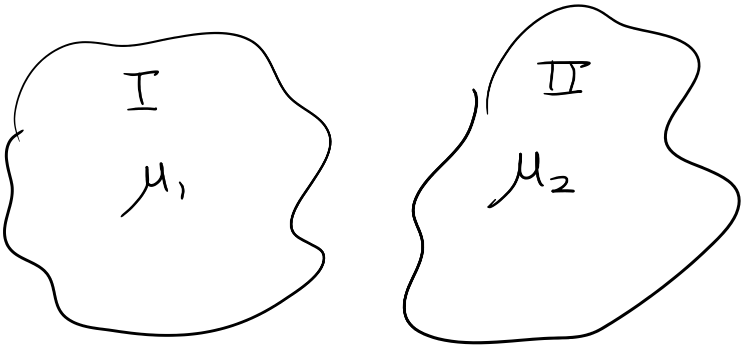
ANOVA

30 May 2023 10:40

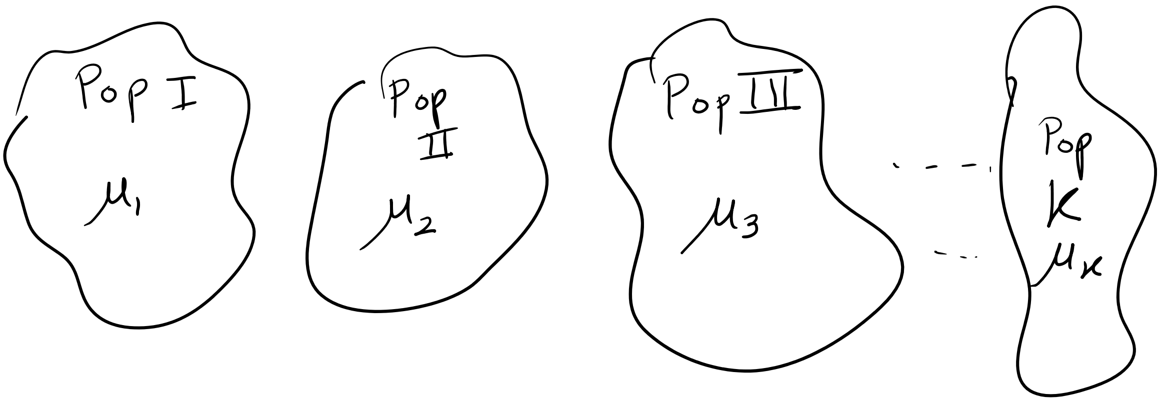
One-Way $\underbrace{\text{ANOVA}}_{\text{Analysis of Variance}}$

- Parametric test which assumes Normal Distribution of all populations

2 Indep Samples



$$\mu_1 \overset{?}{=} \mu_2$$



H_0: $\mu_1 = \mu_2 = \mu_3 = \cdots = \mu_K$

H_1: At least one μ_j is different $j = 1, 2, \ldots K$

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

I	II ✓	III	IV ✓
23.4	34.2	23.8	36.7
24.1	45.2	24.5	39.5
19.6	24.9	29.3	43.2
23.9	40.3	18.3	50.2
29.4	39.4	19.4	47.2
21.9	35.3		34.1
	38.4		

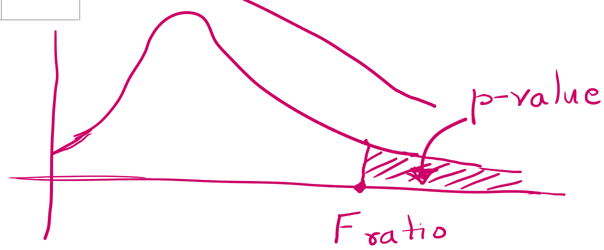
Treatments		Yield
I	$\overline{x_1}$	23.716667
II	$\overline{x_2}$	36.814286
III	$\overline{x_3}$	23.060000
IV	$\overline{x_4}$	41.816667

H_1: At least one μ_j is different

Sources of Variation	Sums of Squares	Degrees of freedom	Mean Square	F Ratio	P-Value

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

F-ratio follows F-Distribution



MSTR : Variation betⁿ Treatments

MSE : Variation within Treatment (Error Mean)

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

< 0.05

∴ We reject Ho at 5% l.o.s.

All 4 treatments might not be having same effect.

grp2 – grp1 p-value 95% Confidence Interval Ho

```
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      18.1  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
```

I < II; I = III; I < IV; II > III; II = IV; III < IV

Every line's Hypothesis test

Ho: group 1 = group 2 $\quad$ H₁: group1 ≠ group2

I < II, IV

III < II, IV

I, III < II, IV

```
In [31]: plant['group'].unique()
Out[31]: array(['ctrl', 'trt1', 'trt2'],
dtype=object)
```

$H_0: \mu_{ctrl} = \mu_{trt1} = \mu_{trt2}$

```
In [35]: ols_plant = ols('weight ~ group', data=plant).fit()
    ...: table = anova_lm(ols_plant, typ=2)
    ...: print(table)
             sum_sq    df          F    PR(>F)
```

```
   ...: table = anova_lm(ols_plant, typ=2)
   ...: print(table)
             sum_sq    df         F    PR(>F)
group       3.76634   2.0  4.846088  0.01591
Residual   10.49209  27.0       NaN       NaN
```

< 0.05

$\therefore$ We reject H_0 at 5% l.o.s.

```
group1  group2  meandiff   p-adj   lower   upper  reject
  ctrl    trt1    -0.371  0.3909 -1.0622  0.3202   False
  ctrl    trt2     0.494   0.198 -0.1972  1.1852   False
  trt1    trt2     0.865   0.012  0.1738  1.5562    True
```

ctrl $\approx$ trt1

ctrl $\approx$ trt2

trt1 $<$ trt2

rejected

grp2 − grp1

tr2 − tr1 > 0

tr2 $>$ tr1

Cars93.csv :-

Price

$H_0: \mu_{\text{no airbag}} = \mu_{\text{Driver airbag}} = \mu_{\text{D \& P airbags}}$

```
In [40]: ols_air = ols('Price ~ AirBags', data=cars93).fit()
    ...: table = anova_lm(ols_air, typ=2)
    ...: print(table)
               sum_sq    df          F        PR(>F)
AirBags   2746.983995   2.0  21.177572  2.901187e-08
Residual  5837.037296  90.0        NaN           NaN
```

< 0.05

$\therefore$ We reject H_0 at 5% l.o.s.

Price Means w.r.t airbags may be different

$\Rightarrow$ Price is influenced by Air Bags

grp2 − grp1

```
              group1        group2  meandiff   p-adj     lower   upper  reject
Driver & Passenger   Driver only   -7.1455  0.0089  -12.7657 -1.5253    True
Driver & Passenger          None  -15.1952     0.0  -21.0136 -9.3768    True
         Driver only          None   -8.0497  0.0001  -12.4542 -3.6453    True
```

1. Drive Only $<$ Drive P
2. None $<$ Drive P
3. None $<$ Drive only

$\Rightarrow$ None $<$ Drive only $<$ Drive Pass.

$H_0: \mu_{USA} = \mu_{\text{non-USA}}$

```
In [43]: ols_org = ols('Price ~ Origin', data=cars93).fit()
    ...: table = anova_lm(ols_org, typ=2)
    ...: print(table)
               sum_sq    df        F    PR(>F)
Origin      87.050054   1.0  0.93228  0.33683
Residual  8496.971236  91.0      NaN      NaN
```

> 0.05

∴ We don't reject Ho at 5% l.o.s.

Price is unaffected by Origin

$$H_0: \mu_s = \mu_c = \mu_m = - -$$

```
In [45]: ols_type = ols('Price ~ Type', data=cars93).fit()
    ...: table = anova_lm(ols_type, typ=2)
    ...: print(table)
               sum_sq    df          F        PR(>F)
Type      3421.435111   5.0  11.531618  1.476999e-08
Residual  5162.586180  87.0        NaN           NaN
```

< 0

∴ We reject Ho at 5% l.o.s.

Price is affected by Type.

grp2 - grp1

group1	group2	meandiff	p-adj	lower	upper	reject
Compact	Large	6.0875	0.341	-2.7052	14.8802	False
Compact	Midsize	9.0057	0.0078	1.6298	16.3816	True
Compact	Small	-8.0458	0.0265	-15.4953	-0.5964	True
Compact	Sporty	1.1804	0.9983	-7.0351	9.3958	False
Compact	Van	0.8875	0.9998	-8.4662	10.2412	False
Large	Midsize	2.9182	0.908	-5.3716	11.208	False
Large	Small	-14.1333	0.0001	-22.4887	-5.778	True
Large	Sporty	-4.9071	0.6131	-13.9521	4.1378	False
Large	Van	-5.2	0.664	-15.29	4.89	False
Midsize	Small	-17.0515	0.0	-23.9002	-10.2028	True
Midsize	Sporty	-7.8253	0.043	-15.5002	-0.1505	True
Midsize	Van	-8.1182	0.0934	-17.0008	0.7645	False
Small	Sporty	9.2262	0.0102	1.4806	16.9718	True
Small	Van	8.9333	0.0504	-0.0105	17.8772	False
Sporty	Van	-0.2929	1.0	-9.8841	9.2984	False

midsize > Compact > small

large > small

midsize > Sporty

sporty > small