One-Way Analysis of Variance

A One-Way Analysis of Variance is a way to test the equality of three or more means at one time by using variances. We will show how to test for these assumtptions in a fortcoming class.

Assumptions

- The populations from which the samples were obtained must be normally or approximately normally distributed. (*Any Test for Normality*).
- The samples must be independent.
- The variances of the populations must be equal. (*Bartlett Test for Homogeneity of Variances*).

Hypotheses

- The null hypothesis will be that all population means are equal

$$\mu_1 = \mu_2 = \ldots = \mu_k$$

- The alternative hypothesis is that at least one mean is different.

ANOVA: One Way ANOVA `R` Example

Assume that we have three fertilizers to be tested. We wish to determine if there is any difference is the mean yields for the three different types of fertilizer.

Fertilizer A	5.6	6.4	6.6	5.8
Fertilizer B	5.1	6.2	6.4	5.7
Fertilizer C	5.0	6.1	5.8	5.5

The following `R` output is a One-Way ANOVA procedure for testing multiple means.

```
Fert <- c("A", "A", "A", "A", "B", "B", "B", "B",
"C", "C", "C", "C")
Yield <- c(5.6, 6.4, 6.6, 5.8, 5.1, 6.2, 6.4, 5.7,
5, 6.1, 5.8, 5.5)

ModelA=aov(Yield~Fert)
```

Hypotheses

H$_0$ μ_A, μ_B & μ_C are all equal.

H$_1$ At least one of the means is different from the rest.

```
> summary(aov(Yield~Fert))
              Df Sum Sq Mean Sq F value Pr(>F)
Fert          2   0.50  0.2500   0.957   0.42
Residuals     9   2.35  0.2611
```

Further to this test, we conclude that there are no differences between the three fertilizer types, based on the high $p-$value.

Two-Way ANOVA

The two-way analysis of variance is an extension to the one-way analysis of variance. There are two independent variables (hence the name two-way).

Factors

The two independent variables in a two-way ANOVA are called factors. The idea is that there are two variables, factors, which affect the dependent variable. Each factor will have two or more levels within it, and the degrees of freedom for each factor is one less than the number of levels.

Assumptions

- The populations from which the samples were obtained must be normally or approximately normally distributed.
- The samples must be independent.
- The variances of the populations must be equal.
- The groups must have the same sample size.

Treatment Groups

- Treatement Groups are formed by making all possible combinations of the two factors.
- For example, if the first factor has 3 levels and the second factor has 2 levels, then there will be $3 \times 2 = 6$ different treatment groups.
- There may be more than one observation per treatment group.

Hypotheses

There are two (sometimes three) sets of hypothesis with the two-way ANOVA. The null hypotheses for each of the sets are given below.

i) The population means of the first factor are equal. This is like the one-way ANOVA for the row factor.

ii) The population means of the second factor are equal. This is like the one-way ANOVA for the column factor.

iii) There is no interaction between the two factors. This is similar to performing a test for independence with contingency tables.

Important It is only possile to test for an interaction effect in the presence of replicate measurements for each **treatment group**.

Main Effect

- The main effect involves the independent variables separately.
- The interaction is ignored for this part. Just the rows or just the columns are used, not mixed.
- This is the part which is similar to the one-way analysis of variance.
- Each of the variances calculated to analyze the main effects are like the between variances

Interaction Effect

The interaction effect is the effect that one factor has on the other factor. The degrees of freedom here is the product of the two degrees of freedom for each factor.

Corn Seed Example

As an example, let's assume we're planting corn. The type of seed and type of fertilizer are the two factors we're considering in this example.

- This example has 15 treatment groups.
- There are $3 - 1 = 2$ degrees of freedom for the type of seed, and $5 - 1 = 4$ degrees of freedom for the type of fertilizer.
- There are $2 \times 4 = 8$ degrees of freedom for the interaction between the type of seed and type of fertilizer.

The data that actually appears in the table are samples. In this case, **two** samples from each treatment group were taken. (Important: It is possible to compute the interaction effect.)

	Fert I	Fert II	Fert III	Fert IV	Fert V
Seed A	106, 110	95, 100	94, 107	103, 104	100, 102
Seed B	110, 112	98, 99	100, 101	108, 112	105, 107
Seed C	94, 97	86, 87	98, 99	99, 101	94, 98

Within Variation

The Within variation is the sum of squares within each treatment group. You have one less than the sample size (remember all treatment groups must have the same sample size for a two-way ANOVA) for each treatment group. The total number of treatment groups is the product of the number of levels for each factor. The within variance is the within variation divided by its degrees of freedom.

F-Tests

There is an F-test for each of the hypotheses, and the F-test is the mean square for each main effect and the interaction effect divided by the within variance. The numerator degrees of freedom come from each effect, and the denominator degrees of freedom is the degrees of freedom for the within variance in each case.

Two-Way ANOVA Table (Without Interaction)

Table 13.2 Summary Table for Two-Way Analysis of Variance with One Observation per Cell (Randomized Block Design)

Source of variation	Degrees of freedom (df)	Sum of squares (SS)	Mean square (MS)	F ratio
Among treatment groups (A)	$K-1$	$SSA = \sum_{k=1}^{K} \frac{T_k^2}{n_k} - \frac{T^2}{N}$	$MSA = \frac{SSA}{K-1}$	$F = \frac{MSA}{MSE}$
Among treatment groups, or blocks (B)	$J-1$	$SSB = \frac{1}{K}\sum_{j=1}^{J} T_j^2 - \frac{T^2}{N}$	$MSB = \frac{SSB}{J-1}$	$F = \frac{MSB}{MSE}$
Sampling error (E)	$(J-1)(K-1)$	$SSE = SST - SSA - SSB$	$MSE = \frac{SSE}{(J-1)(K-1)}$	
Total (T)	$N-1$	$SST = \sum_{j=1}^{J}\sum_{k=1}^{K} X^2 - \frac{T^2}{N}$		

Two-Way ANOVA Table (With Interaction)

Table 13.3 Summary Table for Two-Way Analysis of Variance with More than One Observation per Cell

Source of variation	Degrees of freedom (df)	Sum of squares (SS)	Mean square (MS)	F ratio
Among treatment groups (A)	$K-1$	$SSA = \sum_{k=1}^{K} \frac{T_k^2}{nJ} - \frac{T^2}{N}$	$MSA = \frac{SSA}{K-1}$	$F = \frac{MSA}{MSE}$
Among treatment groups (B)	$J-1$	$SSB = \sum_{j=1}^{J} \frac{T_j^2}{nK} - \frac{T^2}{N}$	$MSB = \frac{SSB}{J-1}$	$F = \frac{MSB}{MSE}$
Interaction (between) factors (A and B) (I)	$(J-1)(K-1)$	$SSI = \frac{1}{n}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\sum_{i=1}^{n} X\right)^2 - SSA - SSB - \frac{T^2}{N}$	$MSI = \frac{SSI}{(J-1)(K-1)}$	$F = \frac{MSI}{MSE}$
Sampling error (E)	$JK(n-1)$	$SSE = SST - SSA - SSB - SSI$	$MSE = \frac{SSE}{JK(n-1)}$	
Total (T)	$N-1$	$SST = \sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{K} X^2 - \frac{T^2}{N}$		

Given Information for Two-Way ANOVA Table

The following pieces of information will be given in an exam question for Two -Way ANOVA (without Replication). With this information , you should be able to construct the ANOVA table.

$$\text{MS}_{Trt} = c \times S_R^2$$

$$\text{MS}_{Block} = r \times S_C^2$$

- r and c are the numbers of rows and columns respectively.
- S_R^2 is the variance of the Row means
- S_C^2 is the variance of the column means.

Take care when working with calculations that involve r and c: r is the number of rows, which is the number of subgroups for each of the treatments (which are arranged along the columns). c is the number of columns, which is the number of subgroups for each of the blocks (which are arranged along the rows).

ANOVA: Worked Example with `R`

- A standard solution was prepared, containing 16.00% (by weight) of chloride. Three titration methods, each with a different technique of end-point determination, were used to analyse the standard solution.
- The procedure was carried out by four different clinical analysts. The order of the experiments was randomized. The results for the chloride found (% w/w) are shown below:

	Analyst 1	Analyst 2	Analyst 3	Analyst 4
Method A	16.03	16.05	16.02	16.12
Method B	16.13	16.13	15.94	15.97
Method C	16.09	16.15	16.12	16.1

- Here the treatment is the titration method and we are interested in determining if there is uniformity between each method. Four analysts performed an experiment using each of the titration methods. This allows the analysts to remove any effect due to the analysts.

- To construct the model using `R`, we use the `aov()` command, specifying the treatment factor and the blocking. Importantly we express the model additively (i.e ". . Meth+Anlt..").

```
> Model=aov(Titr ~ Meth + Anlt)
>
> summary(Model)
Df  Sum Sq   Mean Sq  F value Pr(>F)
Meth           2  0.01202  0.006008 1.279   0.345
Anlt           3  0.01109  0.003697 0.787   0.543
Residuals      6  0.02818  0.004697

---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

- This additive model is the appropriate specification if an interaction is not assumed.
- However - It is ususal for an interaction effect to be accounted for in such experiments. A variant of this `R` implementation, that does specify an interaction term, will be discussed in due course.
- Our conclusion of this procedure is that there is no difference in the titration method (i.e. there is no effect due to method) based on the high $p-$values (0.345).
- We also conclude that there is no effect due to whcih analyst is perfoming the experiment (p-value = 0.543).

Two Way ANOVA table for Crop Example

Source of Variation	SS	df	MS	F	P-value	*F-crit*
Seed	512.8667	2	256.4333	28.283	0.000008	*3.682*
Fertilizer	449.4667	4	112.3667	12.393	0.000119	*3.056*

From the above results, we can see that the main effects are both significant, but the interaction between them isn't. That is, the types of seed aren't all equal, and the types of fertilizer aren't all equal, but the type of seed doesn't interact with the type of fertilizer.

Completing a Two-Way ANOVA Table (with Replication)

If you are required to complete a two-way table, based on partial data, be mindful of the following.

Source	SS	df	MS	F
Main Effect A		A, a-1	SS / df	MS(A) / MS(W)
Main Effect B		B, b-1	SS / df	MS(B) / MS(W)
Interaction Effect		A*B, (a-1)(b-1)	SS / df	MS(A*B) / MS(W)
Within		N - ab, ab(n-1)	SS / df	
Total	sum of others	N - 1, abn - 1		

ANOVA: Two Way Table with Blocking

Blocking: *The block is a factor. The main aim of blocking is to reduce the unexplained variation of an experimental design (compared to non-blocked design). We are not interested in the block effect per se , rather we block when we suspect the the background "noise" would counfound the effect of the actual factor.*
We will arrange blocks along the columns (they will be Factor B)

After examining the previous statistical output, an agricultural analysts points out that there should a blocking to account for the different types of field.

- Field 1 is very boggy,
- Field 2 and Field 3 are reasonably good,
- Field 4 is full of rocks.

The procedure was carried out again. Three plots of land in each field were given the fertilizer treatment.

	Block 1	Block 2	Block 3	Block 4
Fertilizer A	5.6	6.4	6.6	5.8
Fertilizer B	5.1	6.2	6.4	5.7
Fertilizer C	5.0	6.1	5.8	5.5

The following `R` output is a Two-Way ANOVA procedure for analysing this data.

```
Fert <- c("A", "A", "A", "A", "B", "B", "B", "B",
"C", "C", "C", "C")
Yield <- c(5.6, 6.4, 6.6, 5.8, 5.1, 6.2, 6.4, 5.7,
5, 6.1, 5.8, 5.5)
Block <- c("Bk1","Bk2","Bk3","Bk4","Bk1","Bk2","Bk3","Bk4",
"Bk1","Bk2","Bk3","Bk4")

ModelB=aov(Yield~Fert+Block)
```

This output informs us that there is a difference between fertilizer treatments, contrary to what was previously thought. (There is a very important difference between fields also, but not of interest)

```
> summary(ModelB)
            Df Sum Sq Mean Sq F value   Pr(>F)
Fert         2 0.5000  0.2500   10.23 0.011667 *
Block        3 2.2033  0.7344   30.05 0.000519 ***
Residuals    6 0.1467  0.0244
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Conclusion : While there is an effect for Ferilizer. There is aslo a significant effect for the blocks. The yeild rate does depend on the type of field that the experiment was carried out in.

ANOVA: Exercise with `R`

Four laboratory technicians performed six determination of C of 2,4 dinitrophenol in water, according to the same specified procedure. The results in $C/\mu M$ are as follows

Analyst A	Analyst B	Analyst C	Analyst D
701	550	511	613
677	545	523	623
680	573	540	649
660	532	542	632
654	529	559	614
648	534	554	626

The analysis of variance procedure is used to determine if there is a significiant difference between the mean of the determinations makde by the four investigators.

```
summary(aov(Det~Anlt))
            Df Sum Sq Mean Sq  F value   Pr(>F)
Anlt         3  99942   33314   .....    4.64e-12 ***
Residuals   ..   6918     346
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

(a) The value for the degrees of freedom for residuals has been removed from the output. What is this value?

(b) The value for the test statistics (`F value`) has been removed from the output. What is this value?

(c) State the null and alternative hypothesis for this procedure.

(d) Based on the $p-$value, what is your conclusion for this procedure.