Analysis of simple experiments using least squares

Documentation of Methods

Analysis of variance

Experiments involve exposing experimental units to a range of discrete levels of one or more categorical variables or factors. A factor can be a fertilization treatment with four levels corresponding to four different quantities of fertilizer.

An analysis of variance (ANOVA) is conducted to determine whether there are differences between factor levels (treatment means). When the means of only two treatments are compared, we can use a simple t-test. If we compare three or more means, we use an ANOVA. Regression and ANOVA are essentially identical approaches except that the explanatory variables for regression are continuous while they are categorical for ANOVA.

In the following, the principles of one-way and two-way ANOVA are illustrated using experiments with a completely randomized design.

Completely Randomized design

In a completely randomized design (CRD), homogeneous experimental units are located and treatments are randomly assigned to the treatment units. The variable of interest is measured for each experimental unit. No blocking, which can reduce experimental error variation, is used.

Notation and data organization

The notation used is based on Kuehl (2000).

Cell means model: $y_{ij} = \mu_i + e_{ij}$

Treatment Effects Model: $y_{ij} = \mu + \tau_i + e_{ij}$

y_{ij} = response variable; *jth* observation from *ith* treatment group

i = 1,2,…,t treatment groups

j = 1,2,…,r observations per treatment group (replications)

N = rt

μ = grand or overall mean regardless of treatment

μ_i = mean of the *ith* treatment population

τ_i = treatment effect = $\mu_i - \mu$

e_{ij} = experimental error = $y_{ij} - \mu_j$; experimental error variance σ^2 is the variance of e_{ij}

For easy calculations by hand, the data could be organized in columns as:

Observations	Treatments, i = 1,2,…,t					
j = 1,2,…,r *	1	2	3	…	t	
1	y_{11}	y_{21}	y_{31}	…	y_{t1}	
2	y_{12}	y_{22}	y_{32}	…	y_{t2}	
3	y_{13}	y_{23}	y_{33}	…	y_{t3}	
…	…	…	…	…	…	
r	y_{1r}	y_{2r}	y_{3r}	…	y_{tr}	
Sum	$y_{1.}$	$y_{2.}$	$y_{3.}$	…	$y_{t.}$	$y_{..}$
Averages	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\bar{y}_{3.}$		$\bar{y}_{t.}$	$\bar{y}_{..}$

* NOTE: The number of observations for each treatment may not be the same. In that case j = 1,2,…,r_i.

With sums calculated as:

$$y_{i.} = \sum_{j=1}^{r} y_{ij} = y_{i1} + y_{i2} + ... + y_{ir} = \text{total of the observations for the } ith \text{ treatment}$$

$$y_{..} = \sum_{i=1}^{t} \sum_{j=1}^{r} y_{ij} = \text{total of all observations}$$

and averages calculated as:

$$\bar{y}_{i.} = \frac{y_{i.}}{r} \qquad \frac{y_{..}}{N} = \bar{y}_{..}$$

Estimation of model parameters with least squares

$$\hat{\mu} = \bar{y}_{..} = \frac{y_{..}}{N} \qquad \hat{\mu}_i = \frac{y_{i.}}{r} = \bar{y}_{i.} \qquad \hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} \qquad \hat{e}_{ij} = y_{ij} - \bar{y}_{i.}$$

Sum of squares

1. Sum of squared differences between the observed values and the overall mean (SS Total):

$$SS\ Total = \sum_{i=1}^{t}\sum_{j=1}^{r}\left(y_{ij} - \bar{y}_{..}\right)^2 \qquad df = N - 1$$

2. Sum of squared differences between the treatment means and the overall mean, weighted by the number of experimental units in each treatment (SS Treatment = SST):

$$SS\ Treatment = \sum_{i=1}^{t}\sum_{j=1}^{r}\left(\bar{y}_{i.} - \bar{y}_{..}\right)^2 = \sum_{i=1}^{t} r_i\left(\bar{y}_{i.} - \bar{y}_{..}\right)^2 \qquad df = t - 1$$

3. Sum of squared differences between the observed values for each experimental unit and the treatment means (SS Error = SSE):

$$SS\ Error = \sum_{i=1}^{t}\sum_{j=1}^{r}\hat{e}_{ij}^2 = \sum_{i=1}^{t}\sum_{j=1}^{r}\left(y_{ij} - \bar{y}_{i.}\right)^2 \qquad df = N - t$$

Error partition: SS Error = SS Total – SS Treatment

One-way analysis of variance

When only one factor with more than two levels is applied as treatment, a one-way ANOVA is used for the analysis of the data.

Test for differences in treatment means

Main question: Are the treatment means different?

$H_0: \mu_1 = \mu_2 = ... = \mu_t$ vs. $H_1: \mu_i \neq \mu_k$ for some $i \neq k$ → at least one μ_i different

OR

$H_0: \tau_1 = \tau_2 = ... = \tau_t$ vs. $H_1:$ not all τ_i equal to 0

If the treatment does not account for any of the variance in the response variable, then treatment effects are likely all equal to zero and all the treatment means are likely to be the same.

ANOVA table

Source	df	SS	MS	F	p-value
Treatment	t-1	SS Treatment (SST)	$MST = \frac{SST}{t-1}$	$F_0 = \frac{MST}{MSE}$	$F_0 > F_{\alpha,(t-1),(N-t)}$
Error	N-t	SS Error (SSE)	$MSE = \frac{SSE}{N-t}$		
Total	N-1	SS Total			

MSE is an unbiased estimate of the experimental error variance σ^2. In other words, the expected value of MSE is equal to σ^2: $E\left[MSE\right] = \sigma^2$.

The expected value of MST is: $E\left[MST\right] = \sigma^2 + r\theta_t^2$, where $\theta_t^2 = \frac{\sum_{i=1}^{t}\left(\mu_i - \bar{\mu}_.\right)^2}{(t-1)}$ is the variance among the treatment means.

Test statistic: $F_0 = \frac{(SST)/(t-1)}{(SSE)/(N-t)} = \frac{MST}{MSE}$

If the test statistic is larger than the critical value ($F_0 > F_{\alpha,(t-1),(N-t)}$), we reject H_0 and accept H_1, concluding that at least one of the means is significantly different from others. If the test statistic is less than the critical value, we fail to reject the null hypothesis and the associated p-value indicates that differences in treatment means could have arisen by chance alone.

NOTE: The F-test is only a good test if the assumptions of the analysis of variance have been met!

Assumptions

For the estimated means from an experiment to be unbiased estimates of the means in the population, and the MSE to be an unbiased estimate of the variance within each experimental unit, the following assumptions must be met:

1. **Random sampling** → can do nothing to rectify non-random sampling after the event of sampling
2. **Independence of errors**
   a. <u>Remedies:</u> minimize by (i) using block design with each treatment combination applied in every block; (ii) have high number of replication; (iii) thorough randomization
3. **Equal variances**
   a. <u>Test assumption</u> with (i) residual plot or (ii) standard tests for equal variance (e.g., Bartlett, Levene)
   b. <u>Remedies:</u> variance stabilizing transformations (e.g., log-transformation)
4. **Normal distribution of errors**
   a. <u>Test assumption</u> with (i) a normal probability plot for residuals or (ii) standard normality tests (e.g., Shapiro-Wilk, Anderson-Darling, Cramer-von Mises, Kolmogorov-Smirnov)
   b. <u>Causes</u> for non-normality: skewness, kurtosis, multimodality
   c. <u>Remedies:</u> skewness and kurtosis can often be corrected for by the same kind of transformations used to improve homogeneity of variance; arcsine transformation to transform proportions into normally distributed variables; rank transformation
5. **Additivity of treatment effects:** effects of treatments are additive → when response of one factor depends upon level of another factor (interaction present), use factorial experiments; if multiplicative, take log to make effects additive

<u>Process:</u>

- Do analysis with measured response
- If assumptions of the error term are not met, transform response variable
- Redo the analysis and check assumptions; if still not met, try another transformation
- May have to switch to another method (e.g., generalized linear models)

If we fail to reject H_0, we are done; if we reject H_0, we need to find out which levels differ.

Simultaneous Comparisons

If differences among treatment means were detected, we need to find out which means differ or which combinations of means differ. Methods for this include planned contrasts among treatment groups, regression response curves for quantitative treatment factors, selection of the best subset of treatments, comparison of treatments to the control, and all pairwise comparisons. All these methods require a set of simultaneous decisions which affects the statistical errors of inference. There are many different methods that 'preserve' the alpha level used to test all the means together when multiple comparisons are conducted, e.g.:

- Tukey's method for all pairwise comparisons (honestly significant difference = HSD)
- Bonferroni's t statistic for simultaneous inference (adjust alpha level used by dividing by the number of pairs)
- Scheffé's test for simultaneous inference

Comparisonwise error rate (α_C) = significance level or probability of Type I error for a single test

Experimentwise error rate (α_E) = significance level or probability of Type I error associated with a family of comparisons

1. Tukey's HSD

Used for pairwise comparison of all treatment means

Based on the Studentized range statistic $q = \dfrac{\overline{y}_{\text{largest}} - \overline{y}_{\text{smallest}}}{\sqrt{s^2 / r}}$

For a group of *t* treatment means simultaneous confidence intervals for all pairwise comparisons based on Tukey's HSD can be computed as follows:

$$\left|\overline{y}_i - \overline{y}_j\right| \pm q_{\alpha_E, t, \nu} * \sqrt{\frac{s^2}{2} * \left(\frac{1}{r_i} + \frac{1}{r_j}\right)}$$

2. Bonferroni's t statistic for simultaneous inference

One of the simplest and best known methods

Reduces alpha in proportion to the number of comparisons made

Mostly used for only few comparisons as it becomes increasingly conservative as the number of comparisons increases

When k comparisons are made: $\alpha_C = \alpha_E /k$ and the critical t-value for a two-sided t-test is: $t_{\alpha_E/2k,\nu}$,

where ν is the degrees of freedom.

Simultaneous confidence intervals can be obtained by:

$$\left|\overline{y}_i - \overline{y}_j\right| \pm t_{\alpha_E/2k,\nu} * \sqrt{s^2 * \left(\frac{1}{r_i} + \frac{1}{r_j}\right)}$$

<u>Example of calculating α_C:</u>

Number of possible pairs of means for t = 5 treatment groups: $\binom{t}{2} = \binom{5}{2} = \frac{5!}{3!2!} = 10$

For t = 5 treatment groups and $\alpha_E = 0.05$, $\alpha_C = 0.05/10=0.005$ when all possible pairs of means are compared simultaneously.

3. Scheffé's test for simultaneous inference

Conservative, generally used for unplanned contrasts or contrasts suggested by the data

Can test any pair of means or other comparisons

When you consider any contrast, $c = \sum_{i=1}^{t} k_i \overline{y}_{i.}$, among t treatment means with standard error

$s_c = \sqrt{s^2 \left[\sum_{i=1}^{t} \frac{k_i^2}{r_i}\right]}$, simultaneous confidence intervals for all possible contrasts can be computed

as: $\quad c \pm \sqrt{(t-1)F_{\alpha_E,(t-1),\nu}} * s_c$

Two-way analysis of variance

When two or more factors are applied as treatment combinations the design is referred to as factorial treatment design. In the following, the basic factorial treatment design in a completely randomized experiment is introduced with the example of a two-way analysis of variance. The methods can easily be expanded to more than two factors.

A two-factorial design allows looking at the response to one factor A at different levels of another factor B. The effect of a factor is the change in the measured response caused by the change in the level of that factor. There are three effects that are of interest in a factorial treatment design:

- *Simple effects:* contrasts between levels of one factor at a single level of another factor; if factors are not independent, interpretations should be based on simple effect contrasts
- *Main effects:* effects of a factor – contrasts of one factor averaged over all levels of another factor; main effects can be used to interpret the effects of factors separately if no interaction is present, and
- *Interaction effects:* measures differences between simple effects of one factor at different levels of the other factor

Interaction

In factorial treatment designs, the interpretations of effects of one treatment must take into account the effects of the other treatment factors. An interaction is present when the response to factor A depends on the level of factor B. Interaction plots can be used for determining whether there is an interaction between factors (lines are not parallel). In order to allow the investigation of interactions between factors, there must be replication at each combination of factor levels. If interaction is present there are no main effects and interpretation should be based on simple effects.

Notation

Cell means model: $y_{ijk} = \mu_{ij} + e_{ijk}$

Treatment Effects Model: $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ij}$

y_{ij} = response variable; *jth* observation from *ith* treatment group

i = 1,2,…,a levels of factor A

j = 1,2,…,b levels of factor B

k = 1,2,…,r observations per treatment group (replications)

N = rt

μ = grand or overall mean regardless of treatment

μ_{ij} = mean of the treatment combination A_iB_j

α_i = effect of the ith level of A, $\sum_{i=1}^{a} \alpha_i = 0$

βj = effect of the jth level of B, $\sum_{j=1}^{b} \beta_j = 0$

$(\alpha\beta)_{ij}$ = interaction effect between the ith level of A and the jth level of B,

$$\sum_{i=1}^{a} (\alpha\beta)_{ij} = \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0$$

e_{ijk} = random experimental error with mean 0 and variance σ^2

Sum of squares

1. Sum of squared differences between the observed values and the overall mean (SS Total):

$$SS\ Total = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}\left(y_{ijk} - \bar{y}_{...}\right)^2 \quad df = abr - 1$$

2. Sum of squared differences among marginal means for A (SSA):

$$SSA = br\sum_{i=1}^{a}\left(\bar{y}_{i..} - \bar{y}_{...}\right)^2 \quad df = a - 1$$

3. Sum of squared differences among marginal means for B (SSB):

$$SSB = ar\sum_{j=1}^{b}\left(\bar{y}_{.j.} - \bar{y}_{...}\right)^2 \quad df = b - 1$$

4. Sum of squares for interaction (SS(AB)):

$$SS(AB) = r\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{...}\right)^2 \quad df = (a-1)(b-1)$$

5. Sum of squared differences between the observed values for each experimental unit and the treatment means (SS Error = SSE):

$$SS\ Error = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}\hat{e}_{ijk}^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}\left(y_{ijk} - \bar{y}_{ij.}\right)^2 \qquad df = ab(r-1)$$

Error partition: SS Total = SS Treatment + SS Error,

with SS Treatment = SSA + SSB + SS(AB)

ANOVA table

Source	**df**	**SS**	**MS**	**F**	**p-value**
Factor A	a-1	SSA	$MSA = \frac{SSA}{a-1}$	$F_0 = \frac{MSA}{MSE}$	$F_0 > F_{\alpha,(a-1),ab(r-1)}$
Factor B	b-1	SSB	$MSB = \frac{SSB}{b-1}$	$F_0 = \frac{MSB}{MSE}$	$F_0 > F_{\alpha,(b-1),ab(r-1)}$
Interaction A:B	(a-1)(b-1)	SS(AB)	$MS(AB) = \frac{SS(AB)}{(a-1)(b-1)}$	$F_0 = \frac{MS(AB)}{MSE}$	$F_0 > F_{\alpha,(a-1)(b-1),ab(r-1)}$
Error	ab(r-1)	SSE	$MSE = \frac{SSE}{ab(r-1)}$		
Total	abr-1	SS Total			

The expected mean squares are as follows: $E[MSA] = \sigma^2 + rb\theta_a^2$, $E[MSB] = \sigma^2 + ar\theta_b^2$,

$E[MS(AB)] = \sigma^2 + r\theta_{ab}^2$, and $E[MSE] = \sigma^2$

The inferences about factors A and B depend upon the presence or absence of interaction. Therefore, start the interpretation of the ANOVA table with looking at the interaction between factors rather than looking at the main effects.

Hypotheses for the interaction:

If there is no interaction: $(\alpha\beta)_{ij} = \mu_{ij} - \overline{\mu}_{i.} - \overline{\mu}_{.j} + \overline{\mu}_{..} = 0$ and $\theta_{ab}^2 = 0$

$H_0 : (\alpha\beta)_{ij} = \mu_{ij} - \overline{\mu}_{i.} - \overline{\mu}_{.j} + \overline{\mu}_{..} = 0$ for all i, j

vs. $H_1 : (\alpha\beta)_{ij} = \mu_{ij} - \overline{\mu}_{i.} - \overline{\mu}_{.j} + \overline{\mu}_{..} \neq 0$ for some i, j

Hypotheses for differences among marginal means for A:

If there are no differences among the marginal means for A: $\alpha_i = \overline{\mu}_{i.} - \overline{\mu}_{..} = 0$ and $\theta_a^2 = 0$

$H_0 : \overline{\mu}_{1.} = \overline{\mu}_{2.} = ... = \overline{\mu}_{a.}$

vs. $H_1 : \overline{\mu}_{i.} \neq \overline{\mu}_{k.}$ for some i, k

Hypotheses for differences among marginal means for B:

If there are no differences among the marginal means for B: $\beta_i = \overline{\mu}_{.j} - \overline{\mu}_{..} = 0$ and $\theta_b^2 = 0$

$H_0 : \overline{\mu}_{.1} = \overline{\mu}_{.2} = ... = \overline{\mu}_{.b}$

vs. $H_1 : \overline{\mu}_{.j} \neq \overline{\mu}_{.m}$ for some j, m

References

Crawley, M.J. 2002. Statistical computing. An introduction to data analysis using S-Plus. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, 761p.

Faraway, J.J. 2002. Practical regression and ANOVA in R. Available online @ http://www.maths.bath.ac.uk/~jjf23/book/

Kuehl, R.O. 2000. Design of experiments: statistical principals of research design and analysis. Duxbury Press, New York, 666p.