Lecture 8

8.1. Inferences about Population Variances

Summary

In this lecture we presented statistical procedures that can be used to make inferences aboul population variances. In the process we introduced two new probability distributions: the chi-square distribution and the *F* distribution. The chi-square distribution can be used as the basis for interval estimation and hypothesis tests about the variance of a normal population

We illustrated the use of the *F* distribution in hypothesis tests about the variances of the normal populations. In particular, we showed that with independent simple random samples of sizes n_x and n_2 selected from two normal populations with equal variances $\sigma = \sigma$, the sampling distribution of the ratio of the two sample variances *s2/s2* has an *F* distribution with n_1 – 1 degrees of freedom for the numerator and n_2 — 1 degrees of freedom for the denominator.

Inferences about the difference between two population means were then discussed for the matched sample design. In the matched sample design each element provides a pair of data values, one from each population. The difference between the paired data values is then used in the statistical analysis. The matched sample design is generally preferred to the independent sample design because the matched-sample procedure often improves the precision of the estimate.

Finally, interval estimation and hypothesis testing about the difference between two population proportions were discussed. Statistical procedures for analyzing the difference between two population proportions are similar to the procedures for analyzing the difference between two population means.

Key formulas

Interval Estimate of a Population Variance

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}} \tag{8.1}$$

Test Statistic for Hypothesis Tests about a Population Variance

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \tag{8.2}$$

Test Statistic for Hypothesis Tests about Population Variances with $\sigma_1^2 = \sigma_2^2$

$$F = \frac{s_1^2}{s_2^2} \tag{8.3}$$

Lecture supplementary material

FIGURE 11 1 EXAMPLES OF THE SAMPLING DISTRIBUTION OF $(n-1)s^2/\sigma^2$ (A CHI-SQUARE DISTRIBUTION)

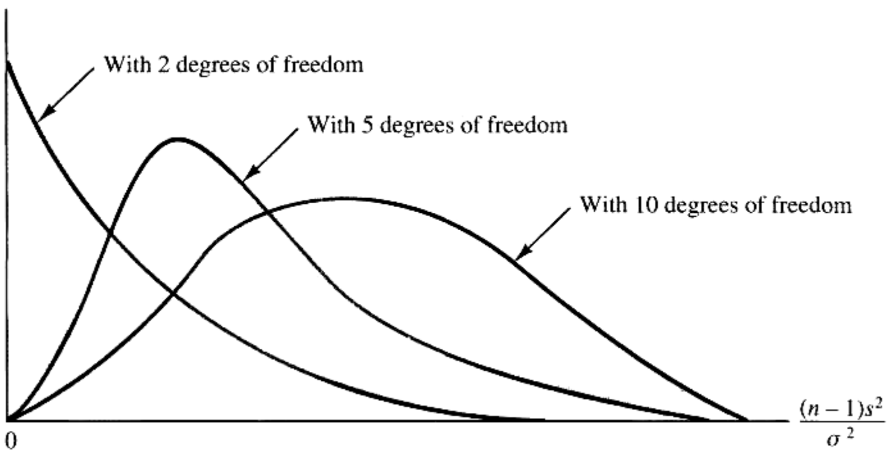
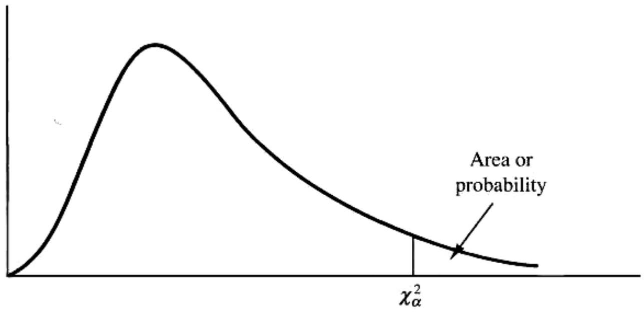


TABLE 11.1 SELECTED VALUES FROM THE CHI-SQUARE DISTRIBUTION TABLE*



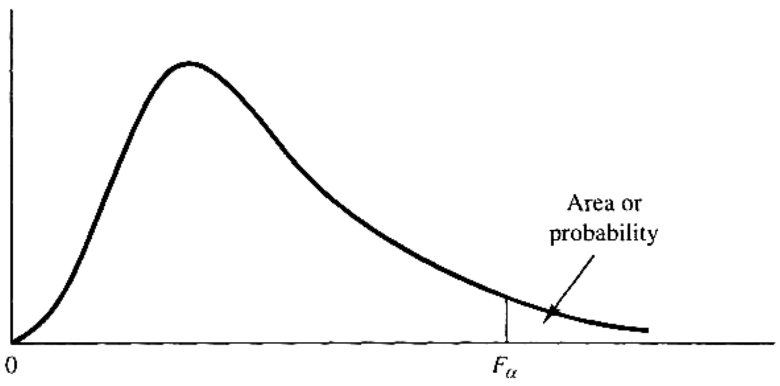
Degrees of Freedom	Area in Upper Tail .99	.975	.95	.90	.10	.05	.025	.01
1	.000	.001	.004	.016	2.706	3.841	5.024	6.635
2	.020	.051	.103	.211	4.605	5.991	7.378	9.210
3	.115	.216	.352	.584	6.251	7.815	9.348	11.345
4	.297	.484	.711	1.064	7.779	9.488	11.143	13.277
5	.554	.831	1.145	1.610	9.236	11.070	12.832	15.086
6	.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.041	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.878	14.573	16.151	18.114	36.741	40.113	43.195	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892
40	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691
60	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379
80	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329
100	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807

*Note: A more extensive table is provided as Table 3 of Appendix B.

Table . Summary of hypothesis test about a population variance

	Lower Tail Test	**Upper Tail Test**	**Two-Tailed Test**
Hypotheses	$H_0: \sigma^2 \geq \sigma_0^2$ $H_a: \sigma^2 < \sigma_0^2$	$H_0: \sigma^2 \leq \sigma_0^2$ $H_a: \sigma^2 > \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_a: \sigma^2 \neq \sigma_0^2$
Test Statistic	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$
Rejection Rule: p-Value Approach	Reject H_0 if p-value$\leq\alpha$	Reject H_0 if p-value$\leq\alpha$	Reject H_0 if p-value$\leq\alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $\chi^2 \leq \chi^2_{(1-\alpha)}$	Reject H_0 if $\chi^2 \geq \chi^2_{\alpha}$	Reject H_0 if $\chi^2 \leq \chi^2_{(1-\alpha/2)}$ or if $\chi^2 \geq \chi^2_{\alpha/2}$

TABLE 11.3 SELECTED VALUES FROM THE F DISTRIBUTION TABLE*



Denominator Degrees of Freedom	Area in Upper Tail	Numerator Degrees of Freedom 10	15	20	25	30
10	.10	2.32	2.24	2.20	2.17	2.16
	.05	2.98	2.85	2.77	2.73	2.70
	.025	3.72	3.52	3.42	3.35	3.31
	.01	4.85	4.56	4.41	4.31	4.25
15	.10	2.06	1.97	1.92	1.89	1.87
	.05	2.54	2.40	2.33	2.28	2.25
	.025	3.06	2.86	2.76	2.69	2.64
	.01	3.80	3.52	3.37	3.28	3.21
20	.10	1.94	1.84	1.79	1.76	1.74
	.05	2.35	2.20	2.12	2.07	2.04
	.025	2.77	2.57	2.46	2.40	2.35
	.01	3.37	3.09	2.94	2.84	2.78
25	.10	1.87	1.77	1.72	1.68	1.66
	.05	2.24	2.09	2.01	1.96	1.92
	.025	2.61	2.41	2.30	2.23	2.18
	.01	3.13	2.85	2.70	2.60	2.54
30	.10	1.82	1.72	1.67	1.63	1.61
	.05	2.16	2.01	1.93	1.88	1.84
	.025	2.51	2.31	2.20	2.12	2.07
	.01	2.98	2.70	2.55	2.45	2.39

Note: A more extensive table is provided as Table 4 of Appendix B.

Table . Summary of hypothesis tests about two population variances

	Upper Tail Test	**Two-Tailed Test**
Hypotheses	$H_0: \sigma_1^2 \le \sigma_2^2$ $H_a: \sigma_1^2 > \sigma_2^2$	$H_0: \sigma_1^2 = \sigma_2^2$ $H_a: \sigma_1^2 \ne \sigma_2^2$ Note: Population I has the lager sample variance
Test Statistic	$F = \frac{s_1^2}{s_2^2}$	$F = \frac{s_1^2}{s_2^2}$
Rejection Rule: p-Value Approach	Reject H_0 if p-value≤α	Reject H_0 if p-value≤α
Rejection Rule: Critical Value Approach	Reject H_0 if $F \ge F_\alpha$	Reject H_0 if $F \ge F_\alpha$

8.2. Tests of Goodness of Fit and Independence

Summary

In this lecture we introduced the goodness of fit test and the test of independence, both of which are based on the use of the chi-square distribution. The purpose of the goodness of fit test is to determine whether a hypothesized probability distribution can be used as a model for a particular population of interest. The computations for conducting the goodness of fit test involve comparing observed frequencies from a sample with expected frequencies when the hypothesized probability distribution is assumed true. A chi-square distribution is used to determine whether the differences between observed and expected frequencies are large enough to reject the hypothesized probability distribution. We illustrated the goodness of fit test for multinomial, Poisson, and normal distributions.

A test of independence for two variables is an extension of the methodology employed in the goodness of fit test for a multinomial population. A contingency table is used to determine the observed and expected frequencies. Then a chi-square value is computed. Large chi-square values, caused by large differences between observed and expected frequencies, lead to the rejection of the null hypothesis of independence.

Glossary

Multinomial population A population in which each element is assigned to one and only one of several categories. The multinomial distribution extends the binomial distribution from two to three or more outcomes.

Goodness of fit test A statistical test conducted to determine whether to reject a hypothesized probability distribution for a population.

Contingency table A table used to summarize observed and expected frequencies for a test of independence.

Key formulas

Test Statistic for Goodness of Fit

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} \tag{8.4}$$

Expected Frequencies for Contingency Tables Under the Assumption of Independence

$$e_{ij} = \frac{(Row\ i\ Total)(Column\ \ j\ Total)}{Sample\ Size} \tag{8.5}$$

Test Statistic for Independence

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \tag{8.6}$$

Lecture supplementary material

MULTINOMIAL DISTRIBUTION GOODNESS OF FIT TEST: A SUMMARY

1. State the null and alternative hypotheses.

   H_o: The population follows a multinomial distribution with specified probabilities for each of the k categories

   H_a: The population does not follow a multinomial distribution with the specified probabilities for each of the k categories

2. Select a random sample and record the observed frequencies f_i for each category.
3. Assume the null hypothesis is true and determine the expected frequency e_i in each category by multiplying the category probability by the sample size.
4. Compute the value of the test statistic.

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

5. Rejection rule:

   p-value approach: Reject H_o if p-value < α

   Critical value approach: Reject H_o if $\chi^2 > \chi^2_\alpha$

where α is the level of significance for the test and there are $k - 1$ degrees of freedom.

TEST OF INDEPENDENCE: A SUMMARY

1. State the null and alternative hypotheses

   H_o: The column variable in independent of the row variable
   H_a: The column variable in not independent of the row variable
2. Select a random sample and record the observed frequencies for each cell of the contingency table.
3. Use equation $e_{ij} = \dfrac{(Row\ i\ Total)(Column\ \ j\ Total)}{Sample\ Size}$ (8.5) to compute the expected frequency for each cell.
4. Use equation $\chi^2 = \sum_i \sum_j \dfrac{(f_{ij} - e_{ij})^2}{e_{ij}}$ (8.6) to compute the value of the test statistic.
5. Rejection rule:

   p-value approach: Reject H_0 if p-value$\leq\alpha$

   Critical Value Approach: Reject H_0 if $\chi^2 \geq \chi^2_\alpha$

   where α is the level of significance, with n rows and m columns providing (n-1)(m-1) degrees of freedom.

POISSON DISTRIBUTION GOODNESS OF FIT TEST: A SUMMARY

1. State the null and alternative hypotheses.
   H_o: The population has a Poisson distribution
   H_a: The population doesn't have Poisson distr.
2. Select a random sample and
   a. Record the observed frequency f_t for each value of the Poisson random variable.
   b. Compute the mean number of occurrences μ.
3. Compute the expected frequency of occurrences e_i for each value of the Poisson random variable. Multiply the sample size by the Poisson probability of occurrence for each value of the Poisson random variable. If there are fewer than five expected occurrences for some values, combine adjacent values and reduce the number of categories as necessary.
4. Compute the value of the test statistic.

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

5. Rejection rule:
   p-value approach: Reject H_0 if p-value < a
   Critical value approach: Reject H_0 if $\chi^2 > \chi^2_\alpha$

where α is the level of significance and there are $k - 2$ degrees of freedom.

NORMAL DISTRIBUTION GOODNESS OF FIT TEST: A SUMMARY

1. State the null and alternative hypotheses.
   H_o: The population has a normal distribution
   H_a: The population does not have a normal distribution
2. Select a random sample and
   a. Compute the sample mean and sample standard deviation.
   b. Define intervals of values so that the expected frequency is at least five for each interval. Using equal probability intervals is a good approach.
   c. Record the observed frequency of data values f_i in each interval defined.
3. Compute the expected number of occurrences e_i for each interval of values defined in step 2(b). Multiply the sample size by the probability of a normal random variable being in the interval.
4. Compute the value of the test statistic.

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

5. Rejection rule:
   p-value approach: Reject H_o if p-value < a
   Critical value approach: Reject H_0 if $\chi^2 > \chi^2_\alpha$

where α is the level of significance and there are $k - 3$ degrees of freedom.