ANOVA

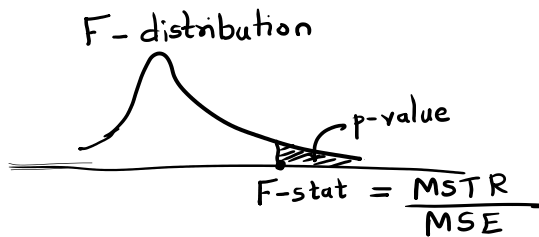r: no. of treatments
n: Total obs.

Sources of Variation	Sums of Squares	Degrees of freedom	Mean Square	F Ratio F-stat	P-Value
Treatment	SSTR	$r-1$	MSTR=SSTR / $(r-1)$	MSTR/MSE	
Error	SSE	$n-r$	MSE = SSE / $(n-r)$		
Total	SST	$n-1$			

MSTR: measure of Between Variation
MSE: measure of within variation

If MSTR > MSE $\Rightarrow$ Trt unequal
o.w.   Trt equal

F-distribution

F-stat $= \dfrac{MSTR}{MSE}$

p-value

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$
H_1: Not all μ_i (i = 1,2,3,4) are equal

```
             sum_sq     df          F      PR(>F)
Treatments  1551.607762  3.0  18.293252  0.000006   < 0.05
Residual     565.457238  20.0       NaN       NaN
```

$\therefore$ We reject H_0 at 5% l.o.s.
Conclusion:- Treatments may not be homogeneous

H_0: Group2 $\overset{?}{=}$ Group 1

```
group1 group2   meandiff    p-adj     lower      upper   reject
    I      II    13.0976    0.0014    4.8177    21.3775   True
    I     III    -0.6567    0.9969   -9.6685     8.3552   False
    I      IV    18.1000    0.0001    9.5075    26.6925   True
   II     III   -13.7543    0.0014  -22.4686    -5.0399   True
   II      IV     5.0024    0.3541   -3.2775    13.2823   False
  III      IV    18.7567    0.0001    9.7448    27.7685   True
```

$\rightarrow \mu_{II} > \mu_I$
$\rightarrow \mu_{III} = \mu_I$
$\rightarrow \mu_{IV} > \mu_I$
$\rightarrow \mu_{II} > \mu_{III}$
$\rightarrow \mu_{II} = \mu_{IV}$
$\rightarrow \mu_{IV} > \mu_{III}$

$\mu_{II}, \mu_{IV} > \mu_I, \mu_{III}$

```
In [51]: agr.groupby('Treatments')['Yield'].mean()
Out[51]:
Treatments
I       23.716667
II      36.814286
III     23.060000
IV      41.816667
```

A college is trying to determine if there is a significant difference in the mean GMAT score of students from different undergraduate backgrounds who apply to the MBA program. The Excel file *GMAT Scores* contains data from a sample of students. What conclusion can be reached using ANOVA?

	sum_sq	df	F	PR(>F)
Major	2983.945344	2.0	14.947815	0.00002
Residual	3493.423077	35.0	NaN	NaN

group2 - group1

group1	group2	meandiff	p-adj	lower	upper	reject
Business	Liberal Arts	-12.0769	0.0371	-23.5391	-0.6147	True	→ LA < Bus
Business	Sciences	11.4231	0.0093	2.5240	20.3222	True	→ Sci > B
Liberal Arts	Sciences	23.5000	0.0000	12.6092	34.3908	True	→ Sci > LA

Sci > B > LA

Conclusion: Science mean GMAT > Business mean GMAT > Liberal Art Mean GMAT

- *A magazine reports percentage returns and expense ratios for stock and bond funds.* The data FUNDS.csv are the expense ratios for 10 midcap stock funds, 10 small-cap stock funds, 10 hybrid stock funds, and 10 specialty stock funds.

- Test for any significant difference in the mean expense ratio among the four types of stock funds.

	sum_sq	df	F	PR(>F)
Fund	2.603	3.0	2.94346	0.045936
Residual	10.612	36.0	NaN	NaN

group1	group2	meandiff	p-adj	lower	upper	reject
Hybrid	Midcap	-0.32	0.5578	-0.9739	0.3339	False
Hybrid	Small-cap	0.02	0.9998	-0.6339	0.6739	False
Hybrid	Specialty	0.40	0.3659	-0.2539	1.0539	False
Midcap	Small-cap	0.34	0.5074	-0.3139	0.9939	False
Midcap	Specialty	0.72	0.0262	0.0661	1.3739	True
Small-cap	Specialty	0.38	0.4108	-0.2739	1.0339	False

Spec > Midcap