15.1

16.1: Single Factor Experimental and Observational Studies

Single-factor experimental and observational studies are the most basic form of comparative studies. An example would be whether or not different dish detergents lead to cleaner plates, with the factor being detergent and the levels being different detergent brands.

If our single factor has r levels, we can use a linear model as a predictor, with $r-1$ indicator variables. Mathematically,

$$Y_{ij} = \beta_0 + \beta_1 X_{ij,1} + \ldots \beta_{r-1} X_{ij,r-1} + \varepsilon_{ij}$$

Where

$$X_{ij,r-1} = \begin{cases} 1 & \text{if treatment r-1} \\ 0 & \text{otherwise} \end{cases}$$

This is the **ANOVA model**. Note, we have i treatments, and replicate the experiment j times, hence the subscripts.

16.2: Relation between Regression and Analysis of Variance

2 major differences: in ANOVA, the explanatory/predictor variables may be qualitative. And 2, if the explanatory variables are quantitative we make no assumptions between them and the response variable, unlike in regression where we have a few major assumptions (assume linearity R^2 is a measure, F-test for lack of fit, assume no multicollinearity, i.e. the errors are not correlated use the variance inflation factor test, assume multivariate normal, i.e. the residuals/error are normal, check with a histogram of residuals or qq plot, t-tests are more thorough, can perhaps fix with Box-Cox transformation, and assume the residuals are homeoscedastic, i.e. they do not "fan out" and the variance is constant the Brown Forsythe test or F-test). Of course, there are tests to check all of those assumptions.

16.3 Singe-Factor ANOVA Model

The ANOVA model assumes that

1. Each probability distribution is normal.
2. Each probability distribution has the same variance.
3. The responses for each factor level are random selections from the corresponding probability distribution and are independent of the responses for any other factor level.

Essentially, what we do is determine whether or not the factor level means are the same, and if they do differ, we examine how they differ and what the implications are.

First, we explain the notation we shall use. r denotes the number of levels of the factor under study, and the index of these levels we denote by i. The number of cases for the ith factor level is denoted by n_i, and the total number of cases in the study is denoted by n_T, where

$$n_T = \sum_{i=1}^{r} n_i \tag{1}$$

So i is the not the trial, but the level (weird). Instead, we use the index j for the specific trial/case. Thus, we define the ANOVA model as

$$Y_{ij} = \mu_i + \varepsilon_{ij} \tag{2}$$

Y_{ij} is the value of the response variable in the jth trial for the ith factor level or treatment, μ_i are the parameters, and ε_{ij} are independent $n(0, \sigma^2)$. This is called the **cell means model**. Note,

1. The observed value of Y in the jth trial for the ith factor level or treatment is the sum of two components, a constant term μ_i and a random error term ε_{ij}.

2. Since $E(\varepsilon_{ij}) = 0$, then $E(Y_{ij}) = \mu_i$, thus all responses or observations Y_{ij} for the ith factor level have the same expectation μ_i, and this parameter is the mean response for the ith factor level or treatment.

3. Since μ_i is a constant, it follows that (from some thm):
$$\sigma^2(Y_{ij}) = \sigma^2(\varepsilon_{ij}) = \sigma^2$$

4. Since each ε_{ij} is normally distributed, so is each Y_{ij}.

5. The error terms are assumed to be independent. Hence, the error term for the outcome of any one trial has no effect on the error term for the outcome of any other trial for the same factor level or for a different factor level. Since ε_{ij} are independent, so are the responses Y_{ij}.

6. In view of these features, ANOVA model can be restated as follows
$$Y_{ij} \quad \text{are independent} \quad n(\mu_i, \sigma^2) \tag{3}$$

The ANOVA model is essentially a linear model, $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. The $\boldsymbol{\beta}$ vector contains the means of the "cells", here the factor levels, hence the name "cell-means model".

In an observational study, the factor level means μ_i correspond to the means for the different factor level populations. In experimental, the mean μ_i stands for the mean response that would be obtained if the ith treatment were applied to all units in the population of experimental units about which inferences are to be drawn, and similarly, the variance σ^2 refers to the variability of responses if any given experimental treatment were applied to the entire population of experimental units.

16.4:Fitting of ANOVA Model

The total of the observations for the ith factor level denoted by $Y_i.$ is:

$$Y_i. = \sum_{j=1}^{n_i} Y_{ij} \tag{4}$$

Note that the dot in $Y_i.$ indicates an aggregation over the j index, reference table 1 The

Table 1: The j units represent different stores, whereas the i represents the different levels, i.e. the different package designs.

Design	j=1	j=2	j=3	j=4	j=5	Total	Mean	Number of Stores
i	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}	$Y_i.$	$\overline{Y}_i.$	n_i
1	11	17	16	14	15	73	14.6	5
2	12	10	15	19	11	67	13.4	5
3	23	20	18	17		78	19.5	4
4	27	33	22	26	28	136	27.2	5
All designs						$Y.. = 354$	$\overline{Y}..$=18.63	19

sample mean for the ith factor level is denoted by $\overline{Y}_i.$.

$$\overline{Y}_i. = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i} = \frac{Y_i.}{n_i} \tag{5}$$

The total of all observations:

$$Y.. = \sum_{i=1}^{r} \sum_{j=1}^{n_i} Y_{ij} \tag{6}$$

where the two dots indicate aggregation over both the j and i indexes. Finally, the overall mean for all responses is:

$$\overline{Y}.. = \frac{\sum_i \sum_j Y_{ij}}{n_T} = \frac{Y..}{n_T} \tag{7}$$

We can write this as a weighted average of factor level means:

$$\overline{Y}.. = \sum_{i=1}^{r} \frac{n_i}{n_T} \overline{Y}_i.$$

For least squares and maximum likelihood estimators, we wanna minimize Q

$$Q = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \mu_i)^2 \tag{8}$$

We can minimize Q by minimizing each component sum of i separately. The sample mean minimizes the sum of squared error, so

$$\boxed{\hat{\mu}_i = \overline{Y}_i.} \tag{9}$$

Thus, the fitted value for the observation Y_{ij}, denoted by $\hat{Y}_{ij}$ for regression models, is simply the corresponding factor level mean here:

$$\hat{Y}_{ij} = \overline{Y}_{i\cdot}$$

The same estimators are obtained by the method of maximum likelihood.

An important property of the residuals for the ANOVA model is that they sum to zero for each factor level i

$$\sum e_{ij} = 0 \qquad i = 1, \dots, r$$

16.5 Analysis of Variance

As in regression, our goal is to partition total sum of squares into the regression sum of squares and the error sum of squares

$$\underbrace{Y_{ij} - \overline{Y}_{\cdot\cdot}}_{\text{total deviation}} = \underbrace{\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}}_{\substack{\text{deviation of} \\ \text{estimated factor level} \\ \text{mean around overall mean}}} + \underbrace{Y_{ij} - \overline{Y}_{i\cdot}}_{\substack{\text{deviation around} \\ \text{estimated factor level mean}}} \tag{10}$$

The total deviation can be viewed as the sum of two components, the deviation of estimated factor level mean around the overall mean and the deviation of Y_{ij} around its estimated factor level mean, which is simply the residual e_{ij}. When we square and then sum we find

$$\sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{\cdot\cdot})^2 = \sum_i n_i (\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2 + \sum_i \sum_j (Y_{ij} - \overline{Y}_{i\cdot})^2 \tag{11}$$

where with SSTO being the total sum of squares, the SSTR being the treatment sum of squares, and SSE the error sum of squares:

$$\text{SSTO} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{\cdot\cdot})^2 \tag{12}$$

$$\text{SSTR} = \sum_i n_i (\overline{Y}_{i\cdot} 0 \overline{Y}_{\cdot\cdot})^2 \tag{13}$$

$$\text{SSE} = \sum_i \sum_j (Y_{ij} - \overline{Y}_{i\cdot})^2 = \sum_i \sum_j e_{ij}^2 \tag{14}$$

and

$$\boxed{\text{SSTO} = \text{SSTR} + \text{SSE}} \tag{15}$$

The SSE is a measure of random variation of the observations around the respective estimated factor level means. The less variation among the observations for each factor level, the smaller is SSE. The more the observations for each factor level differ among themselves, the larger will be SSE. The SSTR is a measure of the extent of difference between the estimated factor level means, based on the deviations of the estimated factor level means $\overline{Y}_{i\cdot}$ around the overall mean $\overline{Y}_{\cdot\cdot}$.. The more the estimated factor level means differ, the larger will be SSTR.

We now examine the degrees of freedom. SSTO has $n_T - 1$ degrees of freedom, because 1 DOF is lost bc the deviations must sum to 0 and thus are not independent. SSTR has r-1 degrees of freedom, r estimated factor level mean deviations, but one degree of freedom is lost because the deviations are not independent because the weighted sum must equal zero. SSE has $n_T - r$ dof. We get this by looking at the component of SSE for the ith factor level

$$\sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i\cdot})^2$$

This expression is the equivalent of a total sum of squares considering only the ith factor level. Hence, there $n_i - 1$ dof associated with this sum of squares. Since SSE is a sum of component sum of squares, we sum the r component degrees of freedom, yielding $n_T - r$. The mean squares are the sum of squares over respective dof.

$$\text{MSE} = \frac{\text{SSE}}{n_T - r} \tag{16}$$

$$\text{MSTR} = \frac{\text{SSTR}}{r - 1} \tag{17}$$

Where MSTR is the treatment mean square and MSE is the mean square error. Expected values are

$$E(MSE) = \sigma^2 \tag{18}$$

$$E(MSTR) = \sigma^2 + \frac{\sum n_i(\mu_i - \mu_\cdot)^2}{r - 1} \tag{19}$$

where

$$\mu_\cdot = \frac{\sum n_i \mu_i}{n_T}$$

MSE is unbiased estimator of σ^2, the variance of error terms ε_{ij}, whether or not the factor level means are equal. When the factor level means are equal, E(MSTR)=σ^2. In table 2, we summarize what we've done so far in the ANOVA table for single factor study:

Table 2: ANOVA table.

Source of Variation	SS	df	MS	E(MS)
Between treatments	SSTR=$\sum n_i(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2$	r-1	MSTR=$\frac{\text{SSTR}}{r-1}$	$\sigma^2 + \frac{\sum n_i(\mu_i - \mu_\cdot)^2}{r-1}$
Error (within treatments)	SSE=$\sum\sum(Y_{ij} - \overline{Y}_{i\cdot})^2$	$n_T - r$	MSE=$\frac{\text{SSE}}{n_T - r}$	σ^2
Total	SSTO=$\sum\sum(Y_{ij} - \overline{Y}_{\cdot\cdot})^2$	$n_T - 1$		

16.6: F-test for Equality of Factor level means

Alternative conclusions we wish to consider:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_r$$
$$H_a : \text{not all } \mu_i \text{ are equal}$$

The test statistic to be used for choosing between the alternatives is

$$F^* = \frac{\text{MSTR}}{\text{MSE}} \tag{20}$$

When all treatment means μ_i are equal, each response Y_{ij} has the same expected value. In view of the additivity of sums of squares and degrees of freedom, Cochran's theorem then implies when H_0 holds, SSE/σ^2 and SSTR/σ^2 are independent χ^2 variables. It follows in the same fashion for regression, that when H_0 holds, F^* is distributed as $F(r-1, n_T - r)$. When H_a holds, i.e. when the μ_i are not all equal, then F^* does not follow the F-distribution. Usually, the risk of making a type I error is controlled in constructing the decision rule. Our decision rule is

$$\text{If} \quad F^* \le F(1-\alpha; r-1, n_T - r) \quad \text{conclude } H_0 \tag{21}$$
$$\text{If} \quad F^* \ge F(1-\alpha; r-1, n_T - r) \quad \text{conclude } H_a \tag{22}$$

where the α represent how wide our F-distribution is as always, i.e. how many deviations from center, or how far from the center in the corresponding F distribution would we have to be to have α % of the curve filled.

Alternative Formulations of Model

Factor effects model:

$$\mu_i \equiv \mu. + (\mu_i - \mu.) \tag{23}$$

where $\mu.$ is a constant that can be defined to fit the purpose of the study. We shall denote the difference by $\mu_i - \mu.$ by τ_i, so

$$\mu_i \equiv \mu. + \tau_i \tag{24}$$

The difference τ_i is called the ith factor level effect or the ith treatment effect. The ANOVA model can now be stated as

$$Y_{ij} = \mu. + \tau_i + \varepsilon_{ij} \tag{25}$$

The splitting up of the factor level means μ_i into two components, an overall constant $\mu.$ and a factor level or treatment effect τ_i, depends on the definition of $\mu.$, which can be defined in many ways. We can define $\mu.$ as an unweighted mean:

$$\mu. = \frac{\sum_{i=1}^r \mu_i}{r} \implies \sum_{i=1}^r \tau_i = 0 \tag{26}$$

We can also define as a weighted mean:

$$\mu. = \sum_{i=1}^{r} w_i \mu_i \qquad \text{where } \sum_{i=1}^{r} w_i = 1 \implies \sum_{i=1}^{r} w_i \tau_i = 0 \tag{27}$$

The test for equality of factor level means for the cell means model is

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_r$$
$$H_a : \text{not all } \mu_i \text{ are equal}$$

but is now for the factor effects model

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_r = 0 \tag{28}$$
$$H_a : \text{not all } \tau_i \text{ equal zero} \tag{29}$$

In a regression case, we find

$$E(Y_{ij}) = \mu. + \tau_i$$

The multiple linear regression model is

$$Y_{ij} = \mu. + \tau_1 X_{ij,1} + \ldots + \tau_{r-1} X_{ij,r-1} + \varepsilon_{ij} \qquad \text{Full Model} \tag{30}$$

where

$$X_{ij,r-1} = \begin{cases} 1 & \text{if case from factor level r-1} \\ -1 & \text{if case from factor level r} \\ 0 & \text{otherwise} \end{cases}$$

The least squares estimator of $\mu.$ is the average of the cell sample means:

$$\hat{\mu}. = \frac{\sum_{i=1}^{r} \overline{Y}_i.}{r}$$

The least squares estimator of the ith factor effect is:

$$\hat{\tau}_i = \overline{Y}_i. - \hat{\mu}.$$

The reduced model is

$$Y_{ij} = \mu. + \varepsilon_{ij}$$

And we use the usual test statistic to test whether or not there is a regression relation:

$$F^* = \frac{\text{MSR}}{\text{MSE}}$$

However, when we use the weighted mean, the weighted model becomes

$$Y_{ij} = \mu. + \tau_1 X_{ij,1} + \ldots + \tau_{r-1} X_{ij,r-1} + \varepsilon_{ij} \qquad \text{full model} \tag{31}$$

Where, because

$$\mu. = \sum_{i=1}^{r} w_i \mu_i = \sum_{i=1}^{r} \frac{n_i}{n_T} \mu_i \implies \sum_{i=1}^{r} \frac{n_i}{n_T} \tau_i = 0$$

which means

$$\tau_r = -\frac{n_1}{n_r}\tau_1 - \ldots - \frac{n_{r-1}}{n_r}\tau_{r-1}$$

so:

$$X_{ij,r-1} = \begin{cases} 1 & \text{if case from factor level r-1} \\ -\frac{n_{r-1}}{n_r} & \text{if case from factor level r} \\ 0 & \text{otherwise} \end{cases}$$

This leads to identical results for the general linear test

$$H_0 : \tau_1 = \ldots = \tau_{r-1}$$
$$H_a : \text{not all equal}$$

The cell means model is the easiest case in the regression context

$$Y_{ij} = \mu_1 X_{ij,1} + \ldots + \mu_r X_{ij,r} + \varepsilon_{ij} \qquad \text{full model} \tag{32}$$

where there are r indicator variables because of no constraint and thus

$$X_r = \begin{cases} 1 & \text{if case from factor level r} \\ 0 & \text{otherwise} \end{cases}$$

Randomization Tests

Don't assume ε_{ij} to be independent normal random variables with mean zero and constant variance σ^2, rather we now consider each ε_{ij} to be a fixed effect associated with the experimental unit. In this framework, we view the n_T experimental units to be a finite population, and associated with each unit is the unit specific effect ε_{ij}. When the randomization assigns this experimental unit to treatment i, the observed response will be $Y_{ij} = \mu. + \tau_i + \varepsilon_{ij}$. The response is still a random variable, but under the randomization view the randomness arises because the treatment effect τ_i is the result of random assignment of the experimental unit to treatment i.

If there are no treatment effects, that is if all $\tau_i = 0$, then the response $Y_{ij} = \mu. + \varepsilon_{ij}$ depends only on the experimental unit. Since the randomization the experimental unit is equally likely to be assigned to any treatment, the observed response Y_{ij}, if there are no treatment effects, could with equal likelihood have been observed for any of the treatments. Thus, when there are no treatment effects, randomization will lead to an assignment of the finite population of n_T observations Y_{ij} to the treatments such that all treatment combinations of observations are equally likely. This, in turn, leads to an exact sampling distribution of the test statistic under $H_0 : \tau_i \equiv 0$, sometimes termed the randomization distribution of the test statistic. Percentiles of the randomization distribution can then be used to test for the presence of factor effects. This use of the randomization distribution provides the basis of a nonparametric test for treatment effects.

Planning of Sample Sizes with Power Approach

We want to plan sample sizes for protection against both type I and type II errors. By the power of F test for single factor study, we refer to the probability that the decision rule will lead to conclusion H_a, that the treatment means differ, when in fact H_a holds. Specifically, the power is given by the following expression for the cell means model (16.2):

$$\text{Power} = P(F^* > F(1-\alpha; r-1, n_T - r)|\phi) \tag{33}$$

where ϕ is the *noncentrality parameter*, that is, a measure of how unequal the treatment means μ_i are:

$$\phi = \frac{1}{\sigma}\sqrt{\frac{\sum n_i(\mu_i - \mu.)^2}{r}} \tag{34}$$

and

$$\mu. = \frac{\sum n_i \mu_i}{n_T}$$

Power probabilities are determined by utilizing the noncentral F distribution since this is the sampling distribution of F^* when H_a holds. Use table B.11 in Applied Linear Statistical Models textbook, where the vast majority of these notes are from. To use the table, note $\nu_1 = r-1$ and $\nu_2 = n_T - r$, and α is the risk you control at.

Using the table for single factor studies: We denote the minimum range by

$$\Delta = \max(\mu_i) - \min(\mu_i)$$

The following three specifications need to be made in using table B.12:

1. The level α at which risk of making a type I error is to be controlled.
2. The magnitude of the minimum range Δ of the μ_i which is important to detect with high probability. The magnitude of σ, the standard deviation of the probability distributions of Y, must also be specified since entry into table B.12 is in terms of the ratio Δ/σ.
3. The level β at which the risk of making a type II error is to be controlled for the specification given in 2. Entry into table B.12 is in terms of the power $1-\beta$.

Finally, we have

$$\min \sum_{i=1}^{r} (\mu_i - \mu.)^2 = \frac{\delta^2}{2}$$

16.11: Planning of Sample Sizes to Find "Best" Treatment

Table B.13, developed by Bechhofer, allows us to determine necessary sample sizes so that the probability $1-\alpha$ the highest (or lowest) estimated treatment mean is from the treatment with the highest (or lowest) population mean. We must specify the probability $1-\alpha$, the standard deviation σ, and the smallest difference λ between the highest (lowest) and second highest (second lowest) treatment means that it is important to recognize. The entry into the table is $\lambda\sqrt{n}/\sigma$, where λ is the difference between highest and second highest factor level means. We can use the table with known λ and desired σ to calculate minimum n.

17.3:Estimation and Testing of Factor Level Means

Inferences for Single Factor Level Mean

Estimation: An unbiased point estimator of the factor level mean is

$$\hat{\mu}_i = \overline{Y}_{i\cdot}$$

Which has mean and variance

$$E(\overline{Y}_{i\cdot}) = \mu_i \qquad \sigma^2(\overline{Y}_{i\cdot}) = \frac{\sigma^2}{n_i}$$

the second term is proven because we have n_i identical variables with variance σ^2. But, the mean $\overline{Y}_{i\cdot} = \frac{\sum Y_{i\cdot}}{n}$, and in general for independent variables X_i, $\text{Var}(\sum X_i/n) = \sum \text{Var}(X_i)/n^2$ by properties of the variance and the independence of the X_i's.

The estimated variance is

$$s^2(\overline{Y}_{i\cdot}) = \frac{\text{MSE}}{n_i} \tag{35}$$

It can be shown that $\frac{\overline{Y}_{i\cdot}-\mu_i}{s(\overline{Y}_{i\cdot})}$ is distributed as $t(n_T - r)$ for ANOVA model. The degrees of freedom are those associated with the MSE. Further, for the ANOVA model, SSE/σ^2 is distributed as χ^2 with $n_T - r$ degrees of freedom, as we saw earlier, and is independent of $\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{r\cdot}$. Then, it follows that our $1-\alpha$ confidence limits are

$$\overline{Y}_{i\cdot} \pm t(1-\alpha/2; n_T - r)s(\overline{Y}_{i\cdot}) \tag{36}$$

The confidence interval can be used to test the hypothesis:

$$H_0 : \mu_i = c \tag{37}$$
$$H_a : \mu_i \neq c \tag{38}$$

where c is an appropriate constant. We conclude H_0, at level of significance α, when c is contained in the confidence interval, and conclude H_a otherwise, as is typical. Alternatively, and importantly equivalently (wow a lot of ly's there). The test statistic is:

$$t^* = \frac{\overline{Y}_{i\cdot} - c}{s(\overline{Y}_{i\cdot})} \tag{39}$$

This statistic follows a t-distribution with $n_T - r$ degrees of freedom when H_0 is true. We conclude H_0 whenever $|t^*| \leq t(1-\alpha/2; n_T - r)$.

Inferences for Difference Between two Factor Level Means

Define

$$D = \mu_i - \mu_{i'} \tag{40}$$

This is a pairwise comparison. A point estimator of D, denoted by $\hat{D}$ is

$$\hat{D} = \overline{Y}_{i\cdot} - \overline{Y}_{i'} \tag{41}$$

Which is unbiased. The variance is

$$\sigma^2(\hat{D}) = \sigma^2(\overline{Y}_{i\cdot}) + \sigma^2(\overline{Y}_{i'\cdot}) = \sigma^2\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right) \tag{42}$$

and of course

$$s^2(\hat{D}) = \text{MSE}\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)$$

because $\hat{D}$ is a linear combination of independent normal variables, it is normally distributed. Then, $\frac{\hat{D}-D}{s(\hat{D})}$ is distributed as $t(n_T - r)$ for ANOVA model. The $1-\alpha$ confidence limits for D are

$$\hat{D} \pm t(1-\alpha/2; n_T - r)s(\hat{D}) \tag{43}$$

The hypothesis test is

$$H_0 : \mu_i = \mu_{i'} \tag{44}$$

$$H_a : \mu_i \neq \mu_{i'} \tag{45}$$

Alternatively and equivalently we could test whether or not the difference in mean equals zero. The test statistic is

$$t^* = \frac{\hat{D}}{s(\hat{D})} \tag{46}$$

And conclusion is reached if

$$|t^*| \leq t(1-\alpha/2; n_T - r)$$

Inferences for Contrast of Factor Level Means

A **contrast** is a comparison involving two or more factor level means and includes the previous case of a pairwise difference between two factor levels.

$$L = \sum_{i=1}^{r} c_i \mu_i \qquad \text{where} \quad \sum_{i=1}^{r} c_i = 0 \tag{47}$$

An unbiased estimator of a contrast L is

$$\hat{L} = \sum_{i=1}^{r} c_i \overline{Y}_{i\cdot} \tag{48}$$

Since $\overline{Y}_{i\cdot}$ are independent, the variance of $\hat{L}$ is

$$\sigma^2(\hat{L}) = \sum_{i=1}^{r} c_i^2 \sigma^2(\overline{Y}_{i\cdot}) = \sum_{i=1}^{r} c_i^2 \left(\frac{\sigma^2}{n_i}\right) = \sigma^2 \sum_{i=1}^{r} \frac{c_i^2}{n_i} \tag{49}$$

An unbiased estimator of this variance is

$$s^2(\hat{L}) = \text{MSE} \sum_{i=1}^{r} \frac{c_i^2}{n_i} \tag{50}$$

$\hat{L}$ is also normally distributed because it is simply a linear combo if indy normal RV. Therefore, $\frac{\hat{L}-L}{s(\hat{L})}$ is distributed as $t(n_T - r)$ for ANOVA model. And, as we have seen many times already, the confidence limits for L are

$$\hat{L} \pm t(1-\alpha/2; n_T - r)s(\hat{L}) \tag{51}$$

The hypothesis test is of the form:

$$H_0 : L = 0 \tag{52}$$
$$H_a : L \neq 0 \tag{53}$$

The test statistic is

$$t^* = \frac{\hat{L}}{s(\hat{L})} \tag{54}$$

Of course, if $|t^*| \leq t(1-\alpha/2; n_T - r)$, then H_0 is concluded.

A linear combination of factor level means has no restriction on the constants c. There, the hypothesis test is

$$H_0 : \sum c_i \mu_i = c$$
$$H_a : \sum c_i \mu_i \neq c$$

We use the test statistic F^* which is just $(t^*)^2$ and follows the $F(1, n_T - r)$ distribution when H_0 holds.

17.5: Tukey Multiple Comparison Procedure

We now move to simultaneous inferences. We start with the **Tukey multiple comparison procedure**. The family of interest is the set of all pairwise comparisons of factor level means. The hypothesis test is

$$H_0 : \mu_i - \mu_{i'} = 0$$
$$H_a : \mu_i - \mu_{i'} \neq 0$$

When all sample sizes are equal, the family confidence coefficient for the Tukey method is exactly $1-\alpha$ and the family significance level is exactly α. When the sample sizes are not equal, the family confidence coefficient is greater than $1-\alpha$ and the family significance level is less than α. In other words, the Tukey procedure is conservative when the sample sizes are not equal.

The Tukey procedure utilizes the *studentized range distribution*.

$$w = \max(Y_i) - \min(Y_i)$$

The ratio of the range over the estimate of the standard deviation is the *studentized range*

$$q(r, \nu) = \frac{w}{s}$$

We use table B.9 to find this.

The Tukey multiple comparison confidence limits for all pairwise comparisons D with family confidence coefficient of at least $1-\alpha$ are as follows:

$$\hat{D} \pm Ts(\hat{D}) \tag{55}$$

where we saw earlier that

$$s^2(\hat{D}) = \text{MSE}\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)$$

and T, the Tukey multiple, is

$$T = \frac{1}{\sqrt{2}} q(1-\alpha; r; n_T - r) \tag{56}$$

The test is

$$H_0 : \mu_i - \mu_{i'} = 0$$
$$H_a : \mu_i - \mu_{i'} \neq 0$$

The test statistic is

$$q^* = \frac{\sqrt{2}\hat{D}}{s(\hat{D})} \tag{57}$$

If you want to take that equivalent route.

We can also plot this with a *paired comparison plot*. When the sample sizes are equal, we plot intervals around each treatment mean $\overline{Y}_{i\cdot}$ as

$$\hat{Y}_{i\cdot} \pm \frac{1}{2} Ts(\hat{D})$$

Scheffe Multiple Comparison Method

When the family of interest is the set of all possible contrasts among the factor level means, we use the Scheffe procedure. The Scheffe confidence intervals for the family of contrasts L are of the form

$$\hat{L} \pm Ss(\hat{L}) \tag{58}$$

where the estimated variance is $s^2(\hat{L}) = \text{MSE} \sum \frac{c_i^2}{n_i}$ and the Scheffe multiple comes from

$$S^2 = (r-1)F(1-\alpha; r-1; n_T - r) \tag{59}$$

The test statistic when we test whether or not the contrasts equal zero is

$$F^* = \frac{\hat{L}^2}{(r-1)s^2(\hat{L})} \tag{60}$$

If only pairwise comparisons are to be made, the Tukey procedure gives narrower bands. However, the Scheffe procedure has the property that if the F test of factor level equality indicates that the factor level means μ_i are not equal, the corresponding Scheffe multiple comparison procedure will find at least one contrast (out of all possible contrasts) that differs significantly from zero (the confidence interval does not cover zero). It may be, though, that this contrast is not one of those that has been estimated.

Bonferroni Multiple Comparison Procedure

The **Bonferroni multiple comparison procedure** is of use when the family of interest is a particular set of pairwise comparisons, contrasts, or linear combinations that is specified by user in advance of the data analysis.

We shall denote the number of statements in the family by g and treat them all as linear combinations since pairwise comparisons and contrasts are special cases of linear combinations. Then, we have that the Bonferroni inequality implies that the confidence coefficient is at least $1-\alpha$ that the following confidence limits for the g linear combinations L are all correct:

$$\hat{L} \pm Bs(\hat{L}) \tag{61}$$

where

$$B = t(1-\alpha/2g; n_T - r) \tag{62}$$

The hypothesis test is

$$H_0 : L = 0$$
$$H_a : L \neq 0$$

The test statistic, if you do not want to go the route of confidence intervals is

$$t^* = \frac{\hat{L}}{s(\hat{L})} \tag{63}$$

If $|t^*| \leq t(1-\alpha/2g; n_T - r)$, we conclude H_0.

If all pairwise comparisons are of interest, Tukey is better than Bonferroni. Bonferroni is better than Scheffe when number of contrasts of interest is about the same as number of factor levels or less. Compare the multiples.

Analysis of Means

We can use Bonferroni in analysis of means. Recall, $\hat{\tau}_i = \overline{Y}_{i\cdot} - \hat{\mu}_\cdot$ where $i = 1, \dots, r$ and $\hat{\mu}_\cdot = \frac{\sum \overline{Y}_{i\cdot}}{r}$. Also,

$$s^2(\hat{\tau}_i) = \frac{\text{MSE}}{n_i}\left(\frac{r-1}{r}\right)^2 + \frac{\text{MSE}}{r^2}\sum_{h\neq i}\frac{1}{n_h}$$

The limits are

$$\hat{\mu}_\cdot \pm t(1-\alpha/2r; n_T - r)s(\hat{\tau}_i)$$

17.8: Planning Sample Sizes with Estimation Approach

We know the variance of an estimated contrast $\hat{L}$ when $n_i \equiv n$ is:

$$\sigma^2(\hat{L}) = \frac{\sigma^2}{n}\sum c_i^2$$

If we know σ based on past experience, set n, and know which contrasts we have and thus our c_i, then we can anticipate the width of our confidence interval using the Scheffe method for example. If we have *unequal sample sizes*, we can use a similar methodology. Perhaps it is best seen with an example:

We use data from problem 16.12 in the textbook. Suppose primary interest is in estimating the following comparisons:

$$L_1 = \mu_1 - \mu_2 \qquad L_3 = \frac{\mu_1 + \mu_2}{2} - \mu_5$$
$$L_2 = \mu_3 - \mu_4 \qquad L_4 = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

What would be the required sample sizes if the precision of each of the estimated comparisons is not to exceed ± 1.0 day, using the most efficient multiple comparison procedure with a 90 percent family confidence coefficient?

We use the estimated variance of an estimated contrast. Because all the sample sizes are equal, $s^2\left\{\hat{L}_i\right\} = \frac{\text{MSE}}{n}\sum_i c_i^2$. Therefore, using the MSE as our estimate of σ^2,

$$s^2\left\{\hat{L}_1\right\} = \frac{2 \cdot 7.52}{n} = \frac{15.04}{n}$$
$$s^2\left\{\hat{L}_2\right\} = \frac{2 \cdot 7.52}{n} = \frac{15.04}{n}$$
$$s^2\left\{\hat{L}_3\right\} = \frac{1.5 \cdot 7.52}{n} = \frac{11.28}{n}$$
$$s^2\left\{\hat{L}_4\right\} = \frac{1 \cdot 7.52}{n} = \frac{7.52}{n}$$

$\hat{L}_1$ (and equivalently $\hat{L}_2$) has the widest width, so that is all we need to check. We compare the multiples. We do this at $n_T = 100$, the total sample size, to give us an idea as to which comparison procedure is most efficient.

$$B = t(1 - .1/(2*4), 95) = 2.27$$
$$S = \sqrt{(r-1) \cdot F(.9, 4, 95)} = 2.83$$
$$T = \frac{1}{\sqrt{2}} q(0.9; 4, 95) = 2.32$$

Bonferroni is the way to go. Therefore, we find n such that

$$Bs\{\hat{L}_1\} \approx\leq 1 \longrightarrow 2 \cdot \sqrt{15.04} < \sqrt{n}$$

This means $n < 77.4$. We now assume $\boxed{77}$ is the correct answer. This means that $n_T = 5 \cdot 77 = 385$. We thus find that the Bonferroni multiple is $t(.9875, 385) = 2.25$, which yields a precision of 0.988. Using n=76 yields 1.002, slightly outside our window.

For unequal samples sizes:

Refer to problem 16.9. Suppose that primary interest is in comparing the below-average and above-average physical fitness groups, respectively, with the average physical fitness group. The two comparisons are of interest:

$$L_1 = \mu_1 - \mu_2 \qquad L_2 = \mu_3 - \mu_2$$

Assume that a reasonable planning value for the error standard deviation is $\sigma = 4.5$ days.

It has been decided to use equal samples sizes (n) for the below average and above-average groups. If the twice the sample size (2n) were to be used for the average physical fitness group, what would be the required sample sizes if the precision of each pairwise comparison is to be ± 2.5 days, using the Bonferroni procedure and a 90 percent family confidence coefficient?

We note that for both cases, because twice the sample size is used to the average physical fitness group, μ_2, noting that $n_3 = 2n$ and $n_2 = n$, using the equation for variance of an estimated difference $\hat{L}_i$:

$$\sigma^2\left\{\hat{L}_i\right\} = \sigma^2\left(\frac{1}{n_2} + \frac{1}{n_3}\right) = \frac{3\sigma^2}{2n}$$

With $\sigma^2 = 4.5^2$ and $n = 10$, we find $n_T = 10+10+2(10) = 40$, and $B = (1-.9/(2*2), 40-3) = 2.02$. This yields precision (remembering to take square root of $\sigma^2\left\{\hat{L}_1\right\}$) $B\sigma\left\{\hat{L}_1\right\} = 3.5$, which is above our threshold. If we try $n = 20$, $n_T = 80$, and we get $B\sigma\left\{\hat{L}_1\right\} = 2.49$. So $n = 20$, and $n_1 = n_3 = 20$, $n_2 = 40$, and $n_T = 40$.

Analysis of Factor Effects when Factor is Quantitative

When the factor under investigation is quantitative, the analysis of factor effects can be carried beyond the point of multiple comparisons to include a study of the anture of the response function.

Start with ANOVA model

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

The hypothesis test is

$$H_0 : \mu_1 = \ldots = \mu_r$$
$$H_a : \text{not all } \mu_i \text{ are equal}$$

The test statistic is

$$F^* = \frac{\text{MSTR}}{\text{MSE}}$$

We compare to $F^* \leq F(1-\alpha; r-1, n_T - r)$ we conclude the null.

We can also use a lack of fit test to check if our regression function (if we run one) is valid. Say we fit a quadratic to data. The hypothesis test for lack of fit is:

$$H_0 : E(Y) = \beta_0 + \beta_1 x + \beta_{11} x^2$$
$$H_a : E(Y) \neq \beta_0 + \beta_1 x + \beta_{11} x^2$$

and the test statistic is

$$F^* = \frac{\text{MSLF}}{\text{MSPE}}$$

where SSLF=SSE-SSPE, SSLF is the lack of fit error, and SSPE is the pure error. In general,

$$\text{SSPE} = \sum (Y_{ij} - \overline{Y}_j)^2$$

and

$$\text{SSLF} = \sum(\overline{Y}_j - \hat{Y}_{ij})^2 = \sum_j n_j(\overline{Y}_j - \hat{Y}_{ij})^2$$

SSLF has r-2 dof, and SSPE has $n-r$ degrees of freedom, which we divide by to find the mean.

18: ANOVA Diagnostics and Remedial Measures

Goals of this chapter include to examine whether the proposed model is appropriate for the set of data at hand. If the proposed model is not appropriate, consider remedial measures, such as transformation of the data or modification of the model. After review of the appropriateness of the model and completion of any necessary remedial measures and an evaluation of their effectiveness, inferences based on the model can be undertaken.

18.1 Residual Analysis

Recall, our residuals for cell means model are

$$e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \overline{Y}_i.$$

The semi-studentized residuals are

$$e^*_{ij} = \frac{e_{ij}}{\sqrt{\text{MSE}}} \tag{64}$$

The studentized residuals are

$$r_{ij} = \frac{e_{ij}}{s(e_{ij})} \tag{65}$$

where

$$s(e_{ij}) = \sqrt{\frac{\text{MSE}(n_i - 1)}{n_i}} \tag{66}$$

and finally, the studentized deleted residuals become here:

$$t_{ij} = e_{ij}\left[\frac{n_T - r - 1}{\text{SSE}\left(1 - \frac{1}{n_1}\right) - e^2_{ij}}\right]^{1/2} \tag{67}$$

We consider residual plots to diagnose the following departures from the ANOVA model.

1. Non-constancy of error variance. The ANOVA model requires that error terms ε_{ij} have constant variance for all factor levels. When the samples sizes are not large and do not differ greatly, the appropriateness of the assumption can be studied by using the residuals, semistudentized residuals, or studentized residuals. Plots of residuals against fitted values or dot plots of residuals are helpful. When sample sizes are large, histograms/boxplots are preferable.

2. Nonindependence of error terms. Residual sequence plot, when data obtained in time sequence, is a good idea to check whether or not the error terms are serially correlated. This can be repeated if we have other logical orderings.

3. Outliers.

   The detection of outliers is facilitated by various plots of the studentized deleted residuals. Residual plots against fitted values, residual dot plots, box plots, and stem-and-leaf plots are nice. The test for outliers in regression is applicable here, with the appropriate Bonferroni critical value being $t(1-\alpha/2n_T; n_T - r - 1)$. If the largest absolute studentized deleted residual exceeds this critical value, that case should be considered an outlier. Occasionally, a test for an outlier is suggested in advance of the analysis, as when a substitute operator is used for one of the production runs in a manufacturing experiment Concern about the validity of this response observation might lead to an outlier test. In this case, the Bonferroni critical value would be $t(1-\alpha/2; n_T - r - 1)$.

4. Omission of important explanatory variables. Use plots to see if single factor is appropriate, for example if experiment differs according to gender of subject, perhaps a second factor would be appropriate.

5. Non-normality of error terms.

   Histograms, dot plots, box plots, and normal probability plots are appropriate diagnostic tools here.

Tests for Constancy of Error Variance

Hartley Test

The test considers r normal populations; the variance of the ith population is denoted by σ_i^2. Independent samples of equal size are selected from the r populations: the sample variance for the ith population is denoted by s_i^2 and the common number of degrees of freedom associated with each sample variance is denoted by df. The hypothesis test is

$$H_0 : \sigma_1^2 = \ldots = \sigma_r^2 \tag{68}$$

$$H_a : \text{not all } \sigma_i^2 \text{ are equal} \tag{69}$$

The Hartley test statistic, denoted by H^* is based solely on the largest sample variance, denoted by $\max(s_i^2)$, and the smallest sample variance, $\min(s_i^2)$, so

$$H^* = \frac{\max(s_i^2)}{\min(s_i^2)} \tag{70}$$

The appropriate decision rule for controlling the risk of making a type I error at α is if $H^* \leq H(1-\alpha, r, df)$ conclude H_0, where df is the degrees of freedom associated with each sample variance and $H(1-\alpha; r, df)$ is the $1-\alpha)100$ percentile of the distribution of H^* when H_0 holds for r populations and df degrees of freedom for each sample variance.

When the Hartley test is used for single-factor ANOVA model with equal sample sizes, i.e. $n_i \equiv n$, we have $df = n-1$. The r normal populations are the normal probability distributions of the Y observations for the r factor levels. The sample variance s_i^2 is the variance of the n_i observations Y_{ij} for the ith factor level or equivalently the variance of the n_i residuals e_{ij}; for $n_i \equiv n$, s_i^2 becomes

$$s_i^2 = \frac{\sum_{j=1}^{n}(Y_{ij} - \overline{Y}_{i\cdot})^2}{n-1} = \frac{\sum_{j=1}^{n} e_{ij}^2}{n-1} \tag{71}$$

Brown-Forsythe Test

Recall, we used the Brown-Forsythe test for constancy of error variance in regression. Like the Hartley test, we can us BF to study the equality of r population variances. Unlike the Hartley test, the BF-test is robust against departures from normality, which often occur together with unequal variances. This test also does not require equal sample sizes, which is nice.

To test the alternatives

$$\begin{aligned} H_0 &: \sigma_1^2 = \ldots = \sigma_r^2 \\ H_a &: \text{not all } \sigma_i^2 \text{ are equal} \end{aligned}$$

using the BF-test, we first compute the absolute deviations of the Y_{ij} observations about their respective factor level medians $\overline{Y}_{i\cdot}$. The BF-test then determines whether or not the expected values of the absolute deviations for the r treatments are equal. If the r error variances σ_i^2 are equal, so will the expected values of the absolute deviations. The BF-test statistic is simply the ordinary F^* statistic for testing differences in treatment means, but now based on the absolute deviations:

$$F_{BF}^* = \frac{\text{MSTR}}{\text{MSE}}$$

where:

$$\text{MSTR} = \frac{\sum n_i(\overline{d}_{i\cdot} - \overline{d}_{\cdot\cdot})^2}{r-1} \tag{72}$$

$$\text{MSE} = \frac{\sum\sum(d_{ij} - \overline{d}_{i\cdot})^2}{n_T - r} \tag{73}$$

$$\overline{d}_{i\cdot} = \frac{\sum_j d_{ij}}{n_i} \tag{74}$$

$$\overline{d}_{\cdot\cdot} = \frac{\sum\sum d_{ij}}{n_T} \tag{75}$$

If the error terms have constant variance and the factor level sample sizes are not extremely small, F_{BF}^* follows approximately an F distribution with $r-1$ and $n_T - r$ degrees of freedom. Large F_{BF}^* values indicate that the error terms do not have constant variance.

Weighted Least Squares

When the errors are normally distributed but their variances are not the same for the different factor levels, cell means model becomes

$$Y_{ij} = \mu_i + \varepsilon_{ij} \tag{76}$$

where ε_{ij} are independent $n(0, \sigma^2)$. Weighted least squares is a standard remedial measure here, like in regression. Since the factor level variances σ_i^2 are usually unknown, they must be estimated. This is ordinarily done by means of the sample variances s_i^2, in which case the weight w_{ij} for the jth case of the ith factor level is

$$w_{ij} = \frac{1}{s_i^2} \tag{77}$$

The test for the equality of the factor level means is now conducted by general linear test approach described in regression case. The full model is fitted, using the weights, and the error sum of squares is obtained. Next, the reduced model under H_0 is fitted and the error sum of squares is obtained. We denote these $\text{SSE}_w(F)$ and $\text{SSE}_w(R)$ respectively.

The general linear test statistic here is:

$$F_w^* = \frac{\text{SSE}_w(R) - \text{SSE}_w(F)}{r-1} \nabla \cdot \frac{\text{SSE}_w(F)}{n_T - r} \tag{78}$$

Since the weights are based on the estimated variances s_i^2, the distribution of F_w^* under H_0 is only approximately an F-distribution with $r-1$ and $n_T - r$ degrees of freedom. When the factor level sample sizes are reasonably large, the approximation generally is satisfactory.

Transformation of Response Variable

When the variance of the error terms for each factor level, denoted by σ_i^2 is proportional to the factor level mean μ_i, a square root transformation is helpful, i.e. if σ_i^2 is proportional to μ_i, then $Y' = \sqrt{Y}$. If σ is proportional to μ_i^2, then $Y' = \ln Y$ is good. If the response if a proportion, then $Y' = 2\arcsin\left(\sqrt{Y}\right)$ is appropriate.

In general, the statistics $s_i^2/\overline{Y}_{i\cdot}, s_i/\overline{Y}_{i\cdot}, s_i/\overline{Y}_{i\cdot}^2$ should be calculated for each factor level, where s_i^2 is the sample variance of the Y observations for the ith factor level. Approximate constancy of one of the three statistics over all factor levels would suggest the corresponding transformation as useful for stabilizing the error variance and making the error distribution more nearly normal.

The Box-Cox procedure can help us identify a power transformation of the type Y^λ to correct for both normality and non-constancy of the error variance. We do not have to pick the max likelihood if there is an easier to interpret λ value that may not be the max estimate but is close to it and the area of the max estimate is relatively flat.

18.6: Effects of Departure from Model

For the fixed ANOVA model I, lack of normality is not an important matter, provided the departure from normality is not extreme. *Kurtosis* of the error distribution (either more or less peaked than a normal distribution) is more important than skewness of the distribution in terms of effects on inferences. The point estimators of factor level means and contrasts are unbiased whether or not the populations are normal. The F test for the equality of factor level means is but little affected by lack of normality, either in terms of the level of significance or power of the test. Hence, the F test is a robust test against departures from normality. For instance, while the specified level of significance might be 0.05, the actual level for a non-normal error distribution might be slightly higher or lower. Typically, the achieved level of significance in the presence of non-normality is slightly higher than the specified one, and the achieved power of the test is slightly less than the calculated one.

When the error variances are unequal, the F test for the equality of means with the fixed ANOVA model is only slightly affected if all factor level sample sizes are equal or do not differ greatly. Specifically, unequal error variances then raise the actual level of significance slightly higher than the specified level. Similarly, the Scheffe multiple comparison procedure based on the F distribution is not affected to any substantial effect by unequal variances when sample sizes are approximately equal. Thus, the F test and related analyses are robust against unequal variances when the sample sizes are roughly equal. Single comparisons between factor level means, on the other hand, can be substantially affected by unequal variances, so that the actual and specified confidence coefficients may differ in this case, *bigly*. The use of equal sample sizes for all factor levels not only tends to minimize the effects on unequal variances on inferences with the F distribution but also simplifies calculational procedures.

Lack of independence of the error terms can have serious effects on inferences in the analysis of variance, both for fixed and random ANOVA models. Difficult to correct, design good experiments.

18.7: Nonparametric Rank F Test

When transformations arent enough. The Y_{ij} observations need to be ranked in ascending order from 1 to n_T. We shall let R_{ij} denote the rank of Y_{ij}. In the case of ties, each of the tied observations is given the mean of the ranks involved. To test whether the treatment means are equal, the usual F^* test statistic is based on the ranks R_{ij}. The test statistic is now denoted by F_R^*:

$$F_R^* = \frac{\text{MSTR}}{\text{MSE}} \tag{79}$$

where:

$$\text{MSTR} = \frac{\sum n_i(\overline{R}_{i\cdot} - \overline{R}_{\cdot\cdot})^2}{r-1} \tag{80}$$

$$\text{MSE} = \frac{\sum\sum(R_{ij} - \overline{R}_{i\cdot})^2}{n_T - r} \tag{81}$$

$$\overline{R}_{i\cdot} = \frac{\sum_j R_{ij}}{n_i} \tag{82}$$

$$\overline{R}_{\cdot\cdot} = \frac{\sum\sum R_{ij}}{n_T} = \frac{n_T+1}{2} \tag{83}$$

Note that $\overline{R}_{\cdot\cdot}$, the overall mean of the ranks, is a constant for any given total number of cases n_T. When the treatment means are the same, the test statistic F_R^* follows approximately the $F(r-1, n_T - r)$ distribution provided that the sample sizes n_i are not very small. To test the alternatives:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_r$$
$$H_a : \text{not all } \mu_i \text{ are equal}$$

The appropriate decision rule to control type I error at α is if $F_R^* \leq F(1-\alpha; r-1, n_T - r)$ we conclude H_0, and H_a otherwise.

For multiple pairwise testing procedure,

$$(\overline{R}_{i\cdot} - \overline{R}_{i'\cdot}) \pm B\left[\frac{n_T(n_T+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)\right]^{1/2} \tag{84}$$

where

$$B = z(1-\alpha/2g) \tag{85}$$

and

$$g = \frac{r(r-1)}{2} \tag{86}$$

19 Two-Factor Studies

For example, we can study gender and age on time spent on learning a task. Factors: Factor A is gender with 2 levels, male=1 and female=2, and B is age with levels young=1, middle=2, and old=3. We note that μ_{ij} is the true treatment mean when the level of factor A is i and that of factor B is j, where $i = 1, \ldots, a; j = 1, \ldots, b$. We also note that $\mu_{i\cdot} = \sum_j \mu_{ij}/b$ the row average when $i = 1, \ldots, a$. $\mu_{\cdot j} = \sum_i \mu_{ij}/a$: the column average $j = 1, \ldots, b$. $\mu_{\cdot\cdot} = \sum_i \sum_j \mu_{ij}/b$: the overall mean. The main effects are defined as (for the two factor study):

$$\alpha_i = \mu_{i\cdot} - \mu_{\cdot\cdot} \quad \text{main effect for factor A at ith level} \tag{87}$$

$$\beta_j = \mu_{\cdot j} - \mu_{\cdot\cdot} \quad \text{main effect for factor B at jth level} \tag{88}$$

Note that $\sum_{i=1}^a \alpha_i = 0$ and $\sum_{j=1}^b \beta_j = 0$. Reference table 3

Table 3: Mean learning time (in minutes)

	Factor B-Age			
Factor A-Gender	j=1(young)	j=2(middle)	j=3(old)	Row average
i=1 Male	9(μ_{11})	11 (μ_{12})	16(μ_{13})	12($\mu_{1\cdot}$)
i=2 Female	9(μ_{21})	11 (μ_{22})	16(μ_{23})	12($\mu_{2\cdot}$)
Column Average	9($\mu_{\cdot 1}$)	11 ($\mu_{\cdot 2}$)	16($\mu_{\cdot 3}$)	12($\mu_{\cdot\cdot}$)

We see that the main gender effects are $\alpha_1 = \mu_{1\cdot} - \mu_{\cdot\cdot} = 12 - 12 = 0$ and $\alpha_2 = \mu_{2\cdot} - \mu_{\cdot\cdot} = 12 - 12 = 0$. For the age effects, we have

$$\begin{aligned}\beta_1 &= \mu_{\cdot 1} - \mu_{\cdot\cdot} = 9 - 12 = -3\\ \beta_2 &= \mu_{\cdot 2} - \mu_{\cdot\cdot} = 11 - 12 = -1\\ \beta_3 &= \mu_{\cdot 3} - \mu_{\cdot\cdot} = 16 - 12 = 4\end{aligned}$$

In this example, the factor effects were additive. Each mean response μ_{ij} can be obtained by taking the respective gender and age main effects to the overall mean $\mu_{\cdot\cdot}$. In general, for additive factor effects:

$$\mu_{ij} = \mu_{\cdot\cdot} + \alpha_i + \beta_j \tag{89}$$

which can be expressed as

$$\mu_{ij} = \mu_{i\cdot} + \mu_{\cdot j} - \mu_{\cdot\cdot} \tag{90}$$

In can also be shown that for each treatment mean μ_{ij} that

$$\mu_{ij} = \mu_{ij'} + \mu_{i'j} - \mu_{i'j'} \quad i \neq i', j \neq j' \tag{91}$$

Graphs and plots

Interaction plots feature our response on the Y axis and on the X axis either factor. Parallel curves indicate no interaction, and the different in height shows the effects of the second factor. Check the picture in figure 1 On the left is the example we did. The factor effects are

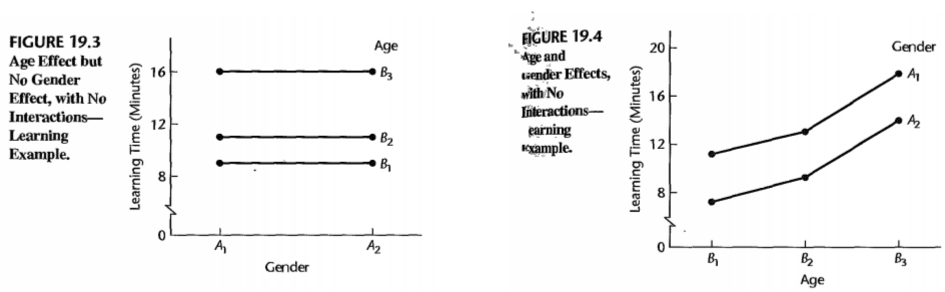

Figure 1: Interaction Plot

additive and the plot shows this. On the right, however, we have age and gender effects , indicated by the plot. The table is given in table 4

Table 4: Mean learning time (in minutes)

	Factor B-Age			
Factor A-Gender	j=1(young)	j=2(middle)	j=3(old)	Row average
i=1 Male	11(μ_{11})	13 (μ_{12})	18(μ_{13})	14($\mu_{1\cdot}$)
i=2 Female	7(μ_{21})	9 (μ_{22})	14(μ_{23})	10($\mu_{2\cdot}$)
Column Average	9($\mu_{\cdot 1}$)	11 ($\mu_{\cdot 2}$)	16($\mu_{\cdot 3}$)	12($\mu_{\cdot\cdot}$)

For factor A, $\alpha_1 = 1; \alpha_2 = -1$ and for factor B, $\beta_1 = -3; \beta_2 = -1; \beta_3 = 4$. The interaction effect is

$$(\alpha\beta)_{ij} = \mu_{ij} - (\mu_{\cdot\cdot} + \alpha_i + \beta_j) \tag{92}$$

which we can equivalently write as

$$(\alpha\beta)_{ij} = \mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{\cdot\cdot}$$

and

$$\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = \sum_i \sum_j (\alpha\beta)_{ij} = 0 \tag{93}$$

Hence, our interaction is

	j=1	j=2	j=3
i=1	-1	0	1
i=2	1	0	-1

Of course, there are other types of interaction effects such as that $\mu_{ij} - \mu_{i'j}$ can be different for different j. In this next example, table 5, $\mu_{11} - \mu_{21} = 0 \neq \mu_{12} - \mu_{22} = 2$, so the gender effects are different across age groups. Similarly, the age effects are different across genders.

Table 5: Mean learning time (in minutes)

	Factor B-Age			
Factor A-Gender	j=1(young)	j=2(middle)	j=3(old)	Row average
i=1 Male	9(μ_{11})	12 (μ_{12})	18(μ_{13})	13($\mu_{1\cdot}$)
i=2 Female	9(μ_{21})	10 (μ_{22})	14(μ_{23})	11($\mu_{2\cdot}$)
Column Average	9($\mu_{\cdot 1}$)	11 ($\mu_{\cdot 2}$)	16($\mu_{\cdot 3}$)	12($\mu_{\cdot\cdot}$)

Fitting of ANOVA Model

The cell means model becomes

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \tag{94}$$

where $i = 1\dots,a$, $j = 1,\dots,b$, and $k = 1,\dots n$.

For the factor effects model,

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \tag{95}$$

where $\mu = \mu_{..}$, the overall mean. We are subject to the following constraints. $\sum_i \alpha_i = 0$, $\sum_j \beta_j = 0$, $\sum_i (\alpha\beta)_{ij} = 0$, and $\sum_j (\alpha\beta)_{ij} = 0$.

For the cell means model, we obtain $\hat{\mu}_{ij}$ by minimizing

$$Q = \sum_i \sum_j \sum_k (Y_{ijk} - \mu_{ij})^2 \tag{96}$$

which yields solutions $\hat{\mu}_{ij} = \overline{Y}_{ij} = \hat{Y}_{ijk}$, the fitted value of the observations of the ijth group. Our residual is

$$e_{ijk} = Y_{ijk} - \hat{Y}_{ijk} = Y_{ijk} - \overline{Y}_{ij} \tag{97}$$

For the factor effects model we have

$$\text{Parameter} \longrightarrow \text{Estimator} \tag{98}$$
$$\mu_{..} \longrightarrow \hat{\mu}_{..} = \overline{Y}_{...} \tag{99}$$
$$\alpha_i = \mu_{i.} - \mu_{..} \longrightarrow \hat{\alpha}_j = \overline{Y}_{i..} - \overline{Y}_{...} \tag{100}$$
$$\beta_j = \mu_{.j} - \mu_{..} \longrightarrow \hat{\beta}_j = \overline{Y}_{.j.} - \overline{Y}_{...} \tag{101}$$
$$(\alpha\beta)_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..} \longrightarrow \widehat{(\alpha\beta)}_{ij} = \overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...} \tag{102}$$

We have the following ANOVA table in table 6 for the two factor design.

Table 6: ANOVA table 2 factor study. The mean squares are the sums divided by their respective degrees of freedom, no need for senseless columns.

Source of Variation	SS	df	E(MS)
Factor A	SSA=$nb\sum\left(\overline{Y}_{i..} - \overline{Y}_{...}\right)^2$	$a-1$	$\sigma^2 + nb\frac{\sum_i(\mu_{i.}-\mu_{..})^2}{a-1}$
Factor B	SSB=$na\sum\left(\overline{Y}_{.j.} - \overline{Y}_{...}\right)$	$b-1$	$\sigma^2 + na\frac{\sum_j(\mu_{.j}-\mu_{..})^2}{b-1}$
AB interactions	SSAB=$n\sum\sum\left(\overline{Y}_{ij.} - \overline{Y}_{j..} - \overline{Y}_{.j.} + \overline{Y}_{...}\right)^2$	$(a-1)(b-1)$	$\sigma^2 + n\frac{\sum_i\sum_j\left(\mu_{ij}-\mu_{i.}-\mu_{.j}+\mu_{..}\right)^2}{(a-1)(b-1)}$
Error	SSE=$\sum_{i=1}^a\sum_{j=1}^b\sum_{k=1}^n\left(Y_{ijk} - \overline{Y}_{ij.}\right)^2$	$ab(n-1)$	σ^2
Total	SSTO=$\sum_{i=1}^a\sum_{j=1}^b\sum_{k=1}^n\left(Y_{ijk} - \overline{Y}_{...}\right)^2$	$nab-1$	

This leads us to our hypothesis tests:

$$H_{i0} : \text{there is no effect} \tag{103}$$
$$H_{ia} : \text{there is an effect} \tag{104}$$

For interaction effects

$$H_0 : \text{all } (\alpha\beta)_{ij} = 0$$
$$H_a : \text{not all } (\alpha\beta)_{ij} \text{ equal zero}$$

The test statistic is

$$F^* = \frac{\text{MSAB}}{\text{MSE}} \tag{105}$$

our rule is if $F^* \leq F\big(1-\alpha;(a-1)(b-1),(n-1)(ab)\big)$, then we conclude H_0.

To test factor A main effects and factor B are respectively

$$H_0 : \alpha_1 = \ldots = \alpha_a = 0 \tag{106}$$
$$H_a : \text{not all } \alpha_i \text{ equal zero} \tag{107}$$

$$H_0 : \beta_1 = \ldots = \beta_b = 0 \tag{108}$$
$$H_a : \text{not all } \beta_j \text{ equal zero} \tag{109}$$

Their test statistics are respectively

$$F_A^* = \frac{\text{MSA}}{\text{MSE}} \qquad F_B^* = \frac{\text{MSB}}{\text{MSE}} \tag{110}$$

For the A case, if $F_A^* \leq F\big(1-\alpha;a-1,(n-1)ab\big)$ we conclude H_0 and similarly if $F_B^* \leq F\big(1-\alpha;b-1,(n-1)ab\big)$ we conclude H_0.