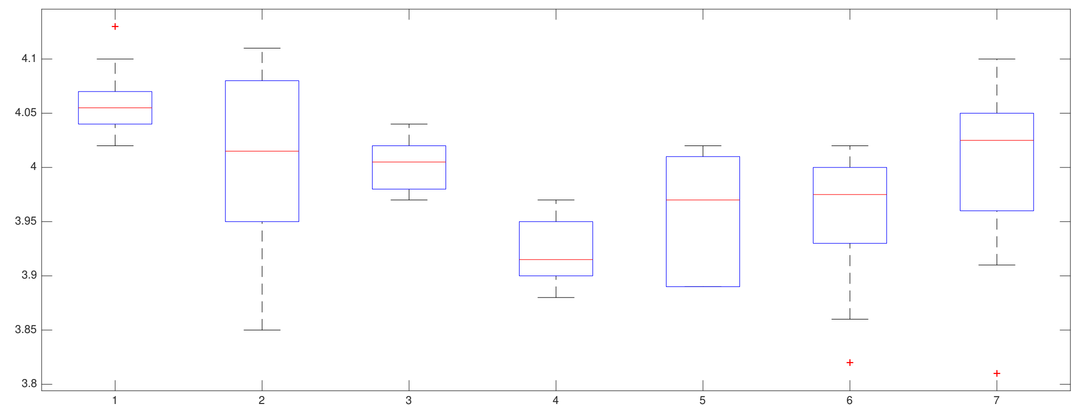
Slides 12: ANOVA one-way layout

- Normal theory model
- Maximum likelihood estimates
- F-tests
- Simultaneous confidence interval
- Bonferroni and Tukey



Suppose the expectation of the response variable is a function of a single main factor having I different levels so that

$$Y(i) = \mu_i + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma), \quad i = 1, \ldots, I.$$

We will use a representation $\mu_i = \mu + \alpha_i$, where $\alpha_1 + \ldots + \alpha_I = 0$,

$\mu = \frac{\mu_1 + \ldots + \mu_I}{I}$ is the overall mean

$\alpha_i = \mu_i - \mu$ is the effect of the main factor at the level i

Given I independent random samples of the same size n

$$(y_{i1}, \ldots, y_{in}), \quad i = 1, \ldots, I$$

we want to develop a test of

$$H_0 : \mu_1 = \ldots = \mu_I, \text{ against } H_1 : \mu_u \neq \mu_v \text{ for some } (u, v).$$

In terms of I different treatments in a comparison study, this null hypothesis claims that the compared treatments have the same effect .

Data: each of $I = 7$ labs made $n = 10$ measurements of chlorpheniramine maleate in tablets with a nominal dosage of 4 mg. See seven boxplots on the first slide.

Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
4.13	3.86	4.00	3.88	4.02	4.02	4.00
4.07	3.85	4.02	3.88	3.95	3.86	4.02
4.04	4.08	4.01	3.91	4.02	3.96	4.03
4.07	4.11	4.01	3.95	3.89	3.97	4.04
4.05	4.08	4.04	3.92	3.91	4.00	4.10
4.04	4.01	3.99	3.97	4.01	3.82	3.81
4.02	4.02	4.03	3.92	3.89	3.98	3.91
4.06	4.04	3.97	3.9	3.89	3.99	3.96
4.10	3.97	3.98	3.97	3.99	4.02	4.05
4.04	3.95	3.98	3.90	4.00	3.93	4.06

Ordered means

Lab i	1	3	7	2	5	6	4
Mean $\hat{\mu}_i$	4.062	4.003	3.998	3.997	3.957	3.955	3.920

Question. Are the observed differences between sample means statistically significant?

With $N = I \cdot n$ independent random variables

$$Y_{ik} = \mu + \alpha_i + \epsilon_{ik}, \quad \epsilon_{ik} \sim \mathrm{N}(0, \sigma),$$

the maximum likelihood approach gives the following point estimates

$$\hat{\mu} = \bar{y}_{..}, \quad \hat{\mu}_i = \bar{y}_{i.}, \quad \hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..},$$

expressed in terms of the sample means

$$\bar{y}_{i.} = \tfrac{1}{n} \sum_k y_{ik}, \quad \bar{y}_{..} = \tfrac{1}{I} \sum_i \bar{y}_{i.} = \tfrac{1}{N} \sum_i \sum_k y_{ik}.$$

The observed responses can be represented as

$$y_{ik} = \hat{\mu} + \hat{\alpha}_i + \hat{\epsilon}_{ik}, \quad \hat{\epsilon}_{ik} = y_{ik} - \bar{y}_{i.},$$

where $\hat{\epsilon}_{ik}$ are the so-called residuals.

Question. What is the sum of all $\hat{\alpha}_i$? What is the total sum of residuals?

The ANOVA tests are built around the following observation:

$$\boxed{\text{Decomposition of the total sum of squares: } SS_{\mathrm{T}} = SS_{\mathrm{A}} + SS_{\mathrm{E}}}$$

saying that the total variation in response values is the sum of the between-group variation and the within-group variation.

$SS_{\mathrm{T}} = \sum_i \sum_k (y_{ik} - \bar{y}_{..})^2$ is the total sum of squares,

$SS_{\mathrm{A}} = n \sum_i \hat{\alpha}_i^2$ is the factor A sum of squares, $\mathrm{df}_{\mathrm{A}} = I - 1$,

$SS_{\mathrm{E}} = \sum_i \sum_k \hat{\epsilon}_{ik}^2$ is the error sum of squares, $\mathrm{df}_{\mathrm{E}} = I \cdot (n-1)$.

Define two mean squares

$$MS_{\mathrm{A}} = \frac{SS_{\mathrm{A}}}{\mathrm{df}_{\mathrm{A}}}, \quad MS_{\mathrm{E}} = \frac{SS_{\mathrm{E}}}{\mathrm{df}_{\mathrm{E}}}.$$

where $\mathrm{df}_{\mathrm{A}} = I - 1$ is the number of degrees of freedom in SS_{A}, and $\mathrm{df}_{\mathrm{E}} = I \cdot (n-1)$ is the number of degrees of freedom in SS_{E}.

Question. What is the number of degrees of freedom df_{T} in SS_{T}?

If treated as random variables, the mean squares lead the following formulas for the expected values

$$\mathrm{E}(MS_{\mathrm{A}}) = \sigma^2 + \tfrac{n}{I-1}\sum_i \alpha_i^2, \qquad \mathrm{E}(MS_{\mathrm{E}}) = \sigma^2,$$

which suggest looking for the ratio between the two mean squares

$$F = \frac{MS_{\mathrm{A}}}{MS_{\mathrm{E}}}$$

to find an evidence against the null hypothesis

$$H_0 : \alpha_1 = \ldots = \alpha_I = 0.$$

F-test: reject H_0 for large values of F based on the null distribution

$$F \overset{H_0}{\sim} F_{n_1,n_2}, \text{ where } n_1 = I - 1,\ n_2 = I(n-1).$$

F_{n_1,n_2} is called F-distribution with degrees of freedom (n_1, n_2).

It is the distribution for the ratio $\frac{X_1/n_1}{X_2/n_2} \sim F_{n_1,n_2}$,
where $X_1 \sim \chi^2_{n_1}$ and $X_2 \sim \chi^2_{n_2}$ are independent random variables.

The pooled sample variance is an unbiased estimate of σ^2.

$$s_{\mathrm{p}}^2 = MS_{\mathrm{E}} = \tfrac{1}{I}\sum_{i=1}^{I}\left(\tfrac{1}{n-1}\sum_{k=1}^{n}(y_{ik}-\bar{y}_{i.})^2\right)$$

Example: seven labs

The normal probability plot of residuals $\hat{\epsilon}_{ik}$ supports the normality assumption. Noise size σ is estimated by $s_{\mathrm{p}} = \sqrt{0.0037} = 0.061$. One-way Anova table

Source	df	SS	MS	F	P
Labs	6	.125	.0210	5.66	.0001
Error	63	.231	.0037		
Total	69	.356			

Conclusion: the largest pairwise difference (1 – 4) is significant.

Question. Which of the $\binom{7}{2} = 21$ pairwise differences are significant?

A naiv approach to find significant pairwise differences is to apply a 95% confidence interval for two independent samples (u, v)

$$I_{\mu_u - \mu_v} = (\bar{y}_{u.} - \bar{y}_{v.}) \pm t_{63}(0.025) \cdot \frac{s_{\mathrm{p}}}{\sqrt{5}} = (\bar{y}_{u.} - \bar{y}_{v.}) \pm 0.055,$$

where $t_{63}(0.025) = 2.00$.

This confidence interval formula detects 9 significant differences:

u–v	1–4	1–6	1–5	3–4	7–4	2–4	1–2	1–7	1–3
$\hat{\mu}_u$-$\hat{\mu}_v$	0.142	0.107	0.105	0.083	0.078	0.077	0.065	0.064	0.059

For all other pairs, $|\hat{\mu}_u - \hat{\mu}_v| < 0.055$ and zero is covered by $I_{\mu_u - \mu_v}$.

However, there exists the so called multiple comparison problem.

The above confidence interval formula is aimed at a single difference, and may produce false discoveries. What we need instead, is a simultaneous confidence interval formula taking care of all $c = 21$ pairwise comparisons.

Question. Why do we use df$=$ 63 for the pairwise confidence interval $I_{\mu_u - \mu_v}$?

Think of a statistical test repeatedly applied to c independent samples of size n. The overall result is positive if we get at least one positive result among these c tests.

Bonferroni's correction: to ensure the overall significance level α, each single test is performed at significance level $\alpha_c = \frac{\alpha}{c}$.

Indeed, assuming H_0 is true, the number of positive results among c tests is $X \sim \text{Bin}(c, \alpha_c)$. Thus for small values of α_c,

$$\text{P}(X \geq 1 | H_0) = 1 - (1 - \alpha_c)^c \approx c\alpha_c = \alpha.$$

Bonferroni's $100(1-\alpha)\%$ simultaneous confidence interval for $c = \binom{I}{2}$ pairwise differences

$$B_{\mu_u - \mu_v} = \bar{y}_{u.} - \bar{y}_{v.} \pm t_{\text{df}}(\tfrac{\alpha_c}{2}) \cdot s_{\text{p}} \sqrt{\tfrac{2}{n}}, \quad 1 \leq u < v \leq I.$$

where $\text{df} = I(n-1)$ and $\alpha_c = \frac{2\alpha}{I(I-1)}$.

Question. Are pairwise differences $\mu_u - \mu_v$ independent ?

Pairwise differences $\delta_{u,v} = \mu_u - \mu_v$ are not independent. For example

$$\delta_{1,2} + \delta_{2,3} = \delta_{1,3}$$

To take account of linear dependence between $\delta_{u,v}$, consider

$$Z_i = \bar{Y}_{i\cdot} - \mu_i \sim \mathrm{N}(0, \tfrac{\sigma}{\sqrt{n}}), \quad i = 1, \ldots, I$$

independent and identically distributed random variables. The range

$$R = \max\{Z_1, \ldots, Z_I\} - \min\{Z_1, \ldots, Z_I\}$$

gives the largest pairwise difference between the components of the vector $(Z_1, \ldots, Z_I)$. The corresponding normalised range has a distribution that is free from the parameter σ

$$\frac{R}{S_{\mathrm{p}}/\sqrt{n}} \sim \mathrm{SR}(I, \mathrm{df}), \quad \mathrm{df} = I(n-1).$$

The so-called studentised range distribution SR has two parameters: the number of samples and the number of df used in the variance estimate s_{p}^2.

Tukey's $100(1-\alpha)\%$ simultaneous confidence interval is built using an appropriate quantile $q_{I,\mathrm{df}}(\alpha)$ of the studentised range distribution

$$T_{\mu_u-\mu_v} = \bar{y}_{u\cdot} - \bar{y}_{v\cdot} \pm q_{I,\mathrm{df}}(\alpha) \cdot \frac{s_{\mathrm{p}}}{\sqrt{n}}$$

Bonferroni method gives slightly wider intervals compared to the Tukey method.

Example: seven labs

Bonferroni's 95% (using $t_{63}(0.0012) = 3.17$)

$$B_{\mu_u-\mu_v} = (\bar{y}_{u\cdot} - \bar{y}_{v\cdot}) \pm t_{63}(\tfrac{.025}{21}) \cdot \tfrac{s_{\mathrm{p}}}{\sqrt{5}} = (\bar{y}_{u\cdot} - \bar{y}_{v\cdot}) \pm 0.086,$$

detects 3 significant differences between labs (1,4), (1,5), (1,6).

Tukey's 95%

$$T_{\mu_u-\mu_v} = \bar{y}_{u\cdot} - \bar{y}_{v\cdot} \pm q_{7,63}(0.05) \cdot \tfrac{0.061}{\sqrt{10}} = \bar{y}_{u\cdot} - \bar{y}_{v\cdot} \pm 0.083,$$

brings four significant pairwise differences: (1,4), (1,5), (1,6), (3,4).