F Distribution

Fisher–Snedecor

F -Distribution

χ^2 was useful in testing hypotheses about a single population variance.

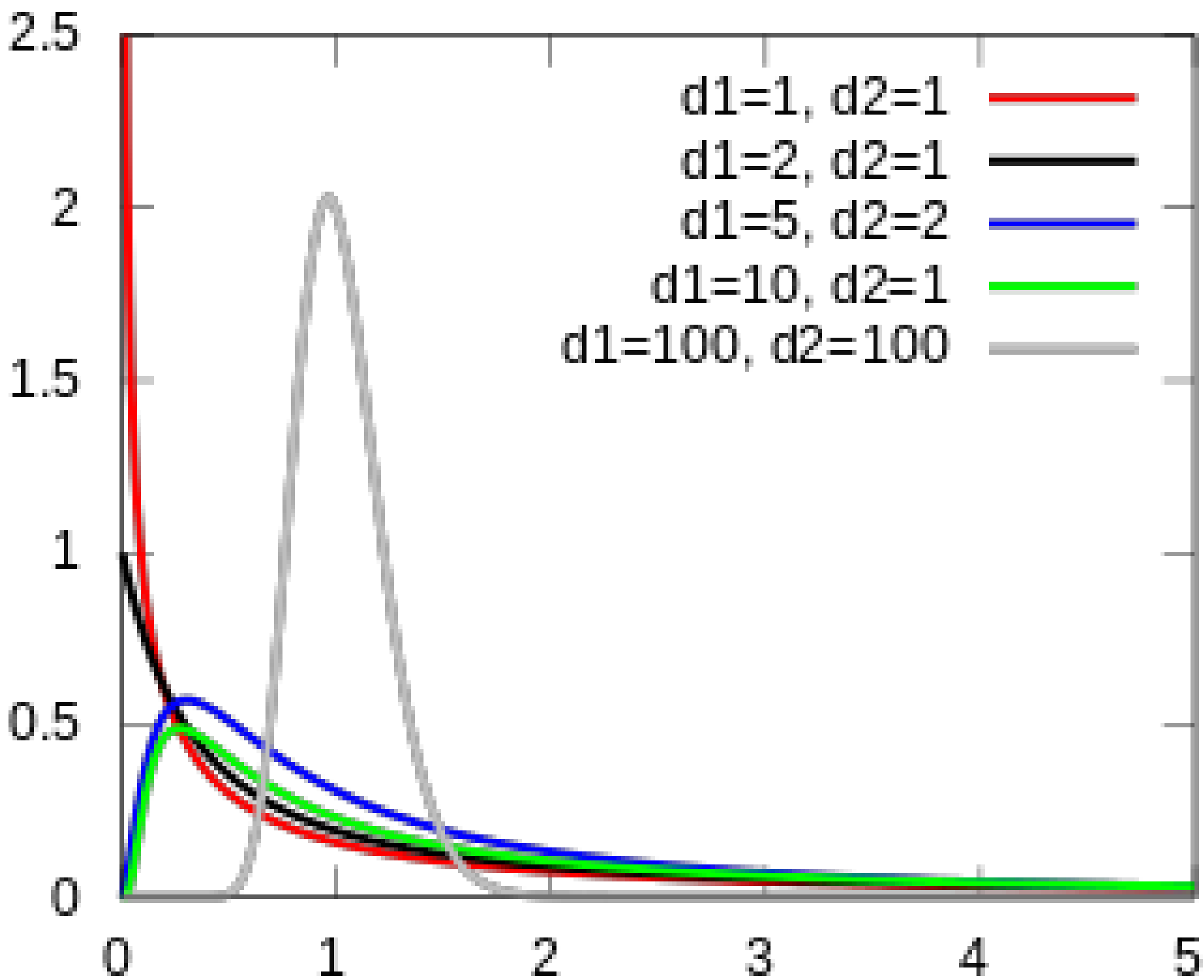
- Sometimes we want to test hypotheses about difference in variances of two populations:
  - Is the variance of 2 stocks the same?
  - Do parts manufactured in 2 shifts or on 2 different machines or in 2 batches have the same variance or not?
  - Is the powder mix for tablet granulations homogeneous?
  - Is there variability in assayed drug blood levels in a bioavailability study?
  - Is there variability in the clinical response to drug therapy of two samples?

F Distribution

- Ratio of 2 variance estimates: $F = \frac{s_1^2}{s_1^2} = \frac{est.\sigma_1^2}{est.\sigma_2^2}$
- Ideally, this ratio should be about 1 if 2 sample come from the same population or from 2 population with same variance, but sampling errors cause variation.
- $Recall\ \chi^2 = \frac{(n-1)s^2}{\sigma^2}$. So, F is also a ratio of 2 chi-square, each divided by its degree of freedom, i.e.,

$$F = \frac{\frac{\chi_{v1}^2}{v_1}}{\frac{\chi_{v2}^2}{v_2}}$$

F distribution



Hypothesis test for 2 sample variance

A machine produces metal sheets with 22mm thickness. There is variability in thickness due to machines, operators, manufacturing environment, raw material, etc. The company wants to know the consistency of two machines and randomly samples 10 sheets from machine 1 and 12 sheets from machine 2. Thickness measurements are taken. Assume sheet thickness is normally distributed in the population.

- The company wants to know if the variance from each sample comes from the same population variance (population variances are equal) or from different population variances (population variances are unequal).
- How do you test this?

Hypothesis test for 2 sample variances

Data

Machine 1		Machine 2	
22.3	21.9	22.0	21.7
21.8	22.4	22.1	21.9
22.3	22.5	21.8	22.0
21.6	22.2	21.9	22.1
21.8	21.6	22.2	21.9
		22.0	22.1
$s_1^2 = 0.11378$	$n = 10$	$s_2^2 = 0.02023$	$n = 12$

Ratio of sample variances, $F = \frac{s_1^2}{s_2^2} = \frac{0.11378}{0.02023} = 5.62$

Hypothesis test for 2 sample variances

What are null and alternate hypotheses ?

$H_0 : \sigma_1^2 = \sigma_2^2;$

$H_1 : \sigma_1^2 \neq \sigma_2^2;$

Is it a one tailed test or a two tailed test?

Two-tailed

What are numerator and denominator degrees of freedom?

$\nu_1 = 10 - 1 = 9; \nu_2 = 12 - 1 = 11$

Hypothesis test for 2 sample variances

Reading an F-table

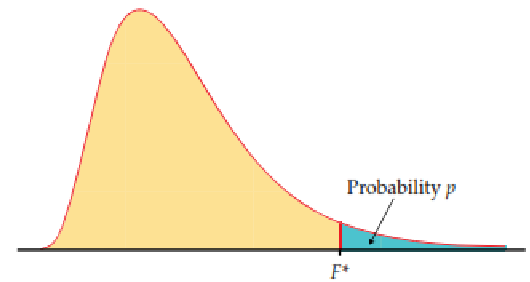


Table entry for p is the critical value F^* with probability p lying to its right.

TABLE E

F critical values (continued)

of freedom in the denominator	p	1	2	3	4	5	6	7	8	9
		Degrees of freedom in the numerator								
8	.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
	.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
	.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
	.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
	.001	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.77
9	.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
	.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
	.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
	.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
	.001	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11
10	.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
	.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
	.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
	.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
	.001	21.04	14.91	12.55	11.28	10.48	9.93	9.52	9.20	8.96
11	.100	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
	.050	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
	.025	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
	.010	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
	.001	19.69	13.81	11.56	10.35	9.58	9.05	8.66	8.35	8.12
12	.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
	.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
	.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
	.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
	.001	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48
13	.100	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
	.050	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
	.025	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31

$$F_{0.025,9,11} = 3.5879;$$

$$F_{0.975,9,11} = \frac{1}{F_{0.025,9,11}}$$

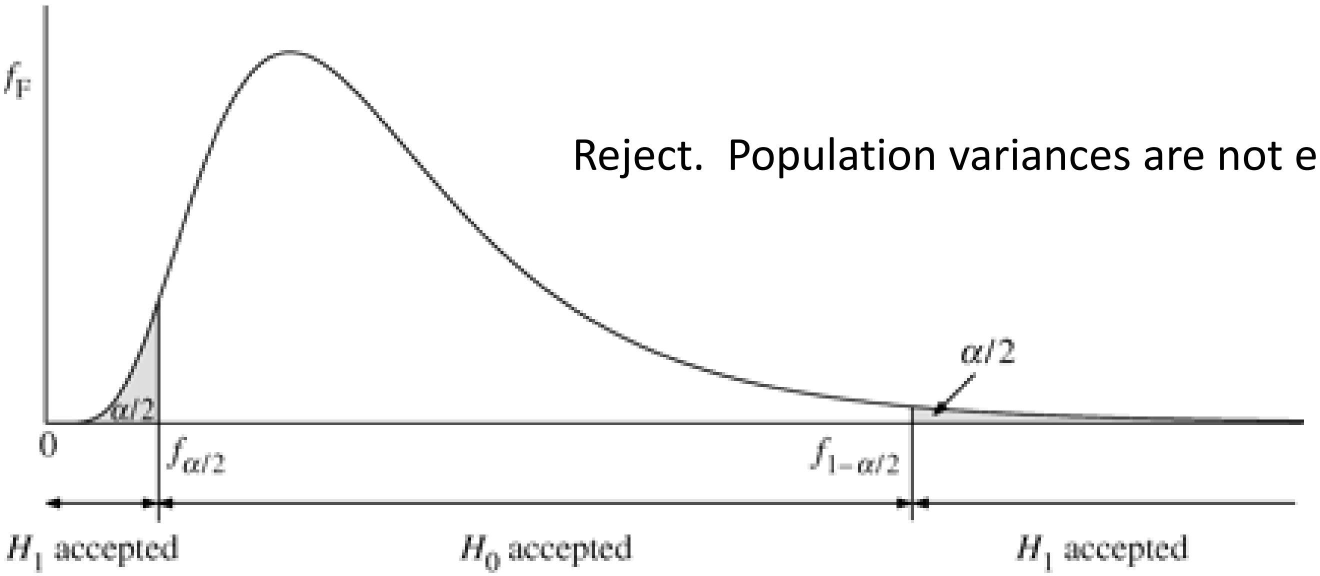
$$= 0.2787$$

Hypothesis test for 2 sample variances

$$F_{0.025,9,11} = 3.5879; F_{0.975,9,11} = \frac{1}{F_{0.025,9,11}} = 0.2787;$$

$$F_{observed} = 5.62$$

Will you reject the null hypothesis or not?



Reject. Population variances are not equal.

Hypothesis test for 2-sample variances

What are the business implications?

Variance in machine 1 is higher than in machine 2.
Machine 1 needs to be inspected for any issues.

Application of F Distribution

- Test for equality of variances.
- Test for differences of means in ANOVA.
- Test for regression models (slopes relating one continuous variable to another, e.g., Entrance exam scores and GPA)

Relations among Distributions – Children of the Normal

- χ^2 is drawn from the normal – N(0,1) deviates squared and summed.
- *F* is the ratio of 2 chi-squares, each divided by its df.
- A χ^2 divided by its *df* is a variance estimate, i.e., a sum of squares divided by the degrees of freedom.
- $F = t^2$. If you square *t*, you get an *F* with 1 *df* in the numerator, i.e., $t_\nu^2 = F_{(1,\nu)}$

(see http://www-ist.massey.ac.nz/dstirlin/CAST/CAST/HsimpleAnova/simpleAnova3.html for an example using this relationship)

Reference