F- distribution (Fisher - Snedecor dist)

→ The F-distribution is a probability distribution that is useful in context of comparing variances of two or more samples.

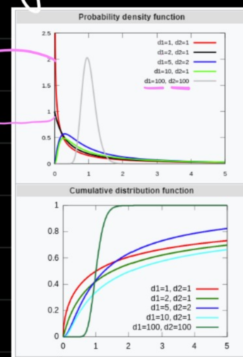
→ It is right skewed and takes only non negative values.

$$\text{pdf}(f\ x; d_1, d_2) = f(x; d_1, d_2) = \frac{\sqrt{\frac{(d_1 x)^{d_1}\, d_2^{d_2}}{(d_1 x + d_2)^{d_1+d_2}}}}{x\, \underbrace{\mathrm{B}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)}_{\text{Beta } f^n}}$$

$$\mathrm{B}(m,n) = \frac{(m-1)!\,(n-1)!}{(m+n-1)!} = \frac{m+n}{mn} \Big/ \binom{m+n}{m}.$$

$d_1 = 1$ $d_2 = 1$

$d_1 = 2$
$d_2 = 1$



The F dist^n with d_1 & d_2 degree of freedom is the distribution of $X = \dfrac{S_1/d_1}{S_2/d_2}$ where S_1 & S_2 are independent random variables with chi-square d^n and degree of freedom d_1 & d_2 respectively.

$$\underline{\text{F statistics}} = \left.\frac{S_1^2}{S_2^2}\right\}$$

(variance ratio test)

$$\frac{S_1^2}{S_2^2}$$

σ – populat std
S – sample std

F distribution critical value landmarks

Table entries are critical values for *F** with probably *p* in right tail of the distribution.

Figure of *F* distribution (like in Moore, 2004, p. 656) here.

Degrees of freedom in denominator (df2)		Degrees of freedom in numerator (df1)										
	p	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**12**	**24**	**1000**
1	0.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	60.71	62.00	63.30
	0.050	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	243.9	249.1	254.2
	0.025	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.6	976.7	997.3	1017.8
	0.010	4052	4999	5404	5624	5764	5859	5928	5981	6107	6234	6363
	0.001	405312	499725	540257	562668	576496	586033	593185	597954	610352	623703	636101
2	0.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.41	9.45	9.49
	0.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.41	19.45	19.49
	0.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.41	39.46	39.50
	0.010	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.42	99.46	99.50
	0.001	998.38	998.84	999.31	999.31	999.31	999.31	999.31	999.31	999.31	999.31	999.31
3	0.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.22	5.18	5.13
	0.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.74	8.64	8.53
	0.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.34	14.12	13.91
	0.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.05	26.60	26.14
	0.001	167.06	148.49	141.10	137.08	134.58	132.83	131.61	130.62	128.32	125.93	123.52
4	0.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.90	3.83	3.76
	0.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.91	5.77	5.63
	0.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.75	8.51	8.26
	0.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.37	13.93	13.47
	0.001	74.13	61.25	56.17	53.43	51.72	50.52	49.65	49.00	47.41	45.77	44.09
5	0.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.27	3.19	3.11
	0.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.68	4.53	4.37
	0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.52	6.28	6.02
	0.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	9.89	9.47	9.03
	0.001	47.18	37.12	33.20	31.08	29.75	28.83	28.17	27.65	26.42	25.13	23.82
6	0.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.90	2.82	2.72
	0.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.00	3.84	3.67
	0.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.37	5.12	4.86
	0.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.72	7.31	6.89
	0.001	35.51	27.00	23.71	21.92	20.80	20.03	19.46	19.03	17.99	16.90	15.77
7	0.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.67	2.58	2.47
	0.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.57	3.41	3.23
	0.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.67	4.41	4.15
	0.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.47	6.07	5.66
	0.001	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	13.71	12.73	11.72
8	0.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.50	2.40	2.30
	0.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.28	3.12	2.93
	0.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.20	3.95	3.68
	0.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.67	5.28	4.87
	0.001	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.19	10.30	9.36
9	0.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.38	2.28	2.16
	0.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.07	2.90	2.71
	0.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	3.87	3.61	3.34
	0.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.11	4.73	4.32
	0.001	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	9.57	8.72	7.84

Critical values computed with Excel 9.0

Degrees of freedom in denominator (df2)		Degrees of freedom in numerator (df1)										
	p	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**12**	**24**	**1000**
10	0.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.28	2.18	2.06
	0.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.91	2.74	2.54
	0.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.62	3.37	3.09
	0.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.71	4.33	3.92
	0.001	21.04	14.90	12.55	11.28	10.48	9.93	9.52	9.20	8.45	7.64	6.78
12	0.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.15	2.04	1.91
	0.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.69	2.51	2.30
	0.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.28	3.02	2.73
	0.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.16	3.78	3.37
	0.001	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.00	6.25	5.44
14	0.100	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.05	1.94	1.80
	0.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.53	2.35	2.14
	0.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.05	2.79	2.50
	0.010	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	3.80	3.43	3.02
	0.001	17.14	11.78	9.73	8.62	7.92	7.44	7.08	6.80	6.13	5.41	4.62
16	0.100	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	1.99	1.87	1.72
	0.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.42	2.24	2.02
	0.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	2.89	2.63	2.32
	0.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.55	3.18	2.76
	0.001	16.12	10.97	9.01	7.94	7.27	6.80	6.46	6.20	5.55	4.85	4.08
18	0.100	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	1.93	1.81	1.66
	0.050	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.34	2.15	1.92
	0.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.77	2.50	2.20
	0.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.37	3.00	2.58
	0.001	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.13	4.45	3.69
20	0.100	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.89	1.77	1.61
	0.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.28	2.08	1.85
	0.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.68	2.41	2.09
	0.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.23	2.86	2.43
	0.001	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	4.82	4.15	3.40
30	0.100	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.77	1.64	1.46
	0.050	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.09	1.89	1.63
	0.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.41	2.14	1.80
	0.010	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	2.84	2.47	2.02
	0.001	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.00	3.36	2.61
50	0.100	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.68	1.54	1.33
	0.050	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	1.95	1.74	1.45
	0.025	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.22	1.93	1.56
	0.010	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.56	2.18	1.70
	0.001	12.22	7.96	6.34	5.46	4.90	4.51	4.22	4.00	3.44	2.82	2.05
100	0.100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.61	1.46	1.22
	0.050	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.85	1.63	1.30
	0.025	5.18	3.83	3.25	2.92	2.70	2.54	2.42	2.32	2.08	1.78	1.36
	0.010	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.37	1.98	1.45
	0.001	11.50	7.41	5.86	5.02	4.48	4.11	3.83	3.61	3.07	2.46	1.64
1000	0.100	2.71	2.31	2.09	1.95	1.85	1.78	1.72	1.68	1.55	1.39	1.08
	0.050	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.76	1.53	1.11
	0.025	5.04	3.70	3.13	2.80	2.58	2.42	2.30	2.20	1.96	1.65	1.13
	0.010	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.20	1.81	1.16
	0.001	10.89	6.96	5.46	4.65	4.14	3.78	3.51	3.30	2.77	2.16	1.22

Use StaTable, WinPepi > WhatIs, or other reliable software to determine specific *p* values

F-test (Variance Ratio test)

* The following date is about the no of bulbs produced daily by two workers A and B.

A	B
40	39
30	38
38	41
41	33
38	32
35	39
	40
	34

$\alpha = 0.05$

Can we consider based on the date worker B is more Stable and effecient ?

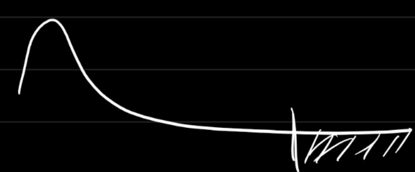
* Why not mean here can be used to test ?

⟶ Mean is same for both the samples.
So we will compare variances.

① $H_0 : S_1^2 = S_2^2$, $H_A = S_1^2 \neq S_2^2$

② F test , One tail test, $\alpha = 0.05$.

③ $F_{statistic} = \frac{S_1^2}{S_2^2}$



worker A.

X_1	$\bar{X}_1$	$(X_i - \bar{X}_1)^2$
40	37	9
30	37	49
38	37	1
41	37	16
38	37	1
35	37	4

$\bar{X}_1 = 37$ $\quad \Sigma (X_i - \bar{X}_1)^2 = 80$

$$S_1^2 = \frac{80}{n-1} = \frac{80}{6-1} = \frac{80}{5} = 16$$

worker B

X_2	$\bar{X}_2$	$(X_2 - \bar{X}_2)^2$
39	37	4
38	37	1
41	37	16
33	37	16
32	37	25
39	37	4
40	37	9
34	37	9

$\bar{X}_2 = 37$ $\quad \Sigma X_2 - \bar{X}_2 = 84$

$$S_2^2 = \frac{84}{8-1} = \frac{84}{7} = 12$$

$$F_{statistics} = \frac{16}{12} = 1.33$$

④ $F_{critical}$ $\alpha = 0.05$, $dof_1 = 5$ $\quad$ $dof_2 = 7$

dof_1 → in Numerator in f table

dof_2 → Denominator in f-table

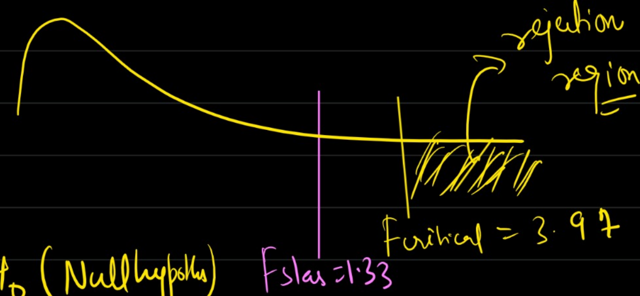
$$F_{critical\ \alpha = 0.05,\ dof(5,7)} = 3.97$$

⑤ Conclusion

Fstatistics is greater than 1.33,

$1.33 < 3.97$

We fail to reject the H_0 (Null hypothesis)



Worker B is not more stable | effective as compared to A.

Types of ANOVA (3 types)

① **One way ANOVA** : One factor with atleast two levels, and levels are independent.

eg. Medicine

3 factors → 0mg 10mg 100mg } dosage

eg. music

high medium → low.

eg. Stress level of three departments

Dep A	Dep B	Dep C
8	1	6
5	2	5
6	3	7
2	4	8

② **Repeated measures ANOVA** – One factor with atleast two levels, but levels are dependents.

eg 1 No of hours studied

Day 1	Day 2	Day 3	Day 4
10	5	4	10

eg. Gym

Day 1	Day 2	Day 3	Day 4
5	1 hr	0.5	1

③ Factorial ANOVA :- Two or more factors (each of which with atleast 2 levels). Levels can be either dependent, independent or both.

→ Two way Anova → It deals with only two factors.

medicine (factors)

		0 mg	10 mg	100 mg → 3 levels
Gender (factor)	Male	2	3	3
		3	4	2
		4	1	1
		5	5	6
	Female	1	2	7
		2	3	8
		1	1	6
		3	4	5

Running
Brand of shoes ← factor

	Brand A	Brand B	Brand C
Monday	10	1	—
Wednesday	15	2	—
Saturday	10	5	—

ANOVA (Analysis of Variance) and its Assumption.

Defn : ANOVA is a statistical method used to compare the means of 2 or more groups

Anova Terms

* Factors / variables / criteria
* Levels.

ex-1

Medicine ⟹ factors/variable/criteria

0mg 10mg 100mg. (Dosage)

↓

Levels

ex-2

Music ⟹ factors/variable

High Sound medium Sound low sound

↓

Levels.

ex-3

Payment mode ⟶ factors

NEFT IMPS PhonePay Paytm.

Levels

* Assumptions of ANOVA

→ The population from which samples are drawn should be Normally distributed.

→ The sample should be Independent of each other/random.

→ Absence of outlier.

→ Homogeneity of variance :- Homogeneity means that the variance among the groups should be approximately equal.

$S_1 \quad S_2 \quad S_3$

	Pop	Sample
St.dev	σ	s
Var.	σ^2	s^2

$$S_1^2 = S_2^2 = S_3^2$$

ANOVA test

Hypothesis testing in ANOVA (Partitioning of variance in ANOVA)

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

H_A : Atleast one of the sample mean is not equal

$\mu_1 = \mu_2 = \mu_3 \cdots\cdots \neq \mu_K$

* Test statistics = $\dfrac{\text{Variance between Samples}}{\text{Variance within Samples.}}$

$H_0 : \bar{x}_1 = \bar{x}_2 = \bar{x}_3$

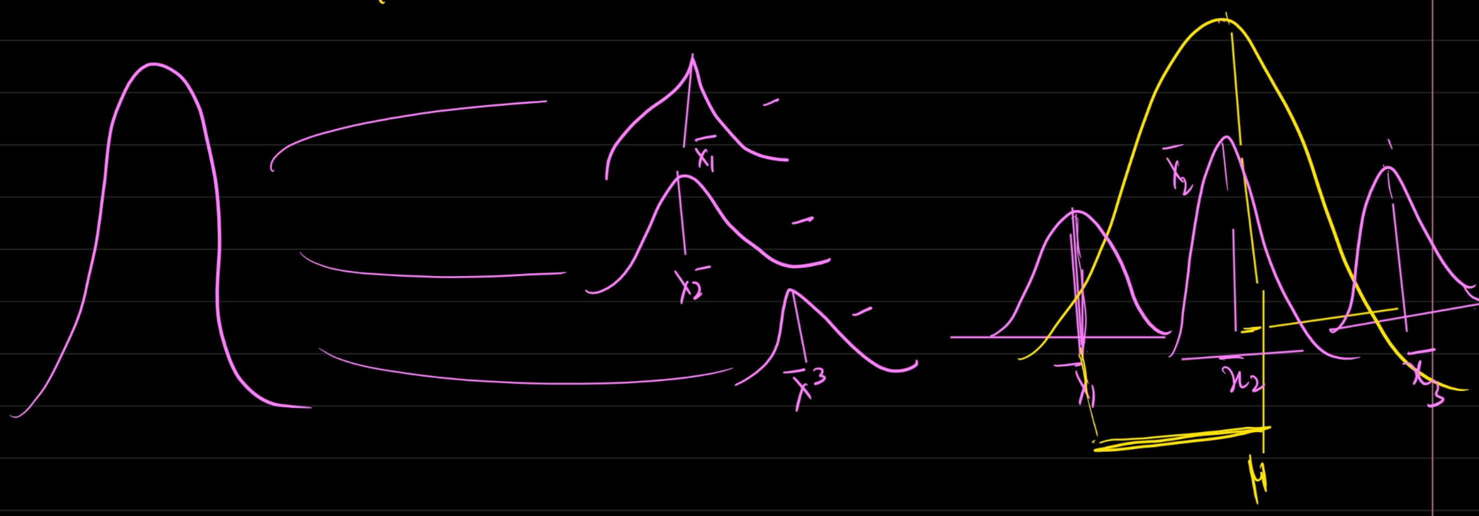
H_A : Atleast one sample mean is not equal.

s^2

variance between sample

variance within sample

x_1	x_2	x_3
1	5	9
2	6	10
3	7	11
4	8	12



One way ANOVA

→ one factor with atleast levels

Q There are three dosage of a medicine given to three sample of the patients. rating if headache is reduced (1-10) Are there differences in the three condition. $\alpha = 0.05$?

0 mg	10 mg	100 mg
9	7	4
8	6	3
7	6	2
8	7	3
8	8	4
9	7	3
8	6	2

① $H_0 : \mu_{0mg} = \mu_{10mg} = \mu_{100mg}$
H_A : not all are equal.

② $\alpha = 0.05$, one tail test

③ Calculate F statistics.

$$F_{statistics} = \frac{\text{Variance b/w sample}}{\text{Variance within sample}}$$

	Sum of squares (SS)	df	MS (SS/df)	F
Between the sample				
Within sample				
total				

0 mg	10 mg	100 mg
9	7	4
8	6	3
7	6	2
8	7	3
8	8	4
9	7	3
8	6	2

* Sum of squares b/w sample $= \sum \frac{(\sum a_i)^2}{n} - \frac{T^2}{N}$

0mg : $9+8+7+8+8+9+8 = 57$
10mg : $7+6+6+7+8+7+6 = 47$
100mg : $4+3+2+3+4+3+2 = 21$

$$= \frac{57^2+47^2+21}{7} - \frac{(\underbrace{57+47+21}_{125})^2}{21} = 98.67$$

* Sum of square within the group.

0 mg	10 mg	100 mg
9	7	4
8	6	3
7	6	2
8	7	3
8	8	4
9	7	3
8	6	2

$$SS_{within} = \sum y^2 - \sum \frac{(\sum a_i)^2}{n}$$

$\sum y^2 = 9^2+8^2+7^2+8^2 \dots +7^2+6^2+6^2+ \dots 4^2+3^2 - +2^2$
$= 853$

$$= 853 - \left(\frac{57^2+47^2+21^2}{7}\right) = 10.29.$$

	Sum of Squares (SS)	df	MS (SS/df)	F
Between the Sample	98.67	2	49.34	
within sample	10.29	18	0.54	
total	108.96	20		

dof

no. of samples − 1

$3 - 1 = 2$

degree of freedom within Sample

$df_{within} = N - a$ (no. of groups)

$= 21 - 3 = 18$

$\Rightarrow N - 1$

$\Rightarrow 21 - 1$

$\Rightarrow 20$

$$F_{statistic} = \frac{MS_{between}}{MS_{within}} \left(\frac{\text{var b/w sample}}{\text{var within sample}}\right)$$

$$F = \frac{49.34}{0.54} = 86.56$$

Step-4

$F_{critical}$ for $\alpha = 0.05$

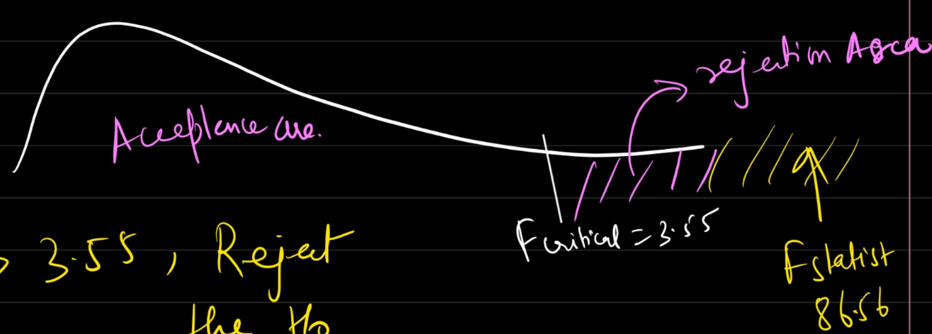
$df_{between} = 3 - 1 = 2 \leftarrow df_1$

$df_{within} = 21 - 3 = 18 \leftarrow df_2$

F-table

$F_{critical\ \alpha=0.05(2,18)} = 3.55$

Step-5



$86.56 > 3.55$, Reject the H_0