STAT 3119

Week2: 9/3/2019 @GWU

Outline Week 2 (Chapter 16)

In the following 14 weeks, we will start to go over different types of designs, models and analysis of the data.

- A1. Design of single factor experimental studies
- A2. Single factor ANOVA model, estimation and testing
- A3. Alternative formulation of the ANOVA model
- A4. Data analysis example with R
- B1. Relationship: ANOVA vs. Regression
- B2. A nonparametric randomization test
- B3. Sample size and power planning in single factor studies

2. SINGLE-FACTOR Experimental studies

Single-Factor ANOVA Model (Ch 16.3)

The basic elements of the ANOVA model for a single-factor study are quite simple. Corresponding to each factor level, there is a probability distribution of responses.

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.

The analysis of the sample data proceeds in two steps:

1. Determine whether or not the factor level means are the same.
2. If the factor level means differ, examine how they differ and what the implications of the differences are.

1st ANOVA model: Cell Means Model

Notation:

r = the number of levels of the factor under study ($i = 1, ..r$ levels)

n_i = The number of cases for the i th factor level

n_T = the number of cases in the study, so $n_T = \sum_{i=1}^{r} n_i$.

Y_{ij} = the value of the response variable in the jth cases for the ith factor level ($j = 1, ..n_i$).

The ANOVA model can now be stated as follows:

$$Y_{ij} = \mu_i + \epsilon_{ij} \qquad (16.2)$$

where ${}_i$ are mean response for the i th factor level or treatment,

ϵ_{ij} (error terms) are independent $N(0, \sigma^2)$, $i = 1, ..r$; $\quad j = 1, ..n_i$.

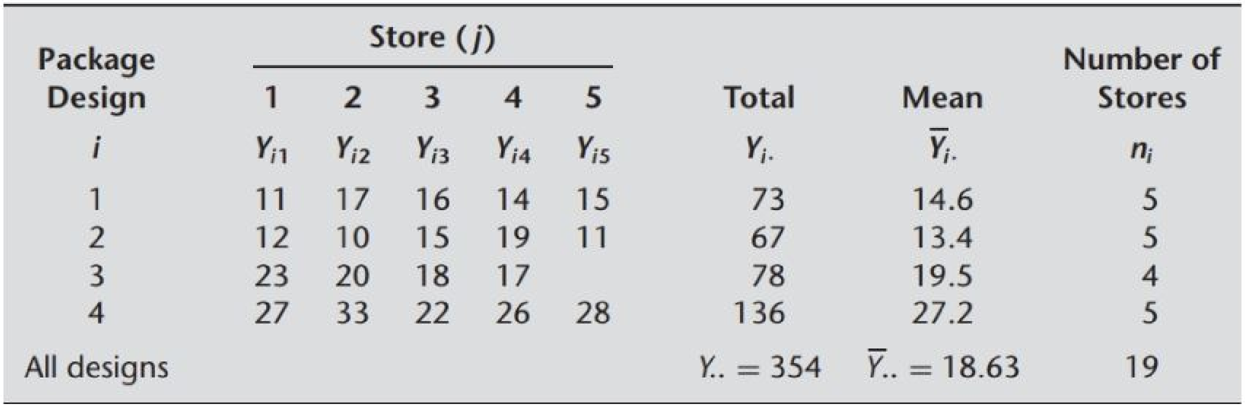
Note: This model implies: $E(Y_{ij}) = \mu_i$, $\quad var(Y_{ij}) = var(\epsilon_{ij}) = \sigma^2$ (all the observations have same variance, regardless of factor levels); $\quad Y_{ij}$ are independent with $N(\mu_i, \sigma^2)$.

Fitting of ANOVA Model (Ch 16.4)

We need to estimate the parameters in model (16.2) such as μ_i and σ^2. For the normal error models, the least square method or the method of maximum likelihood lead to the same estimators for the model parameters.

Example 16.4:

The Kenton Food Company wished to test four different package designs for a new breakfast cereal. Twenty stores, with approximately equal sales volumes, were selected as the experimental units. Each store was randomly assigned one of the package designs, with each package design assigned to five stores. A fire occurred in one store during the study period, so this store had to be dropped from the study. The stores were chosen to be comparable in location and sales volume. Other relevant conditions that could affect sales, such as price, amount and location of shelf space, and special promotional efforts, were kept the same for all of the stores in the experiment. **Sales** (Y), in number of cases, were observed for the study period, and the results are recorded in Table 16.1. This study is a **completely randomized design** with package design as the **single, four-level factor**.



Package Design	Store (j)					Total	Mean	Number of Stores
	1	2	3	4	5			
i	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}	$Y_{i.}$	$\bar{Y}_{i.}$	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						$Y_{..} = 354$	$\bar{Y}_{..} = 18.63$	19

Estimate the mean levels for the factor

Notation:

Factor level total = $Y_{i.} = \sum_{j=1}^{n_i} Y_{ij}$,

Sample mean for ith factor level = $\overline{Y_{i.}} = \sum_{j=1}^{n_i} Y_{ij}/n_i = Y_{i.}/n_i$,

Note the dot in $Y_{i.}$ indicates an aggregation over the j index for a given level i.

Overall total = $Y_{..} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} Y_{ij}$,

Overall mean= $\overline{Y_{..}} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} Y_{ij}/n_T = Y_{..}/n_T = \sum_{i=1}^{r} n_i \overline{Y_{i.}}/n_T$ (a weighted average of the factor level means)

Note the two dots indicate aggregation over both the j and i indexes.

It follows from LS or MLE methods, the best estimators of the mean response μ_i is $\overline{Y_{i.}}$, the sample mean for i factor level.

Fitting data with R

Read data and add labels

```
#Ex16 <- read.table(file.choose(), header=F) # find the file location from file browser
Ex16 <- read.table(url("https://raw.githubusercontent.com/npmldabook/Stat3119/master/Week2/CH16_TA01.txt"))
head(Ex16, 2) # show 2 lines of data
```

```
##   V1 V2 V3
## 1 11  1  1
## 2 17  1  2
```

```
names(Ex16)<- c("sales", "package", "stores")
Ex16$package<- as.factor(Ex16$package)
str(Ex16) #19 obs
```

```
## 'data.frame':    19 obs. of  3 variables:
##  $ sales  : int  11 17 16 14 15 12 10 15 19 11 ...
##  $ package: Factor w/ 4 levels "1","2","3","4": 1 1 1 1 1 2 2 2 2 2 ...
##  $ stores : int  1 2 3 4 5 1 2 3 4 5 ...
```

```
head(Ex16, 2)
```

```
##   sales package stores
## 1    11       1      1
## 2    17       1      2
```

Data checking with basic plots (R)

Check the factor levels:

```
levels(Ex16$package)
```

```
## [1] "1" "2" "3" "4"
```

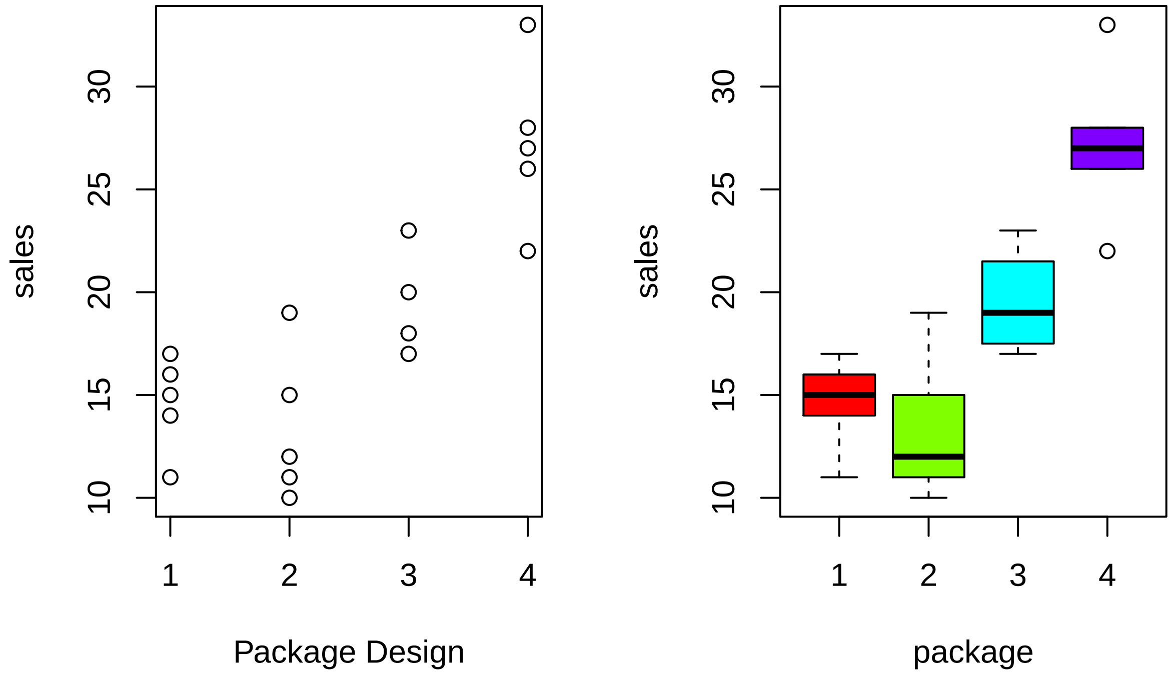
Check the sample size for each level:

```
table(Ex16$package)
```

```
##
## 1 2 3 4
## 5 5 4 5
```

Plots of response Y_{ij} by the r factor levels using a stripe-chart or boxplots:

```
par(mfrow=c(1,2))
stripchart(sales ~ package, vertical = TRUE,  pch=1, data = Ex16, xlab="Package Design")
boxplot(sales ~ package, data = Ex16, col=rainbow(4))
```



Estimate mean responses (R)

- Method 1: get sample mean for each given factor level (same as Table 16.1)

```r
with(Ex16,  by( sales, package, mean))
```

```
## package: 1
## [1] 14.6
## --------------------------------------------------------
## package: 2
## [1] 13.4
## --------------------------------------------------------
## package: 3
## [1] 19.5
## --------------------------------------------------------
## package: 4
## [1] 27.2
```

- Use a model: useful later for other parameterization, note: package is a factor variable (using numeric will result in an error message)

```r
fit <- aov(sales ~ package, data = Ex16)
predict(fit, newdata = data.frame(package = factor(1:4)))
```

```
##    1    2    3    4
## 14.6 13.4 19.5 27.2
```

- Residuals= $Y_{ij} - \overline{Y_{i.}}$, can be generated directly for the 19 observations (same as Table 16.2), which will be useful for the model diagnostics in later chapters.

```r
fit$residuals
```

```
##    1    2    3    4    5    6    7    8    9   10   11   12   13   14   15
## -3.6  2.4  1.4 -0.6  0.4 -1.4 -3.4  1.6  5.6 -2.4  3.5  0.5 -1.5 -2.5 -0.2
##   16   17   18   19
##  5.8 -5.2 -1.2  0.8
```

Testing in One-way ANOVA Model

The question we are answering with ANOVA is: are the variations between the sample means $\overline{Y_{i.}}$ due to true differences about the populations factor level means or just due to sampling variability?

To determine whether or not, the factor level means μ_i are equal, we set up the following hypotheses:

$$H_0:\ \mu_1 = \mu_2 = ... = \mu_r, \quad vs.$$

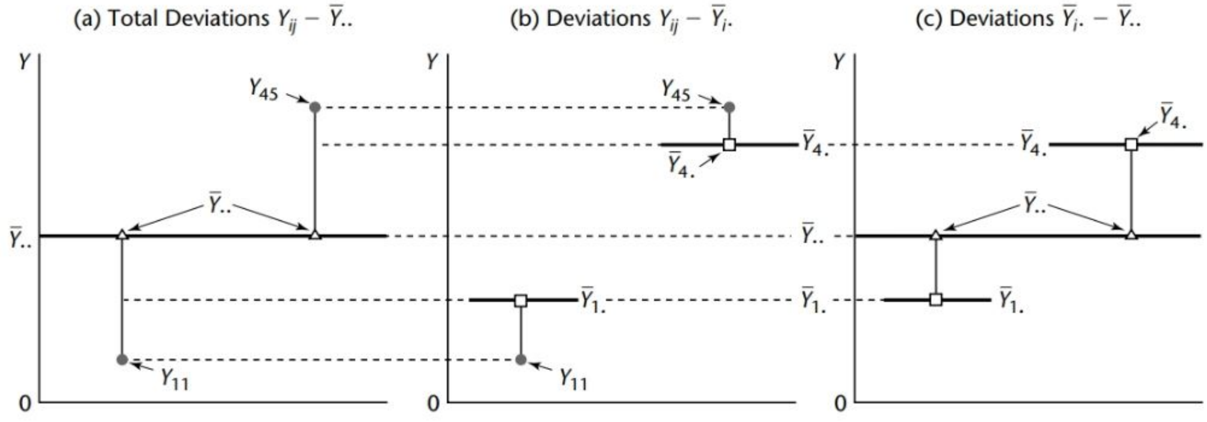
$$H_a: \text{not all} \quad \mu_i \quad \text{are equal.}$$

To test such hypotheses, we construct a F-test to test such hypotheses by variance partition to compare "the variation among the sample means (between-group deviations)" with "variation within groups (within-group deviations)".

Variance Partition (1)

1. Total deviation Yij − Y.. can be viewed as the sum of two components:

$$\underbrace{Y_{ij} - \overline{Y}_{..}}_{\substack{\text{Total}\\ \text{deviation}}} = \underbrace{\overline{Y}_{i.} - \overline{Y}_{..}}_{\substack{\text{Deviation of}\\ \text{estimated}\\ \text{factor level}\\ \text{mean around}\\ \text{overall mean}}} + \underbrace{Y_{ij} - \overline{Y}_{i.}}_{\substack{\text{Deviation}\\ \text{around}\\ \text{estimated}\\ \text{factor}\\ \text{level mean}}} \qquad \textbf{(16.25)}$$


(a) Total Deviations $Y_{ij} - \overline{Y}_{..}$
(b) Deviations $Y_{ij} - \overline{Y}_{i.}$
(c) Deviations $\overline{Y}_{i.} - \overline{Y}_{..}$

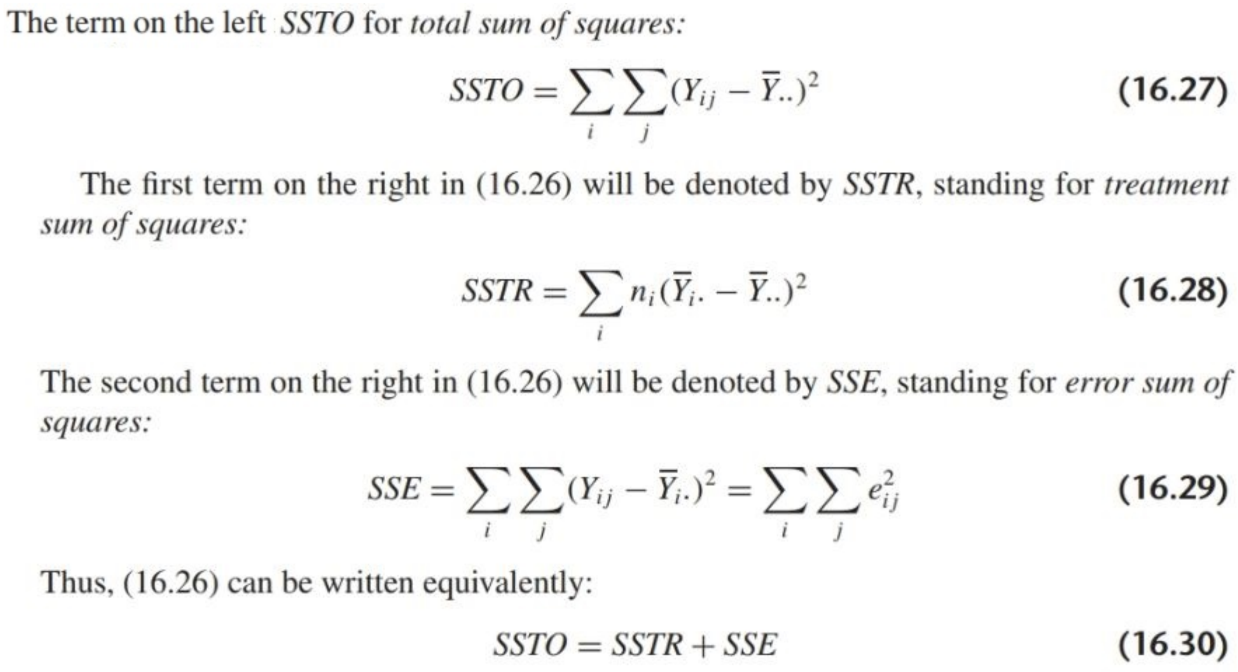
2. To square both sides and sum up, the cross-product = 0 then drop out (see (16.33)),tgus we have

$$\sum_i \sum_j (Y_{ij} - \overline{Y}_{..})^2 = \sum_i n_i(\overline{Y}_{i.} - \overline{Y}_{..})^2 + \sum_i \sum_j (Y_{ij} - \overline{Y}_{i.})^2 \qquad \textbf{(16.26)}$$

Variance Partition (2)

$$\sum_i \sum_j (Y_{ij} - \overline{Y}_{..})^2 = \sum_i n_i(\overline{Y}_{i.} - \overline{Y}_{..})^2 + \sum_i \sum_j (Y_{ij} - \overline{Y}_{i.})^2 \qquad \textbf{(16.26)}$$

Define SSTO, SSTR, SSE:



The term on the left *SSTO* for *total sum of squares:*

$$SSTO = \sum_i \sum_j (Y_{ij} - \overline{Y}..)^2 \qquad \textbf{(16.27)}$$

The first term on the right in (16.26) will be denoted by *SSTR*, standing for *treatment sum of squares:*

$$SSTR = \sum_i n_i (\overline{Y}_i. - \overline{Y}..)^2 \qquad \textbf{(16.28)}$$

The second term on the right in (16.26) will be denoted by *SSE*, standing for *error sum of squares:*

$$SSE = \sum_i \sum_j (Y_{ij} - \overline{Y}_i.)^2 = \sum_i \sum_j e_{ij}^2 \qquad \textbf{(16.29)}$$

Thus, (16.26) can be written equivalently:

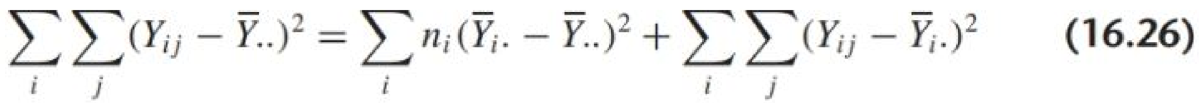
$$SSTO = SSTR + SSE \qquad \textbf{(16.30)}$$

Note: The total sum of squares for the ANOVA model = two components

1. **SSE**: A measure of the 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. If SSE = 0, the observations for any given factor level are all the same, and this holds for all factor levels.

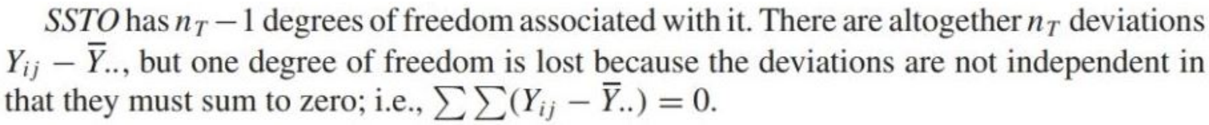
2. **SSTR**: A measure of the extent of differences between the estimated factor level means, based on the deviations of the estimated factor level means $\overline{Y_{i.}}$ around the overall mean $\overline{Y_{..}}$. If all estimated factor level means $\overline{Y_{i.}}$ are the same, then SSTR = 0. The more the estimated factor level means differ, the larger will be SSTR.

Breakdown of degrees of freedom

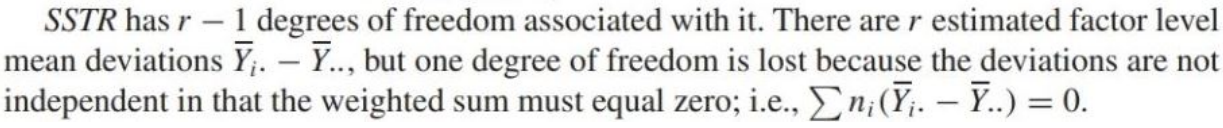


$$\sum_i \sum_j (Y_{ij} - \overline{Y}..)^2 = \sum_i n_i (\overline{Y}_i. - \overline{Y}..)^2 + \sum_i \sum_j (Y_{ij} - \overline{Y}_i.)^2 \qquad \textbf{(16.26)}$$

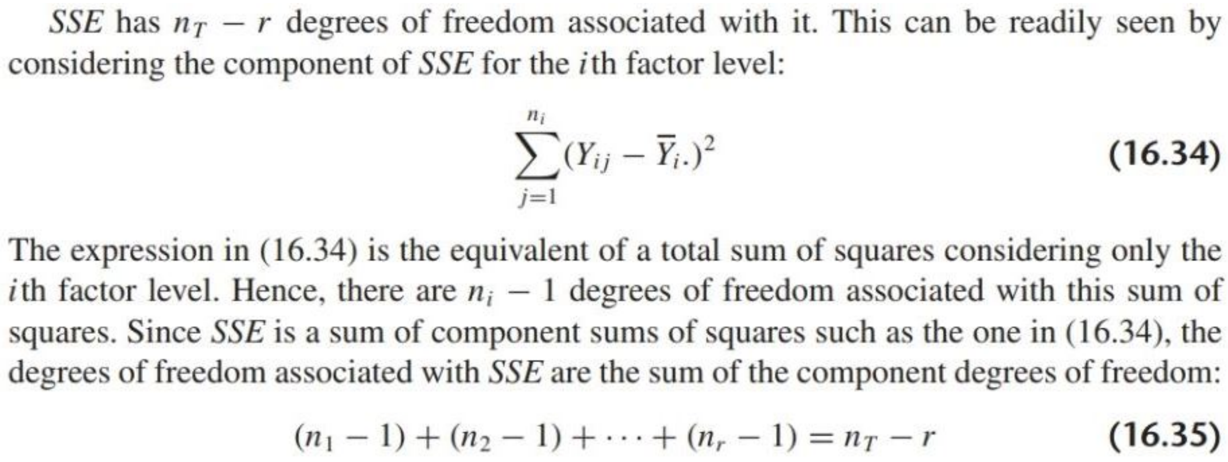
The breakdown of the dfs:



SSTO has $n_T - 1$ degrees of freedom associated with it. There are altogether n_T deviations $Y_{ij} - \overline{Y}..$, but one degree of freedom is lost because the deviations are not independent in that they must sum to zero; i.e., $\sum \sum (Y_{ij} - \overline{Y}..) = 0$.



SSTR has $r - 1$ degrees of freedom associated with it. There are r estimated factor level mean deviations $\overline{Y}_i. - \overline{Y}..$, but one degree of freedom is lost because the deviations are not independent in that the weighted sum must equal zero; i.e., $\sum n_i (\overline{Y}_i. - \overline{Y}..) = 0$.



SSE has $n_T - r$ degrees of freedom associated with it. This can be readily seen by considering the component of *SSE* for the ith factor level:

$$\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\cdot})^2 \qquad \textbf{(16.34)}$$

The expression in (16.34) is the equivalent of a total sum of squares considering only the ith factor level. Hence, there are $n_i - 1$ degrees of freedom associated with this sum of squares. Since *SSE* is a sum of component sums of squares such as the one in (16.34), the degrees of freedom associated with *SSE* are the sum of the component degrees of freedom:

$$(n_1 - 1) + (n_2 - 1) + \cdots + (n_r - 1) = n_T - r \qquad \textbf{(16.35)}$$

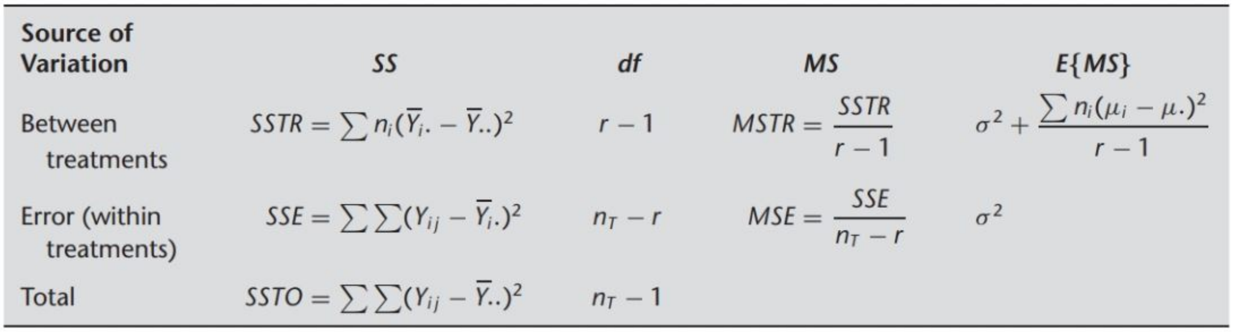
Mean squares

Define Mean squares= SS/df:

$$\textit{Treatment Mean Square} = MSTR = SSTR/(r-1)$$
$$\textit{Error Mean Square} = MSE = SSE/(n_T - r)$$

Expected Mean Squares E(MS) and ANOVA table:

TABLE 16.3 ANOVA Table for Single-Factor Study.



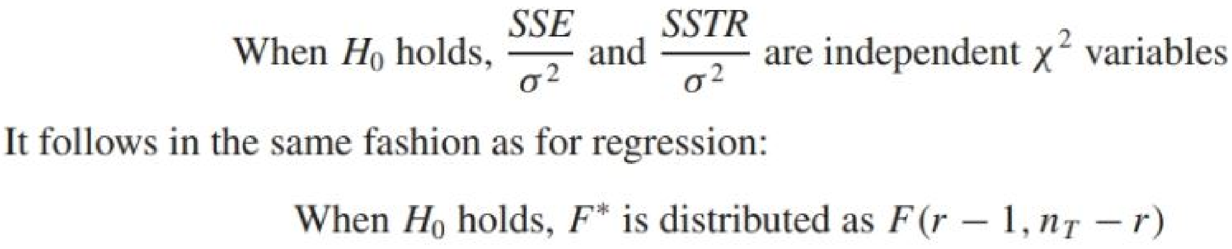
Source of Variation	SS	df	MS	E{MS}
Between treatments	$SSTR = \sum n_i(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2$	$r-1$	$MSTR = \frac{SSTR}{r-1}$	$\sigma^2 + \frac{\sum n_i(\mu_i - \mu_\cdot)^2}{r-1}$
Error (within treatments)	$SSE = \sum\sum(Y_{ij} - \overline{Y}_{i\cdot})^2$	$n_T - r$	$MSE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

Note:

1. MSE is an unbiased estimator of σ^2, the variance of the error terms ij , whether or not the factor level means i are equal.
2. When all factor level means i are equal and hence equal to the weighted mean ., then E{MSTR}=σ^2. When, however, the factor level means are not equal, MSTR tends on the average to be larger than MSE, since the second term will then be positive.
3. Therefore, if MSTR and MSE are of the same order of magnitude, this is taken to suggest that the factor level means i are equal. If MSTR is substantially larger than MSE, this is taken to suggest that the i are not equal.

F test for Equality of Factor Level Means

$$\text{Test statistic } F^* = MSTR/MSE$$



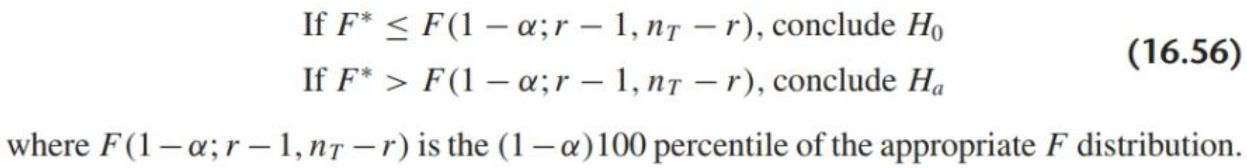
$$\text{When } H_0 \text{ holds, } \frac{SSE}{\sigma^2} \text{ and } \frac{SSTR}{\sigma^2} \text{ are independent } \chi^2 \text{ variables}$$

It follows in the same fashion as for regression:

$$\text{When } H_0 \text{ holds, } F^* \text{ is distributed as } F(r-1, n_T - r)$$

- **Large values of F^* support H_a, since MSTR will tend to exceed MSE when H_a holds. Values of F^* near 1 support H_0, since both MSTR and MSE have the same expected value when H_0 holds. Hence, this F test is an upper-tail one.**

- **Decision rule: The appropriate decision rule to control the level of significance at α is**



$$\begin{aligned} &\text{If } F^* \leq F(1-\alpha; r-1, n_T - r), \text{ conclude } H_0 \\ &\text{If } F^* > F(1-\alpha; r-1, n_T - r), \text{ conclude } H_a \end{aligned} \qquad \textbf{(16.56)}$$

where $F(1-\alpha; r-1, n_T - r)$ is the $(1-\alpha)100$ percentile of the appropriate F distribution.

Cautionary note(!): In hypothesis testing, we don't usually conclude H_0 unless this is study designed with adequate power. Otherwise, the study might have lower statistical power to reject null and establish the alternative hypothesis or a treatment difference. We would say that we can't reject the null hypothesis (for low F values).

ANOVA analysis (R)

1. Fit data and generate the ANOVA table for SS, MS, F-test, P-value is simple in R.

```
fit <- aov(sales ~ package, data = Ex16)
summary(fit)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## package      3  588.2  196.07   18.59 2.58e-05 ***
## Residuals   15  158.2   10.55                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Questions: Can you verify those results of SS, MS, F-test from the formulas that we are given?

2. Following the steps that we learned so far, it is straightforward to calcuate these quantities in R or with a calculator.

- Calculate the total deviation, between group and within group deviation:

```
Overall.mean   = mean(Ex16$sales)
Total.dev = Ex16$sales- Overall.mean

Factor.mean   =   rep(as.numeric(with(Ex16,  by( sales, package, mean))), table( Ex16$package))
Tr.dev = Factor.mean - Overall.mean
Error.dev = Ex16$sales -Factor.mean
round(data.frame(sales=Ex16$sales, Overall.mean, Factor.mean, Total.dev, Tr.dev, Error.dev),2)
```

```
##    sales Overall.mean Factor.mean Total.dev Tr.dev Error.dev
## 1     11        18.63        14.6     -7.63  -4.03      -3.6
## 2     17        18.63        14.6     -1.63  -4.03       2.4
## 3     16        18.63        14.6     -2.63  -4.03       1.4
## 4     14        18.63        14.6     -4.63  -4.03      -0.6
## 5     15        18.63        14.6     -3.63  -4.03       0.4
## 6     12        18.63        13.4     -6.63  -5.23      -1.4
## 7     10        18.63        13.4     -8.63  -5.23      -3.4
## 8     15        18.63        13.4     -3.63  -5.23       1.6
## 9     19        18.63        13.4      0.37  -5.23       5.6
## 10    11        18.63        13.4     -7.63  -5.23      -2.4
## 11    23        18.63        19.5      4.37   0.87       3.5
## 12    20        18.63        19.5      1.37   0.87       0.5
## 13    18        18.63        19.5     -0.63   0.87      -1.5
## 14    17        18.63        19.5     -1.63   0.87      -2.5
## 15    27        18.63        27.2      8.37   8.57      -0.2
## 16    33        18.63        27.2     14.37   8.57       5.8
## 17    22        18.63        27.2      3.37   8.57      -5.2
## 18    26        18.63        27.2      7.37   8.57      -1.2
## 19    28        18.63        27.2      9.37   8.57       0.8
```

- Calculate SS, MS and F in the Table :

```
SSO = sum(Total.dev^2)
SSTR = sum(Tr.dev^2)
SSE = sum(Error.dev^2)

MSTR= sum(Tr.dev^2)/(4-1)
MSE= sum(Error.dev^2)/(19-4)

Fstat=  MSTR/MSE

round(unlist(list(SSO=SSO,SSTR=SSTR, SSE=SSE, MSTR=MSTR, MSE=MSE, Fstat=Fstat)),2)
```

```
##    SSO   SSTR    SSE   MSTR    MSE  Fstat
## 746.42 588.22 158.20 196.07  10.55  18.59
```

- To get the critical value of $F(A = (1-\alpha); \textit{numerator df}, \textit{den. df})$, you can use Table B.4 in the appendix.

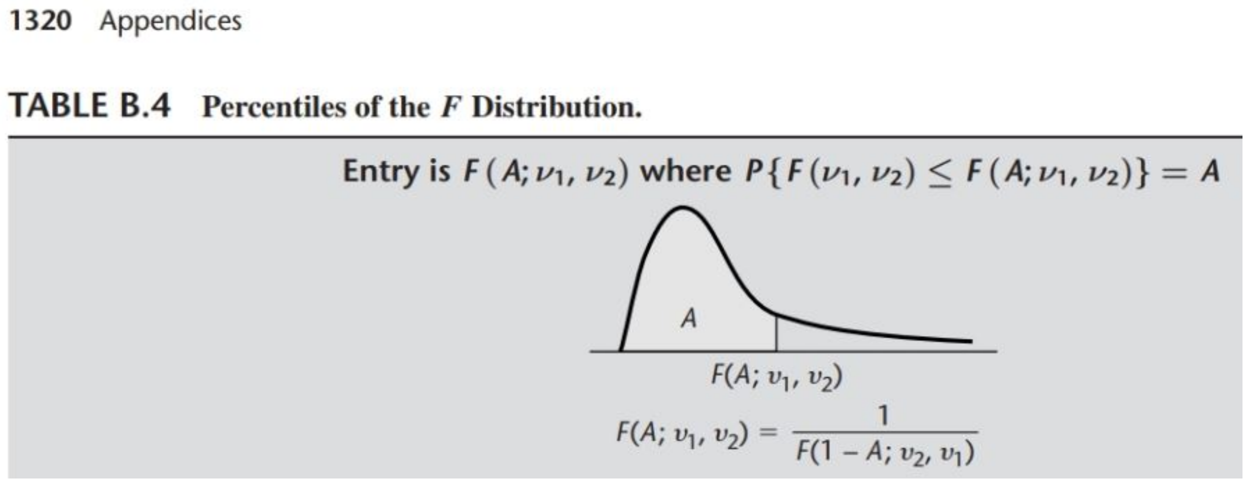

1320 Appendices

TABLE B.4 Percentiles of the F Distribution.

Entry is $F(A;\nu_1,\nu_2)$ where $P\{F(\nu_1,\nu_2) \leq F(A;\nu_1,\nu_2)\} = A$

$$F(A;\upsilon_1,\upsilon_2) = \frac{1}{F(1-A;\upsilon_2,\upsilon_1)}$$

- Or it is much easier to use R to get the critical value of the test, $F_{(1-\alpha,df1,df2)}$, we can compare F^* to make decision about our test.

```
# For alpha=0.05
qf(.95, df1=3, df2=15)
```

```
## [1] 3.287382
```

```
# For alpha=0.01
qf(.99, df1=3, df2=15)
```

```
## [1] 5.416965
```

- Finaly, we calculate the p-value= $Pr(F > F^*) = 1 - P(F \leq F^*)$ to verify ANOVA table.

```
1- pf(18.59,df1=3, df2=15)
```

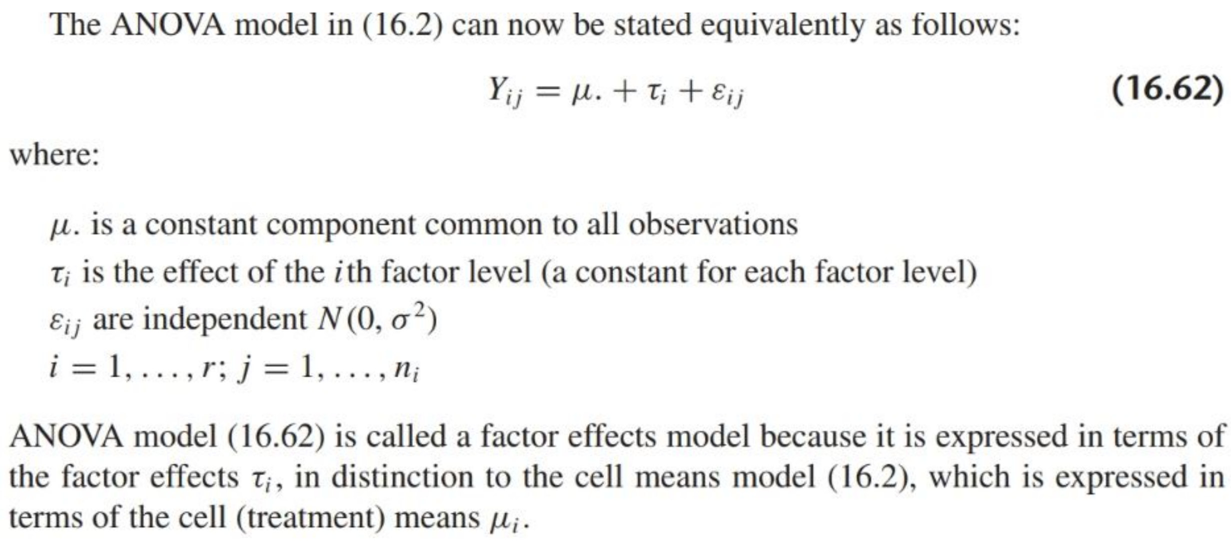
```
## [1] 2.585817e-05
```

Conclusion: From the ANOVA table, we can see $F^* > F_{(1-\alpha)}(\textit{the critical value for a given}\, \alpha)$, or we can just look at the p-value, $P = 0.00003 < \alpha$ (a commonly used significant level, e.g. 0.05 or 0.01). This suggests the data from the experiment do not support that all designs having the same effect on sales volume, or the sale volumes are significantly different for the four package designs. (We will discuss later the additional analysis to study the further comparisons of the factor level means.)

Alternative formulation of one-factor (one-way) ANOVA

Factor Effects Model is alternative formulation. We rewrite,

$$\mu_i \equiv \mu_. + (\mu_i - \mu_.) \equiv \mu_. + \tau_i$$



The ANOVA model in (16.2) can now be stated equivalently as follows:

$$Y_{ij} = \mu. + \tau_i + \varepsilon_{ij} \qquad \textbf{(16.62)}$$

where:

$\mu.$ is a constant component common to all observations
τ_i is the effect of the ith factor level (a constant for each factor level)
ε_{ij} are independent $N(0, \sigma^2)$
$i = 1, \ldots, r; j = 1, \ldots, n_i$

ANOVA model (16.62) is called a factor effects model because it is expressed in terms of the factor effects τ_i, in distinction to the cell means model (16.2), which is expressed in terms of the cell (treatment) means μ_i.

Note: Although both forms of the model are useful, the effects model is more widely encountered in the experimental design literature. It has some intuitive appeal in that $\mu_.$ is a constant and the treatment effects τ_i represent deviations from this overall mean when the specific treatments are applied.

Define overall population mean $\mu_.$ and estimate τ_i

In the **factor effects model**, we break the ith treatment mean μ_i into two components such that $\mu_i = \mu_. + \tau_i$. We usually think of

$$\mu_. = \sum \mu_i / r$$

as an overall mean. This implies a constraint in τ_i, i.e.

$$\sum_{i=1}^{r} \tau_i = 0$$

Consequently, an equivalent way to write the hypotheses for testing equality of the factor means can be written in terms of the treatment effects τ_i, say

$$H_0 \tau_1 = \tau_2 = \mathring{u}\mathring{u}\mathring{u} = \tau_r = 0$$

$$H_a \tau_i 0 \;\; \textit{for at least one i}$$

Thus, it is equivalent to state that all factor level means μ_i are equal or that all factor level effects τ_i equal zero.

Estimate τ_i (R)

We continue the previous R program to calculate the τ_i as the difference between cells means and overall mean .

```
cell.means<- as.numeric(with(Ex16,  by( sales, package, mean)))
Overall.mean <- mean(cell.means)
tau <- cell.means - Overall.mean
list(cell.means=cell.means, Overall.mean=Overall.mean , tau=tau)
```

```
## $cell.means
## [1] 14.6 13.4 19.5 27.2
##
## $Overall.mean
## [1] 18.675
##
## $tau
## [1] -4.075 -5.275  0.825  8.525
```

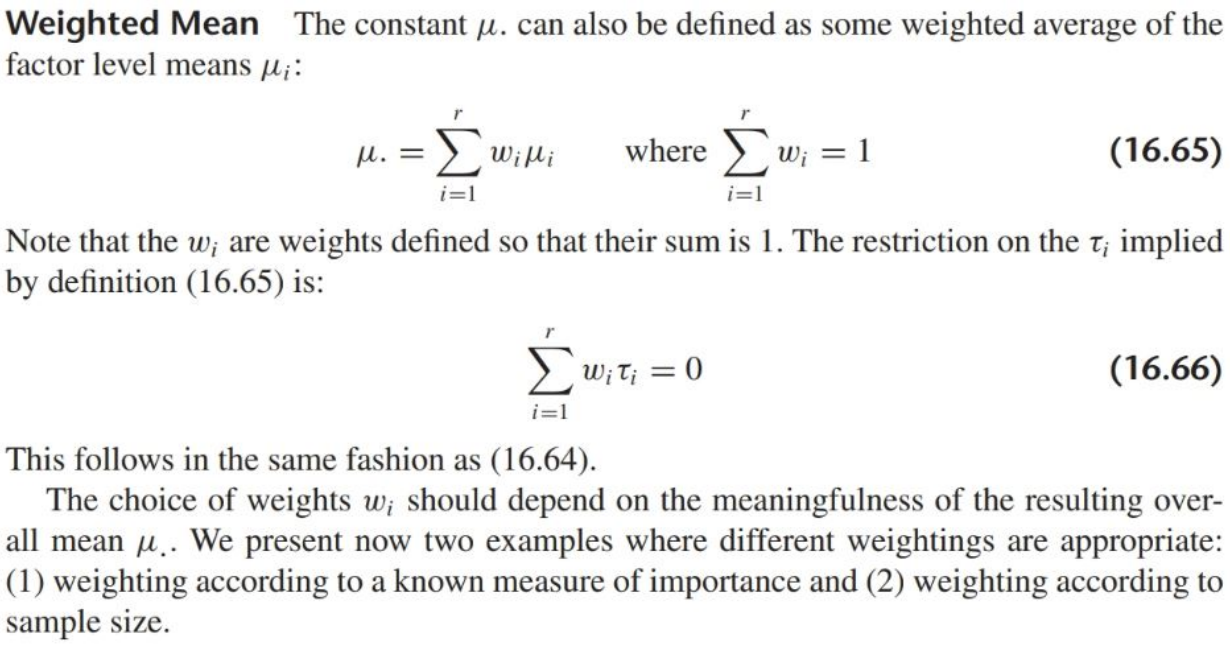
Alternatively, we can also get the estimated overall mean and $\tau_i = \mu_i - \mu_.$ directly from model output. We need to set the correct contrast `contr.sum` is used to set sum of τ_i to be zero.

```
op <- options()
options(contrasts = c("contr.sum", "contr.poly"))
fit2 <- aov(sales ~ package, data = Ex16)
dummy.coef(fit2)
```

```
## Full coefficients are
##
## (Intercept):     18.675
## package:              1      2      3      4
##                  -4.075 -5.275  0.825  8.525
```

```
options(op)     # reset (all) initial options
```

Other definition of $\mu_.$ and τ_i



Weighted Mean The constant $\mu.$ can also be defined as some weighted average of the factor level means μ_i:

$$\mu. = \sum_{i=1}^{r} w_i \mu_i \qquad \text{where} \sum_{i=1}^{r} w_i = 1 \qquad \textbf{(16.65)}$$

Note that the w_i are weights defined so that their sum is 1. The restriction on the τ_i implied by definition (16.65) is:

$$\sum_{i=1}^{r} w_i \tau_i = 0 \qquad \textbf{(16.66)}$$

This follows in the same fashion as (16.64).

The choice of weights w_i should depend on the meaningfulness of the resulting overall mean $\mu_.$. We present now two examples where different weightings are appropriate: (1) weighting according to a known measure of importance and (2) weighting according to sample size.

Choices of the weights

1. User defined meaninful weights, so the overall mean is an weight average. e.g. A car rental firm wanted to estimate the average fuel consumption (in miles per gallon) for its large fleet of cars, which consists

of 50 percent compacts, 30 percent sedans, and 20 percent station wagons. Here $w_1 = 0.5$, $w_2 = 0.3$, $w_3 = 0.2$.

2. When the group sample sizes (for given factor levels) are different, we may use weights as $w_i = n_i/n_T$.
3. Given the weights, the estimated overall mean is

$$\hat{\mu}_. = \sum_i w_i \overline{Y_{i.}}$$

Summary of Today's Lesson

- One factor (one-way) ANOVA model
- Estimation and hypothesis testing
- Data fitting with R
- Reading: Chapter 16.1-16.7 , Appendix A on F-distribution;
- Homework for Today (no need to turn-in) : Rerun the example (Table 16.1) for data checking and ANOVA analysis and reproduce the results given in the Textbook and lecture note.