ANOVA

Example 1. One-way ANOVA

Ms Rachael Khanna the brand manager of ENZO detergent powder at the 'one stop' retail was interested in understanding whether the price discounts have any impact on the sales quantity of ENZO. To test whether the price discounts had any impact, price discounts of 0% (no discount), 10% and 20% were given on randomly selected days. The quantity (in kilograms) of ENZO sold in a day under different discount levels is shown in Table below. Conduct a one-way ANOVA to check whether discount had any significant impact on the average sales quantity at $\alpha = 0.05$.

Table: Sales of ENZO at different price discounts

No Discount (0% discount)									
39	32	25	25	37	28	26	26	40	29
37	34	28	36	38	38	34	31	39	36
34	25	33	26	33	26	26	27	32	40
10% Discount									
34	41	45	39	38	33	35	41	47	34
47	44	46	38	42	33	37	45	38	44
38	35	34	34	37	39	34	34	36	41
20% Discount									
42	43	44	46	41	52	43	42	50	41
41	47	55	55	47	48	41	42	45	48
40	50	52	43	47	55	49	46	55	42

Solution:

In this case, the number of groups

$k = 3;\ n_1 = n_2 = 30;\ \mu_1 = 32,\ \mu_2 = 38.77,\ \mu_3 = 46.4;\ and\ \mu = 39.05.$

The sum of squares of between groups variation (SSB) is given by

$$SSB = \sum_{i=1}^{k} n_i \times (\mu_i - \mu)^2 = 30 \times [(32 - 39.05)^2 + (38.77 - 39.05)^2$$

$$+\ (46.4 - 39.05)^2] = 3114.156$$

So

$$MSB = \frac{SSB}{k-1} = \frac{3114.156}{2} = 1557.078$$

The sum of squares of within the group variation is given by

$$SSW = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \mu_i)^2 = \sum_{j=1}^{30}(Y_{1j} - 32)^2 + \sum_{j=1}^{30}(Y_{2j} - 38.77)^2$$

$$+ \sum_{j=1}^{30}(Y_{3j} - 46.4)^2 = 2056.567$$

$$MSW = \frac{SSW}{n-k} = \frac{2056.567}{90-3} = 23.63$$

The F-statistic value is

$$F_{2,87} = \frac{MSB}{MSW} = \frac{1557.078}{23.6387} = 65.86$$

The critical F-value with degrees of freedom (2, 87) for $\alpha = 0.05$ is 3.101 [Excel function FINV(0.05, 2, 87) or F.INV.RT(0.05, 2, 87)]. The p-value for $F_{2,87} = 65.86$ is 3.82×10^{-18} [using Excel function FDIST(65.86, 2, 87) or F.DIST.RT(65.86, 2, 87)]. Since the calculated F-statistic is much higher than the critical F-value, we reject the null hypothesis and conclude that the mean sales quantity values under different discounts are different. The Excel output of ANOVA is shown in Table below.

Table: One-way ANOVA excel output

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
No Discount	30	960	32	27.17241		
10% Discount	30	1163	38.76667	20.46092		
20% Discount	30	1392	46.4	23.28276		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Between Groups	3114.15556	2	1557.078	65.86986	3.82E-18	3.101296
Within Groups	2056.56667	87	23.6387			
Total	5170.72222	89				

Example 2. One-way ANOVA

Share Raja Khan (SRK) is a top stockbroker and believes that the average annual stock return depends on the industrial sector. To validate his belief, SRK collected annual return of shares from three different industrial sectors − consumer goods, services, and industrial goods. The annual return of shares in 2015−2016 for different sectors is shown in Table below.

Table: Annual return of stocks under different industrial sector

Annual return on 30 consumer goods stocks									
6.32%	14.73%	11.95%	12.36%	10.28%	3.81%	10.15%	11.06%	6.29%	5.15%
8.44%	14.28%	8.89%	5.98%	6.96%	11.62%	5.22%	5.34%	5.93%	7.10%
10.91%	8.20%	10.19%	9.04%	8.61%	9.39%	2.63%	2.77%	4.76%	9.60%
Annual return on 30 services stocks									
13.70%	3.58%	1.36%	17.41%	10.01%	10.88%	15.63%	−0.04%	10.32%	7.40%
11.48%	9.71%	11.19%	8.21%	1.64%	1.45%	10.12%	13.85%	−10.27%	5.26%
12.05%	4.47%	8.71%	5.59%	10.02%	7.65%	10.03%	7.87%	6.59%	13.60%
Annual return on 30 industrial goods stocks									
6.74%	7.11%	5.69%	2.48%	5.42%	8.00%	2.55%	8.34%	4.99%	3.39%
8.73%	13.85%	5.29%	9.06%	2.84%	5.82%	7.66%	4.12%	9.10%	8.76%
10.77%	1.48%	4.71%	10.66%	0.44%	2.94%	6.55%	2.84%	3.90%	7.28%

Solution:

In this case, the number of groups

$k = 3;\ n_1 = n_2 = 30;\ \mu_1 = 0.082,\ \mu_2 = 0.079,\ \mu_3 = 0.0605;\ and\ \mu = 0.0743$

.

The sum of squares of between groups variation (SSB) is given by

$$SSB = \sum_{i=1}^{k} n_i \times (\mu_i - \mu)^2 = 30 \times [(0.082 - 0.0743)^2 + (0.079 - 0.0743)^2$$

$$+ (0.0605 - 0.0743)^2] = 0.0087$$

Therefore

$$MSB = \frac{SSB}{k-1} = \frac{0.0087}{2} = 0.0043$$

The sum of squares of within the group variation is given by

$$SSW = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \mu_i)^2 = \sum_{j=1}^{30}(Y_{1j} - 0.082)^2 + \sum_{j=1}^{30}(Y_{2j} - 0.079)^2$$

$$+ \sum_{j=1}^{30}(Y_{3j} - 0.0605)^2 = 0.1463$$

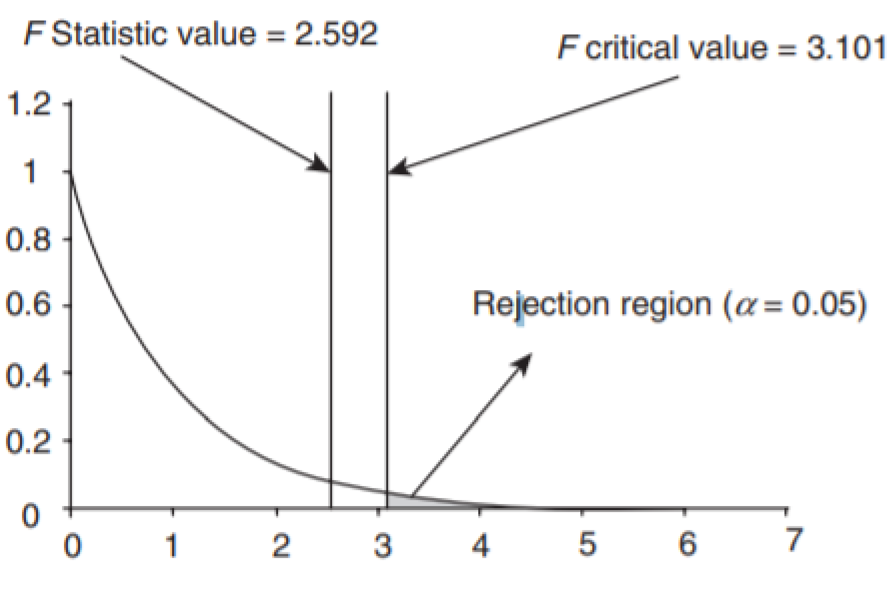
So

$$MSW = \frac{SSW}{n-k} = \frac{0.1463}{90-3} = 0.0016$$

The F-statistic value is

$$F_{2,\,87} = \frac{MSB}{MSW} = \frac{0.0043}{0.0016} = 2.592$$

The critical F-value with degrees of freedom (2, 87) for α = 0.05 is 3.101 [Excel function FINV(0.05, 2, 87) or F.INV.RT(0.05, 2, 87)]. The P-value for $F_{2,\,87}$ = 2.592 is 0.0805 [using Excel function FDIST(2.592, 2, 87) or F.DIST.RT(2.592, 2, 87)]. Since the calculated F-statistic is less than the critical F-value, we retain the null hypothesis and conclude that the average annual returns under industrial sectors consumer goods, services, and industrial goods are not different (the below figure shows the F-critical value and F-statistic value for an F-distribution with degrees of freedom 2 and 87 for numerator and denominator, respectively).



F-distribution with critical value

The Excel output of ANOVA is shown in Table below.

Table: Microsoft excel ANOVA table

ANOVA: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Consumer Goods	30	2.4796	0.082653	0.00101		
Services	30	2.3947	0.079823	0.003073		
Industrial Goods	30	1.8151	0.060503	0.000963		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*p-value*	*F critical*
Between Groups	0.008722	2	0.004361	2.59294	0.080572	3.101296
Within Groups	0.146317	87	0.001682			
Total	0.155039	89				

Example 3. Two-way ANOVA

Table below shows the sales quantity of detergents at different discount values and different locations collected over 20 days. Conduct a two-way ANOVA at $\alpha = 0.05$ to test the effects of discounts and location on the sales.

Table: Sales quantity at different locations under different discount rates

	Location 1		Location 2		
	Discount		Discount		
0%	10%	20%	0%	10%	20%
20	28	32	20	19	20
16	23	29	21	27	31
24	25	28	23	23	35
20	31	27	19	30	25
19	25	30	25	25	31
10	24	26	22	21	31
24	28	37	25	33	31
16	23	33	21	26	23
25	26	27	26	22	22
16	25	31	22	28	32
18	22	37	25	24	22
20	24	28	23	23	29
17	26	25	23	26	25
26	28	23	24	16	34
16	21	26	20	30	30
21	27	33	23	22	25
24	25	28	18	16	39
19	20	30	19	25	32
19	26	30	19	34	29
21	26	26	30	23	22

The two-way ANOVA with replication (since the data in Table above is repeated for locations) output from Microsoft Excel is shown in Table below.

Table: Two-way ANOVA with replication excel output

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Sample (Location)	7.008333	1	7.008333	0.443898	0.506593	3.92433
Columns (Discount)	1240.317	2	620.1583	39.27997	1.06E-13	3.075853
Interaction	84.81667	2	42.40833	2.686085	0.07246	3.075853
Within	1799.85	114	15.78816			
Total	3131.992	119				

In Table above, the sample stands for the row factor (which in this case is location), column stands for the column factor (discount in this case), and interaction stands for interaction effect (location × discount). The p-value for locations (data in rows) is 0.5065, thus it is not statistically significant (we retain the null hypothesis that the locations have no statistically significant influence on sales), whereas for discount rates (data in column) the p-value is 1.06×10^{-13}, so we reject the null hypothesis (that the discount rate has influence on sales). The p-value for the interaction effect is 0.0724 and is not significant. That is, only the factor discount is statistically significant at $\alpha = 0.05$.