Mathematics 231

Lecture 27

Liam O'Brien

1

Announcements

- Reading
  - Today M&M 7.1 428-435
    M&M 7.3 474-477
  - Next class M&M 8.2 505-515
    M&M 12.0 637

2

Topics

- Hypothesis testing for comparing two variances
- Hypothesis testing for matched pairs

3

Testing Variance Equality

- When we have a 2-sample t-test with independent samples we need to decide if we have equal variances.
- What we can do?
- Rule-of-thumb: If the difference is less than 10% between s_1 and s_2, then assume equality.
- Formal way: Hypothesis test

4

Test of Equality of Variances

$H_0 : \sigma_1 = \sigma_2 \qquad H_A : \sigma_1 \neq \sigma_2$

We have an estimate of each of these from our two samples and will construct a test statistic from these.

$F = \frac{s_1^2}{s_2^2}$ where $s_1 > s_2$

Under H_0 this has an F-distribution with $\mathrm{n}_1 - 1$ numerator, and $\mathrm{n}_2 - 1$ denominator degrees of freedom.

5

The F Statistic

The F statistic has two different degrees of freedom terms: $\mathrm{n}_1 - 1$ in the numerator, and $\mathrm{n}_2 - 1$ in the denomiator.

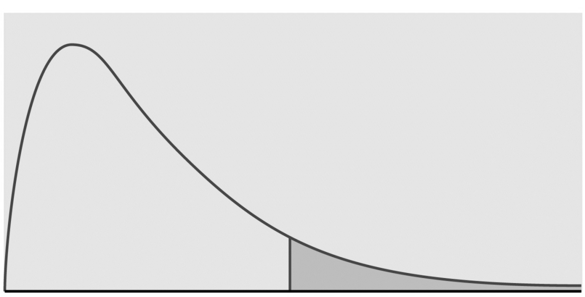
Note: df corresond to numerator and denominator of F.

F distribution cannot assume negative values and is skewed to the right.

Its shape depends on the degrees of freedom.

6

F Distribution



7

Test of Equality of Variances

$H_0 : \sigma_1 = \sigma_2 \qquad H_A : \sigma_1 \neq \sigma_2$

Reject H_0 if $2P(F > f) < \alpha$. Generally we set α=0.10. Note that the F-table only has particular values for df tabulated like the t-distribution.

Can only get a range for the p-value from Table E.

Stata gives exact probability.

8

TABLE E *F* critical values

Degrees of freedom in the denominator	*p*	1	2	3	4	5	6	7	8	9
		Degrees of freedom in the numerator								
	.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
	.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
1	.025	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28
	.010	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5
	.001	405284	500000	540379	562500	576405	585937	592873	598144	602284
	.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
2	.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
	.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
	.001	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37	999.39
	.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
3	.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
	.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
	.001	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62	129.86
	.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
4	.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
	.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
	.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47
	.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
	.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
5	.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
	.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
	.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	27.24
	.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
	.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
6	.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
	.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
	.001	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03	18.69
	.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
	.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
7	.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
	.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
	.001	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	14.33

Finding F Probabilities in Stata

- If you have *numdf* numerator df, and *dendf* denominator df, type:

  display Ftail(numdf, dendf, f)

  to get upper tail probability (multiply by 2 for variance test p-value).

Test of Equality of Variances: Example

$H_0: \sigma_1 = \sigma_2 \qquad H_A: \sigma_1 \neq \sigma_2$

From healthy/failed firms example:

$s_1 = 0.639; \quad s_2 = 0.481$

$$F = \frac{s_1^2}{s_2^2} = \frac{0.639^2}{0.481^2} = 1.76$$

. display Ftail(67,32,1.76)

.04044562

p-value=2(0.04)=0.08 < 0.10

Reject H_0, and variances are significantly different at the 10% level.

Paired Samples

- Defining characteristic of paired data is that for each observation in the first group, there is a corresponding observation in the second.
- Example 1: Study of a single population with observations both before and after an intervention.
- Example 2: Study of two populations of subjects who are matched on important characteristics.
- Pairing (a form of blocking) helps control known sources of variation.

Paired Samples

Given an SRS of size n, we want to test $H_0 : \delta = 0$ against $H_A : \delta \neq 0$ (two-sided).

Use test statistic: $t = \dfrac{\bar{d} - \delta}{s_d / \sqrt{n}} = \dfrac{\bar{d}}{s_d / \sqrt{n}}$

When $H_0 : \delta = 0$ is true, this statistic has a t-distribution with n-1 degrees of freedom.

This is called the paired t-test.

Example: MLA Listening Scores for French Teachers

- 20 French teachers attend summer institute for 4 weeks to improve language skills.
- At beginning, they were given MLA listening test of understanding spoken French.
- After 4 weeks, MLA test given again.
- Has attendance at the summer institute improved French teachers' language skills?

TABLE 7.1 Modern Language Association listening scores for French teachers

Teacher	Pretest	Posttest	Gain	Teacher	Pretest	Posttest	Gain
1	32	34	2	11	30	36	6
2	31	31	0	12	20	26	6
3	29	35	6	13	24	27	3
4	10	16	6	14	24	24	0
5	30	33	3	15	31	32	1
6	33	36	3	16	30	31	1
7	22	24	2	17	15	15	0
8	25	28	3	18	32	34	2
9	32	26	−6	19	23	26	3
10	20	26	6	20	23	26	3

Example: MLA Test

We have the following summary statistics for the differences:

$n = 20$

$\bar{d} = 2.5$

$s_d = 0.647$

To do this in Stata, use the one-sample mean calculator option. Paired data result in a one-sample test.

Example: MLA Test

The test statistic is

$$t = \frac{2.5}{0.647/\sqrt{20}} = 17.3 \sim t_{19}$$

$p < 0.001$

Reject H_0 at the α=0.05 level and conclude that scores significantly increased.

17

Example: MLA Test

```
. ttesti 20 2.5 0.647 0

One-sample t test

------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      20         2.5    .1446736        .647    2.197195    2.802805
------------------------------------------------------------------------------
Degrees of freedom: 19

                               Ho: mean(x) = 0

     Ha: mean < 0              Ha: mean != 0              Ha: mean > 0
       t =  17.2803              t =  17.2803              t =  17.2803
   P < t =   1.0000          P > |t| =   0.0000          P > t =   0.0000
```

18