Serik Sagitov, Chalmers and GU, January 29, 2016

Chapter 12. Analysis of variance

Chapter 11:	$I = 2$ samples	independent samples	paired samples
Chapter 12:	$I \geq 3$ samples of equal size J	one-way layout	two-way layout

1 One-way layout

Consider I independent IID samples $(Y_{11}, \ldots, Y_{1J}), \ldots, (Y_{I1}, \ldots, Y_{IJ})$ measuring I treatment results. We have one main factor (factor A having I levels) as the principle course of variation in the data. The goal is to test

H_0: all I treatments have the same effect, vs H_1: there are systematic differences.

Example (seven labs)

Data: 70 measurements of chlorpheniramine maleate in tablets with a nominal dosage of 4 mg. Seven labs made ten measurements each: $I = 7$, $J = 10$.

Lab	1	3	7	2	5	6	4
Mean	4.062	4.003	3.998	3.997	3.957	3.955	3.920

Normal theory model

Normally distributed observations $Y_{ij} \sim \mathrm{N}(\mu_i, \sigma^2)$ with equal variances (compare to the t-tests).

$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \qquad \sum_i \alpha_i = 0$

obs = overall mean + differential effect + noise, $\epsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$

Sample means as maximum likelihood estimates

$\bar{Y}_{i.} = \frac{1}{J}\sum_j Y_{ij}, \quad \bar{Y}_{..} = \frac{1}{I}\sum_i Y_{i.} = \frac{1}{IJ}\sum_i\sum_j Y_{ij}, \qquad \hat{\mu} = \bar{Y}_{..}, \quad \hat{\mu}_i = \bar{Y}_{i.}, \quad \hat{\alpha}_i = \bar{Y}_{i.} - \bar{Y}_{..}$

$Y_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\epsilon}_{ij}$, where $\sum_i \hat{\alpha}_i = 0$ and $\hat{\epsilon}_{ij} = Y_{ij} - \bar{Y}_{i.}$ are the so-called residuals

Decomposition of the total sum of squares: $\mathrm{SS_T} = \mathrm{SS_A} + \mathrm{SS_E}$.

$\mathrm{SS_T} = \sum_i\sum_j (Y_{ij} - \bar{Y}_{..})^2$ total sum of squares for the pooled sample with $\mathrm{df_T} = IJ - 1$,

$\mathrm{SS_A} = J\sum_i \hat{\alpha}_i^2$ factor A sum of squares (between-group variation) with $\mathrm{df_A} = I - 1$,

$\mathrm{SS_E} = \sum_i\sum_j \hat{\epsilon}_{ij}^2$ error sum of squares (within-group variation) with $\mathrm{df_E} = I(J - 1)$

Mean squares and their expected values

$\mathrm{MS_A} = \frac{\mathrm{SS_A}}{\mathrm{df_A}}, \quad \mathrm{E}(\mathrm{MS_A}) = \sigma^2 + \frac{J}{I-1}\sum_i \alpha_i^2, \qquad \mathrm{MS_E} = \frac{\mathrm{SS_E}}{\mathrm{df_E}}, \quad \mathrm{E}(\mathrm{MS_E}) = \sigma^2$

Pooled sample variance $s_p^2 = \mathrm{MS_E} = \frac{1}{I(J-1)}\sum_i\sum_j (Y_{ij} - \bar{Y}_{i.})^2$ is an unbiased estimate of σ^2.

One-way F-test

Use $F = \frac{\mathrm{MS_A}}{\mathrm{MS_E}}$ as test statistic for $H_0 : \alpha_1 = \ldots = \alpha_I = 0$ against $H_1 : \alpha_u \neq \alpha_v$ for some (u, v).
Reject H_0 for large values of F, since $\mathrm{E}_{H_0}(\mathrm{MS_A}) = \sigma^2$ and $\mathrm{E}_{H_1}(\mathrm{MS_A}) = \sigma^2 + \frac{J}{I-1}\sum_i \alpha_i^2 > \sigma^2$.

Null distribution $F \sim F_{n_1,n_2}$ with degrees of freedom $n_1 = I - 1$ and $n_2 = I(J - 1)$.

If $X_1 \sim \chi^2_{n_1}$ and $X_2 \sim \chi^2_{n_2}$ are two independent random variables, then $\frac{X_1/n_1}{X_2/n_2} \sim F_{n_1,n_2}$.

Example (seven labs)
The normal probability plot of residuals $\hat{\epsilon}_{ij}$ supports the normality assumption. Noise size σ is estimated by $s_p = \sqrt{0.0037} = 0.061$.

One-way Anova table

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

Which of the $\binom{7}{2} = 21$ pairwise differences are significant? Using the 95% CI for a pair of independent samples $(\alpha_u - \alpha_v)$: $(\bar{Y}_{u.} - \bar{Y}_{v.}) \pm t_{63}(0.025) \cdot \frac{s_p}{\sqrt{5}} = (\bar{Y}_{u.} - \bar{Y}_{v.}) \pm 0.055$,
where $t_{63}(0.025) = 2.00$, we find 9 significant differences:

Labs	1–4	1–6	1–5	3–4	7–4	2–4	1–2	1–7	1–3	5–4
Diff	0.142	0.107	0.105	0.083	0.078	0.077	0.065	0.064	0.059	0.047

The multiple comparison problem: the above CI formula is aimed at a single difference, and may produce false discoveries. We need a simultaneous CI formula for all 21 pairwise comparisons.

Bonferroni method
Think of k independent replications of a statistical test. The overall result is positive if we get at least one positive result among these k tests. The overall significance level α is obtained, if each single test is performed at significance level α/k:

indeed, assuming the null hypothesis is true, the number of positive results is $X \sim \text{Bin}(k, \frac{\alpha}{k})$, and due to independence $\text{P}(X \geq 1 | H_0) = 1 - (1 - \frac{\alpha}{k})^k \approx \alpha$ for small values of α.

Warning: $\binom{I}{2}$ pairwise Anova comparisons are not independent as required by Bonferroni method

Simultaneuos $100(1-\alpha)\%$ CI formula for $\binom{I}{2}$ pairwise differences $(\alpha_u - \alpha_v)$:

$$(\bar{Y}_{u.} - \bar{Y}_{v.}) \pm t_{I(J-1)}(\tfrac{\alpha}{I(I-1)}) \cdot s_p \sqrt{\tfrac{2}{J}}$$

Flexibility of the formula: works for different sample sizes as well after replacing $\sqrt{\frac{2}{J}}$ by $\sqrt{\frac{1}{J_u} + \frac{1}{J_v}}$.

Example (seven labs)
The Bonferroni simultaneuos 95% CI for $(\alpha_u - \alpha_v)$

$$(\bar{Y}_{u.} - \bar{Y}_{v.}) \pm t_{63}(\tfrac{.05}{42}) \cdot \tfrac{s_p}{\sqrt{5}} = (\bar{Y}_{u.} - \bar{Y}_{v.}) \pm 0.086,$$

where $t_{63}(0.0012) = 3.17$, detects 3 significant differences between labs (1,4), (1,5), (1,6).

Tukey method
If I independent samples $(Y_{i1}, \ldots, Y_{iJ})$ taken from $\text{N}(\mu_i, \sigma^2)$ have the same size J, then the sample means $\bar{Y}_{i.} \sim \text{N}(\mu_i, \frac{\sigma^2}{J})$ are independent and

$$\frac{\sqrt{J}}{s_p} \max_{u,v} |\bar{Y}_{u.} - \bar{Y}_{v.} - (\mu_u - \mu_v)| \sim \text{SR}(I, I(J-1))$$

Studentized range distribution $\text{SR}(k, \text{df})$ has two parameters: the number of samples k, and the number of degrees of freedom used in the variance estimate s_p^2.

$$\boxed{\text{Tukey's 95\% simultaneuos CI} = (\bar{Y}_{u.} - \bar{Y}_{v.}) \pm q_{I,I(J-1)}(0.05) \cdot \tfrac{s_p}{\sqrt{J}}}$$

Example (seven labs)
Using $q_{7,60}(0.05) = 4.31$ from the SR-distribution table, we find four significant pairwise differences: (1,4), (1,5), (1,6), (3,4), since $(\bar{Y}_{u.} - \bar{Y}_{v.}) \pm q_{7,63}(0.05) \cdot \frac{0.061}{\sqrt{10}} = (\bar{Y}_{u.} - \bar{Y}_{v.}) \pm 0.083$.

Kruskal-Wallis test
A nonparametric test, without assuming normality, for
H_0: all observations are equal in distribution, no treatment effects.
Extending the idea of the rank-sum test, consider the pooled sample of size $N = IJ$. Let R_{ij} be the pooled ranks of the sample values Y_{ij}, so that $\sum_i \sum_j R_{ij} = \frac{N(N+1)}{2}$ and $\bar{R}_{..} = \frac{N+1}{2}$ is the mean rank.

$$\boxed{\text{Kruskal-Wallis test statistic } K = \frac{12J}{N(N+1)} \sum_{i=1}^{I} (\bar{R}_{i.} - \tfrac{N+1}{2})^2}$$

Reject H_0 for large K using the null distribution table. For $I = 3$, $J \geq 5$ or $I \geq 4$, $J \geq 4$, use the approximate null distribution $K \overset{a}{\sim} \chi^2_{I-1}$.

Example (seven labs)
In the table below the actual measurements are replaced by their ranks $1 \div 70$. With the observed test statistic $K = 28.17$ and df $= 6$, using χ^2_5-distribution table we get a P-value ≈ 0.0001.

Labs	1	2	3	4	5	6	7
	70	4	35	6	46	48	38
	63	3	45	7	21	5	50
	53	65	40	13	47	22	52
	64	69	41	20	8	28	58
	59	66	57	16	14	37	68
	54	39	32	26	42	2	1
	43	44	51	17	9	31	15
	61	56	25	11	10	34	23
	67	24	29	27	33	49	60
	55	19	30	12	36	18	62
Means	58.9	38.9	38.5	15.5	26.6	27.4	42.7

2 Two-way layout

Suppose the data values are influenced by two main factors and a noise:
$Y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \epsilon_{ijk}, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J, \quad k = 1, \ldots, K,$
grand mean + main A-effect +main B-effect + interaction + noise.
Factor A has I levels, factor B has J levels, and we have K observations for each combination (i, j).

Normal theory model
Key assumption: all noise components $\epsilon_{ijk} \sim \mathrm{N}(0, \sigma^2)$ are independent and have the same variance.
Parameter constraints and numbers of degrees of freedom
$\mathrm{df_A} = I - 1$, because $\sum_i \alpha_i = 0$,
$\mathrm{df_B} = J - 1$, because $\sum_j \beta_j = 0$,
$\mathrm{df_{AB}} = IJ - I - J + 1 = (I-1)(J-1)$, because $\sum_i \delta_{ij} = 0$, $\sum_j \delta_{ij} = 0$.

Maximum likelihood estimates: $\hat{\mu} = \bar{Y}_{...}$, $\hat{\alpha}_i = \bar{Y}_{i..} - \bar{Y}_{...}$, $\hat{\beta}_j = \bar{Y}_{.j.} - \bar{Y}_{...}$,
$\hat{\delta}_{ij} = \bar{Y}_{ij.} - \bar{Y}_{...} - \hat{\alpha}_i - \hat{\beta}_j = \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}$,
and the residuals $\hat{\epsilon}_{ijk} = Y_{ij.} - \bar{Y}_{ijk}$.

Example (iron retention)
Raw data X_{ijk} is the percentage of iron retained in mice. Factor A: $I = 2$ iron forms, factor B: $J = 3$ dosage levels, $K = 18$ observations for each (iron form, dosage level) combination. From the graphs we see that the raw data is not normally distributed.
However, the transformed data $Y_{ijk} = \ln(X_{ijk})$ produce more satisfactory graphs. The sample means and maximum likelihood estimates for the transformed data

$(\bar{Y}_{ij.}) = \begin{pmatrix} 1.16 & 1.90 & 2.28 \\ 1.68 & 2.09 & 2.40 \end{pmatrix}$ two rows produce two profiles: not parallel - possible interaction,
$\bar{Y}_{...} = 1.92$, $\hat{\alpha}_1 = -0.14$, $\hat{\alpha}_2 = 0.14$,
$\hat{\beta}_1 = -0.50$, $\hat{\beta}_2 = 0.08$, $\hat{\beta}_3 = 0.42$, $(\hat{\delta}_{ij}) = \begin{pmatrix} -0.12 & 0.04 & 0.08 \\ 0.12 & -0.04 & -0.08 \end{pmatrix}$

Sums of squares

$\mathrm{SS_T} = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{...})^2 = \mathrm{SS_A} + \mathrm{SS_B} + \mathrm{SS_{AB}} + \mathrm{SS_E}$, $\mathrm{df_T} = IJK - 1$
$\mathrm{SS_A} = JK \sum_i \hat{\alpha}^2$, $\mathrm{df_A} = I - 1$, $\mathrm{MS_A} = \frac{\mathrm{SS}_A}{\mathrm{df_A}}$, $\mathrm{E(MS_A)} = \sigma^2 + \frac{JK}{I-1} \sum_i \alpha_i^2$
$\mathrm{SS_B} = IK \sum_j \hat{\beta}^2$, $\mathrm{df_B} = J - 1$, $\mathrm{MS_B} = \frac{\mathrm{SS_B}}{\mathrm{df_B}}$, $\mathrm{E(MS_B)} = \sigma^2 + \frac{IK}{J-1} \sum_j \beta_j^2$
$\mathrm{SS_{AB}} = K \sum_i \sum_j \delta_{ij}^2$, $\mathrm{df_{AB}} = (I-1)(J-1)$, $\mathrm{MS_{AB}} = \frac{\mathrm{SS_{AB}}}{\mathrm{df_{AB}}}$, $\mathrm{E(MS_{AB})} = \sigma^2 + \frac{K}{(I-1)(J-1)} \sum_i \sum_j \delta_{ij}^2$
$\mathrm{SS_E} = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2$, $\mathrm{df_E} = IJ(K-1)$, $\mathrm{MS_E} = \frac{\mathrm{SS_E}}{\mathrm{df_E}}$, $\mathrm{E(MS_E)} = \sigma^2$

Pooled sample variance $s_p^2 = \mathrm{MS_E}$ is an unbiased estimate of σ^2.

Three F-tests

Null hypothesis	No-effect property	Test statistics and null distribution
H_A: $\alpha_1 = \ldots = \alpha_I = 0$	$\mathrm{E(MS_A)} = \sigma^2$	$F_\mathrm{A} = \frac{\mathrm{MS_A}}{\mathrm{MS_E}} \sim F_{\mathrm{df_A},\mathrm{df_E}}$
H_B: $\beta_1 = \ldots = \beta_J = 0$	$\mathrm{E(MS_B)} = \sigma^2$	$F_\mathrm{B} = \frac{\mathrm{MS_B}}{\mathrm{MS_E}} \sim F_{\mathrm{df_B},\mathrm{df_E}}$
H_AB: all $\delta_{ij} = 0$	$\mathrm{E(MS_{AB})} = \sigma^2$	$F_\mathrm{AB} = \frac{\mathrm{MS_{AB}}}{\mathrm{MS_E}} \sim F_{\mathrm{df_{AB}},\mathrm{df_E}}$

Reject null hypothesis for large values of the respective test statistic F.
Inspect normal probability plot for the residuals $\hat{\epsilon}_{ijk}$.

Example (iron retention)
Two-way Anova table for the transformed iron retention data. Dosage effect was expected from the beginning. Interaction is not significant.

Source	df	SS	MS	F	P
Iron form	1	2.074	2.074	5.99	0.017
Dosage	2	15.588	7.794	22.53	0.000
Interaction	2	0.810	0.405	1.17	0.315
Error	102	35.296	0.346		
Total	107	53.768			

Significant effect due to iron form. Estimated log scale difference $\hat{\alpha}_2 - \hat{\alpha}_1 = \bar{Y}_{2..} - \bar{Y}_{1..} = 0.28$ yields the multiplicative effect of $e^{0.28} = 1.32$ on a linear scale.

3 Randomized block design

Blocking is used to remove the effects of a few of the most important nuisance variables. Randomization is then used to reduce the contaminating effects of the remaining nuisance variables.

$\boxed{\text{Block what you can, randomize what you cannot.}}$

Experimental design: randomly assign I treatments within each of J blocks.
Test the null hypothesis of no treatment effects using the two-way layout Anova.
The block effect is anticipated and is not of major interest. Examples:

Block	Treatments	Observation
A homogeneous plot of land divided into I subplots	I fertilizers each applied to a randomly chosen subplot	The yield on the subplot (i, j)
A four-wheel car	4 types of tires tested on the same car	tire's life-length
A litter of I animals	I diets randomly assigned to I sinlings	the weight gain

Additive model
If $K = 1$, then we cannot estimate interaction. This leads to the additive model without interaction $Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$. Maximum likelihood estimates

$$\hat{\mu} = \bar{Y}_{..},\ \hat{\alpha}_i = \bar{Y}_{i.} - \bar{Y}_{..},\ \hat{\beta}_i = \bar{Y}_{.j} - \bar{Y}_{..}, \quad \hat{\epsilon}_{ij} = Y_{ij} - \bar{Y}_{..} - \hat{\alpha}_i - \hat{\beta}_i = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}$$

Sums of squares

$$\begin{array}{lll}
\mathrm{SS_T} = \sum_i \sum_j (\bar{Y}_{ij} - \bar{Y}_{..})^2 = \mathrm{SS_A} + \mathrm{SS_B} + \mathrm{SS_E}, & \mathrm{df_T} = IJ - 1 & \\
\mathrm{SS_A} = J \sum_i \hat{\alpha}_i^2, \quad \mathrm{df_A} = I - 1, & \mathrm{MS_A} = \frac{\mathrm{SS}_A}{\mathrm{df_A}} & F_\mathrm{A} = \frac{\mathrm{MS_A}}{\mathrm{MS_E}} \sim F_{\mathrm{df_A}, \mathrm{df_E}} \\
\mathrm{SS_B} = I \sum_j \hat{\beta}_j^2, \quad \mathrm{df_B} = J - 1 & \mathrm{MS_B} = \frac{\mathrm{SS_B}}{\mathrm{df_B}} & F_\mathrm{B} = \frac{\mathrm{MS_B}}{\mathrm{MS_E}} \sim F_{\mathrm{df_B}, \mathrm{df_E}} \\
\mathrm{SS_E} = \sum_i \sum_j \hat{\epsilon}_{ij}^2, \quad \mathrm{df_E} = (I - 1)(J - 1) & \mathrm{MS_E} = \frac{\mathrm{SS_E}}{\mathrm{df_E}} & \mathrm{E(MS_E)} = \sigma^2
\end{array}$$

Example (itching)
Data: the duration of the itching in seconds Y_{ij}, with $K = 1$ observation per cell,
$I = 7$ treatments to relieve itching applied to $J = 10$ male volunteers aged 20-30.

Subject	No Drug	Placebo	Papaverine	Morphine	Aminophylline	Pentabarbital	Tripelennamine
BG	174	263	105	199	141	108	141
JF	224	213	103	143	168	341	184
BS	260	231	145	113	78	159	125
SI	225	291	103	225	164	135	227
BW	165	168	144	176	127	239	194
TS	237	121	94	144	114	136	155
GM	191	137	35	87	96	140	121
SS	100	102	133	120	222	134	129
MU	115	89	83	100	165	185	79
OS	189	433	237	173	168	188	317

Boxplots indicate violations of the assumptions of normality and equal variance. Notice much bigger variance for the placebo group.

Two-way Anova table

Source	df	SS	MS	F	P
Drugs	6	53013	8835	2.85	0.018
Subjects	9	103280	11476	3.71	0.001
Error	54	167130	3096		
Total	69	323422			

Tukey's method of multiple comparison $q_{I,(I-1)(J-1)}(\alpha) \cdot \frac{s_p}{\sqrt{J}} = q_{7,54}(0.05) \cdot \sqrt{\frac{3096}{10}} = 75.8$ reveals only one significant difference: papaverine vs placebo with $208.4 - 118.2 = 90.2 > 75.8$.

Treatment	2	1	6	7	4	5	3
Mean	208.4	191.0	176.5	167.2	148.0	144.3	118.2

Friedman's test

Nonparametric test, when ϵ_{ij} are non-normal, to test H_0: no treatment effects.
Ranking within j-th block: $(R_{1j}, \ldots, R_{Ij})$ = ranks of $(Y_{1j}, \ldots, Y_{Ij})$ so that $R_{1j} + \ldots + R_{Ij} = \frac{I(I+1)}{2}$, implying $\frac{1}{I}(R_{1j} + \ldots + R_{Ij}) = \frac{I+1}{2}$ and $\bar{R}_{..} = \frac{I+1}{2}$.

Test statistic $Q = \frac{12J}{I(I+1)} \sum_{i=1}^{I} (\bar{R}_{i.} - \frac{I+1}{2})^2$ has an approximate null distribution $Q \overset{a}{\sim} \chi^2_{I-1}$.

Since Q is a measure of agreement between J rankings, we reject H_0 for large values of Q.

Example (itching)

From the values R_{ij} and $\bar{R}_{i.}$ below and $\frac{I+1}{2} = 4$, we find the Friedman test statistic $Q = 14.86$. Using the chi-square distribution table with df = 6 we obtain an approximate P-value to be 2.14%. We reject the null hypothesis of no effect even in the non-parametric setting.

Subject	No Drug	Placebo	Papaverine	Morphine	Aminophylline	Pentabarbital	Tripelennamine
BG	5	7	1	6	3.5	2	3.5
JF	6	5	1	2	3	7	4
BS	7	6	4	2	1	5	3
SI	6	7	1	4	3	2	5
BW	3	4	2	5	1	7	6
TS	7	3	1	5	2	4	6
GM	7	5	1	2	3	6	4
SS	1	2	5	3	7	6	4
MU	5	3	2	4	6	7	1
OS	4	7	5	2	1	3	6
$\bar{R}_{i.}$	5.10	4.90	2.30	3.50	3.05	4.90	4.25