# BAN 602: Quantitative Fundamentals

Spring, 2020 Lecture Slides – Week 5

CAL STATE EAST BAY

# Agenda

- Inferences About Proportions Variances
    - Inferences About a Population Variance
    - Inferences About Two Population Variances
- Test of Independence
- Goodness of Fit Test
    - Multinomial Probability Distribution



© Somak Paul

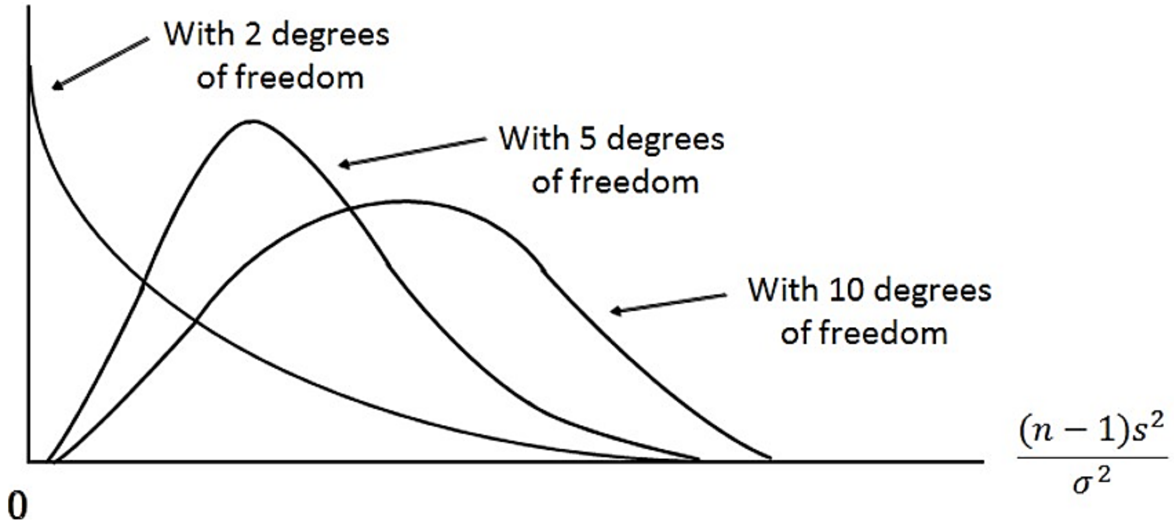
# Inferences About a Population Variance and Chi-Square Distribution

- A variance can provide important decision-making information.
- Consider the production process of filling containers with a liquid detergent product.
- The mean filling weight is important, but also is the variance of the filling weights.
- By selecting a sample of containers, we can compute a sample variance for the amount of detergent placed in a container.
- If the sample variance is excessive, overfilling and underfilling may be occurring even though the mean is correct.
- The chi-square distribution is based on sampling from a normal population.
- The sampling distribution of $\frac{(n-1)s^2}{\sigma^2}$ has a chi-square distribution with $n-1$ degrees of freedom whenever a simple random sample of size $n$ is selected from a normal population.
- We can use the chi-square distribution to develop interval estimates and conduct hypothesis tests about a population variance.



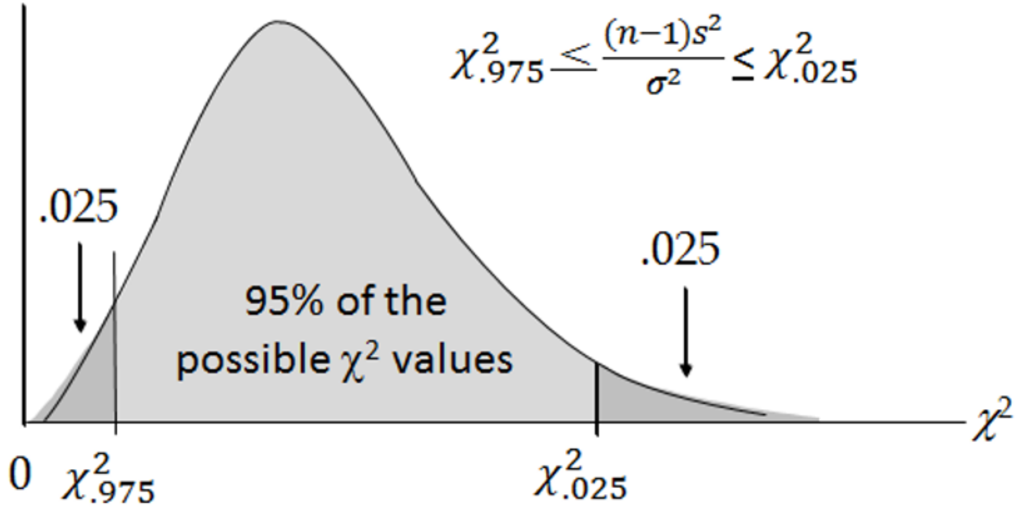
© Somak Paul

# Examples of Sampling Distribution of $\frac{(n-1)s^2}{\sigma^2}$




- The notation $\chi^2_\alpha$ denotes the value of the chi-square distribution that has an area of $\alpha$ to the right it.

- For example, there is a 0.95 probability of obtaining a $\chi^2$ (chi-square) value such that $\chi^2_{0.975} \le \chi^2 \le \chi^2_{0.025}$.






© Somak Paul

# Interval Estimation of $\sigma^2$

There is a $(1-\alpha)$ probability of obtaining a $\chi^2$ value such that

$$\chi^2_{\left(1-\frac{\alpha}{2}\right)} \le \chi^2 \le \chi^2_{\frac{\alpha}{2}}$$

Substituting $\frac{(n-1)s^2}{\sigma^2}$ for $\chi^2$ we get

$$\chi^2_{\left(1-\frac{\alpha}{2}\right)} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{\frac{\alpha}{2}}$$

After algebraic manipulation we get

$$\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{\left(1-\frac{\alpha}{2}\right)}}$$

Where the $\chi^2$ values are based on a chi-square distribution with $n-1$ degrees of freedom and where $1-\alpha$ is the confidence coefficient.

Taking the square root of the upper and lower limits of the variance interval provides the confidence interval for the population standard deviation.

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}} \le \sigma \le \sqrt{\frac{(n-1)s^2}{\chi^2_{\left(1-\frac{\alpha}{2}\right)}}}$$



© Somak Paul

# Interval Estimation of $\sigma^2$

Example: Buyer's Digest (A)

Buyer's Digest rates thermostats manufactured for home temperature control. In a recent test, 10 thermostats manufactured by ThermoRite were selected and placed in a test room that was maintained at a temperature of 68°F. We will use the 10 readings below to develop a 95% confidence interval estimate of the population variance.

| Thermostat | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Temperature | 67.4 | 67.8 | 68.2 | 69.3 | 69.5 | 67.0 | 68.1 | 68.6 | 67.9 | 67.2 |

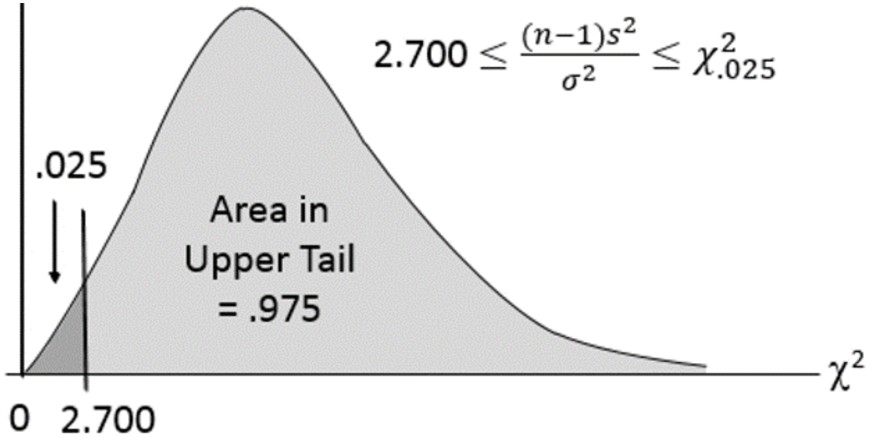


© Somak Paul

# Interval Estimation of $\sigma^2$

Selected Values from the Chi-Square Distribution Table.

For $n - 1 = 10 - 1 = 9$ *df* and $\alpha = 0.05$




| Degrees of Freedom | .99 Area in Upper Tail | .975 Area in Upper Tail | .95 Area in Upper Tail | .90 Area in Upper Tail | .10 Area in Upper Tail | .05 Area in Upper Tail | .025 Area in Upper Tail | .01 Area in Upper Tail |
|---|---|---|---|---|---|---|---|---|
| 5 | 0.554 | 0.831 | 1.145 | 1.610 | 9.236 | 11.070 | 12.832 | 15.086 |
| 6 | 0.872 | 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 |

CAL STATE EAST BAY

© Somak Paul

# Interval Estimation of $\sigma^2$

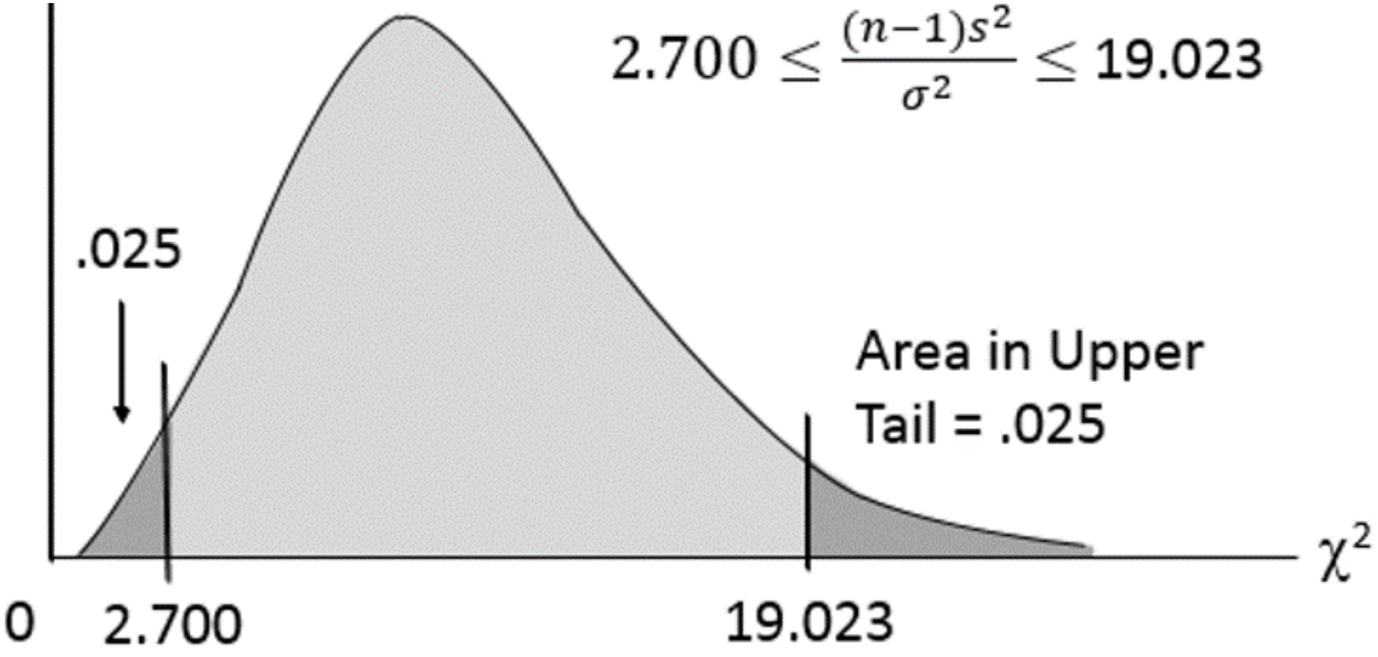
The sample variance $s^2$ provides a point estimate of $\sigma^2$.

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{6.3}{9} = .70$$

A 95% confidence interval for the population variance is given by:

$$\frac{(10-1).70}{19.02} \le \sigma^2 \le \frac{(10-1).70}{2.70}$$

$$.33 \le \sigma^2 \le 2.33$$






© Somak Paul

# Hypothesis Testing About a Population Variance

**Two-tailed test**

$H_0: \sigma^2 = \sigma_0^2$
$H_a: \sigma^2 \neq \sigma_0^2$

Critical value:

Reject $H_0$ if $\chi^2 \leq \chi^2_{\left(1-\frac{\alpha}{2}\right)}$
or if $\chi^2 \geq \chi^2_{\frac{\alpha}{2}}$

$p$-value:

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

**Lower-tail test**

$H_0: \sigma^2 \geq \sigma_0^2$
$H_a: \sigma^2 < \sigma_0^2$

Critical value:

Reject $H_0$ if $\chi^2 \leq \chi^2_{(1-\alpha)}$

$p$-value:

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

**Upper-tail test**

$H_0: \sigma^2 \leq \sigma_0^2$
$H_a: \sigma^2 > \sigma_0^2$

Critical value:

Reject $H_0$ if $\chi^2 \geq \chi^2_{\alpha}$

$p$-value:

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


© Somak Paul

# Hypothesis Testing About a Population Variance

For each type of test,

- $\sigma_0^2$ = the hypothesized value for the population variance.
- the test statistic is $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$.
- the chi-square critical values are based on a chi-square distribution with $n-1$ degrees of freedom.

**Example**: Buyer's Digest (B)

Recall that Buyer's Digest is rating ThermoRite thermostats. Buyer's Digest gives an "acceptable" rating to a thermostat with a temperature variance of 0.5 or less. Using the 10 readings, we will conduct a hypothesis test (with *a* = 0.10) to determine whether the ThermoRite thermostat's temperature variance is "acceptable".

Hypotheses

$$H_0: \sigma^2 \leq 0.5$$
$$H_a: \sigma^2 > 0.5$$

where $\sigma^2$ = the variance of the population of thermostat temperature readings.



© Somak Paul

# Hypothesis Testing About a Population Variance

For $n - 1 = 10 - 1 = 9$ *df* and $\alpha = 0.10$

Selected Values from the Chi-Square Distribution Table

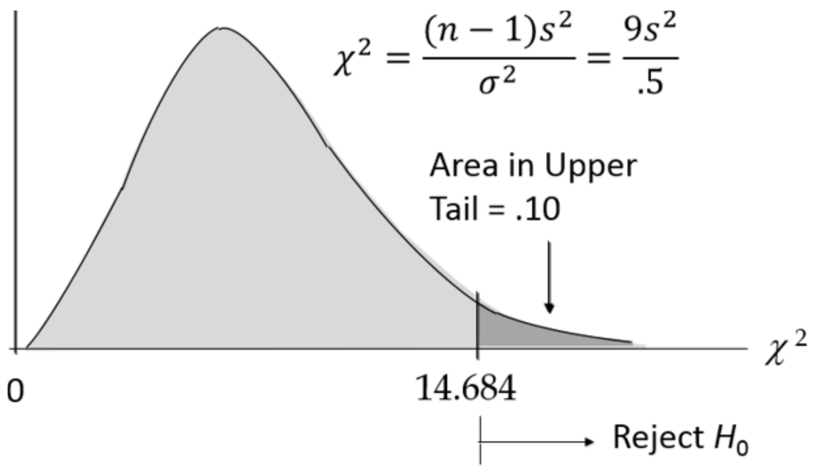
| Degrees of Freedom | .99 Area in Upper Tail | .975 Area in Upper Tail | .95 Area in Upper Tail | .90 Area in Upper Tail | .10 Area in Upper Tail | .05 Area in Upper Tail | .025 Area in Upper Tail | .01 Area in Upper Tail |
|---|---|---|---|---|---|---|---|---|
| 5 | 0.554 | 0.831 | 1.145 | 1.610 | 9.236 | 11.070 | 12.832 | 15.086 |
| 6 | 0.872 | 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 |



© Somak Paul

# Hypothesis Testing About a Population Variance

The sample variance is $s^2 = 0.7$.




Test Statistic: $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(10-1)(0.7)}{0.5} = 12.6$

Conclusion:

Because $\chi^2$ =12.6 is less than 14.684 we cannot reject $H_0$.

The sample variance $s^2 = 0.7$ is insufficient evidence to conclude that the temperature variance for ThermoRite thermostats is unacceptable.

**Using the *p*-Value**

- The rejection region for the ThermoRite thermostat example is in the upper tail; thus the appropriate *p*-value is greater than 0.10.

- Because the *p*-value > $\alpha = 0.10$, we cannot reject the null hypothesis.

- The sample variance $s^2 = 0.7$ is insufficient evidence to conclude that the temperature variance is unacceptable (> 0.5). The exact *p*-value is 0.18156.


© Somak Paul

# Inferences About Two Population Variances

- We may want to compare the variances in:
  - product quality resulting from two different production processes,
  - temperatures for two heating devices, or
  - assembly times for two assembly methods.
- We use data collected from two independent random samples, one from population 1 and another from population 2.
- The two sample variances will be the basis for making inferences about the two population variances.


© Somak Paul

# Hypothesis Testing About a Population Variance

**Two-tailed test**

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

Critical value:

Reject $H_0$ if $F \geq F_{\alpha/2}$

*p*-value:

Reject $H_0$ if *p*-value$\leq \alpha$

**Upper-tail test**

$H_0: \sigma_1^2 \leq \sigma_2^2$
$H_a: \sigma_1^2 > \sigma_2^2$

Critical value:

Reject $H_0$ if $F \geq F_\alpha$

*p*-value:

Reject $H_0$ if *p*-value$\leq \alpha$

- $\sigma_1^2$ = the variance of population 1 and $\sigma_2^2$ = the variance of population 2
- The test statistic is $F = \frac{s_1^2}{s_2^2}$ where $s_1^2$ is the larger sample variance.
- The value of $F_\alpha$ is based on an $F$ distribution with $n_1 - 1$ degrees of freedom in the numerator $n_2 - 1$ degrees of freedom in the denominator.



© Somak Paul

# Hypothesis Testing About the Variances of Two Populations

**Example:** Buyer's Digest (C)

Buyer's Digest has conducted the same test, as described earlier, on another 10 thermostats, this time manufactured by TempKing. We will conduct a hypothesis test with $\alpha = 0.10$ to see if the variances are equal for ThermoRite's thermostats and TempKing's thermostats.

ThermoRite Sample

| Thermostat | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Temperature | 67.4 | 67.8 | 68.2 | 69.3 | 69.5 | 67.0 | 68.1 | 68.6 | 67.9 | 67.2 |

TempKing Sample

| Thermostat | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Temperature | 67.7 | 66.4 | 69.2 | 70.1 | 69.5 | 69.7 | 68.1 | 66.6 | 67.3 | 67.5 |

$H_0: \sigma_1^2 = \sigma_2^2$ TempKing and ThermoRite thermostats have the same temperature variance.

$H_a: \sigma_1^2 \neq \sigma_2^2$ The population variances are not equal.



© Somak Paul

# Hypothesis Testing About the Variances of Two Populations

Rejection Rule

The $F$ distribution table (next slide) shows that with $\alpha/2$ = 0.05, 9 numerator *df* and 9 denominator *df*, $F_{0.05} = 3.18$.

We reject $H_0$ if $F \geq 3.18$.

Selected Values from the $F$ distribution Table

| Denominator Degrees of Freedom | Area in Upper Tail | Numerator Degrees of Freedom 7 | 8 | 9 | 10 | 15 |
|---|---|---|---|---|---|---|
| 8 | .10 | 2.62 | 2.59 | 2.56 | 2.54 | 2.46 |
| | .05 | 3.50 | 3.44 | 3.39 | 3.35 | 3.22 |
| | .025 | 4.53 | 4.43 | 4.36 | 4.30 | 4.10 |
| | .01 | 6.18 | 6.03 | 5.91 | 5.81 | 5.52 |
| 9 | .10 | 2.51 | 2.47 | 2.44 | 2.42 | 2.34 |
| | .05 | 3.29 | 3.23 | 3.18 | 3.14 | 3.01 |
| | .025 | 4.20 | 4.10 | 4.03 | 3.96 | 3.77 |
| | .01 | 5.61 | 5.47 | 5.35 | 5.26 | 4.96 |

CAL STATE EAST BAY

© Somak Paul

# Hypothesis Testing About the Variances of Two Populations

Test Statistic:

TempKing's sample variance is 1.768. ThermoRite's sample variance is 0.7. $F = \frac{s_1^2}{s_2^2} = \frac{1.768}{0.7} = 2.53$

Conclusion:

We cannot reject $H_0$ because $F = 2.53 < F_{0.05} = 3.18$.

There is insufficient evidence to conclude that the population variances differ for the two thermostat brands.

Determining and using the *p*-Value

| | | | | |
|---|---|---|---|---|
| Area in Upper Tail | .10 | .05 | .025 | .01 |
| $F$ Value ($df_1 = 9$, $df_2 = 9$) | 2.44 | 3.18 | 4.03 | 5.35 |

- Because $F$=2.53 is between 2.44 and 3.18, the area in the upper tail of the distribution is between 0.10 and 0.05.
- But this is a two-tailed test; after doubling the upper-tail area, the *p*-value is between 0.20 and 0.10.
- Because $\alpha$ = 0.10, we have *p*-value > $\alpha$ and therefore we cannot reject the null hypothesis.



© Somak Paul

# Test of Independence

1. Set up the null and alternative hypotheses.
   $H_0$: The column variable is independent of the row variable
   $H_a$: The column variable is <u>not</u> independent of the row variable
2. Select a random sample and record the observed frequency, $f_{ij}$, for each cell of the contingency table.
3. Compute the expected frequency, $e_{ij}$, for each cell. $e_{ij} = \frac{(\text{Row } i \text{ Total})(\text{Column } j \text{ Total})}{\text{Sample Size}}$
4. Compute the test statistic. $\chi^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$
5. Determine the rejection rule.

$$\text{Reject } H_0 \text{ if } \chi^2 \geq \chi^2_\alpha$$

Where $\alpha$ is the significance level with $n$ rows and $m$ columns, there are $(n-1)(m-1)$ degrees of freedom.



© Somak Paul

# Test of Independence

**Example:** Finger Lakes Homes

Each home sold by Finger Lakes Homes can be classified according to price and to style. Finger Lakes' manager would like to determine if the price of the home and the style of the home are independent variables.

The number of homes sold for each model and price for the past two years is shown below. For convenience, the price of the home is listed as either *less than \$200,000* or *more than or equal to \$200,000*.

| Price | Colonial | Log | Split-Level | A-Frame |
|---|---|---|---|---|
| < \$200,000 | 18 | 6 | 19 | 12 |
| ≥ \$200,000 | 12 | 14 | 16 | 3 |

Hypotheses

$H_0$: Price of the home is independent of the style of the home that is purchased

$H_a$: Price of the home is not independent of the style of the home that is purchased



© Somak Paul

# Test of Independence

Expected Frequencies

| Price | Colonial | Log | Split-Level | A-frame | Total |
|---|---|---|---|---|---|
| < \$200 K | 18 | 6 | 19 | 12 | 55 |
| ≥ \$200 K | 12 | 14 | 16 | 3 | 45 |
| Total | 30 | 20 | 35 | 15 | 100 |

- Rejection Rule

  With $\alpha = 0.05$ and $(2 – 1)(4 – 1) = 3$ degrees of freedom, $\chi^2_\alpha = 7.815$.

  Reject $H_0$ if the $p$-value$\leq 0.05$ or if $\chi^2 \geq 7.815$.

- Test Statistic $\chi^2 = \dfrac{(18-16.5)^2}{16.5} + \dfrac{(6-11)^2}{11} + \cdots + \dfrac{(3-6.75)^2}{6.75}$

$$= 0.1364 + 2.2727 + \cdots + 2.0833$$

$$= 9.149$$



© Somak Paul

# Test of Independence

Conclusion using the *p*-value approach

| Area in Upper Tail | .10 | .05 | .025 | .01 | .005 |
|---|---|---|---|---|---|
| $\chi^2$ Value (df = 3) | 6.251 | 7.815 | 9.348 | 11.345 | 12.838 |

Because $\chi^2 = 9.145$ is between 7.815 and 9.348, the area in the upper tail of the distribution is between 0.025 and 0.05.

Therefore, the *p*-value $\leq \alpha$. We reject the null hypothesis. The actual *p*-value is 0.0274.

Conclusion Using the Critical Value Approach: $\chi^2 = 9.145 \geq 7.815$

We reject at the 0.05 level of significance, the assumption that the price of the home is independent of the style of home that is purchased.


© Somak Paul

# Goodness of Fit Test: Multinomial Probability Distribution

1. State the null and alternative hypotheses.

   $H_0$: 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 specified probabilities for each of the $k$ categories

2. Select a random sample and record the observed frequency, $f_i$, for each of the $k$ categories.
3. Assuming $H_0$ is true, compute 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.

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

where:
$f_i$ = observed frequency for category $i$
$e_i$ = expected frequency for category $i$
$k$ = number of categories

Note: The test statistic has a chi-square distribution with $k - 1$ degrees of freedom provided that the expected frequencies are 5 or more for all categories.



© Somak Paul

# Goodness of Fit Test: Multinomial Probability Distribution

Rejection Rule:

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

Critical value approach: Reject $H_0$ if $\chi^2 \geq \chi^2_\alpha$

Where $\alpha$ is the significance level and there are $k - 1$ degrees of freedom.

**Example:** Finger Lakes Homes (A)

Finger Lakes Homes manufactures four models of prefabricated homes, a two-story colonial, a log cabin, a split-level, and an A-frame. To help in production planning, management would like to determine if previous customer purchases indicate that there is a preference in the style selected.

The number of homes sold of each model for 100 sales over the past two years is shown below.

| Model | Colonial | Log | Split-Level | A-Frame |
|---|---|---|---|---|
| # Sold | 30 | 20 | 35 | 15 |


© Somak Paul

# Multinomial Distribution Goodness of Fit Test

Hypotheses

$H_0$: $p_C = p_L = p_S = p_A = .25$

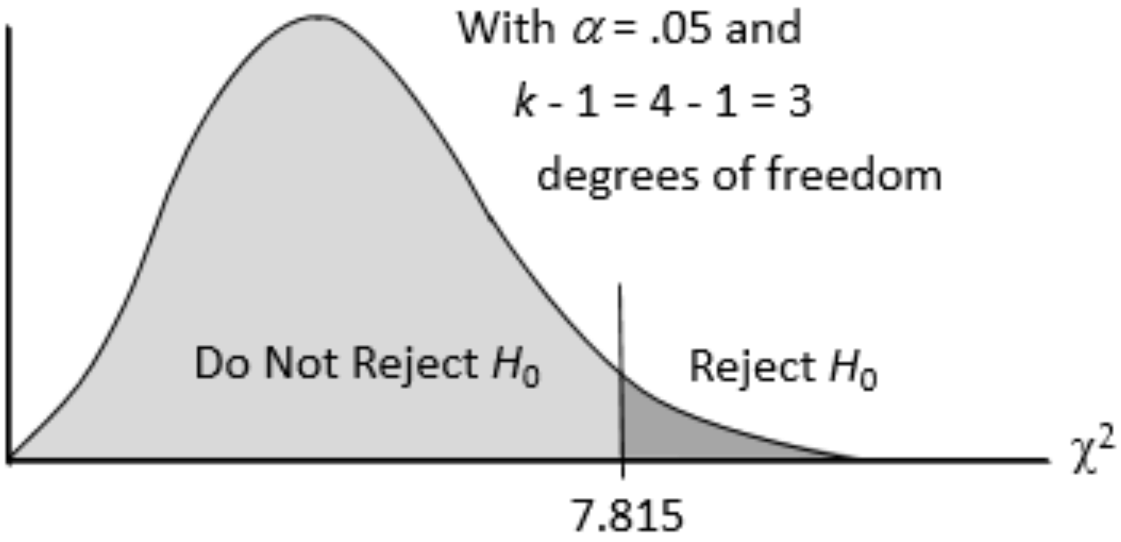
$H_a$: The population proportions are <u>not</u> $p_C = .25$, $p_L = .25$, $p_S = .25$, and $p_A = .25$ where:

$p_C$ = population proportion that purchase a colonial

$p_L$ = population proportion that purchase a log cabin

$p_S$ = population proportion that purchase a split-level

$p_A$ = population proportion that purchase an A-frame




Rejection Rule:

Reject $H_0$ if the $p$-value $\leq 0.05$ or $\chi^2 \geq 7.815$.

# Multinomial Distribution Goodness of Fit Test

Expected Frequencies:

$$e_1 = 0.25(100) = 25$$
$$e_2 = 0.25(100) = 25$$
$$e_3 = 0.25(100) = 25$$
$$e_4 = 0.25(100) = 25$$

Test Statistic:

$$\chi^2 = \frac{(30-25)^2}{25} + \frac{(20-25)^2}{25} + \frac{(35-25)^2}{25} + \frac{(15-25)^2}{25} = 1 + 1 + 4 + 4 = \mathbf{10}$$

Conclusion using the *p*-value approach:

| Area in Upper Tail | .10 | .05 | .025 | .01 | .005 |
|---|---|---|---|---|---|
| $\chi^2$ Value (df = 3) | 6.251 | 7.815 | 9.348 | 11.345 | 12.838 |

Because $\chi^2 = 10$ is between 9.348 and 11.345, the area in the upper tail of the distribution is between 0.01 and 0.025.

Therefore, the *p*-value $\leq \alpha$. We reject the null hypothesis.

Conclusion Using the Critical Value Approach:

$$\chi^2 = 10 \geq 7.815$$

Reject $H_0$, at the 0.05 level of significance



© Somak Paul