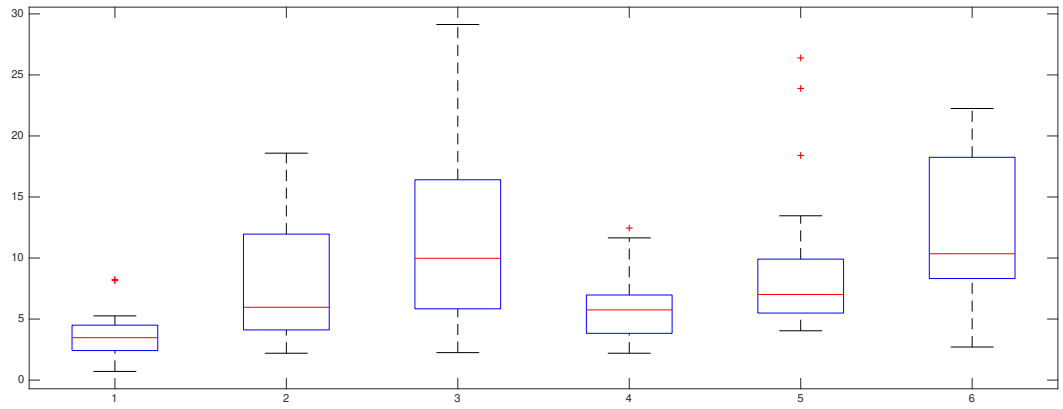
Slides 13: ANOVA two-way layout

- Normal theory model
- Maximum likelihood estimates
- Three F-tests
- F-distribution table
- Randomised block design
- Additive model
- Course topics



Fe^{3+} (10.2)	Fe^{3+} (1.2)	Fe^{3+} (0.3)	Fe^{2+} (10.2)	Fe^{2+} (1.2)	Fe^{2+} (0.3)
.71	2.2	2.25	2.2	4.04	2.71
1.66	2.93	3.93	2.69	4.16	5.43
2.01	3.08	5.08	3.54	4.42	6.38
2.16	3.49	5.82	3.75	4.93	6.38
2.42	4.11	5.84	3.83	5.49	8.32
2.42	4.95	6.89	4.08	5.77	9.04
2.56	5.16	8.5	4.27	5.86	9.56
2.6	5.54	8.56	4.53	6.28	10.01
3.31	5.68	9.44	5.32	6.97	10.08
3.64	6.25	10.52	6.18	7.06	10.62
3.74	7.25	13.46	6.22	7.78	13.8
3.74	7.9	13.57	6.33	9.23	15.99
4.39	8.85	14.76	6.97	9.34	17.9
4.5	11.96	16.41	6.97	9.91	18.25
5.07	15.54	16.96	7.52	13.46	19.32
5.26	15.89	17.56	8.36	18.4	19.87
8.15	18.3	22.82	11.65	23.89	21.6
8.24	18.59	29.13	12.45	26.39	22.25

Percentages of iron retained in mice. See boxplots on the first slide.

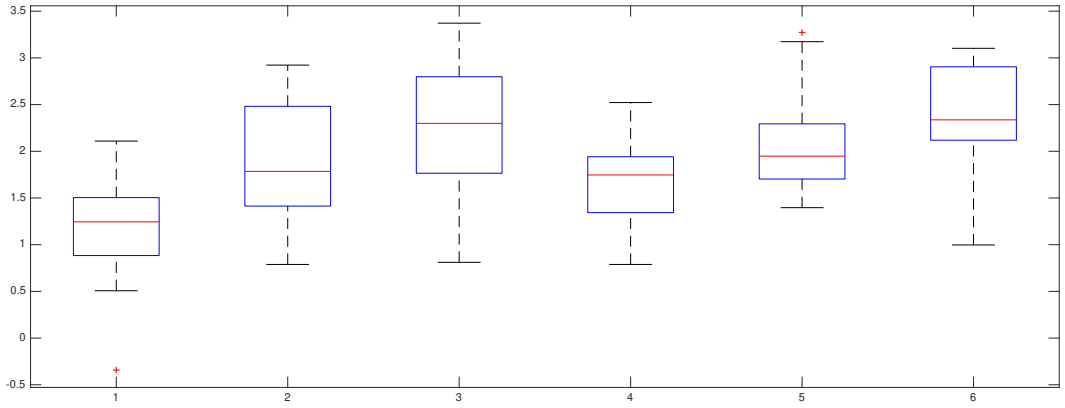
Factor A has two levels $I = 2$ representing two iron forms, factor B has three levels $J = 3$ representing dosage concentrations, $n = 18$.

Assume that the data is generated in the following way

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \epsilon_{ijk}, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J, \quad k = 1, \ldots, n,$$

where $\epsilon_{ijk} \sim \mathrm{N}(0, \sigma)$ are independent and have the same variance. Two main effects (α_i, β_j) and interaction δ_{ij}.

The raw data $\{z_{ijk}\}$ of the case study is not normally distributed. The transformed data $y_{ijk} = \ln(z_{ijk})$ produce more satisfactory boxplots.



Question. Is there a significant difference between Fe^{2+} and Fe^{3+}? Is there a significant interaction between two main factors?

The maximum likelihood estimates

$$\hat{\mu} = \bar{y}_{...} = \frac{1}{IJn} \sum_i \sum_j \sum_k y_{ijk},$$

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

$$\hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...},$$

$$\hat{\delta}_{ij} = \bar{y}_{ij.} - \bar{y}_{...} - \hat{\alpha}_i - \hat{\beta}_j = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...},$$

where

$$\bar{y}_{i..} = \frac{1}{Jn} \sum_j \sum_k y_{ijk}, \quad \bar{y}_{.j.} = \frac{1}{In} \sum_i \sum_k y_{ijk}, \quad \bar{y}_{ij.} = \frac{1}{n} \sum_k y_{ijk}.$$

The observed responses have the following decomposition

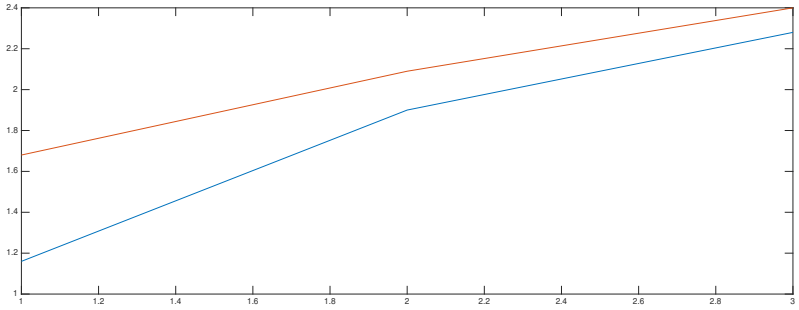
$$y_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\delta}_{ij} + \hat{\epsilon}_{ijk},$$

involving residuals

$$\hat{\epsilon}_{ijk} = y_{ijk} - \bar{y}_{ij.}$$

The six sample means for the transformed data ($\bar{y}_{ij.}$)

	10.2	1.2	0.3	Level mean
Fe^{3+}	1.16	1.90	2.28	1.78
Fe^{2+}	1.68	2.09	2.40	2.06
Level mean	1.42	2.00	2.34	1.92



Two profiles for iron forms are not parallel, which indicates possible interaction.

Maximum likelihood estimates

$$\hat{\mu} = 1.92,\ (\hat{\alpha}_1, \hat{\alpha}_2) = (-0.14, 0.14),\ (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = (-0.50, 0.08, 0.42),$$

and interaction terms

$$(\hat{\delta}_{ij}) = \begin{pmatrix} -0.12 & 0.04 & 0.08 \\ 0.12 & -0.04 & -0.08 \end{pmatrix}$$

Question. How many are degrees of freedom in the interaction terms?

Null hypotheses of interest

H_{A}: $\alpha_1 = \ldots = \alpha_I = 0$, no main effect A

H_{B}: $\beta_1 = \ldots = \beta_J = 0$, no main effect B

H_{AB}: all $\delta_{ij} = 0$, no interaction between A and B

ANOVA-2 is based on a decomposition of the sums of squares

$$SS_{\mathrm{T}} = SS_{\mathrm{A}} + SS_{\mathrm{B}} + SS_{\mathrm{AB}} + SS_{\mathrm{E}},$$

where

$$SS_{\mathrm{T}} = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{\ldots})^2, \qquad \mathrm{df}_{\mathrm{T}} = IJn - 1$$

$$SS_{\mathrm{A}} = Jn \sum_i \hat{\alpha}_i^2, \qquad \mathrm{df}_{\mathrm{A}} = I - 1$$

$$SS_{\mathrm{B}} = In \sum_j \hat{\beta}_j^2, \qquad \mathrm{df}_{\mathrm{B}} = J - 1$$

$$SS_{\mathrm{AB}} = n \sum_i \sum_j \hat{\delta}_{ij}^2, \qquad \mathrm{df}_{\mathrm{AB}} = (I-1)(J-1)$$

$$SS_{\mathrm{E}} = \sum_i \sum_j \sum_k \hat{\epsilon}_{ijk}^2, \qquad \mathrm{df}_{\mathrm{E}} = IJ(n-1)$$

The mean sums of squares and their expected values

$$MS_{\mathrm{A}} = \frac{SS_A}{\mathrm{df_A}}, \qquad \mathrm{E}(MS_{\mathrm{A}}) = \sigma^2 + \frac{Jn}{I-1}\sum_i \alpha_i^2$$
$$MS_{\mathrm{B}} = \frac{SS_{\mathrm{B}}}{\mathrm{df_B}}, \qquad \mathrm{E}(MS_{\mathrm{B}}) = \sigma^2 + \frac{In}{J-1}\sum_j \beta_j^2$$
$$MS_{\mathrm{AB}} = \frac{SS_{\mathrm{AB}}}{\mathrm{df_{AB}}}, \qquad \mathrm{E}(MS_{\mathrm{AB}}) = \sigma^2 + \frac{n}{(I-1)(J-1)}\sum_i\sum_j \delta_{ij}^2$$
$$MS_{\mathrm{E}} = \frac{SS_{\mathrm{E}}}{\mathrm{df_E}}, \qquad \mathrm{E}(MS_{\mathrm{E}}) = \sigma^2$$

Pooled sample variance $s_{\mathrm{p}}^2 = MS_{\mathrm{E}}$ is an unbiased estimate of σ^2.

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

Reject null hypothesis for large values of the respective test statistic.

Inspect normal probability plot for the residuals $\hat{\epsilon}_{ijk}$.

ANOVA-2 table

A two-way Anova table for the transformed iron retention data:

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			

According to the rightmost column

- the dosage effect is significant, as expected,
- interaction is not statistically significant,
- there is a significant effect due to iron form (compare to the previous analysis of two samples).

The log scale difference $\hat{\alpha}_2 - \hat{\alpha}_1 = \bar{y}_{2\cdot\cdot} - \bar{y}_{1\cdot\cdot} = 0.28$ yields the multiplicative effect of $e^{0.28} = 1.32$ on the original scale.

The retention percentage of Fe^{2+} is 1.32 times higher than that of Fe^{3+}.

F-distribution table

$F_{4,df}$-distribution table.

	5	9	10	15	18	20	25	27	30	35	36	40	45	50
0.100	3.5202	2.6927	2.6053	2.3614	2.2858	2.2489	2.1842	2.1655	2.1422	2.1128	2.1079	2.0909	2.0742	2.0608
0.050	5.1922	3.6331	3.4780	3.0556	2.9277	2.8661	2.7587	2.7278	2.6896	2.6415	2.6335	2.6060	2.5787	2.5572
0.025	7.3879	4.7181	4.4683	3.8043	3.6083	3.5147	3.3530	3.3067	3.2499	3.1785	3.1668	3.1261	3.0860	3.0544
0.010	11.3919	6.4221	5.9943	4.8932	4.5790	4.4307	4.1774	4.1056	4.0179	3.9082	3.8903	3.8283	3.7674	3.7195
0.001	31.0850	12.5603	11.2828	8.2527	7.4593	7.0960	6.4931	6.3261	6.1245	5.8764	5.8362	5.6981	5.5639	5.4593

Question. Why do I use only *df* which are either multiple of 5 or 9?

Answer.

1. If we use $F_{4,df}$ as $F_{\mathrm{df_A},\mathrm{df_E}}$, then $4 = \mathrm{df_A} = I - 1$ and

$$df = \mathrm{df_E} = IJ(n-1) \text{ must be a multiple of } I = 5$$

2. If we use $F_{4,df}$ as $F_{\mathrm{df_{AB}},\mathrm{df_E}}$, then $4 = \mathrm{df_{AB}} = (I-1)(J-1)$ and

$$df = \mathrm{df_E} = IJ(n-1) \text{ must be a multiple of } IJ.$$

Two options:

(a) $4 = (I-1)(J-1) = 1 \cdot 4$, then $IJ = 2 \cdot 5 = 10$

(b) $4 = (I-1)(J-1) = 2 \cdot 2$, then $IJ = 3 \cdot 3 = \ \ 9$

Blocking is used to remove the effects of the most important nuisance variable. Randomisation is then used to reduce the contaminating effects of the remaining nuisance variables.

Block what you can, randomise what you cannot.

Experimental design: randomly assign I treatments within each of J blocks. Test the null hypothesis of no treatment effect using the two-way layout Anova.

Examples:

Blocking: a homogeneous plot of land is divided into I subplots. Treatments: I fertilizers each applied to a randomly chosen subplot. Response: the yield on the subplot (i, j)

Blocking: a four-wheel car. Treatments: 4 types of tires tested on the same car. Response: tire's life-length

Blocking: a litter of I animals. Treatments: I diets randomly assigned to I siblings. Response: the weight gain.

Question. Why is the block effect usually not of a particular interest?

Consider a two-way layout ANOVA with $n = 1$. With only one replication per cell, we cannot estimate interaction. This restricts us to the additive model without interaction

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

For the given data (y_{ij}), find the maximum likelihood estimates and residuals

$$\hat{\mu} = \bar{y}_{..}, \qquad \hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}, \qquad \hat{\beta}_i = \bar{y}_{.j} - \bar{y}_{..},$$
$$\hat{\epsilon}_{ij} = y_{ij} - \bar{y}_{..} - \hat{\alpha}_i - \hat{\beta}_i = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..},$$

yields a representation

$$y_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\epsilon}_{ij}.$$

Question. How do you check the normality assumption using a normal probability plot? What if the normality assumption is clearly violated?

The decomposition takes a reduced form

$$SS_{\mathrm{T}} = SS_{\mathrm{A}} + SS_{\mathrm{B}} + SS_{\mathrm{E}},$$

$$\begin{aligned}
SS_{\mathrm{T}} &= \textstyle\sum_i \sum_j (\bar{y}_{ij} - \bar{y}_{..})^2, & \mathrm{df}_{\mathrm{T}} &= IJ - 1 \\
SS_{\mathrm{A}} &= J \textstyle\sum_i \hat{\alpha}_i^2, & \mathrm{df}_{\mathrm{A}} &= I - 1 \\
SS_{\mathrm{B}} &= I \textstyle\sum_j \hat{\beta}_j^2, & \mathrm{df}_{\mathrm{B}} &= J - 1 \\
SS_{\mathrm{E}} &= \textstyle\sum_i \sum_j \hat{\epsilon}_{ij}^2, & \mathrm{df}_{\mathrm{E}} &= (I-1)(J-1) \\
MS_{\mathrm{A}} &= \tfrac{SS_A}{\mathrm{df}_{\mathrm{A}}}, & \mathrm{E}(MS_{\mathrm{A}}) &= \sigma^2 + \tfrac{J}{I-1} \textstyle\sum_i \alpha_i^2 \\
MS_{\mathrm{B}} &= \tfrac{SS_{\mathrm{B}}}{\mathrm{df}_{\mathrm{B}}}, & \mathrm{E}(MS_{\mathrm{B}}) &= \sigma^2 + \tfrac{I}{J-1} \textstyle\sum_j \beta_j^2 \\
MS_{\mathrm{E}} &= \tfrac{SS_{\mathrm{E}}}{\mathrm{df}_{\mathrm{E}}}, & \mathrm{E}(\mathrm{MS}_{\mathrm{E}}) &= \sigma^2
\end{aligned}$$

We can apply two F-tests for two different null hypotheses

$$H_{\mathrm{A}} : \alpha_1 = \ldots = \alpha_I = 0, \qquad F_{\mathrm{A}} = \frac{MS_{\mathrm{A}}}{MS_{\mathrm{E}}} \overset{H_{\mathrm{A}}}{\sim} F_{\mathrm{df}_{\mathrm{A}},\mathrm{df}_{\mathrm{E}}},$$

$$H_{\mathrm{B}} : \beta_1 = \ldots = \beta_J = 0, \qquad F_{\mathrm{B}} = \frac{MS_{\mathrm{B}}}{MS_{\mathrm{E}}} \overset{H_{\mathrm{B}}}{\sim} F_{\mathrm{df}_{\mathrm{B}},\mathrm{df}_{\mathrm{E}}}.$$

Example: itching

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

Papa = Papaverine, Amin = Aminophylline, Pent = Pentabarbital, Trip = Tripelennamine

Subject	No Drug	Placebo	Papa	Morphine	Amin	Pent	Trip
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

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			

Significant treatment effect.

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

List of course topics

Statistical inference vs probability theory. Statistical models.
Population distribution. Population mean and standard deviation, population proportion.
Randomisation.
Sampling with replacement, random (iid) sample.
Sampling without replacement, simple random sample.

Point estimate, sampling distribution.
Mean square error, systematic error and random (sampling) error.
Unbiased point estimate, consistent point estimate.
Sample mean, sample variance, sample standard deviation, sample proportion.
Finite population correction.
Standard error of the sample mean and sample proportion.
Approximate confidence interval for the mean.
Stratified random sampling. Optimal allocation of observations, proportional allocation.

Parametric models, population parameters.
Binomial, geometric, Poisson, discrete uniform models.
Continuous uniform, exponential, gamma models.
Normal distribution, central limit theorem, continuity correction.
Method of moments for point estimation.
Maximum likelihood estimate (MLE). Likelihood function.
Normal approximation for the sampling distribution of MLE.
Sufficient statistics for population parameters.
Exact confidence intervals for the mean and variance. Chi-squared and t-distributions.

Statistical hypotheses, simple and composite, null and alternative.
Rejection region. Two types of error.
Significance level, test power.
P-value of the test, one-sided and two-sided p-values.

Large-sample test for the proportion. Small-sample test for the proportion.
Large-sample test for the mean. One-sample t-test.
Nested hypotheses, generalised likelihood ratio test.
Chi-squared test of goodness of fit, its approximate nature. Multinomial distribution.

Bayes formulas for probabilities and densities.
Prior and posterior distributions.
Loss function, posterior risk, 0-1 loss function and squared error loss.
Conjugate priors. Normal-normal model.
Beta and Dirichlet distributions. Beta-binomial model and Dirichlet-multinomial model.
Bayesian estimation, MAP and PME. Credibility interval.
Posterior odds. Bayesian hypotheses testing.

Empirical cumulative distribution function. Empirical variance.
Survival function and hazard function. Weibull distribution. Empirical survival function.
Kernel density estimate. Steam-and-leaf plot.
Population quantiles. Ordered sample and empirical quantiles.
QQ-plots, normal probability plot.
Coefficient of skewness and kurtosis. Light tails and heavy tails of probability distributions.
Leptokurtic and platykurtic distributions.

Population mean, mode, and median. Sample median, outliers.
Sign test and non-parametric confidence interval for the median.
Trimmed means.
Sample range, quartiles, IQR and MAD. Boxplots.

Two independent versus paired samples.
Approximate confidence interval and large sample test for the mean difference.
Two-sample t-test, pooled sample variance.
Exact confidence interval for the mean difference. Transformation of variables.

Ranks vs exact measurements. Rank sum test. Signed rank test.
Approximate confidence interval for the difference $p_1 - p_2$.
Large sample test for two proportions.
Fisher's exact test.
Double-blind randomised controlled experiments.
Confounding factors, Simpson's paradox.

One-way ANOVA, sums of squares and mean squares.
Normal theory model, F-test, F-distribution.
Normal probability plots for the residuals.
Multiple comparison or multiple testing problem.
Simultaneous CI, Bonferroni's method and Tukey's method.
Two-way ANOVA, main effects and interaction. Three F-tests.
Additive model. Randomised block design.
Kruskal-Wallis test. Friedman's test.

Categorical data.
Chi-squared tests of homogeneity and independence.
Prospective and retrospective studies. Matched-pairs design, McNemar's test. Odds ratio.

Simple linear regression model. Normal equations. Least squares estimates.
Sample correlation coefficient, sample covariance.
Corrected MLE of the noise variance. Coefficient of determination.
Confidence intervals and hypotheses testing for the intercept and slope. Model utility test.
Prediction interval for a new observation.
Standardised residuals.
Linear regression and ANOVA.
Multiple regression. Design matrix.
Coefficient of multiple determination. Adjusted coefficient of multiple determination.
Collinearity problem.