```
from scipy.stats import mannwhitneyu
stats, p = mannwhitneyu(df1.Design1, df1.Design2)
print(stats, p)
```

9.0 0.9641

$p > 0.05$ → accept H_0

(Both design1 & design2 have same sales).

Kruskal-Wallis Test

```
df2 = pd.read_excel('kruskal wallis.xlsx', sheet_name=0)
df2
```

```
from scipy.stats import kruskal
stats, p = kruskal(df2.Design1, df2.Design2, df2.Design3)
print(stats, p)
```

2.734 0.254

$p > 0.05$

∴, H_0 → accepted

(design 1/2/3 are same, there is no diff. in sales.)

Chi-square

```
df3 = pd.read_excel('chisquaretest.xlsx', sheet_name=0)
df4 = df3.dropna()
```

↳ removes all Null value & store in df4.

(When Both var. are categorical).

```
from scipy.stats import chi2_contingency.
chitable = pd.crosstab(df4.Gender, df4.Smoking).
chitable

stats, p, dof, expected = chi2_contingency(chitable)
print(stats, p)
```

3.1712 0.2048

$p > 0.05$ (we accept H_0)

No dependency on gender & smoking.

Practical Implementation of Parametric Test

One-Sample Test.

```
df5 = pd.read_excel('one sample.xlsx', sheet_name = 0)
df5.head()

from scipy.stats import ttest_1samp        → popln mean.
stats, p = ttest_1samp(df5.Height, 65)
print(stats, p)
```

11.498 1.087.e-26 (exponential e-26)

$P < 0.05$ ∴, reject H_0 accept H_1

TWO

1 Sample paired ttest

```
df6 = pd.read_excel('Pairedsample.xlsx', sheet_name=0)
df6.head()
```

```
from scipy.stats import ttest_rel
stats, p = ttest_rel(df6.English, df6.Math)
print(stats, p)
```

36.3156 → 3.0717-e-128 (exponential) → P ≤ 0.05 (accept H1)

Two sample independent ttest .

```
df7 = pd.read_excel('IndependentSample.xlsx', sheet_name = 3)
df7.head()
```

```
from scipy.stats import ttest_ind
stats, p = ttest_ind(df7.Nmathlete, df7.athelete)
print(stats, p)
```

13.368 → 4.116 e-33 → P < 0.55 (accept H1)

(there is difference b/w time of athlete & non-athlete.)

One-sample f test (one way ANOVA)

→

- ANOVA is used to compare 2 or more sample mean.

4 Friday
March 2011

March 2011

Wk	Mo	Tu	We	Th	Fr	Sa	Su
9		1	2	3	4	5	6
10	7	8	9	10	11	12	13
11	14	15	16	17	18	19	20
12	21	22	23	24	25	26	27
13	28	29	30	31			

Week 9 ▪ 63-302

	Statistical Technique	Dependent variable	Independn variable	Purpose of Technique.
1.	ANOVA (analysis of variance).	continous	categorical (also called factors)	It is used to compare ~~the mean of~~ two or more sample mean
2.	ANCOVA (analysis of covariance)	continous	continous & categorical	It is used to compare of covariance of sample mean.

Ex

- Class A v/s Class B (Between comparision).
- within your class itself (within class A). (within comparision).

ANOVA is used in Between as well as within comparision.

{ depn var → continous
{ indepn categorical var → factors. (also called)

- oneway ANOVA (1 depn var & 1 independn var.)
- Two way ANOVA (1 depn var & 2 independn var)
- multi way ANOVA (1 depn var. & more than 2 indepn. var.)

Metric Dependent Variable.

- Variable
  - Binary.
    - t Test.
- Independent Variable.
  - Categorical Factorial
    - Analysis of variance
      - One factor
        - One-Way Analysis of variance
      - More than one Factor
        - N-Way Analysis of variance.
  - categorical & continous
    - Anaysis of covariance.
  - continous
    - Regression

10
11
12
1
2
3
4
5
6

- When Both variable are continous → Regression
- Indepn var → categorical → ANOVA
- Indepn var → categorial + continous → ANCOVA

One way Analysis of Variance.

- eta(η^2) → deciding factor. → how much indepn. var is going to affect dependent. var.

$0 < \eta^2 < 1$

$\eta^2 = 0.725$ → 72.5% indepn var is affecting depn var.

- **F statistics.**

indepn var → categorical.

Null hypo. that category means are equal in popul^n is tested by an F statistic based on ratio of mean sq. related to x & mean sq. related to error.

- **Mean square.**

Sum of sq. is divided by appropriate degree of freedom.

- **$SS_{between}$ (SS_x)**

represents variation b/w categories of X or portion of sum of sq. in Y related to X

- **SS_{within} (SS_{error})**

. variation in Y due to variation with each of categories of X.

- **SS_y** → total variation in q. Y

(1) Indentify the Dependent (continous) & Independent (categorical) Var.

Steps involved in ANOVA.

(2) Decompose Total Variation
(3) Measure effects
(4) Test the significance
(5) Interpret result.

$SS_y = SS_{between} + SS_{within}$.

$SS_y = SS_x + SS_{error}$. $\longrightarrow \sum_{j}^{c}\sum_{i}^{n}(y_{ij} - \bar{y}_j)^2$

$\sum_{i=1}^{n}(y_i - \bar{y})^2$ $\qquad \sum_{j=1}^{c} n(\bar{y}_j - \bar{y})^2$

~~where y_i~~

9
10
11
12
1
2

$\eta^2 = \frac{SS_x}{SS_y} = \frac{SS_y - SS_{error}}{SS_y}$.

varies b/w 0 and 1.

$s_y^2 = \frac{SS_x}{C-1}$ or $s_y^2 = \frac{SS_{error}}{(N-C)}$

$H_0 : \mu_0 = \mu_1 = \mu_2 \cdots = \mu_c$

$H_1 :$ ⊗

= mean sq. due to x

$= MS_x$

= mean sq. due to error

$= MS_{error}$

$F = \frac{SS_x / C-1}{SS_{error} / N-C}$ → No. of categories

F stats: $F = \frac{MS_x}{MS_{error}}$

compare calculated F stats & F value by table.

9 Wednesday
March 2011

March 2011

Wk	Mo	Tu	We	Th	Fr	Sa	Su
9		1	2	3	4	5	6
10	7	8	9	10	11	12	13
11	14	15	16	17	18	19	20
12	21	22	23	24	25	26	27
13	28	29	30	31			

Week 10 ▪ 68-297

Day 13.

Output variable (Target Var.) ⟶ Dependent Var.
Input var. ⟶ Independents var.

- One Way → one Dependent + 1 Indepen.

Ex Sales (depen^) V/S Promotion of item (indepn.)

~~In Python,~~

~~F < 0.05~~ ~~E~~ ~~Prob.~~

- Two way. → 1 dependⁿ + 2 Independⁿ.

Ex

Sales V/S Promotion, coupon
dependⁿ Indepn.

ANCOVA.

Analysis of covariance
Used when depn → continous
indepⁿ var ⟶ continous + categorical

Ex DV → Sales
IDV → Promotion, coupon, client Rating.

- Python Implementation.

```
df = pd.read_excel('ANCOVA.xlsx', sheet_name = 0)
df.head()
```

ONE WAY ANOVA.

import statsmodels.api as sm
from statsmodels.formula.api import ols → ordinary least square

model = ols('Sales ~ C(Promotion)', df).fit()
↓ DV ↓ categorical var. ↓ IDV

DV & IDV are differentiated by ~

- typ=3 (discriminant analysis)
- typ=1 (Regression)

standard rule for anova & ancova

onewayanova = sm.stats.anova_lm(model, typ=2)
print(onewayanova).

	sum_sq	df	F	PR(>F)
C(Promn)	106.0667	2.0	17.94	0.000011
Residual	79.800	27.0	NaN	NaN

(SSx) # O/P: — df ↑ c-1 — Prob.value. — P<0.05

B/w variation ← C(Promn)

Residual → within variation (SSerror) → N-C

F value ~ calculated F value.

BCZ of promotion, there is significant impact on sales.

TWOWAY ANOVA.

model = ols('sales ~ C(Promotion) + C(coupon)', df).fit()
twoway = sm.stats.anova_lm(model, typ=2).
twoway.

#	sum_sq	df	F	PR(>F)
C(Promotion)	106.0667	2.0	52.098	8.032 e-10
C(coupon)	53.333	1.0	52.392	1.0950 e-07
Residual	26.4667	26.0	Nan	NaN

Both are affect sale

Week 10 ▪ 70-295

① ~~Multiway~~ ANCOVA.

```
model = OLS('Sales ~ C(Promotion) + C(Coupon) + ClientRating', df).fit()
ancova = sm.stats.anova_lm(model, typ=2)
```

	PR(>F)	
Promotion	1.301 e-09	} has impact on sales.
coupon	1.471 e-07	
clientRating	3.745 e-01	→ No impact on sales (p>0.05)

Finding eta (η^2).

```
def anova_table(onewayanova):
    onewayanova['eta_sq'] = oneway[:-1]['sum_sq']/
                            sum(onewayanova['sum_sq'])
    cols = ['sum_sq', 'df', 'F', 'PR(>F)', 'eta_sq']
    oneway.anova = onewayanova[cols]
    return onewayanova;

anova_table(onewayanova)
anova_table(twoway).
anova_table(ancova)
```

②