DO 1: 04:00 PM – 04:50 PM
DO 2: 09:40 AM – 11:30 PM

Subject Code

18MEO113T - Design of Experiments

Handled by

S. Murali, Ph.D.,
Associate Professor,
Department of Mechanical Engineering,
SRM IST, Kattankulathur.

Disclaimer

The content prepared in the presentation are from various sources, only used for education purpose. Thanks to all the sources.

Unit 5

Unit 5

1: Introduction and uses of confounding

2: 2^3 factorial experiment with complete confounding

3: 2^3 factorial experiment with partial confounding

4: Confounding in the 2^n series and examples

5: Confounding of 3^n factorial and examples

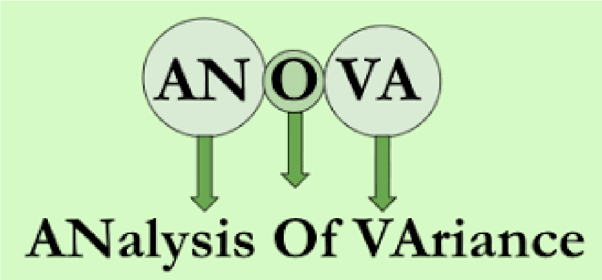
6: ANOVA (One-way and two-way, higher-way ANOVA)

7: MANOVA and ANCOVA overview

8: Solving Case studies on ANOVA with statistics software

9: Regression Models and Regression Analysis

ANOVA



- Analysis of Variance (abbreviated as ANOVA)
- ANOVA was developed by **Fisher in the early 1920s**, and initially applied to agricultural experiments. Used extensively today for industrial experiments.
- The ANOVA technique enables us to perform to examine the significance of the difference amongst more than two sample means at the same time.
- Using this technique, one can infer whether the samples have been drawn from populations with the same mean.

ANOVA

- ANOVA is a procedure for testing the difference among different data groups for homogeneity.
- Variance is an important statistical measure described as the mean of the squares of deviations taken from the mean of the given data series. It is a frequently used measure of variation. The square of standard deviation is called as the variance.

  i.e., Variance = $(\text{Standard deviation})^2$

- There may be variation between samples and also within sample items.

ANOVA

- An ANOVA test is a way to find out if survey or experiment results are significant.
- In other words, ANOVA help us to figure out if there is a need to reject the null hypothesis or accept the alternative hypothesis.
- Basically, we’re testing groups to see if they are different.

ANOVA

- Example:
  - A group of psychiatric patients are trying three different therapies: counselling, medication and biofeedback. We want to see if one therapy is better than the others.
  - A manufacturer has two different processes to make light bulbs. They want to know if one process is better than the other.
  - Students from different colleges take the same exam. You want to see if one college outperforms the other.

ANOVA

- Example:
  - A Manager wants to evaluate the performance of three(or more) employees to see if any performance different from others.
  - A Marketing executives want to see if there is a difference in sales productivity in the 5 company region.
  - A Teacher wants to see if there is a difference in students performance if he use 3 or more approach to teach.

ANOVA

- ANOVA is two types:
  - – One Way ANOVA: Only one factor is investigated
    - One independent variable (With 2 levels)
    - Analysis of Variance could have one independent variable
  - – Two Way ANOVA: Investigate two factors at the same time
    - Two independent variables (can have multiple levels)
    - Analysis of Variance could have two independent variables
  - – Two way ANOVA without replication
  - – Two way ANOVA with replication

ANOVA

- Two way ANOVA without replication
  - We are testing one set of individual before and after they take a medication to see if it works or not.
- Two way ANOVA with replication
  - Two groups, and the members of those groups are doing more than one thing.
  - For example, two groups of patients from different hospitals trying two different therapies.

ANOVA

- Three way ANOVA
  - A three-way ANOVA tests which of three separate variables influence an outcome, and the relationship between the three variables. It is also called a three-factor ANOVA.

ANOVA

- Principle of ANOVA
  - We have to make two estimates of population variance viz., one based on between samples variance and the other based on within samples variance. Then the said two estimates of population variance are compared with F-test, wherein we work out.

$$F = \frac{Estimate\ of\ population\ variance\ based\ on\ between\ samples\ variance}{Estimate\ of\ population\ variance\ based\ on\ within\ samples\ variance}$$

  - This value of F is to be compared to the F-limit for given degrees of freedom. If the F-value we work out is equal or exceed the F-limit value, we may say that there are significant difference between the sample means.

ANOVA

- ANOVA Technique
  - Obtain the mean of each sample
  - Work out the mean of the sample means
  - Calculate sum of squares for variance between the samples (or SS between)
  - Obtain variance or mean square (MS) between samples
  - Calculate sum of squares for variance within samples (or SS within)
  - Obtain the variance or mean square (MS) within samples
  - Find sum of squares of deviation for total variance
  - Finally, find F-ratio.

One-way ANOVA

- A one way ANOVA is used to compare two means from two independent (unrelated) groups using the F-distribution.
- The **null hypothesis** for the test is that the **two means are equal**.
- Therefore, a significant result means that the **two means are unequal**.
- **Alternate hypothesis …? Two means are not equal**..

One-way ANOVA

Note

- The **independent variable** is the categorical variable that defines the compared groups. E.g., instructional methods, grade level, or marital status.
- The **dependent variable** is the measured variable whose means are being compared e.g., level of job satisfaction or text anxiety.

One-way ANOVA

Examples of when to use a one-way ANOVA

- Situation 1: You have a group of individuals randomly split into smaller groups and completing different tasks. For example, you might be studying the effects of tea on weight loss and form three groups: green tea, black tea, and no tea.

One-way ANOVA

Examples of when to use a one-way ANOVA

- Situation 2: Similar to situation 1, but in this case, the individuals are split into groups based on an attribute they possess. For example, you might be studying the leg strength of people according to weight. You could split participants into weight categories (obese, overweight and normal) and measure their leg strength on a weight machine.

One-way ANOVA

Limitation of the one-way ANOVA

- A one-way ANOVA will tell you that at least two groups were different from each other. **But it won't tell you which groups were different**. If your test returns a significant f-statistic, you may need to run an ad hoc test (like the Least Significant Difference test) to tell you exactly which groups had a difference in means.

One-way ANOVA

Assumptions

1. Your **dependent variable** should be measured at the **interval** or **ratio scales** (i.e., they are continuous)

Examples of variable that meet this criterion include, revision time (measured in hours), intelligence (measured using IQ score), exam performance (measured from 0 to 100), weight (measured in kg)

One-way ANOVA

Assumptions

2. Your **independent variable** should consist of two or more categorical, independent groups. Typically, a one-way ANOVA is used when you have three or more categorical, independent groups, but it can be used for just two groups (but an independent-samples t-test is more commonly used for two groups)

Examples independent variables that meet this criterion include ethnicity (e.g., 3 groups: Indian, Chinese, Korean), physical activity level (e.g., 4 groups: sedentary, low, moderate and high), profession (e.g., 5 groups: surgeon, doctor, nurse, dentist, therapist), and so forth.

One-way ANOVA

Assumptions

3. You should have **independence of observations,** meaning there is no relationship between the observations in each group or between the groups themselves.

Example: it is an important assumption of one-way ANOVA. If your study fails this assumption, you will need to use another statistical test instead of the one-way ANOVA (e.g., a repeated measures design)

One-way ANOVA

Assumptions

4. There should be no significant outliers. Outliers are simple single data points within your data that do not follow the usual pattern.

Example: In a study of 100 students' IQ scores, where the mean score was 108 with only a small variation between students, one student had a score of 156, which is very unusual.

The problem with outliers is that they can have a negative effect on the one-way ANOVA, reducing the validity of your results.

One-way ANOVA

Assumptions

5. Your **dependent variable** should be approximately **normally distributed** for each category of the **independent variable**.

One-way ANOVA only requires approximately normal data because it is quite “robust” to violations of normality, meaning that assumption can be a little violated and still provide valid results. The **Kolmogorov–Smirnov** test and the **Shapiro–Wilk** test are most widely used methods to test the normality of the data

One-way ANOVA

Assumptions

6. There needs to be homogeneity of variances.

Source of Variation	Degrees of Freedom	Sum Of Squares	Mean Square (Variance)	F
Among Groups	c - 1	SSA	$MSA = \dfrac{SSA}{c - 1}$	$F_{STAT} = \dfrac{MSA}{MSW}$
Within Groups	n - c	SSW	$MSW = \dfrac{SSW}{n - c}$	c = number of groups n = sum of the sample sizes from all groups
Total	n – 1	SST		

One-way ANOVA Process

- A one-way ANOVA uses the following null and alternative hypotheses:
- **H_0 (null hypothesis):** $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$ (all the population means are equal)
- **H_1 (alternative hypothesis):** at least one population mean is different from the rest

Source	Sum of Squares (SS)	df	Mean Squares (MS)	F	p
Treatment	SSR	df_r	MSR	MSR/MSE	F_{df_r, df_e}
Error	SSE	df_e	MSE		
Total	SST	df_t			

One-way ANOVA Process

- **SSR:** regression sum of squares
- **SSE:** error sum of squares
- **SST:** total sum of squares (SST = SSR + SSE)
- **df_r:** regression degrees of freedom ($df_r = k-1$)
- **df_e:** error degrees of freedom ($df_e = n-k$)
- **df_t:** total degrees of freedom ($df_t = n-1$)
  - **k:** total number of groups
  - **n:** total observations
- **MSR:** regression mean square ($MSR = SSR/df_r$)
- **MSE:** error mean square ($MSE = SSE/df_e$)
- **F:** The F test statistic (F = MSR/MSE)
- **p:** The p-value that corresponds to $F_{dfr,\ dfe}$

If the p-value is less than your chosen significance level (e.g. 0.05), then you can reject the null hypothesis and conclude that at least one of the population means is different from the others.

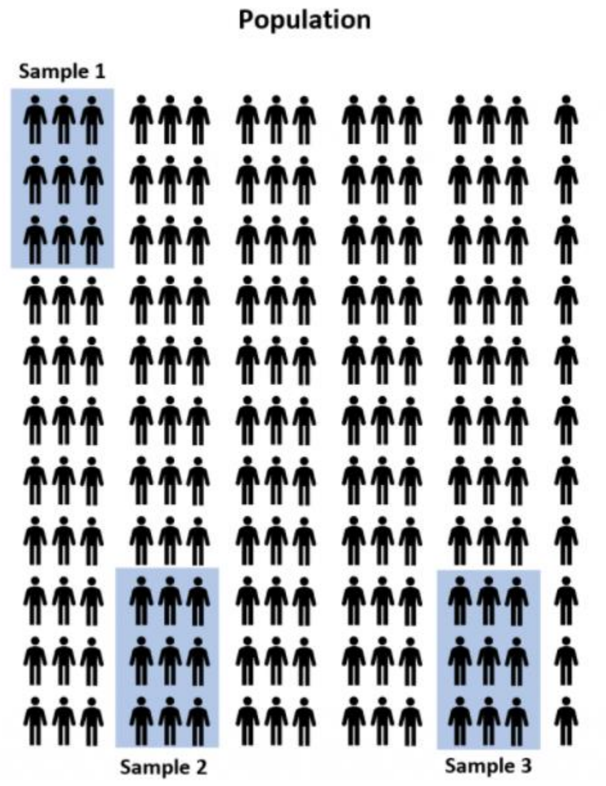
Note: If you reject the null hypothesis, this indicates that at least one of the population means is different from the others, but the ANOVA table doesn't specify which population means are different. To determine this, you need to perform post hoc tests, also known as "multiple comparisons" tests.

One-way ANOVA (Problem 1)

Suppose we want to know whether or not three different exam preparation programs lead to different mean scores on a certain exam. To test this, we recruit 30 students to participate in a study and split them into three groups. The students in each group are randomly assigned to use one of the three exam preparation programs for the next three weeks to prepare for an exam. All students take the same exam at the end of the three weeks.

One-way ANOVA (Problem 1)

Since there are millions of high school students around the country, it would be too time-consuming and costly to go around to each student and let them use one of the exam prep programs. Instead, we might select three random samples of 100 students from the population and allow each sample to use one of the three test prep programs to prepare for the exam. Then, we could record the scores for each student once they take the exam.



One-way ANOVA (Problem 1)

The exam scores for each group are shown below:

Group 1	Group 2	Group 3
85	91	79
86	92	78
88	93	88
75	85	94
78	87	92
94	84	85
98	82	83
79	88	85
71	95	82
80	96	81

Determine if the mean exam score is different between the three groups:

One-way ANOVA (Problem 1)

Step 1: Calculate the group means and the overall mean.

First, we will calculate the mean for all three groups along with the overall mean:

	Group 1	Group 2	Group 3
	85	91	79
	86	92	78
	88	93	88
	75	85	94
	78	87	92
	94	84	85
	98	82	83
	79	88	85
	71	95	82
	80	96	81
Group Means	**83.4**	**89.3**	**84.7**
Overall Mean	**85.8**		

Step 1: Find Group Means and Overall Means

One-way ANOVA (Problem 1)

Step 2: Calculate SSR.

Next, we will calculate the regression sum of squares (SSR) using the following formula:

Step 2: Calculate Sum of Squares of Regression (SSR)

$n\Sigma(X_j - \overline{X}..)^2$

Between Groups

where:

- **n**: the sample size of group j
- **Σ**: a greek symbol that means "sum"
- **X_j**: the mean of group j
- **$\overline{X}$..**: the overall mean

In our example, we calculate that SSR = $10(83.4-85.8)^2 + 10(89.3-85.8)^2 + 10(84.7-85.8)^2$ = **192.2**

SRM INSTITUTE OF SCIENCE & TECHNOLOGY (Deemed to be University u/s 3 of UGC Act, 1956)

One-way ANOVA (Problem 1)

Step 3: Calculate SSE.

Next, we will calculate the error sum of squares (SSE) using the following formula:

$\Sigma(X_{ij} - \overline{X}_j)^2$

Within Groups

where:

- **Σ**: a greek symbol that means "sum"
- $\mathbf{X_{ij}}$: the i^{th} observation in group j
- $\mathbf{\overline{X}_j}$: the mean of group j

One-way ANOVA (Problem 1)

Step 3: Calculate SSE.

Step 3: Calculate Sum of Squares of Error (SSE)

Within Groups

In our example, we calculate SSE as follows:

Group 1: $(85-83.4)^2 + (86-83.4)^2 + (88-83.4)^2 + (75-83.4)^2 + (78-83.4)^2 + (94-83.4)^2 + (98-83.4)^2 + (79-83.4)^2 + (71-83.4)^2 + (80-83.4)^2 =$ **640.4**

Group 2: $(91-89.3)^2 + (92-89.3)^2 + (93-89.3)^2 + (85-89.3)^2 + (87-89.3)^2 + (84-89.3)^2 + (82-89.3)^2 + (88-89.3)^2 + (95-89.3)^2 + (96-89.3)^2 =$ **208.1**

Group 3: $(79-84.7)^2 + (78-84.7)^2 + (88-84.7)^2 + (94-84.7)^2 + (92-84.7)^2 + (85-84.7)^2 + (83-84.7)^2 + (85-84.7)^2 + (82-84.7)^2 + (81-84.7)^2 =$ **252.1**

SSE: 640.4 + 208.1 + 252.1 = **1100.6**

One-way ANOVA (Problem 1)

Step 4: Calculate SST.

Next, we will calculate the total sum of squares (SST) using the following formula:

SST = SSR + SSE

In our example, SST = 192.2 + 1100.6 = **1292.8**

One-way ANOVA (Problem 1)

Step 5: Fill in the ANOVA table.

Now that we have SSR, SSE, and SST, we can fill in the ANOVA table:

Source	Sum of Squares (SS)	df	Mean Squares (MS)	F
Treatment	192.2	2	96.1	2.358
Error	1100.6	27	40.8	
Total	1292.8	29		

Here is how we calculated the various numbers in the table:

- **df treatment:** k-1 = 3-1 = 2
- **df error:** n-k = 30-3 = 27
- **df total:** n-1 = 30-1 = 29
- **MS treatment:** SST / df treatment = 192.2 / 2 = 96.1
- **MS error:** SSE / df error = 1100.6 / 27 = 40.8
- **F:** MS treatment / MS error = 96.1 / 40.8 = 2.358

Note: *n = total observations, k = number of groups*

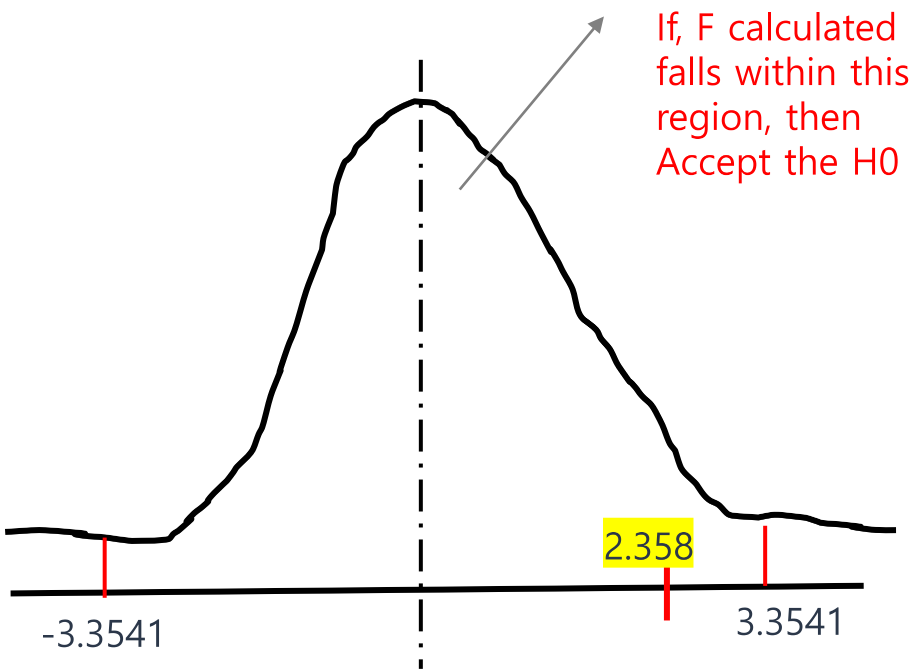
One-way ANOVA (Problem 1)

Step 6: Interpret the results.

The F test statistic for this one-way ANOVA is **2.358**. To determine if this is a statistically significant result, we must compare this to the F critical value found in the F distribution table with the following values:

- α (significance level) = 0.05
- DF1 (numerator degrees of freedom) = df treatment = 2
- DF2 (denominator degrees of freedom) = df error = 27

We find that the F critical value is **3.3541**.



One-way ANOVA (Problem 1)

Step 6: Interpret the results.

Since the F test statistic in the ANOVA table is less than the F critical value in the F distribution table, we fail to reject the null hypothesis. This means we don't have sufficient evidence to say that there is a statistically significant difference between the mean exam scores of the three groups.

F critical > F calculated → Accept the null hypothesis

One-way ANOVA

F – Table with alpha = 0.05 (95% confidence interval)

	DF1	α = 0.05																	
DF2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	Inf
1	161.45	199.5	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.1	251.14	252.2	253.25	254.31
2	18.513	19	19.164	19.247	19.296	19.33	19.353	19.371	19.385	19.396	19.413	19.429	19.446	19.454	19.462	19.471	19.479	19.487	19.496
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.572	8.5494	8.5264
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.041	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.717	5.6877	5.6581	5.6281
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.365
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.099	4.06	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6689
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.866	3.787	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	5.3177	4.459	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	4.9646	4.1028	3.7083	3.478	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.913	2.845	2.774	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.948	2.8962	2.8536	2.7876	2.7186	2.6464	2.609	2.5705	2.5309	2.4901	2.448	2.4045
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.341	2.2962
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.671	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5342	2.463	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	4.494	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	4.4513	3.5915	3.1968	2.9647	2.81	2.6987	2.6143	2.548	2.4943	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.104	2.0584	2.0107	1.9604
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.308	2.2341	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.878
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.599	2.514	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.366	2.321	2.2504	2.1757	2.096	2.054	2.0102	1.9645	1.9165	1.8657	1.8117
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.938	1.8894	1.838	1.7831
23	4.2793	3.4221	3.028	2.7955	2.64	2.5277	2.4422	2.3748	2.3201	2.2747	2.2036	2.1282	2.0476	2.005	1.9605	1.9139	1.8648	1.8128	1.757
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.1834	2.1077	2.0267	1.9838	1.939	1.892	1.8424	1.7896	1.733
25	4.2417	3.3852	2.9912	2.7587	2.603	2.4904	2.4047	2.3371	2.2821	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.711
26	4.2252	3.369	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1479	2.0716	1.9898	1.9464	1.901	1.8533	1.8027	1.7488	1.6906
27	4.21	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6717
28	4.196	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.236	2.19	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	4.183	3.3277	2.934	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6376
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	4.0847	3.2317	2.8387	2.606	2.4495	2.3359	2.249	2.1802	2.124	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.097	2.0401	1.9926	1.9174	1.8364	1.748	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.175	2.0868	2.0164	1.9588	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.429	1.3519	1.2539
Inf	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.394	1.318	1.2214	1

One-way ANOVA (Problem 1 – Second Method)

The exam scores for each group are shown below:

Group 1	Group 2	Group 3
85	91	79
86	92	78
88	93	88
75	85	94
78	87	92
94	84	85
98	82	83
79	88	85
71	95	82
80	96	81

Determine if the mean exam score is different between the three groups:

One-way ANOVA (Problem 1 – Second Method)

Step 1: Calculate Grand Total

Step 2: Count number of observations

Step 3: Calculate Correction Factor (C)

Group 1	Group 2	Group 3			
85	91	79			
86	92	78			
88	93	88			
75	85	94			
78	87	92			
94	84	85			
98	82	83			
79	88	85			
71	95	82			
80	96	81			
			Grand Total	2574	1
			Total Observation	30	2
			Correction Factor = (Grand Total^2/Total Observation)	220849.2	3

One-way ANOVA (Problem 1 – Second Method)

Step 4: SS Total

$$SS\ Total = \sum\sum x_{ij}^2 - C$$

Step 5: SS Between group

$$SS\ between\ group = \sum_{i=1}^{k} \frac{T_i^2}{n_i} - C$$

Step 6: SS Within group (Error)

= SS total – SS between group

Step 7: Make F-distribution table and complete

One-way ANOVA (Problem 1 – Second Method)

Step 4: SS Total

$$SS\ Total = \sum\sum x_{ij}^2 - C$$

= (85^2 + 86^2 + 88^2 + …..+ 82^2 + 81^2) - C

= 222142 – 220849 = 1292.8

Group 1	Group 2	Group 3
85	91	79
86	92	78
88	93	88
75	85	94
78	87	92
94	84	85
98	82	83
79	88	85
71	95	82
80	96	81

One-way ANOVA (Problem 1 – Second Method)

Step 5: SS Between group

$$SS\ between\ group = \sum_{i=1}^{k} \frac{T_i^2}{n_i} - C$$

$$= \frac{834^2}{10} + \frac{893^2}{10} + \frac{847^2}{10} - C$$

= 221041 – 220849 = 192.2

	Group 1	Group 2	Group 3
	85	91	79
	86	92	78
	88	93	88
	75	85	94
	78	87	92
	94	84	85
	98	82	83
	79	88	85
	71	95	82
	80	96	81
Sub Total	834	893	847

One-way ANOVA (Problem 1 – Second Method)

Step 6: SS Within group (Error)

= SS total – SS between group

= 1292.8 - 192.2

SS within group (Error) = 1100.6

Step 7: Make F-distribution table

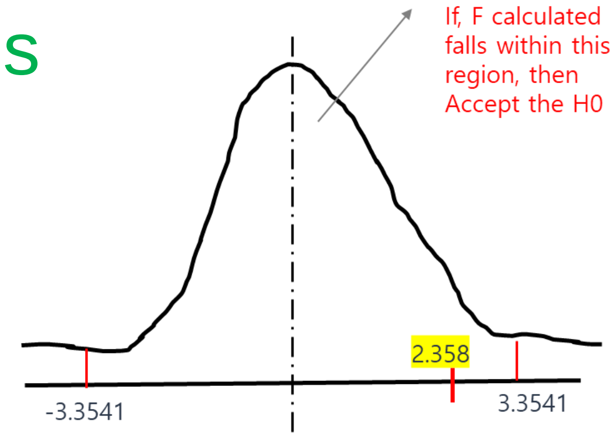
Source of Variation	*SS*	*df*	*MS*	*F*	*F crit*
Between Groups	192.2	2	96.1	2.357532	3.3541
Within Groups	1100.6	27	40.76296		
Total	1292.8	29			

One-way ANOVA (Problem 1 – Second Method)

Step 8: Write recommendation or summary

Since the F test statistic in the ANOVA table is less than the F critical value in the F distribution table, we fail to reject the null hypothesis. This means we don't have sufficient evidence to say that there is a statistically significant difference between the mean exam scores of the three groups.

F critical > F calculated → Accept the null hypothesis



One-way ANOVA (Problem 2)

The exam scores for each group are shown below:

Group 1	Group 2	Group 3
51	23	56
45	43	76
33	23	74
45	43	87
67	45	56

Determine if the mean exam score is different between the three groups:

One-way ANOVA (Problem 2)

The exam scores for each group are shown below:

	Group 1	Group 2	Group 3
	51	23	56
	45	43	76
	33	23	74
	45	43	87
	67	45	56
Group Means	48.2	35.4	69.8
Overall Mean	51.13333		

Step 1: Find Group Means and Overall Means

Between Groups

Step 2: Calculate Sum of Squares of Regression (SSR)

SSR	5*(48.2-51.13)^2+5*(35.4-51.13)^2+5*(69.8-51.13)^2
SSR	3022.933

$n\Sigma(X_j - \bar{X}..)^2$

Within Groups

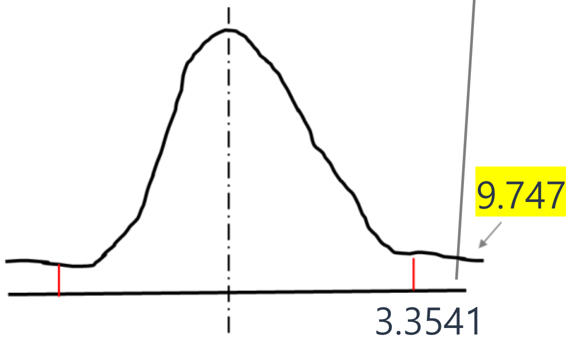
Step 3: Calculate Sum of Squares of Error (SSE)

SSE for Group 1	612.8	
SSE for Group 2	515.2	
SSE for Group 3	732.8	
SSE	1860.8	
SST	SSR+SSE	4883.733

$\Sigma(X_{ij} - \bar{X}_j)^2$

SSE for Group 1	=(51-48.2)^2+(45-48.2)^2+(33-48.2)^2+(45-48.2)^2+(67-48.2)^2

If, F calculated falls outside the curve, then Reject the H0



From F-Table

Source	Sum of Squares (SS)	df	Mean Squ	F		F-Critical (for 2, 12)
Treatment	3022.933333	2	1511.467	9.747206	>	3.8853
Error	1860.8	12	155.0667			
Total	4883.733333					

One-way ANOVA (Problem 3)

A paper manufacturer makes grocery bags. They are interested in increasing the tensile strength of their product. It is thought that strength is a function of the hardwood concentration in the pulp. An investigation is carried out to compare four levels of hardwood concentration: 5%, 10%, 15% and 20%. Six test specimens are made at each level and all 24 specimens are then tested in random order. The results are shown below:

Hardwood Concentration (%)	Tensile strength (psi)						Mean	Standard Deviation
5	7	8	15	11	9	10	10.00	2.83
10	12	17	13	18	19	15	15.67	2.81
15	14	18	19	17	16	18	17.00	1.79
20	19	25	22	23	18	20	21.17	2.64
All							15.96	4.72

Source: Applied Statistics and Probability for Engineers - Montgomery and Runger

Do all our groups come from populations with the same mean?

One-way ANOVA (Problem 3)

Group 1	Group 2	Group 3	Group 4
7	12	14	19
8	17	18	25
15	13	19	22
11	18	17	23
9	19	16	18
10	15	18	20

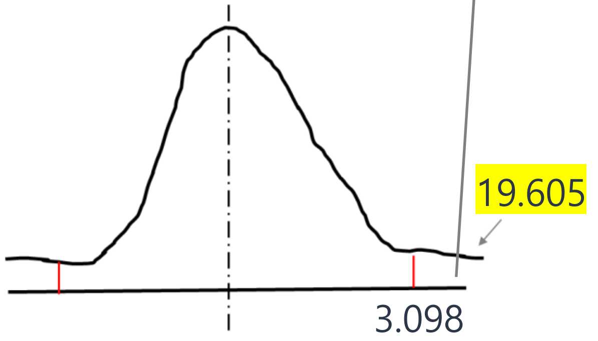
Step 1: Find Group Means and Overall Means

Step 2: Calculate Sum of Squares of Regression (SSR) $n\Sigma(X_j - \overline{X}..)^2$

Step 3: Calculate Sum of Squares of Error (SSE) $\Sigma(X_{ij} - \overline{X}_j)^2$

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	382.7917	3	127.5972	19.60521	3.59E-06	3.098391
Within Groups	130.1667	20	6.508333			
Total	512.9583	23				

If, F calculated falls outside the curve, then Reject the H0



One-way ANOVA (Problem 3)

The ANOVA table: Tensile strength of paper

Source of variation	Sum of squares	Degrees of freedom	Mean square	F	p
Between groups	382.79	3	127.60	$\frac{127.60}{6.51} = 19.61$	$p < 0.001$
Within groups	130.17	20	6.51		
Total	512.96	23			

From our results we can say that there is strong evidence that the mean tensile strength varies with hardwood concentration. Although the F-test does not specify the nature of these differences it is evident from our results that, as the hardwood concentration increases, so does the tensile strength of the paper. It is possible to test more specific hypotheses — for example, that there is an increasing or decreasing trend in the means — but these tests will not be covered in this leaflet.

One-way ANOVA (Problem 4)

A pharmaceutical company conducts an experiment to test the effect of a new cholesterol medication. The company selects 15 subjects randomly from a larger population. Each subject is randomly assigned to one of three treatment groups. Within each treament group, subjects receive a different dose of the new medication. In Group 1, subjects receive 0 mg/day; in Group 2, 50 mg/day; and in Group 3, 100 mg/day. The treatment levels represent all the levels of interest to the experimenter, so this experiment used a fixed-effects model to select treatment levels for study. After 30 days, doctors measure the cholesterol level of each subject. The results for all 15 subjects appear in the table below:

One-way ANOVA (Problem 4)

In conducting this experiment, the experimenter had two research questions:

Does dosage level have a significant effect on cholesterol level?

How strong is the effect of dosage level on cholesterol level?

To answer these questions, the experimenter intends to use one-way analysis of variance.

Dosage		
Group 1, 0 mg	Group 2, 50 mg	Group 3, 100 mg
210	210	180
240	240	210
270	240	210
270	270	210
300	270	240

One-way ANOVA (Problem 4)

Group 1	Group 2	Group 3
210	210	180
240	240	210
270	240	210
270	270	210
300	270	240

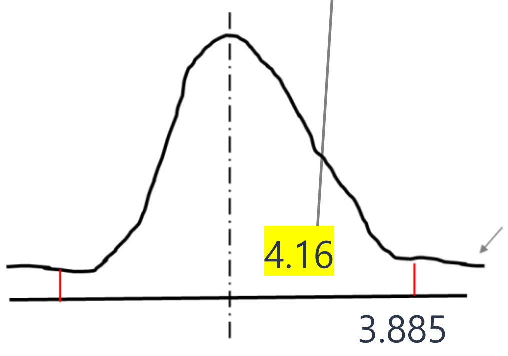
Step 1: Find Group Means and Overall Means

Step 2: Calculate Sum of Squares of Regression (SSR) $n\Sigma(X_j - \bar{X}_{..})^2$

Step 3: Calculate Sum of Squares of Error (SSE) $\Sigma(X_{ij} - \bar{X}_j)^2$

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	6240	2	3120	4.16	0.042418	3.885294
Within Groups	9000	12	750			
Total	15240	14				

If, F calculated falls inside the region, then accept the H0



One-way ANOVA (Problem 5)

A trucking company wishes to test the average life of each of the four brands of types. The company uses all brands on randomly selected trucks. The records showing the lives (thousands of miles) are as given. Test the hypothesis that the average life for each brand of types is the same at 5% level of significance.

Brand 1	20	23	18	17	-
Brand 2	19	15	17	20	16
Brand 3	21	19	20	17	16
Brand 4	15	17	16	18	-

One-way ANOVA (Problem 5)

	Brand 1	Brand 2	Brand 3	Brand 4
	20	19	21	15
	23	15	19	17
	18	17	20	16
	17	20	17	18
	-	16	16	-
Group means	19.5	17.4	18.6	16.5
Overall Means	18			

Remember, n- means the number of observations in each group. Here, Brand 1 and Brand 4 have observations of 4. But Brand 2 and Brand 3 have observations of 5. So, when substituting 'n' to calculate SSR, we need to use n=4 for Brand 1 & 4; n = 5 for Brand 2 & 3.

Step 1: Find Group Means and Overall Means

Step 2: Calculate Sum of Squares of Regression (SSR)

$n\Sigma(X_j - \bar{X}_{..})^2$

Step 3: Calculate Sum of Squares of Error (SSE)

$\Sigma(X_{ij} - \bar{X}_j)^2$

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	21.6	3	7.2	1.668874	0.219095	3.343889
Within Groups	60.4	14	4.314286			
Total	82	17				

Two-way ANOVA

- The two-way ANOVA compares the mean differences between groups that have been split on two independent variables (called factors).
- The primary purpose of a two-way ANOVA… is to understand if there is an interaction between the two independent variables on the dependent variable.
- For example, you may want to determine whether there is an interaction between physical activity level(IV) and gender(IV) on blood cholesterol concentration(DV) in children.

Two-way ANOVA

Assumptions

- Assumption #1: Your dependent variable should be measured at the continuous level (i.e., they are interval or ratio variables).
- Assumption #2: Your two independent variables should each consist of two or more categorical, independent groups.
- Assumption #3: You should have independence of observations, which means that there is no relationship between the observations in each group or between the groups themselves.
- Assumption #4: There should be no significant outliers. Outliers are data points within your data that do not follow the usual pattern
- Assumption #5: Your dependent variable should be approximately normally distributed for each combination of the groups of the two independent variables.
- Assumption #6: There needs to be homogeneity of variances for each combination of the groups of the two independent variables.

Two-way ANOVA (with interaction) method 1

ANOVA Summary Table for two-way classification				
Source of variation	**Sum of squares**	**Degree of freedom**	**Mean squares**	**F Value**
Sum of squares Between Columns	SSC	$c-1$	$MSC = \frac{SSC}{c-1}$	$F_{Treatment} = \frac{MSC}{MSE}$
Sum of squares Between rows	SSR	$r-1$	$MSR = \frac{SSR}{r-1}$	$F_{Block} = \frac{MSR}{MSE}$
Sum of squares interaction	SSI	$(c-1)(r-1)$	$MSI = \frac{SSI}{(c-1)(r-1)}$	$F_{Interaction} = \frac{MSI}{MSE}$
Sum of squares of errors	SSE	$rc(n-1)$	$MSE = \frac{SSE}{rc(n-1)}$	
Total	SST	$N-1$		

Two-way ANOVA (without interaction) method 2

Two-Way ANOVA Summary Table				
Source of variation	**Sum of squares**	**Degree of freedom**	**Mean squares**	**F Value**
Sum of squares Between Columns	SSC	$c-1$	$MSC = \frac{SSC}{c-1}$	$F_{Treatment} = \frac{MSC}{MSE}$
Sum of squares Between rows	SSR	$r-1$	$MSR = \frac{SSR}{r-1}$	
Sum of squares of errors	SSE	$(c-1)(r-1)$	$MSE = \frac{SSE}{(c-1)(r-1)}$	$F_{Block} = \frac{MSR}{MSE}$
Total	SST	$n-1$		

Two-way ANOVA (Problem 1)

An engineer is studying methods for improving the ability to detect targets on a radar scope. Two factors she considers to be important are the amount of background noise, or "ground clutter," on the scope and the "type of filter" placed over the screen. The response variable is intensity level. It is experienced that the ground clutter can be categorized into three levels, ie., low, medium, and high and two filter types are available in the market.

Factor	Filter Types							
Ground Clutter	Type - 1				Type 2			
Low (1)	90	96	100	92	86	84	92	81
Medium (2)	102	106	105	96	97	90	97	80
High (3)	114	112	108	98	93	91	95	83

Two-way ANOVA (Problem 1)

Two factors: Ground Clutter type & Filter Type

Response variable: Intensity level

DF for clutter = a – 1 = 3-1 = 2

DF for Filter = b – 1 = 2 – 1 = 1

DF for interaction = (a -1) (b-1) = 2

DF for errors = ab(n-1) = 2 * 3 *(4-1) =18

DF for total = N – 1 = 24 – 1 = 23

a = number of levels in factor 1 (clutter)
b = number of levels in factor 2 (filter)
n = number of replicates in each condition
N = abn

Factor	Filter Types							
Ground Clutter	Type - 1				Type 2			
Low (1)	90	96	100	92	86	84	92	81
Medium (2)	102	106	105	96	97	90	97	80
High (3)	114	112	108	98	93	91	95	83

Two-way ANOVA (Problem 1)

Step 1: Calculate Row, Column, and Grand Total

Factor	Filter Types									
Ground Clutter	Type - 1				Type 2					
Low (1)	90	96	100	92	86	84	92	81		721
Medium (2)	102	106	105	96	97	90	97	80		773
High (3)	114	112	108	98	93	91	95	83		794
	306	314	313	286	276	265	284	244		**2288**

Grand Total

Step 2: Count the total number of observations = N = 24

Step 3: Calculate Correction Factor (C) = (Grand Total^2) / N = 2288^2/24 = 218122.7

Two-way ANOVA (Problem 1)

Step 4: SS Total = Sum of squares of all factors responses – Correction Factor (C)

$=(90^2+96^2+100^2+92^2+86^2+84^2+92^2+81^2+102^2+106^2+105^2+96^2+97^2+90^2+97^2+80^2+114^2+112^2+108^2+98^2+93^2+91^2+95^2+83^2) - 218122.7$

SS Total = 1985.3333

Step 5: SS Between Clutter = $\frac{1}{bn} * Sum\ of\ square\ of\ factor\ 1 - Correction\ factor$

$$= \left(\frac{721^2}{2*4} + \frac{773^2}{2*4} + \frac{794^2}{2*4}\right) - 218122.7 = 353.0833$$

Factor	Filter Types									
Ground Clutter	Type - 1				Type 2					
Low (1)	90	96	100	92	86	84	92	81		721
Medium (2)	102	106	105	96	97	90	97	80		773
High (3)	114	112	108	98	93	91	95	83		794

Two-way ANOVA (Problem 1)

Step 6: SS Between Filter = $\frac{1}{an} * Sum\ of\ square\ of\ factor\ 2\ - Correction\ factor$

$= \frac{1}{3*4} *$ [(306+314+313+286) ^{2}+(276+265+284+244)^2] - 218122.7 = 937.5

Factor	Filter Types							
Ground Clutter	Type - 1				Type 2			
Low (1)	90	96	100	92	86	84	92	81
Medium (2)	102	106	105	96	97	90	97	80
High (3)	114	112	108	98	93	91	95	83
	306	314	313	286	276	265	284	244

Two-way ANOVA (Problem 1)

Step 7: SS Interaction = $\frac{1}{n} * Sum\ of\ square\ of\ both\ factor\ 2\ - Correction\ factor$ – SS Clutter – SS Filter

$$= \frac{1}{4}(378^2 + 409^2 + 432^2 + 343^2 + 364^2 + 362^2) - 218122.7 - 353.08333 - 937.5$$

SS Interaction = 81.25

Factor	Filter Types									
Ground Clutter	Type - 1					Type 2				
Low (1)	90	96	100	92	378	86	84	92	81	343
Medium (2)	102	106	105	96	409	97	90	97	80	364
High (3)	114	112	108	98	432	93	91	95	83	362

Two-way ANOVA (Problem 1)

Step 8: SS Error = SS Total – SS Clutter – SS Filter – SS Interaction

= 1985.333 – 353.0833 – 937.5 – 81.25

SS Error = 613.5

Step 9: Make ANOVA Table

Sources of variation	SS	DOF	MS	F-Value		F- Critical Value from Table	
Ground Clutter Type	353.08333	2	176.5417	5.179707	>	F(2,23) = 3.4224	Significant / Reject H0
Filter Type	937.5	1	937.5	27.50611	>	F(1,23) = 4.2793	Significant / Reject H0
Interaction	81.25	2	40.625	1.191932	<	F(2,23) = 3.4224	Non-Significant / Accept H0
Error	613.5	18	34.08333				
Total	1985.333	23	86.31883				

Two-way ANOVA (Problem 2)

Suppose you want to determine whether the brand of laundry detergent used and the temperature affects the amount of dirt removed from your laundry. To this end you buy two detergents with different brand (“Super” and “Best”) and choose three different temperature levels (“cold”, “warm” and “hot”). Then you divide your laundry randomly into “4*r” pile of equal size and assign each ‘r’ piles into the combination of (“super” and “Best”) and (“cold”, “warm” and “hot”). In this example, we are interested in testing Null Hypothesis.

H_{0D} = The amount of dirt removed does not depend on the type of detergent.

H_{0T} = The amount of dirt removed does not depend on the temperature.

Two-way ANOVA (Problem 2)

The example has two factors(factor detergent, factor temperature) at a=2(Super and Best) and b=3(cold, warm and hot) levels. Thus, there are a*b = 3*2=6 different combination of detergent and temperature with each combination. There are r=4 loads. (r is called the number of replicates). This sums up to “n=a*b*r”=24=2*3*4 loads in total.

Two-way ANOVA (Example problem)

The amounts of Y(ijk) of dirt removed when washing sub pile k(k=1,2,3,4) with detergent i(i=1,2) at temperature j(j=1,2,3) are recorded in table below:-

	Cold	Warm	Hot
Super	4,5,6,5	7,9,8,12	10,12,11,9
Best	6,6,4,4	13,15,12,12	12,13,10,13

	cold	warm	hot
	4	7	10
	5	9	12
	6	8	11
Super	5	12	9
	6	13	12
	6	15	13
	4	12	10
Best	4	12	13

Factor	Super					Best								
Cold	4	5	6	5	20	6	6	4	4	20		40		
Warm	7	9	8	12	36	13	15	12	12	52		88		
Hot	10	12	11	9	42	12	13	10	13	48		90		
	21	26	25	26		31	34	26	29				Grand Total	218

Two-way ANOVA (Example problem)

Solution:

Calculate mean

	Cold	Warm	Hot	m_D
Super	4,5,6,5 (5)	7,9,8,12 (9)	10,12,11,9 (10)	8
Best	6,6,4,4 (5)	13,15,12,12 (13)	12,13,10,13 (12)	10
m_T	5	11	11	9

	cold	warm	hot	M(d) [Y(i)]
	4	7	10	
	5	9	12	
	6	8	11	
Super	5	12	9	
	mean(Yij)=5	mean(Yij)=9	mean(Yij)=10.5 ~10	8
	6	13	12	
	6	15	13	
	4	12	10	
Best	4	12	13	
	mean(Yij)=5	mean(Yij)=13	mean(Yij)=12	10
M(t)[Y(j)]	5	11	11	9

- We have calculated all the means like detergent mean(Md), temperature mean(Mt) and mean of every group combination.
- Now what we only have to do is calculate the sum of squares(ss) and degree of freedom(df) for temperature, detergent and interaction between factor and levels.

Two-way ANOVA (Example problem)

Step 1:

Calculate of SS (within) and df(within) is:

$$SS_{within} = \sum_{i=1}^{2}\sum_{j=1}^{3}\sum_{k=1}^{4}\left(Y_{ijk} - \bar{Y_{ij\cdot}}\right)^2$$

$$\begin{aligned}
&= (4-5)^2 + (5-5)^2 + (6-5)^2 + (5-5)^2 \\
&\quad +(7-9)^2 + (9-9)^2 + (8-9)^2 + (12-9)^2 \\
&\quad \cdots\cdots \\
&\quad +(12-12)^2 + (13-12)^2 + (10-12)^2 + (13-12)^2 \\
&= 38 \\
df_{within} &= (r-1) * a * b = 3 * 2 * 3 = 18 \\
MS_{within} &= SS_{within} \Big/ df_{within} = 38/18 = 2.1111
\end{aligned}$$

Two-way ANOVA (Example problem)

Step 2:

Calculate of SS (detergent) and df(detergent) and MS(detergent)

$$
\begin{aligned}
SS_{detergent} &= r \cdot b \cdot \sum_{i=1}^{2} \left(\bar{Y_{i..}} - \bar{Y_{...}}\right)^2 \\
&= 4 \times 3 \times \left[(8-9)^2 + (10-9)^2\right] = 24 \\
df_{detergent} &= a - 1 = 1 \\
MS_{detergent} &= SS_{detergent} \Big/ df_{detergent} = 24 \Big/ 1 = 24
\end{aligned}
$$

Two-way ANOVA (Example problem)

Step 3:

Calculate of SS (temperature), df(temperature) and MS(temperature)

$$
\begin{aligned}
SS_{temperature} &= r \cdot a \cdot \sum_{j=1}^{3} \left(\bar{Y_{.j.}} - \bar{Y_{...}}\right)^2 \\
&= 4 \times 2 \times \left[(5-9)^2 + (11-9)^2 + (11-9)^2\right] = 24 \\
df_{temparature} &= b - 1 = 2 \\
MS_{temperature} &= SS_{temperature} \Big/ df_{temperature} = 192 \Big/ 2 = 81
\end{aligned}
$$

Two-way ANOVA (Example problem)

Step 4:

Calculate of SS (interaction), df(interaction) and MS(interaction)

$$
\begin{aligned}
SS_{interaction} &= r \times \sum_{i=1}^{2}\sum_{j=1}^{3}\left(\bar{Y_{ij\cdot}} - \bar{Y_{i\cdot\cdot}} - \bar{Y_{\cdot j\cdot}} + \bar{Y_{\cdots}}\right)^2 \\
&= 4 \times \left[(5-8-5+9)^2 + (9-8-11+9)^2 + (110-8-11+9)^2 + \cdots + (12-11-10+9)^2\right] = 12 \\
df_{interaction} &= (a-1)\times(b-1) = 2\times 1 = 2 \\
MS_{interaction} &= SS_{interaction} \Big/ df_{interaction} = 12 \Big/ 2 = 6
\end{aligned}
$$

Two-way ANOVA (Example problem)

Step 5: F-Test

$$MS_{detergent} \big/ MS_{within} \sim F\left(df_{detergent}, df_{within}\right)$$
$$MS_{temperature} \big/ MS_{within} \sim F\left(df_{temperature}, df_{within}\right)$$
$$MS_{interaction} \big/ MS_{within} \sim F\left(df_{interaction}, df_{within}\right)$$

ANOVA						
rce of Varic	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Deterdent	20.16667	1	20.16667	9.810811	0.005758	4.413873
Temp	200.3333	2	100.1667	48.72973	5.44E-08	3.554557
Interactio	16.33333	2	8.166667	3.972973	0.037224	3.554557
Within	37	18	2.055556			
Total	273.8333	23				

Thank you

Reference:

Douglas C. Montgomery. Design of Experiments. Eighth Edition.