PARAMETRIC TEST

1. One-Sample Test-
    - Used for comparing population mean(μ) with sample mean($\bar{x}$).
    - E.g – we want to compare the height students in a school(population) with height of students in a class(sample mean)
    - H0 : There is no significance difference in μ and $\bar{x}$.
      H1: There is significant difference in in μ and $\bar{x}$.
      ```
      import ttest_1samp
      stats,p=ttest_1samp(dataset.Height, μ)
      ```
    - If p<0.05 : reject H0:There is significant difference between mean height of the students of the class and mean height of students of the school.
      If p>=0.05: accept H0
2. Two-sample paired test –
    - Compare mean of 2 paired samples(the samples of same person/object in different situations)
    - E.g - Compare English and maths marks of same class students.
    - H0: There is no significant difference in their English and maths marks.
      H1: There is a significant difference in their English and maths marks.
      ```
      From scipy.py import ttest_rel
      Stats,p = ttest_rel(dataset.english,dataset.math)
      ```

- Apply p-value statistic rules.

3. Two-sampled Independent t-test –
    - Compare mean of 2 Independent samples
    - E.g – Compare marks of 2 different classes
    - H0: There is no significant difference in average marks of class A and class B.
      H1: There is no significant difference in average marks of class A and class B.
    - From scipy.stats import ttest_ind
      Stats,p = ttest_ind(dataset.marks_classA,dataset.marks_classB)

4. Anova and Ancova-

Statistical Technique	Dependent variable(Y)	Independent variable(X)	Purpose
ANOVA(Analysis of variation)	Continous	Categorical	Used to compare 2 or more sample means
ANCOVA (Analysis of Covariance)	Continous	Continous and Categorical	To compare the covariance of sample means.

Anova -

- 1-way anova – 1 Dependent variable, 1 independent variable.
- 2-way anova – 1 Dependent variable, 2 independent variable
- Multi- way anova – 1 Dependent variable, more than 2 independent variable

Between and within comparison- Consider 2 sections A nad B each having 20 students.

- Comparing marks of English and maths of students of class A – Within comparison
- Comparing average English marks of class A and class B – between comparison.

One -way anova –

Data – Sales info of 30 stores with different levels of promotion(Low, Medium, High).i.e 10 stores with low promotion, 10 with Medium and 10 with High.

Because of promotion will there be any variations in sales?

Conducting One-way anova :

1) Identify Dependent and Independent Variable.
    - D.V – Sales value of stores(continuous)
    - ID.V – Promotion(Low, Medium, High)-(categorical)
2) Decompose the Total Variation :

Consider following data for the above example.

	Sales In	Different	Promotion Levels

Store No.	High	Medium	Low
1	10		
2	..		
…Till 10	..values of sales…	..values of sales…	..values of sales…
Sum:	83	62	37
Category mean($\bar{y}j$):	83/10 = 8.3	6.2	3.7

- Grand Mean ($\bar{y}$) : (83+62+37)/30 = 6.067
- Between variation : **SSx =** $\sum_{j=1}^{c} \boldsymbol{n(\bar{y}j - \bar{y})^2}$ **=** 10(8.3-6.067)^2 + 10(6.2-6.067)^2 +10(3,7-6.067)^2 = <u>106.067</u>
- Within variation : **SSerror =** $\sum_{j}^{c} \sum_{i}^{n} \boldsymbol{(yij - \bar{y}j)^2}$ = 79.80 , where : yi: individual observation, $\bar{y}j$:mean of category, ($\bar{y}$) : grand mean, yij : ith observation in jth category,n =10.
- Total variation **: SSy =** $\sum_{i=1}^{N} \boldsymbol{(yi - \bar{y})^2}$ **,** N = total observation=30; So, here:SSy = 185.867
- **<u>SSy = SSx + SSerror</u>**

3) Deciding Factor - η^2(Eeta square)

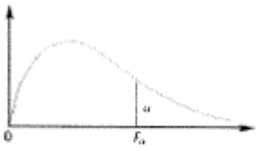
- Measures the strength of effects of Independent var on the D.V .
- Always between 0 to +1.
- η^2= SSx / SSy = between variation/Total variation
- So, η^2= 106.067/185.867 = 0.571

- This means that : 57.1% of variation in Sales(DV) is observed due to variation in promotion levels.

4) F-test :

- F-test is the statistical test used for anova.
- H0: No significant difference in sales due to promotion.
- α=0.05, There will be 2 d.o.f : c-1 and N-c
  So, dof : 3-1=**2**, 30-3=**27.**
- Calculated F- value : $\mathbf{F = \frac{\frac{ss_x}{c-1}}{\frac{SSerror}{N-c}} = MSx/\ MSerror}$
- Find the table value :

Critical Values of the *F* Distribution for $\alpha = 0.05$



Denominator Degrees of Freedom (k_2)	Numerator Degrees of Freedom (k_1)																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.83
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Table value : 3.3541

- Inferences:

If T.V(Table value)> C.V(Calculated value):Accept H0
T.V<C.V: Reject H0

Here, as TV <CV : We reject H0
Hence, the conclusion is that there is a significance difference in sales due to promotion.

Two-way anova-
e.g- Finding the variation in sales due to 2 independent variables say Promotion(Low, medium, High) and Coupon Level(A,B,C).

ANCOVA-

e.g – Finding the variation in sales due to Promotion(Low, medium, High) and Coupon Level(A,B,C) – categorical as well as Client rating(0-5) -continuous independent variables.