Statistical Analysis

Population Variance

JOHNS HOPKINS
CAREY BUSINESS SCHOOL

Today's Lesson

1. The Chi-Square Distribution
2. Statistical Inference for a Population Variance
3. The F Distribution
4. Comparing Two Population Variances by Using Independent Samples

When to use Chi-Square

Chi-Square can be used to test for variance.

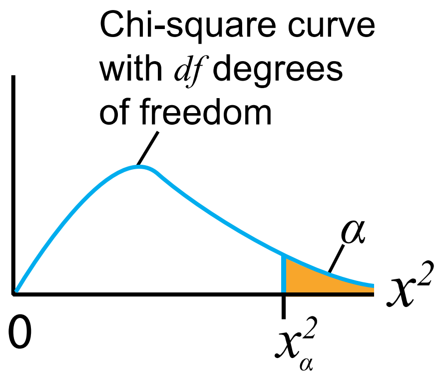
Often in a manufacturing process, the variance can be more important than the sample mean.

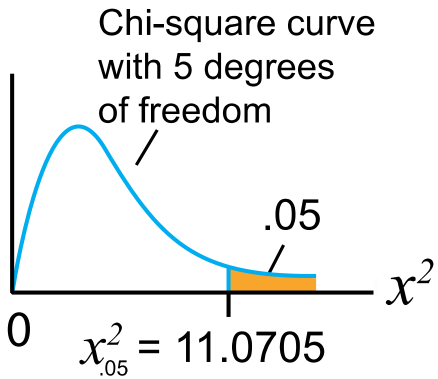
We are assuming a normal distribution in this case. Those as we will see shortly, Chi-Square is used for data which can not be assumed to be a normal distribution.

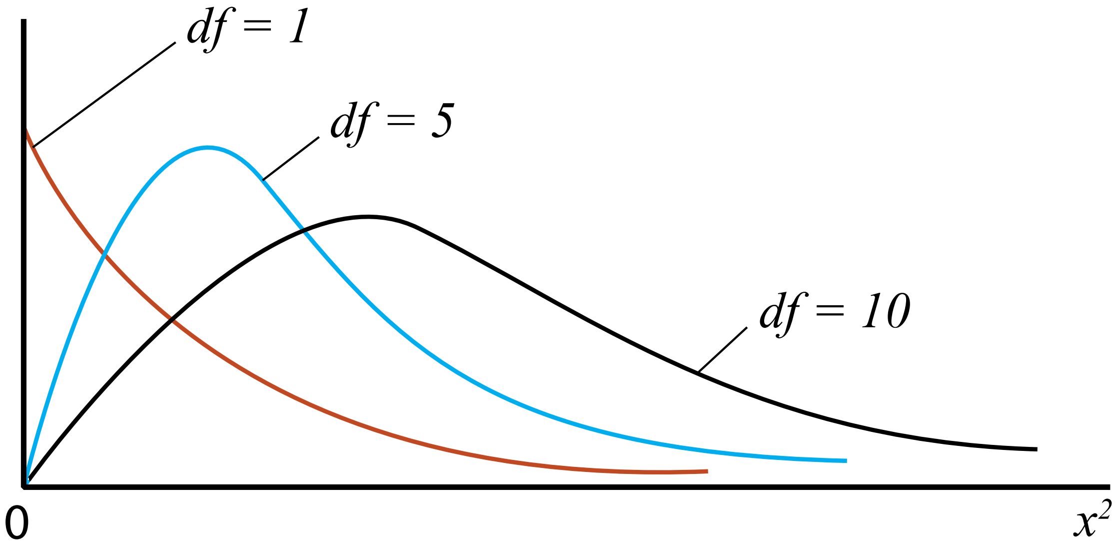
The Chi-Square Distribution

- Sometimes make inferences using the chi-square distribution
  - o Denoted χ^2
- Skewed to the right
- Exact shape depends on the degrees of freedom
  - o Denoted df
- A **chi-square point** χ^2_α is the point under a chi-square distribution that gives right-hand tail area α

The Chi-Square Distribution continued







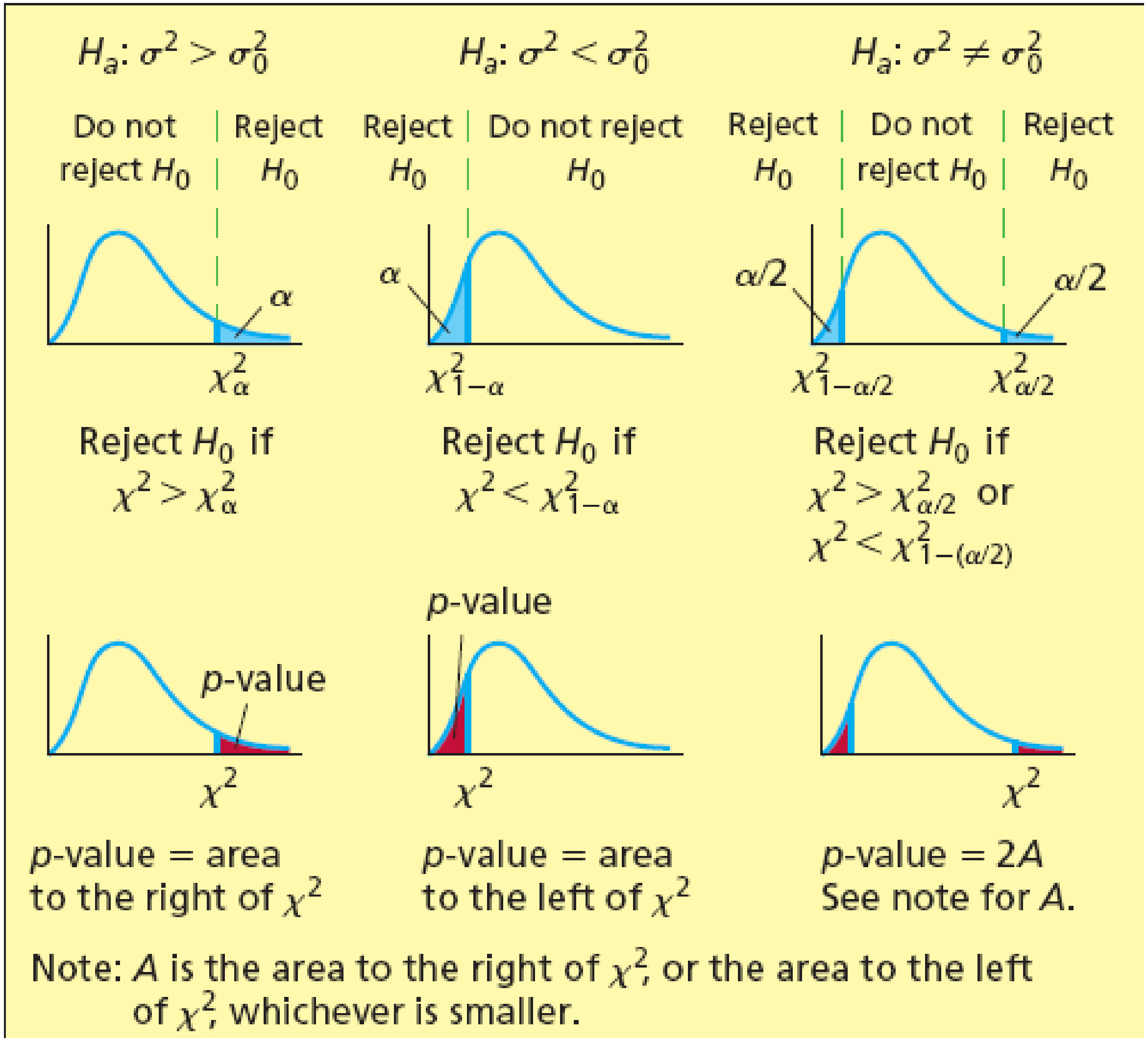
A Portion of the Chi-Square Table

Degrees of Freedom *(df)*	$X^2_{.10}$	$X^2_{.05}$	$X^2_{.025}$	$X^2_{.01}$	$X^2_{.005}$
1	2.70554	3.84146	5.02389	6.63490	7.78944
2	4.60517	5.99147	7.37776	9.21034	10.5966
3	6.25139	7.81473	9.34840	11.3449	12.8381
4	7.77944	9.48773	11.1433	13.2767	14.8602
5	9.23635	11.0705	12.8325	15.0863	16.7496
6	10.6446	12.5916	14.4494	16.8119	18.5476

Statistical Inference for Population Variance

- If s^2 is the variance of a random sample of n measurements from a normal population with variance σ^2
- The sampling distribution of the statistic $(n - 1)\, s^2 / \sigma^2$ is a chi-square distribution with $(n - 1)$ degrees of freedom
- Can calculate confidence interval and perform hypothesis testing

Hypothesis Testing for Population Variance



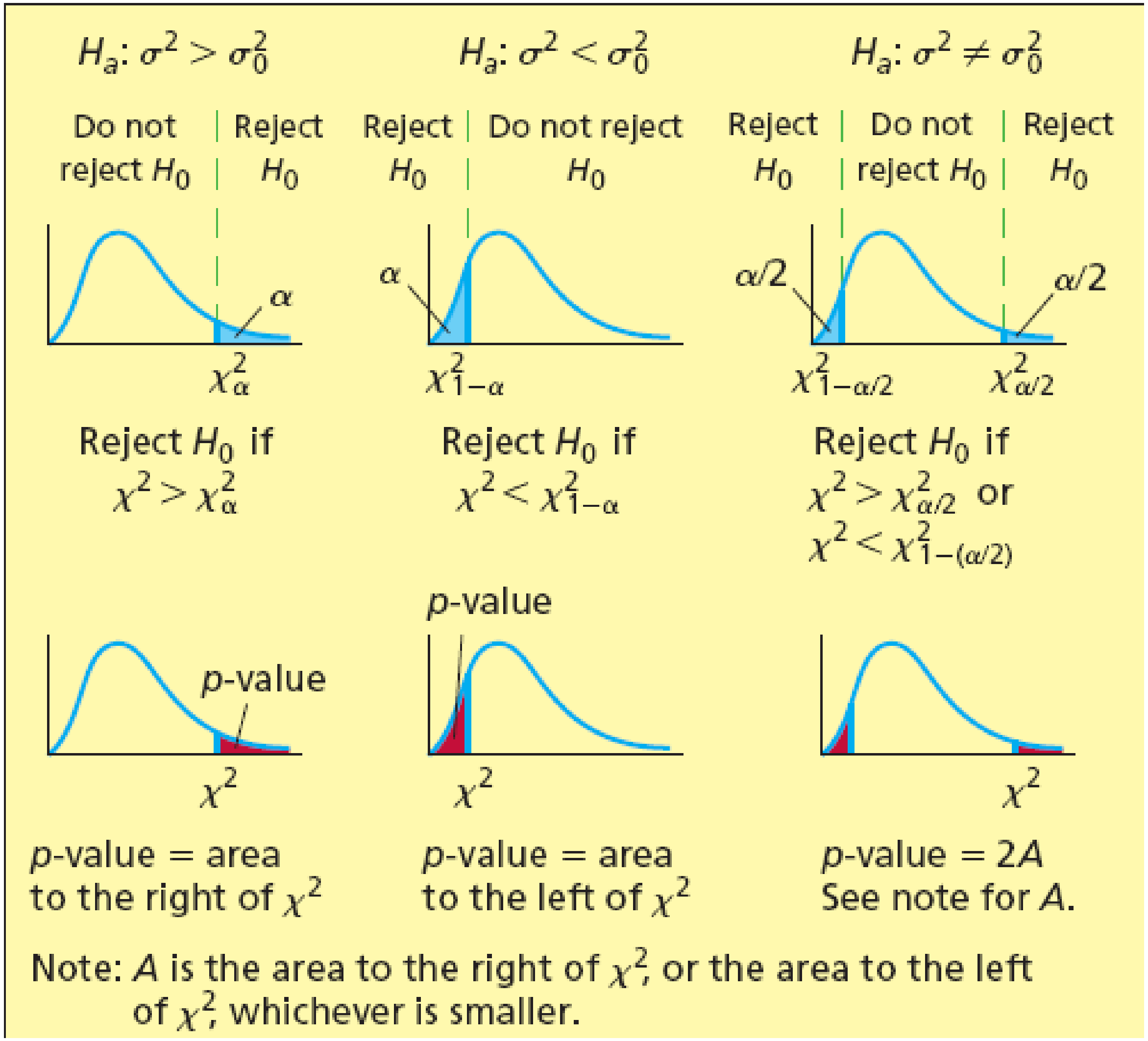
Example

The manufacturer of an expensive nutrition supplement has come under attack for bottles not containing a uniform amount of product. They have told their customers that the variance in weights of product is less than 1.2 ounces squared. A marketing researcher selects 25 bottles and finds a variance equal to 1.7. At the 10% level of significance is the company maintaining the product claim?

$$H_o \quad \sigma^2 \geq 1.2$$

$$H_a \quad \sigma^2 < 1.2$$

Hypothesis Testing for Population Variance



Example solved

$$\chi^2 = \frac{(n-1)\, s^2}{\sigma^2}$$

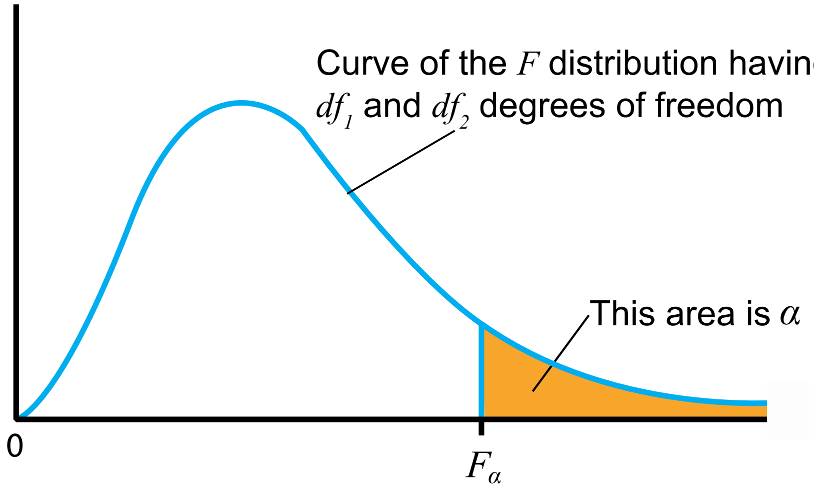
$$= \frac{24(1.7)}{1.2} = 34$$

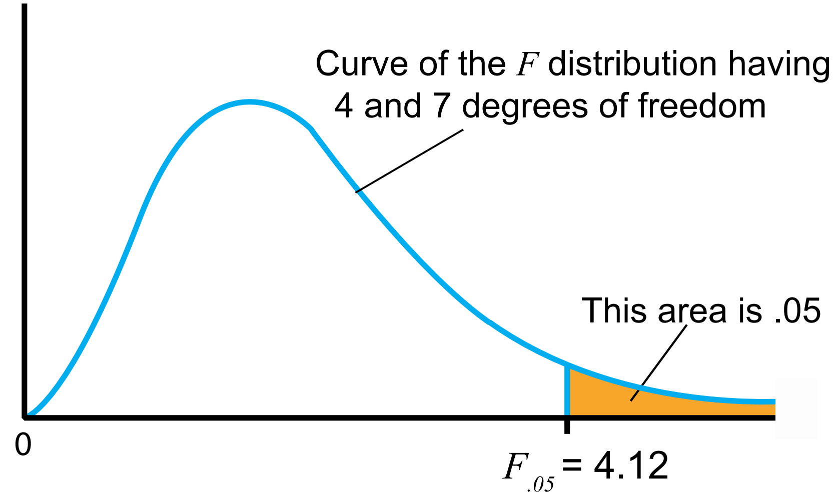
Reading from the Chi Square Table, the test statistic =15.659.

This is a left tail and since the calculated value 34, is greater than 15.659, we do not reject the null. However, results support the variability in the product weight is not less than 1.2.

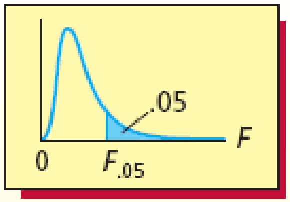
F – Test for two population variances

F Distribution





Portion on an F Table: Values of $F_{.05}$



df_2 \ df_1	Numerator Degrees of Freedom (df_1): 1	2	3	4	5	6	7	8	9	10	12	15	20	24
Denominator Degrees of Freedom (df_2): 1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84
7	5.59	4.71	4.25	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29

Comparing Two Population Variances Using Independent Samples

- Population 1 has variance σ_1^2 and population 2 has variance σ_2^2
- The null hypothesis H_0 is that the variances are the same
    - o $H_0 : \sigma_1^2 = \sigma_2^2$
- The alternative is that one is smaller than the other
    - o That population has less variable measurements
    - o Suppose $\sigma_1^2 > \sigma_2^2$
    - o More usual to normalize
- Test $H_0 : \sigma_1^2/\sigma_2^2 = 1 \text{ vs. } \sigma_1^2/\sigma_2^2 > 1$

Comparing Two Population Variances *continued*

- Reject H_0 in favor of H_a if s_1^2/s_2^2 is significantly greater than *1*
- s_1^2 is the variance of a random of size n_1 from a population with variance σ_1^2
- s_2^2 is the variance of a random of size n_2 from a population with variance σ_2^2
- To decide how large s_1^2/s_2^2 must be to reject H_0, describe the sampling distribution of s_1^2/s_2^2
- The sampling distribution of s_1^2/s_2^2 is the F distribution

Two Tailed Alternative

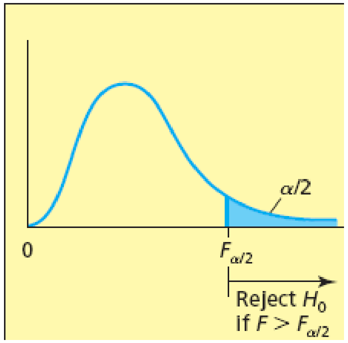
- The null and alternative hypotheses are:

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_a : \sigma_1^2 \neq \sigma_2^2$$

- Test statistic F is the ratio of the larger sample variance divided by the smaller sample variance

- $df_1 = n\text{-}1$ for sample having the larger variance and $df_2 = n\text{-}1$ for smaller variance

- Reject if $F > F_{\alpha/2}$ or p-value $< \alpha$



Example using the F-Distribution

An economist wants to test the variance in salary of women with children and without children and selects ten women in each category.

Salary variance for women with children is $354 square compared to variance for women without children of $452 squared and test at 90% confidence level.

$$H_o \quad \sigma^2{}_1 = \sigma^2{}_2$$
$$H_a \quad \sigma^2{}_1 \neq \sigma^2{}_2$$

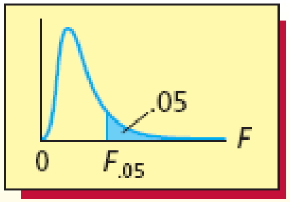
F test is set up as ratio with the larger value as the numerator and the rejection area is a right tail.

F= 452/354 = 1.24.

The table value for F = 3.18.

Interpretation is there is no statistical difference in variance between the two groups.

Portion on an F Table: Values of $F_{.05}$



df_2 \ df_1	Numerator Degrees of Freedom (df_1): 1	2	3	4	5	6	7	8	9	10	12	15	20	24
Denominator Degrees of Freedom (df_2): 1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84
7	5.59	4.71	4.25	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29

Example using the F-Distribution

An economist wants to test the variance in salary of women with children and without children and selects ten women in each category.

Salary variance for women with children is $354 square compared to variance for women without children of $452 squared and test at 90% confidence level.

$$H_o \quad \sigma^2{}_1 = \sigma^2{}_2$$
$$H_a \quad \sigma^2{}_1 \neq \sigma^2{}_2$$

F test is set up as ratio with the larger value as the numerator and the rejection area is a right tail.

F= 452/354 = 1.24.

The table value for F = 3.18.

Interpretation is there is no statistical difference in variance between the two groups.