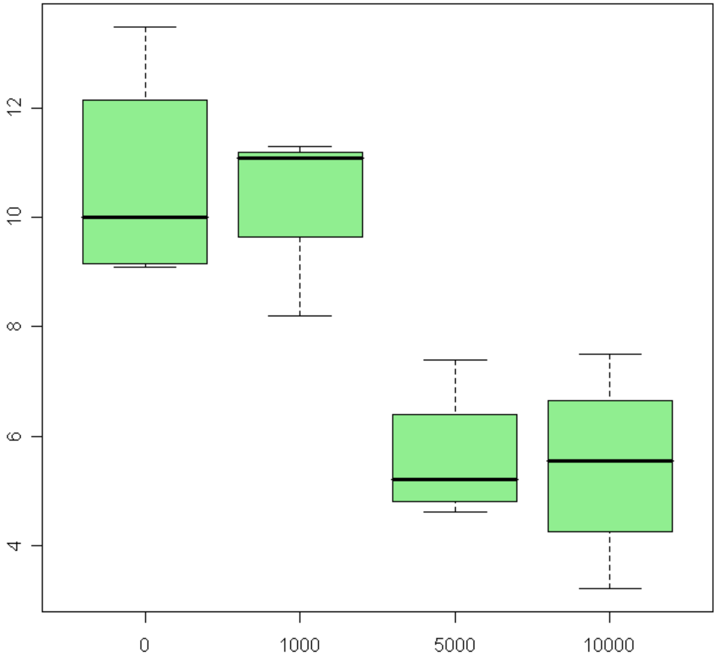
  Largest s_i = 2.053; smallest s_i = 1.244

- 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.



Smoking influence on sleep

A study of the effect of smoking classifies subjects as nonsmokers, moderate smokers, and heavy smokers. The investigators interview a random sample of 200 people in each group and ask “How many hours do you sleep on a typical night?”

1. Study design?

2. Hypotheses?

3. ANOVA assumptions?

4. Degrees of freedom?

The ANOVA table

Source of variation	Sum of squares SS	DF	Mean square MS	F	P value	F crit
Among or between “groups”	$\sum n_i(\bar{x}_i - \bar{x})^2$	I -1	SSG/DFG	MSG/MSE	Tail area above F	Value of F for α
Within groups or “error”	$\sum (n_i - 1)s_i^2$	N - I	SSE/DFE			
Total	SST=SSG+SSE $\sum (x_{ij} - \bar{x})^2$	N – 1				

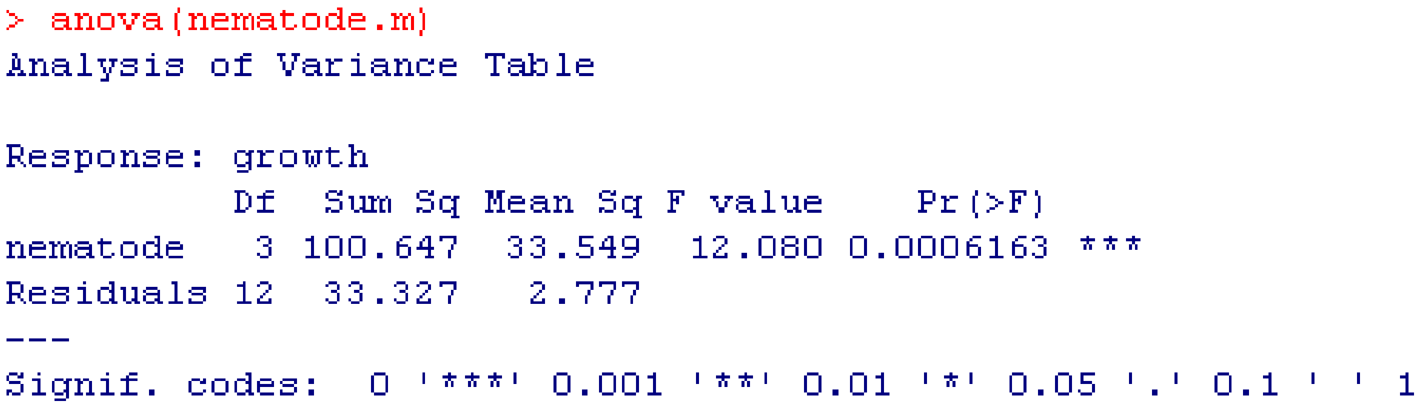
R^2 = SSG/SST
Coefficient of determination

$\sqrt{\text{MSE}} = s_p$
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.

Data (“Total”) = fit (“Groups”) + residual (“Error”)



```
> anova(nematode.m)
Analysis of Variance Table

Response: growth
          Df  Sum Sq Mean Sq F value    Pr(>F)    
nematode   3 100.647  33.549  12.080 0.0006163 ***
Residuals 12  33.327   2.777                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

(or just look at the p-value directly)

Thus, the test is significant at α 5% ➔ Not all mean seedling lengths are the same; nematode amount is an influential factor.

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.

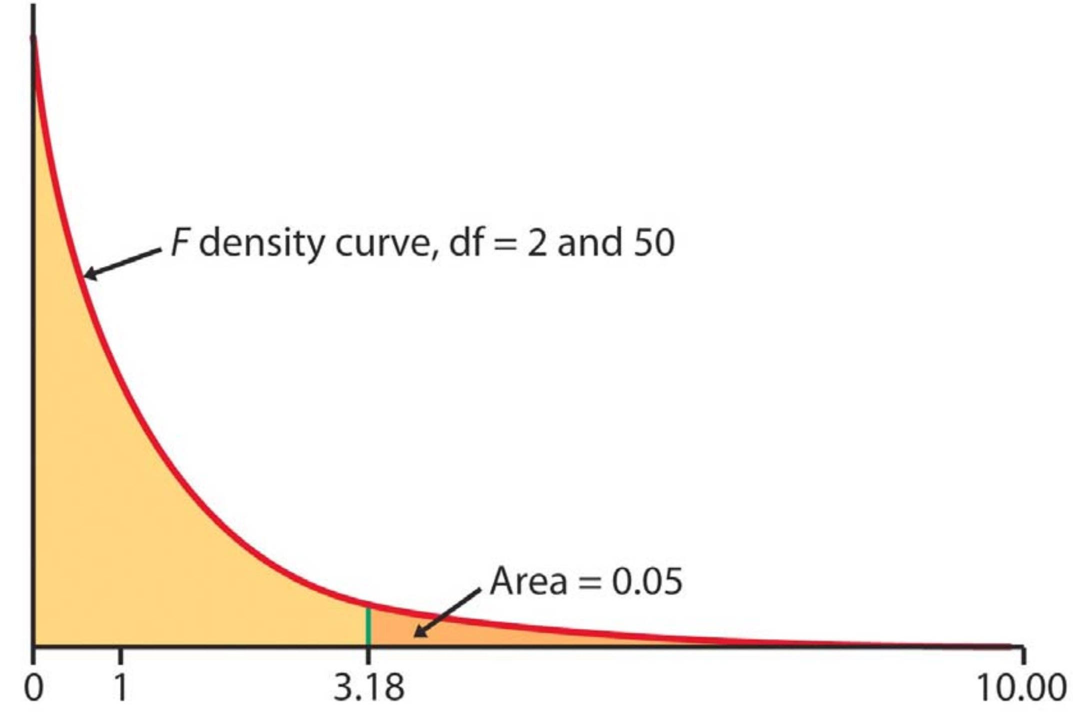


Table E *F* distribution critical values

$df_{num} = I - 1$

For df: 5,4

		Degrees of freedom in the numerator							
Degrees of freedom in the denominator	*p*	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

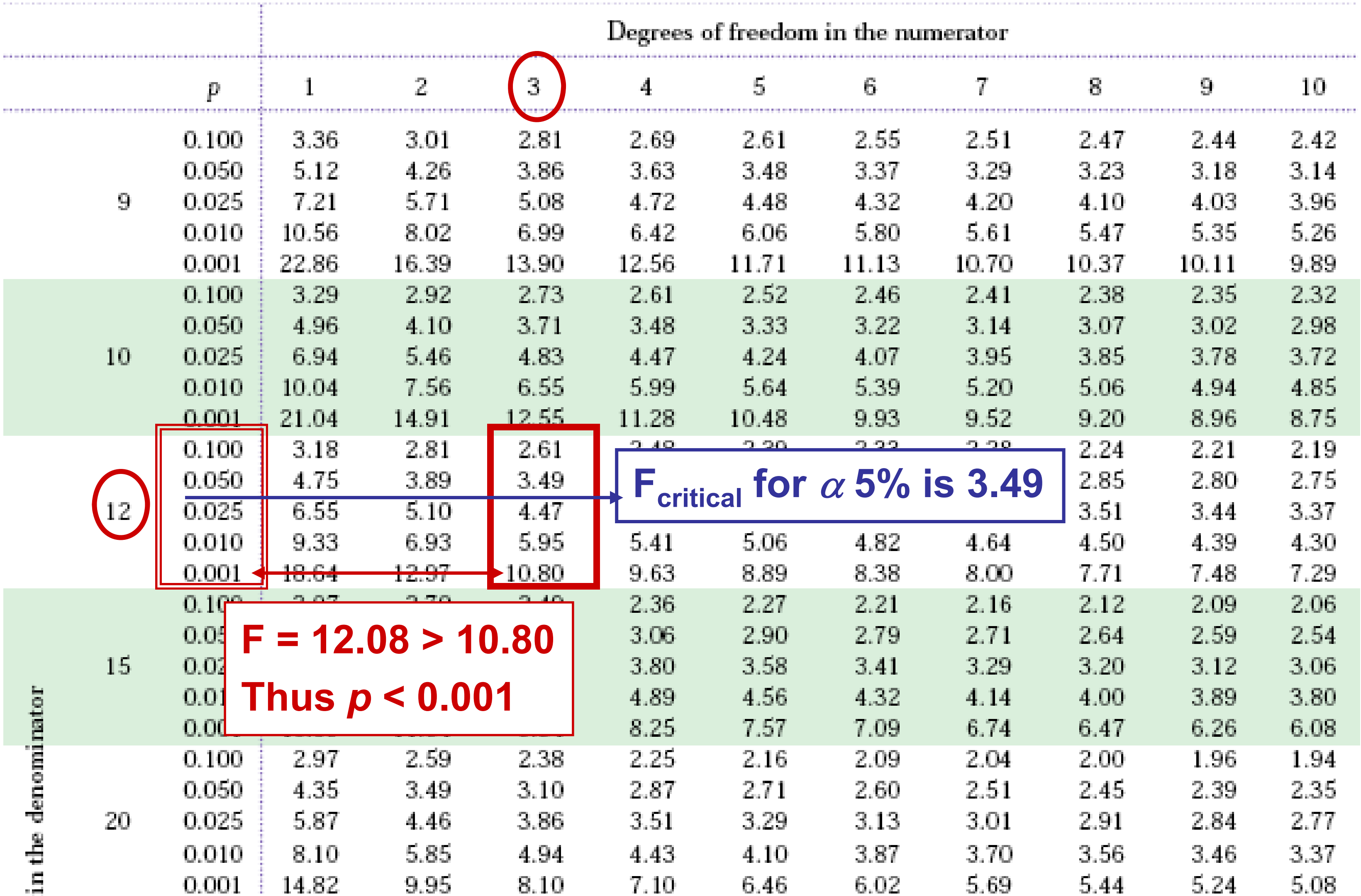
p

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	2.48	2.39	2.33	2.28	2.24	2.21	2.19
	0.050	4.75	3.89	3.49	[illegible]	[illegible]	[illegible]	[illegible]	2.85	2.80	2.75
12	0.025	6.55	5.10	4.47	[illegible]	[illegible]	[illegible]	[illegible]	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	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
	0.05	[illegible]	[illegible]	[illegible]	3.06	2.90	2.79	2.71	2.64	2.59	2.54
15	0.02	[illegible]	[illegible]	[illegible]	3.80	3.58	3.41	3.29	3.20	3.12	3.06
	0.01	[illegible]	[illegible]	[illegible]	4.89	4.56	4.32	4.14	4.00	3.89	3.80
	0.00	[illegible]	[illegible]	[illegible]	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	17.769					

TABLE D F distribution critical values

$df_{num} = I - 1$

Degrees of freedom in the denominator		Degrees of freedom in the numerator							
	p	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

$df_{den} = N - I$

F

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.

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

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). SSG is the sum of squares for error.

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

Note: Two sample *t*–test and ANOVA

A two sample *t*–test assuming equal variance and 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.

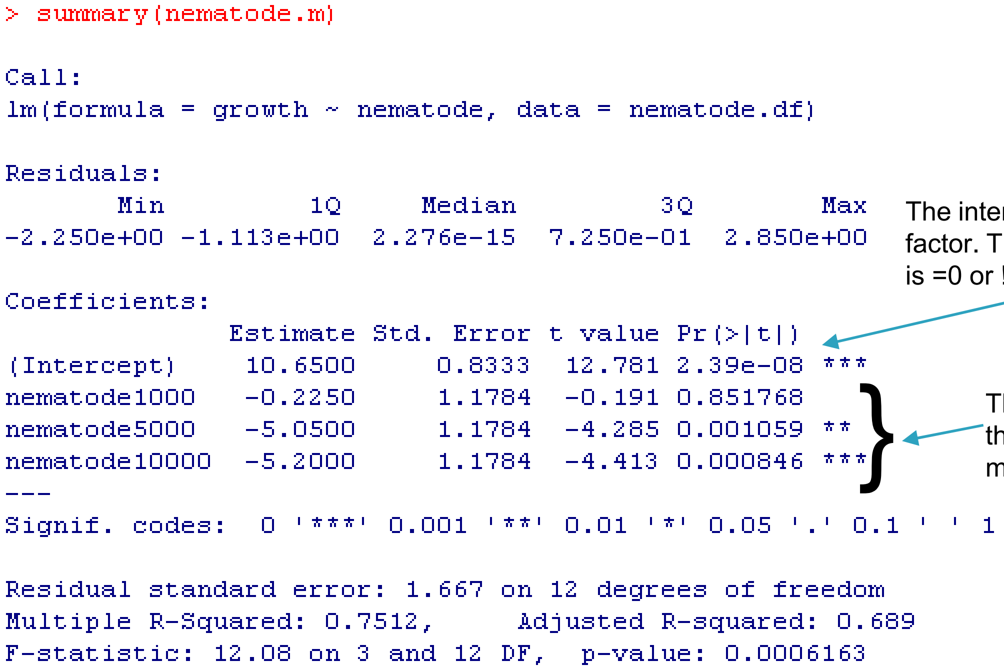
You have calculated a p-value for your ANOVA test. Now what?

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 (boxplots).
- There are several tests of statistical significance designed specifically for multiple tests. You can choose *apriori* **contrasts**, or *aposteriori* **multiple comparisons**.
- You can find the confidence interval for each mean μ_i shown to be significantly different from the others.

R–first indication

- The *summary()* produces the following



```
> summary(nematode.m)

Call:
lm(formula = growth ~ nematode, data = nematode.df)

Residuals:
       Min         1Q     Median         3Q        Max
-2.250e+00 -1.113e+00  2.276e-15  7.250e-01  2.850e+00

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.6500     0.8333  12.781 2.39e-08 ***
nematode1000   -0.2250     1.1784  -0.191 0.851768
nematode5000   -5.0500     1.1784  -4.285 0.001059 **
nematode10000  -5.2000     1.1784  -4.413 0.000846 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.667 on 12 degrees of freedom
Multiple R-Squared: 0.7512,     Adjusted R-squared: 0.689
F-statistic: 12.08 on 3 and 12 DF,  p-value: 0.0006163
```

The intercept always represents the first level of the factor. The test is checking if the mean in this first group is =0 or !=0

The remaining tests for each level are checking if the mean for the particular level is equal to the mean of the first group.

However, these tests are each taken separately. If we want to talk about comparing all the means together we need to make some adjustments.

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:

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

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

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

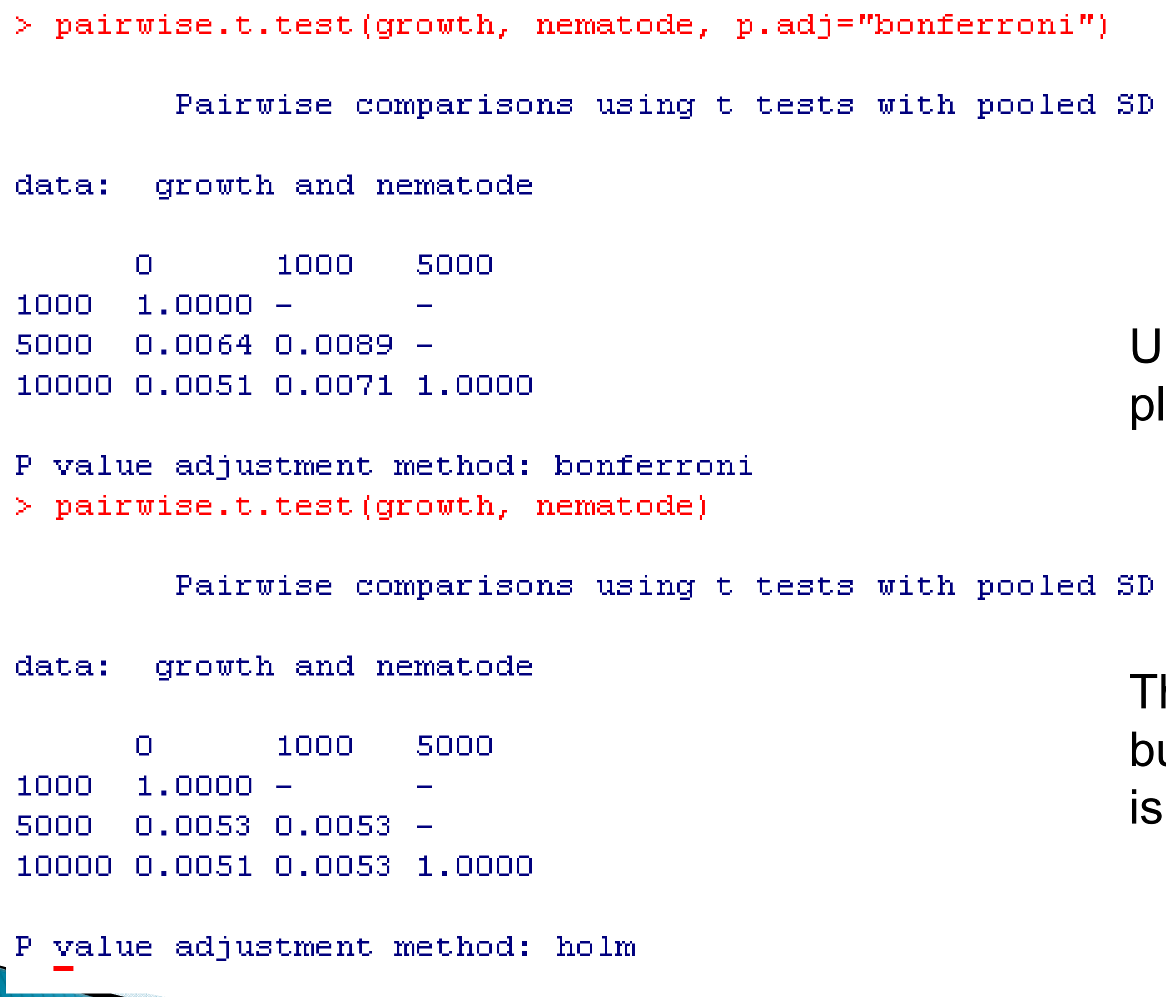
The value of t^{**} depends on which procedure you choose to use.

The Bonferroni procedure

The Bonferroni procedure is the simplest possible approach to the problem of performing many pair-wise tests simultaneously. It multiplies each *p*-value by the number of comparisons made. This ensures that the probability of making *any* false rejection among all comparisons made is no greater than the chosen significance level α.

> Bonferroni procedure tends to be conservative, in the sense that if the Bonferrony procedure tells that there is a significant difference then probably there is indeed one.

R Pairwise test procedure:



```
> pairwise.t.test(growth, nematode, p.adj="bonferroni")

        Pairwise comparisons using t tests with pooled SD

data:  growth and nematode

      0      1000   5000  
1000  1.0000 -      -     
5000  0.0064 0.0089 -     
10000 0.0051 0.0071 1.0000

P value adjustment method: bonferroni
> pairwise.t.test(growth, nematode)

        Pairwise comparisons using t tests with pooled SD

data:  growth and nematode

      0      1000   5000  
1000  1.0000 -      -     
5000  0.0053 0.0053 -     
10000 0.0051 0.0053 1.0000

P value adjustment method: holm
```

Up is Bonferroni for the plants example.

The default is not Bonferroni but a procedure by Holm that is less conservative (down).

Contrasts

- This is a very early idea in statistics back from the time when everything had to be done by hand. The idea is: what if I want to test a particular linear combination of the factor levels.
- It turns out that you can do this easily by hand

- We will not discuss this in detail other than to note that the summary() function gives a so called treatment contrast (all means are compared with the first mean) and to provide the following example.

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 *sp*. The alternative hypothesis can be one- or two-sided.

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

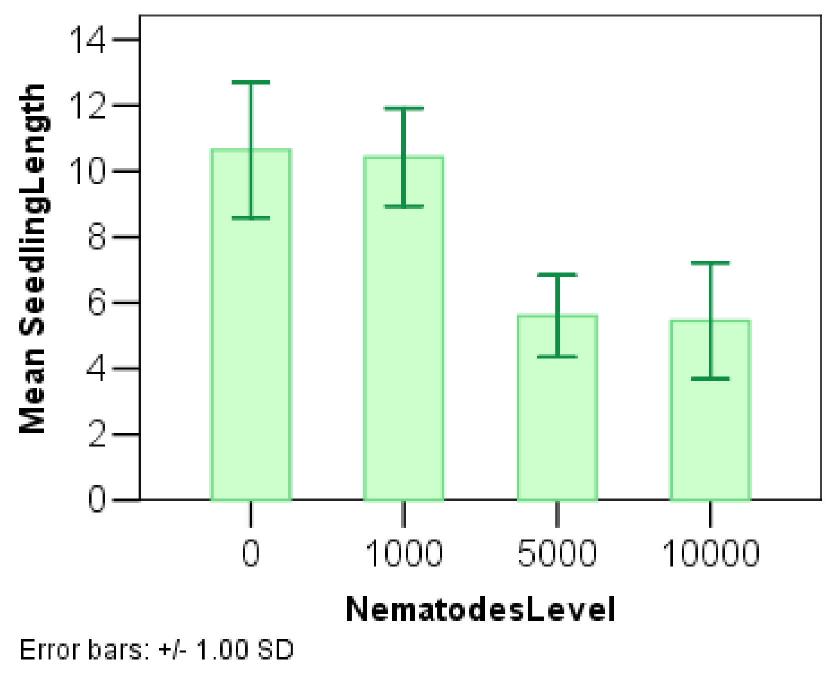
$$c \pm t^* SE_c$$

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

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 nematode 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:

	$\bar{x}_i$	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

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: (+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 R: 1-pt(3.6,12) ≈ 0.002 (*p*-value).

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

Relaxing the variance assumption

- The traditional ANOVA assumes that the variances of all the groups are equal.

- However, there exist an alternative procedure that does not assume this. It is due to Welch and implemented in R in the function:
  - Oneway.test()

- Please follow the R textbook on pages 117–120 for this type of ANOVA as well as nice graphical representations of the results.