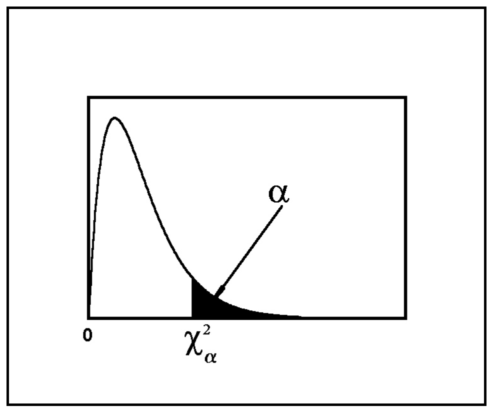
Statistical Inference Theory
Lesson 43
The Chi-Square Distribution
660

The chi-square distribution has many applications in statistical analysis. Before discussing these applications, we will first see how to interpret this distribution using the chi-square table F.

43.1 - What is the Chi-square distribution?

Assume we have a sequence of n independent normally distributed random variables $X_1, X_2, \ldots, X_n$ each with a mean $\mu = 0$ and $\sigma = 1$. We define the chi-square distribution χ^2 as



$$\chi^2 = X_1^2 + X_2^2 + X_3^2 + \ldots + X_n^2$$

Associated with each chi-square distribution is the number of degrees of freedom d = n.

For a given sample size n, each chi-square random variable χ^2 has a distribution represented by the above graph where

1. The total area under the curve is 1.

2. The values for chi-square χ^2 are found on the horizontal axis.

Table F gives the chi-square values for degrees of freedom from 1 to 30 and different values of α.

When the sample size is larger than 30 (N > 30), we can use the fact that the chi-square distribution approximately equals

$$\chi^2 \approx \frac{1}{2}(Z + \sqrt{2d - 1})^2,$$

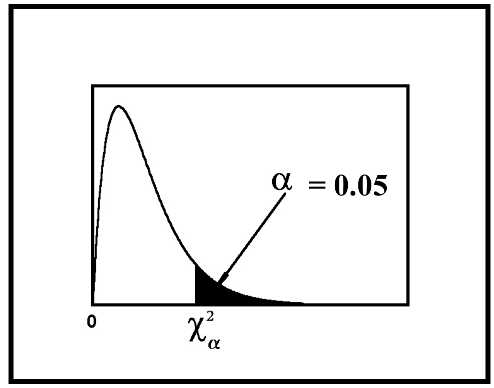
where Z is the standard normal distribution with mean 0 and standard deviation 1.

If $\alpha \leq 0.5$ then $Z \geq 0$.

If $\alpha > 0.5$ then $Z < 0$.

43.1 - Example 1: A sample of size N = 9 is taken. From the chi-square table find χ^2 for $\alpha = 0.05$.

Solution:



Step 1: Since N = 9, the number of degrees of freedom is d = 9.

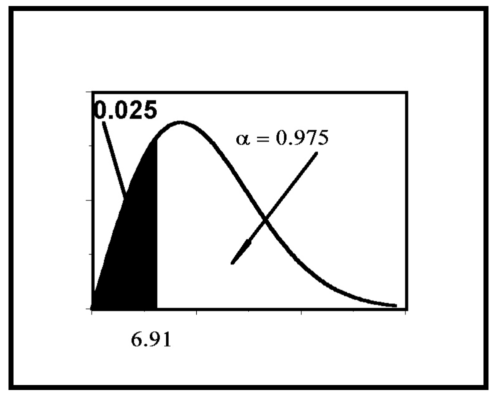
Step 2: Since $\alpha = 0.05$, select from the first row of the table $\chi^2_{0.05}$.

Step 3: Select from the first column the row for d = 9.

Step 4: Match this row with the column of $\chi^2_{0.05}$. The intersection of this column and row is $\chi^2_{0.05} = 16.9$.

43.1 - Example 2: A sample of size N = 16 is taken. From the chi-square table find χ^2 for the shaded area given in the figure.

Solution:



Step 1: Since N = 16, the number of degrees of freedom is d = 16.

Step 2: Since the shaded area is 0.025, we need to compute

$\alpha = 1 - 0.025 = 0.975$.

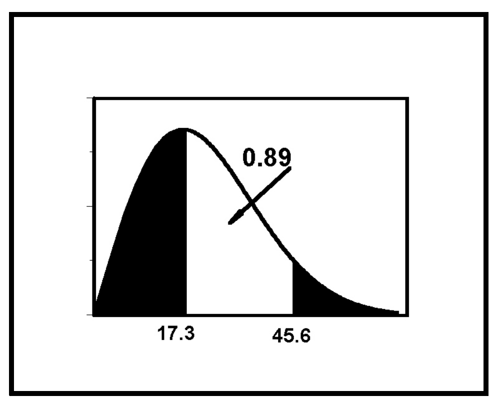
Step 3: Using $\alpha = 0.975$, we use $\chi^2_{0.975}$. From the table we find

$\chi^2_{0.975} = 6.91$.

43.1 - Example 3: A sample of size N = 26 is taken. From the chi-square table find the non-shaded area between

$\chi^2 = 17.3$ and $\chi^2 = 45.6$.

Solution:



Step 1: Since N = 26, the number of degrees of freedom is d = 26.

Step 2: For d = 26, $\chi^2_{0.90} = 17.3$.

Step 3: For d = 26, $\chi^2_{0.01} = 45.6$.

Step 4: From step 2 the left-hand shaded area in the figure is 1 - 0.90 = 0.10 .

Step 5: From step 3 the right-hand shaded area in the figure is 0.01 .

Step 6: Therefore, the non-shaded area is 1 - (0.10 + 0.01) = 0.89 .

543.1 - Example 4: A sample of size N = 100 is taken. Find χ^2 for $\alpha = 0.05$.

Solution:

Since our chi-square table only exists for N $\leq$ 30, we use the formula

$$\chi^2 \approx \frac{1}{2}(Z + \sqrt{2d - 1})^2.$$

where Z is standard normal distribution with mean 0 and standard deviation 1.

Step 1: Since $\alpha = 0.05$, rewrite the above formula as

$$\chi^2_{0.05} \approx \frac{1}{2}(Z_{0.05} + \sqrt{2d - 1})^2.$$

Step 2: d = 100

Step 3: Since Z is a standard normal distribution random variable, look up the area in the standard normal distribution table for $0.50 - \alpha = 0.45$.

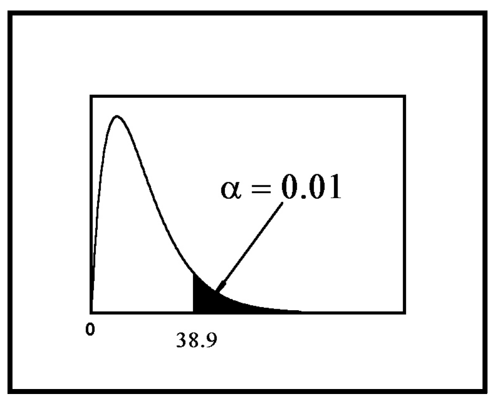
Step 4: From step 3, $Z_{0.05} = 1.64$.

Step 5: $\chi^2_{0.05} \approx \frac{1}{2}(Z_{0.05} + \sqrt{2d - 1})^2 = \frac{1}{2}(1.64 + \sqrt{2(100) - 1})^2 \approx 124$

Solved Problems

43.1 - Solved Problem 1: A sample of size N = 21 is taken. From the chi-square table find χ^2 for $\alpha = 0.01$.

Solution:



Step 1: Since N = 21, the number of degrees of freedom is d = 21.

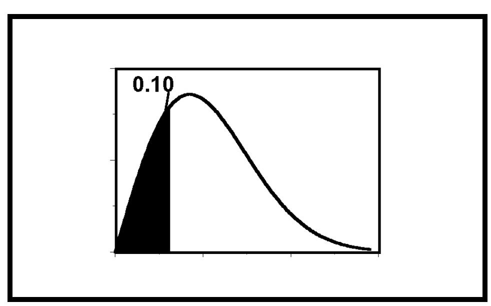
Step 2: Since $\alpha = 0.01$, select from the first row of the table $\chi^2_{0.01}$.

Step 3: Select from the first column the row for d = 21.

Step 4: Match this row with the column of $\chi^2_{0.01}$.

The intersection of this column and row is $\chi^2_{0.01} = 38.9$.

43.1 - Solved Problem 2: A sample of size N = 30 is taken. From the chi-square table, find χ^2 for the shaded area given in the figure.



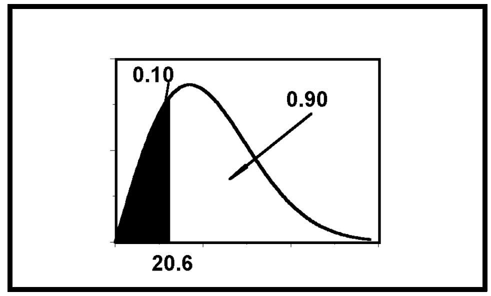
Solution:

Step 1: Since N = 30, the number of degrees of freedom is d = 30 .

Step 2: Since the shaded area is 0.10, we need to compute

$\alpha = 1 - 0.10 = 0.90$.

Step 3: Using $\alpha = 0.90$, we use $\chi^2_{0.90}$. From the table we find $\chi^2_{0.90} = 20.6$.



43.1 - Solved Problem 3: A sample of size N = 4 is taken. From the chi-square table find the area between

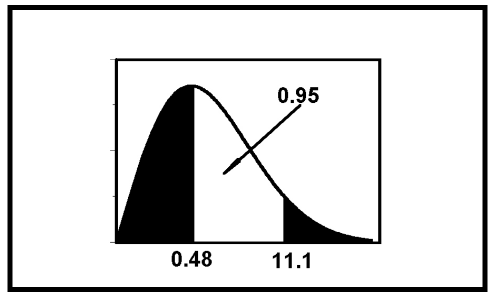
$\chi^2 = 0.48$ and $\chi^2 = 11.1$.

Solution:

Step 1: Since N = 4, the number of degrees of freedom is d = 4.

Step 2: For d = 4, $\chi^2_{.975} = 0.48$

Step 3: For d = 4, $\chi^2_{0.025} = 11.1$



Step 4: From step 2 the left-hand shaded area in the figure is 1 - 0.975 = 0.025 .

Step 5: From step 3 the right-hand shaded area in the figure is 0.025

Step 6: Therefore, the non-shaded area is 1 - (0.025 + 0.025) = 0.95 .

43.1 - Solved Problem 4: A sample of size N = 250 is taken. Find χ^2 for $\alpha = 0.95$.

Solution: Since our chi-square table only exists for $N \leq 30$, we use the formula

$$\chi^2 \approx \frac{1}{2}(Z + \sqrt{2d - 1})^2$$

where Z is standard normal distribution with mean 0 and standard deviation 1.

Step 1: Since $\alpha = 0.95$, rewrite the above formula as

$$\chi^2_{0.95} \approx \frac{1}{2}(Z_{0.95} + \sqrt{2d - 1})^2$$

Step 2: d = 250

Step 3: Since Z is a standard normal distribution random variable, look up the area in the standard normal distribution table for 0.95- 0.50 = 0.45 .

Step 4: From step 3, $Z_{0.95} = -1.64$

Step 5: $\chi^2_{0.95} \approx \frac{1}{2}(Z_{0.95} + \sqrt{2d - 1})^2 = \frac{1}{2}(-1.64 + \sqrt{2(250) - 1})^2 \approx 214$

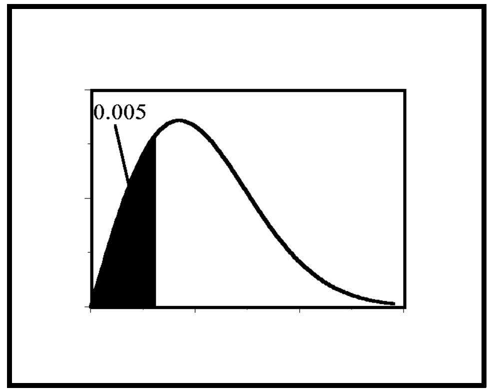
Unsolved Problems with Answers

43.1 - Problem 1: A sample of size N = 21 is taken. From the chi-square table, find χ^2 for $\alpha = 0.005$.

Answer: $\chi^2 = 41.4$

⇑ *Refer back to* **43.1 - Example 1 & 43.1 - Solved Problem 1**.

43.1 - Problem 2: A sample of size N = 10 is taken. From the chi-square table, find χ^2 for the shaded area given the figure.



Answer:

$\chi^2 = 2.16$

⇑ *Refer back to* **43.1 - Example 2 & 43.1 - Solved Problem 2**.

43.1 - Problem 3: A sample of size N = 14 is taken. From the chi-square table, find the area between $\chi^2 = 4.66$ and $\chi^2 = 7.79$.

Answer:

0.08

⇑ *Refer back to* **43.1 - Example 3 & 43.1 - Solved Problem 3**.

43.1 - Problem 4: A sample of size N = 35 is taken. Find χ^2 for $\alpha = 0.90$.

Answer:

$\chi^2 = 24.7$.

⇑ *Refer back to* **43.1 - Example 4 & 43.1 - Solved Problem 4**.

43.2 - Estimating σ^2 and σ using a χ^2 confidence interval.

Assume a random sample of size N is taken from a normal population with variance σ^2. If the variance of the sample is s^2, then it can be shown that $\frac{Ns^2}{\sigma^2}$ has a Chi-square distribution. Therefore, $\chi^2 = \frac{Ns^2}{\sigma^2}$ with d = N -1 degrees of freedom.

From this distribution ,we can derive the following confidence intervals for the variance σ^2 and standard deviation σ of a population:

$$\frac{Ns^2}{\chi^2_{\alpha}} \le \sigma^2 \le \frac{Ns^2}{\chi^2_{1-\alpha}}$$

$$\frac{\sqrt{N}s}{\chi_{\alpha}} \le \sigma \le \frac{\sqrt{N}s}{\chi_{1-\alpha}}$$

43.2 - Example.1: The Sweet Water Bottling company has a machine that fills in bottles 12 ounces of water. Each morning the variance of the machine is set to $\sigma^2 = 0.1$ ounces. During the day, the vibrations in the operation of the machine can significantly change the variance of the machine. To check for significant changes in variance, a sample of N bottles is taken and the sample variance s^2 is recorded.

(a). For N = 16, $s^2 = 0.15$ and a 95% confidence interval, estimate the variance σ^2 and standard deviation σ of the machine.

(b). For N = 100, $s^2 = 0.11$ and a 95% confidence interval, estimate the variance σ^2 and standard deviation σ of the machine.

Solutions:

➤**(a)**.

Step 1: Since we have a 95% confidence interval,

$\alpha = (1 - 0.95)/2 = 0.025$,

$1 - \alpha = 0.975$.

Step 2: For N - 1 = 16 - 1 = 15 degrees of freedom, from the table we have

$\chi^2_{0.025} = 27.50$,

$\chi^2_{0.975} = 6.26$

Step 3: From the above formula:

$$\frac{16(0.15)}{27.50} \le \sigma^2 \le \frac{16(0.15)}{6.26},$$

$$0.09 \le \sigma^2 \le 0.38$$

Step 4: Take the square root of each number in the inequality:

$$0.3 \le \sigma \le 0.62$$

➤**(b)**.

Since N = 100, we use the formula:

$\chi^2 \approx \frac{1}{2}(Z + \sqrt{2(N) - 1})^2.$

Step 1: $\chi^2_{0.025} \approx \frac{1}{2}(Z_{0.025} + \sqrt{2(99) - 1})^2 = (1/2)(1.96 + 14)^2 \approx 127.36$

Step 2: $\chi^2_{0.975} \approx \frac{1}{2}(Z_{0.975} + \sqrt{2(99) - 1})^2 = (1/2)(-1.96 + 14)^2 \approx 72.48$

Step 3: From the above formula:

$$\frac{100(0.11)}{127.36} \leq \sigma^2 \leq \frac{100(0.11)}{72.48},$$

$$0.09 \leq \sigma^2 \leq 0.15$$

Step 4: Taking the square root of both sides gives

$$0.3 \leq \sigma \leq 0.39$$

Solved Problems

43.2 - Solved Problem 1: A local community college recently took two random sample of grade point averages of its graduating students. To check the consistency of the teachers' grading policies, it needs an estimate of the variance σ^2 and standard deviation σ.

(a). If N = 30 and $s^2 = 0.3$, estimate the variance σ^2 and standard deviation σ of the grade point average using a confidence interval of 90%.

(b). If N = 200 and $s^2 = 0.18$, estimate the variance σ^2 and standard deviation σ of the grade point average using a confidence interval of 90%.

Solutions:

➤**(a)**.
Step 1: Since we have a 90% confidence interval,

$\alpha = (1 - 0.90)/2 = 0.05,$

$1 - \alpha = 0.95$

Step 2: For N - 1 = 30 - 1 = 29 degrees of freedom, from the table we have

$\chi^2_{0.05} = 42.6,$

$\chi^2_{0.95} = 17.7$

Step 3: From the above formula:

$$\frac{30(0.3)}{42.60} \le \sigma^2 \le \frac{30(0.3)}{17.70},$$

$$0.21 \le \sigma^2 \le 0.51$$

Step 4: Take the square root of both numbers in the equality gives

$$0.46 \le \sigma \le 0.71$$

➤**(b)**.

Since N = 200, we use the formula:

$$\chi^2 \approx \frac{1}{2}(Z + \sqrt{2d - 1})^2.$$

Step 1: $\chi^2_{0.05} \approx \frac{1}{2}(Z_{0.05} + \sqrt{2(200) - 1})^2 = \frac{1}{2}(1.64 + 19.92)2 \approx 233.6$

Step 2: $\chi^2_{0.95} = \frac{1}{2}(Z_{0.95} + \sqrt{2(200) - 1})^2 = \frac{1}{2}(-1.64 + 19.92)^2 \approx 167.1$

Step 3: From the above formula:

$$\frac{200(0.18)}{233.6} \le \sigma^2 \le \frac{200(0.18)}{167.1},$$

$$0.15 \le \sigma^2 \le 0.22$$

Step 4: Take the square root of both numbers in the above inequality:

$0.38 \le \sigma \le 0.47$

Unsolved Problems with Answers

43.2 - Problem 1: A machine drills holes in steel plates. The drilling error has a mean $\mu = 0.01$ mm with $\sigma^2 = 0.009$. Due to possible vibrations in the machine the drilling accuracy can change. To monitor the accuracy, periodically samples of size N are taken from plates that have been drilled.

(a). If $N = 25$ and $s^2 = 0.0096$, estimate the variance σ^2 and standard deviation σ of the accuracy in drilling using a 99% confidence interval.

(b). If $N = 50$ and $s^2 = 0.0092$, estimate the variance σ^2 and standard deviation σ of the accuracy in drilling using a 99% confidence interval.

Answers:

➤**(a)**. $0.0066 \le \sigma^2 \le 0.023$

$0.08 \le \sigma \le 0.15.$

➤**(b)**. $0.006 \leq \sigma^2 \leq 0.018$

$0.08 \leq \sigma \leq 0.13$

⇑ *Refer back to* **43.2 - Example 1 & 43.2 - Solved Problem 1**.

43.3 - Hypothesis testing for and a population standard deviation σ.

To test an hypothesis for a population standard deviation we need to use the formula:

$$\chi^2 = \frac{Ns^2}{\sigma^2}.$$

43.3 - Example 1: The Sweet Water Bottling company has a machine that fills in bottles 12 ounces of water. Each morning the standard deviation of the machine is set to $\sigma = 0.1$ ounces. During the day, the vibrations in the operation of the machine can significantly change the amount filled in each bottle by the machine. To check for significant changes in variance, a sample of 30 bottles is taken and the standard deviation $s = 0.12$ is recorded.

(a). State H_o and H_a.

(b). After each sample, in order to decide whether the machine should be shut down for adjustments, assume the following decision rule:

D.R.: If $\chi^2 \geq c^*$ *then stop the machine and make appropriate adjustments on the machine.*

Find c^* for a Type I error of $\alpha = 0.05$.

(c). Based on the above decision rule, would the machine be shut down? Explain.

Solutions:

➤**(a)**.
The concern is that the standard deviation will be larger than $\sigma = 0.1$. Therefore,

H_o: $\sigma = 0.1$

H_a: $\sigma > 0.1$

➤**(b)**.
Step 1: Since the alternative hypothesis is $\sigma > 0.1$, we use the right-hand side of the chi-square table for $\alpha = 0.05$.

Step 2: N = 30 , s = 0.12, $\sigma = 0.10$

Step 3: $\chi^2 = \dfrac{Ns^2}{\sigma^2} = \dfrac{(30)(0.12^2)}{0.10^2} = 43.2$

Step 4: For d = 30 - 1 = 29, and $\alpha = 0.05$, the Chi-square table gives $\chi^2_{0.05} = 42.6$.

Step 5: Therefore, c* = 42.6

Step 6: The decision rule is:

D.R.: If $\chi^2 \geq 42.6$ *then stop the machine and make appropriate adjustments on the machine.*

➤**(c)**.
Since $\chi^2 = 43.2 > 42.6$, the machine will be shut down.

Since the value of $\chi^2 = 43.2$ falls in the tail-end of the chi-square distribution, s = 0.12 is significant from $\sigma = 0.10$ and there is only a 5% chance that this difference is caused by random variation.

Solved Problems

43.3 - Problem 1: Recently the United States Defense department is considering to order a new type of anti-tank missile. The manufacturer of these missiles claims that the missiles will hit the area around any enemy tank within a standard deviation of 2 feet. A test sample of 26 missiles against tanks produced a sample standard deviation of 2.5 feet.

(a). State H_o and H_a.

(b). Assume the final decision depends on the result of this sample test. Using the following decision rule:

D.R.: If $\chi^2 \geq c^*$ *then the missiles will not be purchased.*

Find c* for a Type I error of $\alpha = 0.01$.

(c). Based on the above decision rule, would the missiles be purchased? Explain.

Solutions:

➤**(a)**.
The concern is that the standard deviation will be larger than $\sigma = 2$ feet. Therefore,

H_o: $\sigma = 2$

H_a: $\sigma > 2$

➤**(b)**.
Step 1: Since the alternative hypothesis is $\sigma > 2$, we use the right-hand side of the chi-square table for $\alpha = 0.01$.

Step 2: N = 26 , s = 2.5, σ = 2

Step 3: $\chi^2 = \dfrac{Ns^2}{\sigma^2} = \dfrac{(26)(2.5)^2}{2^2} = 40.25$

Step 4: For d = 26 - 1 = 25, and α = 0.01, the Chi-square table gives $\chi^2_{0.01} = 44.3$.

Step 5: Therefore, c* = 44.3

Step 6: The decision rule is : *D.R.: If* $\chi^2 \geq 44.3$ *then the missiles will not be purchased.*

➤**(c)**.
Since $\chi^2 = 40.25 < 44.3$, the missiles will be purchased.

The value s = 2.5 feet can be explained as caused by random variation. Therefore there is no significant difference from the standard deviation of 2 feet.

Unsolved Problems with Answers

43.3 - Problem 1: The CEO of a large railroad company claims that the company's trains arrive in New York city within 1 standard deviation of official arrival time. To substantiate this claim, the company checked the arrival time of 10 trains into New York city. From this sample they computed a standard deviation of 1.8 minutes.

(a). State H_o and H_a.

(b). In supporting or rejecting this claim, use the following decision rule:

D.R.: If $\chi^2 \geq c^*$ *then the claim is rejected.*

Find c^* for a Type I error of α = 0.10 .

(c). Based on the above decision rule, would the claim be rejected? Explain.

Answers:

➤**(a)**. H_o: σ = 1

H_a: σ > 1

➤**(b)**.
The decision rule is :

D.R.: If $\chi^2 \geq 14.7$ *then reject the claim.*

➤**(c)**.
Since $\chi^2 = 32.4 > 14.7$, the value s = 1.8 minutes cannot not be explained as caused by random variation. Therefore, the claim is rejected.

⇑ *Refer back to* **43.3 - Example 1 & 43.3 - Solved Problem 1**.

43.4 - Testing For Goodness of Fit

The Chi-square distribution can be used to test the discrepancy between observed and expected frequencies. This test is given by the chi-square distribution:

$$\chi^2 = \frac{(x_1 - e_1)^2}{e_1} + \frac{(x_2 - e_2)^2}{e_2} + \frac{(x_3 - e_3)^2}{e_3} + \ldots + \frac{(x_N - e_N)^2}{e_N}.$$

where

x_k is the observed frequencies,

e_k is the expected frequencies,

$d = N - 1$, the degrees of freedom.

To test the data for goodness of fit, the null hypothesis is

H_o: The data fits the tested distribution.

H_a: The data does not fit the tested distribution.

The null is rejected if $\chi^2 > \chi^2_\alpha$ for a given value of α.

43.4 - Example 1: Mr. Goodman recently purchased a rare coin. To test if the coin is fair he tossed the coin 500 times. Assuming that he recorded 275 heads and 225 tails .

(a). State H_o and H_a.

(b). Complete the table:

Coin face	Heads	Tails
Data values (x)	275	225
Expected (e)		

(c). Compute χ^2. For $\alpha = 0.05$, would the null hypothesis be rejected? Explain.

Solutions:

➤**(a)**.
H_o: The coin is fair.

H_a: The coin is biased.

➤**(b)**.
To complete the above table, we need to compute each expected values.

Step 1: To compute the expected values we assume H_o is true.

Step 2: For each toss the probability of heads is p = 0.50 and the probability of tails is q = 0.50 .

Step 3: Since the number of tosses is 500, the expected values for both heads and tails are 500(0.50) = 250 .

Step 4:

Coin face	Heads	Tails
Data values (x)	275	225
Expected (e)	250	250

➤**(c)**.
Step 1: For the above formula we set:

$x_1 = 275$, $x_2 = 225$, $e_1 = 250$, $e_2 = 250$, $d = 2 - 1 = 1$.

Step 2: $\chi^2 = \dfrac{(275 - 250)^2}{250} + \dfrac{(225 - 250)^2}{250} = \dfrac{(25)^2}{250} + \dfrac{(25\ )^2}{250} = 5$

Step 3: For d = 2 - 1 = 1, the chi-square table give $\chi^2_{0.05} = 3.84$.

Step 4: Since 5 > 3.84 we reject H_0 and conclude that at a $\alpha = 0.05$ level of significance, we reject the coin as being fair.

43.4 - Example 2: A Las Vegas casino recently purchased an electronic machine that simulates the toss of a pair of dice. To test the randomness of the machine according to the odds for dice, the machine was made to toss a pair of dice 360 times and each sum of the dice were recorded.

(a). State H_o and H_a.

(b). Complete the table:

Sum of two dice	2	3	4	5	6	7	8	9	10	11	12
Data values (x)	9	21	30	30	60	57	53	35	38	26	1
Expected (e)											

(c). Compute χ^2 .

For $\alpha = 0.01$, would the null hypothesis be rejected? Explain.

(d). For $\alpha = 0.10$, would the null hypothesis be rejected? Explain.

Solutions:

➤**(a)**.

H_o: The machine is generating random pairs of numbers according to the odds on dice.

H_a: The machine is biased.

➤**(b)**.

To complete the above table, we need to compute each expected values.

Step 1: To compute the expected values we assume H_o is true.

Step 2: Assume X is the random distribution for the sum of a pair of dice. Therefore,

$P\{X = 2\} = 1/36$, $P\{X = 3\} = 2/36$, $P\{X = 4\} = 3/36$, $P\{X = 5\} = 4/36$,

$P\{X = 6\} = 5/36$, $P\{X = 7\} = 6/36$, $P\{X = 8\} = 5/36$, $P\{X = 9\} = 4/36$,

$P\{X = 10\} = 3/36$, $P\{X = 11\} = 2/36$, $P\{X = 12\} = 1/36$.

Step 3: Since the number of tosses is 360, the expected values are as followings:

$e_1 = 360(1/36) = 10$, $e_2 = 360(2/36) = 20$, $e_3 = 360(3/36) = 30$,

$e_4 = 360(4/36) = 40$, $e_5 = 360(5/36) = 50$, $e_6 = 360(6/36) = 60$,

$e_7 = 360(5/36) = 50$, $e_8 = 360(4/36) = 40$, $e_9 = 360(3/36) = 30$,

$e_{10} = 360(2/36) = 20$, $e_{11} = 360(1/36) = 10$.

Step 4:

Sum of two dice	2	3	4	5	6	7	8	9	10	11	12
Data values (x)	9	21	30	30	60	57	53	35	38	26	1
Expected (e)	10	20	30	40	50	60	50	40	30	20	10

➤**(c)**.

Step 1: For the above formula we set

$x_1 = 9$, $x_2 = 21$, $x_3 = 30$, $x_4 = 30$, $x_5 = 60$, $x_6 = 57$, $x_7 = 53$, $x_8 = 35$,

$x_9 = 38$, $x_{10} = 26$, $x_{11} = 1$

$e_1 = 10$, $e_2 = 20$, $e_3 = 30$, $e_4 = 40$, $e_5 = 50$, $e_6 = 60$, $e_7 = 50$, $e_8 = 40$,

$e_9 = 30$, $x_{10} = 20$, $x_{11} = 10$

Step 2: $\chi^2 = \dfrac{(9 - 10)^2}{10} + \dfrac{(21 - 20)^2}{20} + \dfrac{(30 - 30)^2}{30} + \dfrac{(30 - 40)^2}{40} +$

$\dfrac{(60 - 50)^2}{50} + \dfrac{(57 - 60)^2}{60} + \dfrac{(53 - 50)^2}{50} + \dfrac{(35 - 40)^2}{40} +$

$\dfrac{(38 - 30)^2}{30} + \dfrac{(26 - 20)^2}{20} + \dfrac{(1 - 10)^2}{10} \approx 17.64$

Step 3: For d = 11 - 1 = 10, the chi-square table gives

$\chi^2_{0.01} = 23.2$

Step 4: Since 17.64 < 23.2 we reject H_a and conclude that at an $\alpha = 0.01$ level of significance, we have no reason to believe the machine is not generating numbers according to the odds of dice.

➤**(d)**.
For d = 11 - 1 = 10, the chi-square table gives

$\chi^2_{0.10} = 16$

Step 5: Since 17.64 > 16 we reject H_o and conclude that at an $\alpha = 0.10$ level of significance, we have reason to believe the machine is not generating numbers according to the odds of dice.

43.4 - Example 3: According to the theory of human genetics, in a large population, half the children born are boys and the other half are girls. To test this proportion, 1,000 families each with four children were randomly selected.

(a). State H_o and H_a.

(b). Complete the following table:

k girls	0	1	2	3	4
Number families that have k girls	75	281	355	220	69
Expected number families that have k girls					

(c). compute χ^2.

For $\alpha = 0.025$, would you support the hypothesis that there is an equal proportion of boys and girls born? Explain.

(d). For $\alpha = 0.01$, would you support the hypothesis that there is an equal proportion of boys and girls born? Explain.

Solutions:

➤(a).
H_o: One-half of all children born are boys and other half girls.

H_a: The proportion of girls and boys born are not equal.

➤(b).
Step 1: Assume H_o is true.

Step 2: It is reasonable to assume that the distribution of gender among children is a binomial distribution:

$$P\{X = k\} = \binom{4}{k}(0.5)^k(0.5)^{4-k}$$

k girls per family	0	1	2	3	4
Number families that have k girls	75	281	355	220	69
Expected number families that have k girls: $(1000)\binom{4}{k}(0.5)^k(0.5)^{4-k}$	62	250	375	250	62

➤(c).
Step 1: For the above formula we set:

$x_1 = 75$, $x_2 = 281$, $x_3 = 355$, $x_4 = 220$, $x_5 = 69$,

$e_1 = 62.5$, $e_2 = 250$, $e_3 = 375$, $e_4 = 250$, $e_5 = 62.5$

Step 2: $$\chi^2 = \frac{(75-62.5)^2}{62.5} + \frac{(281-250)^2}{250} + \frac{(355-375)^2}{375} + \frac{(220-250)^2}{250} + \frac{(69-62.5)^2}{62.5} =$$

11.69

Step 3: For d = 5 - 1 = 4, the chi-square table gives

$\chi^2_{0.025} = 11.1$

Step 4: Since 11.1 < 11.69 we reject H_0 and conclude that at a $\alpha = 0.025$ level of significance, we can't conclude the theory that 50% of children born are males.

➤(d).
For d = 5 - 1 = 4, the chi-square table gives:

$\chi^2_{0.01} = 13.3$

Step 5: Since 11.69 < 13.3 we reject H_a and conclude that at a $\alpha = 0.01$ level of significance, we can conclude the theory that 50% of children born are males.

Solved Problems

43.4 - Solved Problem 1: A statistician was hired by a professional basketball team to determine if the team's percentage of winning games significantly depend on the day of the week the games are played on. For over 350 winning games he compiled the following winning games:

Day of the week	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Sun.
Number games won	61	47	55	38	59	35	55

(a). State H_o and H_a.

(b). Complete the table:

Day of the week	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Sun.
Number games won	61	47	55	38	59	35	55
Expected (e)							

(c). Compute χ^2. For $\alpha = 0.10$, would the null hypothesis be rejected? Explain.

Solutions:

➤**(a)**.
H_o: The day of the week has no affect on the team's ability to win.
H_a: The day of the week has an affect on the team's ability to win.

➤**(b)**.
To complete the above table, we need to compute each expected values.

Step 1: To compute the expected values we assume H_o is true.

Step 2: There are 350 games spread over 7 days a week.

Step 3: Since there are 350 games spread over 7 days a week, the expected number of games won on each day of the week is 50.

Step 4:

Day of the week	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Sun.
Number games won	61	47	55	38	59	35	55
Expected (e)	50	50	50	50	50	50	50

➤**(c)**.
Step 1: For the above formula we set

$x_1 = 61, x_2 = 47, x_3 = 55, x_4 = 38, x_5 = 59, x_6 = 35, x_7 = 55$

$e_1 = 50, e_2 = 50, e_3 = 50, e_4 = 50, e_5 = 50, e_6 = 50, e_7 = 50,$

$d = 7 - 1 = 6$

Step 2: $\chi^2 = \frac{(61 - 50)^2}{50} + \frac{(47 - 50)^2}{50} + \frac{(55 - 50)^2}{50} + \frac{(38 - 50)^2}{50} +$

$$\frac{(59 - 50)^2}{50} + \frac{(35 - 50)^2}{50} + \frac{(55 - 50)^2}{50} = 12.60$$

Step 3: For d = 7 - 1 = 6, the chi-square table give $\chi^2_{0.10} = 10.60$.

Step 4: Since 12.60 > 10.6 we reject H_o and conclude that at a $\alpha = 0.10$ level of significance, we conclude that the day of the week affects the teams ability to win.

43.4 - Solved Problem 2: A manufacturer of personal computers purchases disk drives from five different companies. According to the claims of these five companies, the following table lists their percentage of defective hard drives:

Companies	**A**	**B**	**C**	**D**	**E**
Percentage of defective hard drives	7%	9%	5%	4%	3%

To monitor these percentages, the company keeps careful records of the defective hard drives for each of the five companies. The following table lists the total number of defective disks:

Companies	**A**	**B**	**C**	**D**	**E**
Number of hard drives received	1,211	2,198	3,658	2587	5,544
Number of defective hard drives	108	154	170	121	158

(a). State H_o and H_a.

(b). Complete the table:

Companies	**A**	**B**	**C**	**D**	**E**
Number of hard drives	108	154	170	121	158
Number of expected defective hard drives					

(c). Compute χ^2 .

For $\alpha = 0.05$, would the null hypothesis be rejected? Explain.

(d). For $\alpha = 0.01$, would the null hypothesis be rejected? Explain.

Solutions:

➤**(a)**.
H_o: The percentage of defective hard drives claimed by their five manufacturers is correct.

H_a: The claim is not correct.

➤**(b)**.
To complete the above table, we need to compute each expected values.

Step 1: To compute the expected values we assume H_o is true.

Step 2: The expected number of defective hard drives is computed in the table below by multiplying row one times row two:

Companies	A	B	C	D	E
Number of hard drives received	1,211	2,198	3,658	2587	5,544
Percentage of defective hard drives	7%	9%	5%	4%	3%
Expected No. of Defective hard drives	84.77	197.82	182.9	103.48	166.32

Step 3: The table below gives the number of defective hard drives and the expected number of hard drives per company:

Companies	A	B	C	D	E
Number of hard drives defective	108	154	170	121	158
Expected No. of Defective hard drives	84.77	197.82	182.9	103.48	166.32

➤**(c)**.
Step 1: For the above formula we set:

$x_1 = 108,\ x_2 = 154,\ x_3 = 170,\ x_4 = 121,\ x_5 = 158$

$e_1 = 84.77,\ e_2 = 197.82,\ e_3 = 182.9,\ e_4 = 103.48,\ e_5 = 166.32$

Step 2:

$$\chi^2 = \frac{(108 - 84.77)^2}{84.77} + \frac{(154 - 197.82)^2}{197.82} + \frac{(170 - 182.9)^2}{182.9} + \frac{(121 - 103.48)^2}{103.48} + \frac{(158 - 166.32)^2}{166.32}$$

≈ 20.33

Step 3: For d = 5 - 1 = 4, the chi-square table gives:

$\chi^2_{0.05} = 9.49$.

Step 4: Since 9.49 < 20.33 we reject H_0 and conclude that at a $\alpha = 0.05$ level of significance, we have reason to believe the percentages of defective hard drives reported by the five companies are not correct.

➤**(d)**.
For d = 5 - 1 = 4, the chi-square table gives

$\chi^2_{0.01} = 13.3$

Step 4: Since 9.49 < 13.3 we reject H_a and conclude that at a $\alpha = 0.01$ level of significance, we have no reason to doubt the percentage of defective drives reported by the five companies.

43.4 - Problem 3: Mr. Jones is the president of a large railroad. He recently claimed that 70% of the time its express trains from Boston to New York City arrives on time. Everyday there are five such trips made by these trains. To test this claim, the arrival times were recorded over 3,000 days.

(a). State H_o and H_a.

(b). Complete the following table:

Number k of trains that arrived on time per day	**0**	**1**	**2**	**3**	**4**	**5**
Number days that k trains arrived on time	5	90	376	1001	1121	407
Expected Number days that k trains arrived on time						

(c). Compute χ^2 .

For $\alpha = 0.025$, would you support the hypothesis? Explain.

(d). For $\alpha = 0.01$, would you support the hypothesis?

Solutions:

➤**(a)**.
H_o: At least seventy percent of the express trains from Boston to New York City arrive on time.

H_a: Reject the claim that at least seventy percent of the express trains from Boston the New York City arrive on time.

➤**(b)**.
Step 1: Assume H_o is true.

Step 2: It is reasonable to assume that the distribution of arriving on time is a binomial distribution:

$P\{X = k\} = \binom{5}{k}(0.7)^k(0.3)^{5-k}$.

Number days that k trains arrived on time	15	98	412	925	1070	480
Number k of trains that arrived on time per day	0	1	2	3	4	5
Expected Number of days that k trains arrived on time $(3000)\binom{5}{k}(0.7)^k(0.3)^{5-k}$	7.29	85.05	396.9	926.1	1080.45	504.21

➤**(c)**.
Step 1: For the above formula we set:

$x_1 = 15$, $x_2 = 98$, $x_3 = 412$, $x_4 = 925$, $x_5 = 1070$, $x_6 = 480$

$e_1 = 7.29$, $e_2 = 85.05$, $e_3 = 396.9$, $e_4 = 926.1$, $e_5 = 1080.45$, $e_6 = 504.21$

Step 2: $$\chi^2 = \frac{(15 - 7.29)^2}{7.29} + \frac{(98 - 85.05)^2}{85.05} + \frac{(412 - 396.9)^2}{396.9} + \frac{(925 - 926.1)^2}{926.1} + \frac{(1070 - 1080.45)^2}{1080.45} + \frac{(480 - 504.21)^2}{504.21} \approx 11.96$$

Step 3: For d = 6 - 1 = 5, the chi-square table give $\chi^2_{0.025} = 12.8$.

Step 4: Since 11.96 < 12.8, we reject H_a and conclude that at a $\alpha = 0.05$ level of significance, we can't reject the claim that at least 70% of the trains arrive on time.

➤**(d)**.
For d = 6 - 1 = 5, the chi-square table give $\chi^2_{0.1} = 9.24$.

Step 5: Since 9.24 < 11.96, we reject H_o and conclude that at a $\alpha = 0.1$ level of significance, we can conclude the claim that at least 70% of the trains arrive on time is false.

Unsolved Problems with Answers

43.4 - Problem 1: A local trucking company leased 7 copy machines for its office. The warranty for each machine states that each machine will be down no more than 5% of the time. Careful records were kept on each machines' down time over a 1,000 hour time period. The following table summarizes these records:

Machine	A	B	C	D	E	F	G
Number hours down	57	60	55	48	59	61	58

(a). State H_o and H_a.

(b). Complete the table

Machine	A	B	C	D	E	F	G
Number hours down	57	60	55	48	59	61	58
Expected number of hours down							

(c).
Compute χ^2 For $\alpha = 0.10$, would the null hypothesis be rejected? Explain.

Answers:

➤**(a)**. H_o: The machines were down no more than 5% of the time.

H_a: The machines were down more than 5% down time.

➤**(b)**.

Machine	A	B	C	D	E	F	G
Number hours down	57	60	55	48	59	61	58
Expected number of hours down	50	50	50	50	50	50	50

➤**(c)**. $\chi^2 = 8.88$.

Since $8.88 < 10.6$ we reject H_a and at $\alpha = 0.10$ level of significance, we conclude that we have no statistical basis for rejecting the leasing company's claim.

⇑ *Refer back to* **43.4 - Example 1 & 43.4 - Solved Problem 1**.

43.4 - Problem 2: For the Presidential election in 1996, the Democratic party took a random survey, in January, of voters concerns on four issues. The following table was a result of this survey:

Issues	Percentage of voters survey
Crime	52%
Decline in family values	31%
Federal Balanced Budget	11%
Environment	6%

Three months later, a second survey of 1,500 voters was taken to see if the above percentages had significantly changed. The following tables is the results of this latest survey:

Issues	Number of voters in survey
Crime	810
Decline in family values	501
Federal Balanced Budget	144
Environment	45

(a). State H_o and H_a.
(b). Complete the table

Issues	Number of voters in survey	Expected Number of voters in survey
Crime	810	
Decline in family values	501	
Federal Balanced Budget	144	
Environment	45	

(c). Compute χ^2.

For $\alpha = 0.05$, would the null hypothesis be rejected? Explain.

(d). For $\alpha = 0.005$, would the null hypothesis be rejected? Explain.

Answers:

➤**(a)**. H_o: The percentage of voters concern on issues has not changed.

H_a: The percentage of voters concern on issues has changed.

➤**(b)**.

Issues	Number of voters in survey	Expected Number of voters in survey
Crime	810	780
Decline in family values	501	465
Federal Balanced Budget	144	165
Environment	45	90

➤**(c)**. $\chi^2 = 29.75$

Since 29.75 > 7.81, reject H_o. For $\alpha = 0.05$, there has been a significant change of opinion.

➤**(d)**. $\chi^2 = 29.75$

Since 29.75 > 12.8, reject H_0. For $\alpha = 0.005$, there has been a significant change of opinion.

⇑ *Refer back to* **43.4 - Example 2 & 43.4 - Solved Problem 2**.

43.4 - Problem 3: After ten years Mrs. Billings has decided to sell her 6 unit bed and breakfast inn. In advertising her property, she claims that she has a vacancy rate of 35%. To support her claim, she provides the following information to perspective buyers:

Number of rooms vacant per day	0	1	2	3	4	5	6
Number of days vacant	266	880	1244	750	305	52	3

(a). State H_o and H_a.

(b). Complete the following table:

Number of rooms vacant per day	0	1	2	3	4	5	6
Number of days vacant	266	880	1244	750	305	52	3
Expected Number of days vacant							

(c). Compute χ^2.

For $\alpha = 0.05$, would you support the hypothesis? Explain.

(d). For $\alpha = 0.01$, would you support the hypothesis?

Answers:

➤**(a)**.
H_o: Her vacancy rate is 35%.

H_a: Her vacancy rate is not 35%.

➤**(b)**.

Number of rooms vacant per day	0	1	2	3	4	5	6
Number of days vacant	266	880	1244	750	305	52	3
Expected Number of days vacant	262.5	854	1148	822.5	332.5	70	7

➤**(c)**. $\chi^2 \approx 24.45$

Since 24.45 > 12.6, we reject H_o and have a statistical basis for rejecting her claim of a vacancy rate of 35%.

➤**(d)**. Since 24.45 > 16.8 we reject H_o and have a statistical basis for rejecting her claim of a vacancy rate of 35%.

⇑ *Refer back to* **43.4 - Example 3 & 43.4 - Solved Problem 3**.

43.5 - Contingency Tables

A contingency table is made up of r rows and c columns of data. Such tables allow us to compare the dependent relationship between data collected from different populations[1]. To make such a determination we use the chi-square formula:

$$\chi^2 = \frac{(x_1 - e_1)^2}{e_1} + \frac{(x_2 - e_2)^2}{e_2} + \frac{(x_3 - e_3)^2}{e_3} + \dots + \frac{(x_N - e_N)^2}{e_N}$$

where N is the number of cells in the table,
x is the data for each cell,

e is the expected values for each cell.

For chi-square the degrees of freedom is $d = (r - 1)(c - 1)$

For hypothesis we always assume the data from the various populations are independent of each other.

43.5 - Example 1: Ms. Jones teaches an introductory course in Statistics. She has been requested, by the administration, to study the success rate of her students that have at least one year of algebra in comparison to those students that do not have this preparation. She decided to use a contingency table to make this comparison:

	Passed Statistics	**Failed statistics**	**Total**
Students with a year of algebra	37	16	53
Students without a year of algebra	28	19	47
Total	65	35	100

(a). State H_o and H_a.

(b). Assume H_o is true, complete the expected frequencies for the following table:

	Passed Statistics	**Failed statistics**	**Total**
Students with a year of algebra			53
Students without a year of algebra			47
Total	65	35	100

(c). Compute χ^2.

(d). For $\alpha = 0.05$, would you reject H_o?

[1]A statistically dependent relationship between data does not necessary imply a causal relationship.

Solutions:

➤**(a)**.
H_o: A year of algebra and passing statistics are statistically independent.

H_a: A year of algebra and passing statistics are statistically dependent.

➤**(b)**.
Assuming H_o is true, we should assume that 65% (65/100) of all students in her class should pass statistics and 35% should fail.

	Passed Statistics	**Failed statistics**	**Total**
Students with a year of algebra	(0.65)53 = 34.45	(0.35)53 = 18.55	53
Students without a year of algebra	(0.65)47 = 30.55	(0.35)47 = 16.45	47
Total	65	35	100

➤**(c)**.
Step 1: For the formula,

$$\chi^2 = \frac{(x_1 - e_1)^2}{e_1} + \frac{(x_2 - e_2)^2}{e_2} + \frac{(x_3 - e_3)^2}{e_3} + \ldots + \frac{(x_N - e_N)^2}{e_N}$$

Step 2: $x_1 = 37$, $x_2 = 16$, $x_3 = 28$, $x_4 = 19$

$e_1 = 34.45$, $e_2 = 18.55$, $e_3 = 30.55$, $e_4 = 16.45$

Step 3: $$\chi^2 = \frac{(37 - 34.45)^2}{34.45} + \frac{(16 - 18.55)^2}{18.55} + \frac{(28 - 30.55)^2}{30.55} + \frac{(19 - 16.45)^2}{16.45} \approx 1.15$$

Step 4: d = (r - 1)(c - 1) = (2-1)(2-1) = 1

➤**(d)**.
For $\alpha = 0.05$, and 1 degree of freedom,

$\chi^2_{0.05} = 3.84$

Since 1.15 < 3.84, there is no statistical reason to reject $H_{o.}$ Therefore, we conclude that a one year algebra background and passing statistics are independent of each other.

Solved Problems

43.5 - Problem 1: Recently a national survey of 1,000 adults was taken. From this survey, data on each person's annual income and educational level was presented in the following table:

	K1 - K12	College Degree	Graduate School	Total
\$0 - \$9,999	111	27	2	140
\$10,000 - \$29,999	208	41	13	262
\$30,000 - \$59,999	107	197	97	401
\$60,000 and over	51	100	46	197
Total	477	365	158	1000

Assume we wish to use this data to decide if their level of education and income are statistically related.

(a). State H_o and H_a.

(b). Assume H_o is true, complete the expected frequencies for the following table:

	K1 - K12	College Degree	Graduate School	Total
\$0 - \$9,999				140
\$10,000 - \$29,999				262
\$30,000 - \$59,999				401
\$60,000 and over				197
Total	477	365	158	100

(c). Compute χ^2.

(d). For $\alpha = 0.01$, would you reject H_o?

Solutions:

➤**(a)**.
H_o: Level of income and education are statically independent.

H_a: Level of income and education are statically dependent.

➤**(b)**.
Assuming H_o is true, we have

14% $\left(\frac{140}{1000} = 14\%\right)$ of people interviewed have incomes less than \$10,000,

26.2% $\left(\frac{262}{1000} = 26.2\%\right)$ of people interviewed have incomes from \$10,000 - \$29,999

40.1% ($\frac{401}{1000}$ = **40.1%)** of people interviewed have incomes from $30,000 - $59,999

19.7% ($\frac{197}{1000}$ = **19.7%)** of people interviewed have incomes $60,000 or more.

	K1 - K12	College Degree	Graduate School	Total
$0 - $9,999	(0.14)477 = 66.78	(0.14)365 = 51.1	(0.14)158 = 22.12	140
$10,000 - $29,999	(0.262)477 = 124.974	(0.262)365 = 95.63	(0.262)158 = 41.396	262
$30,000 - $59,999	(0.401)477 = 191.277	(0.401)365 = 146.365	(0.401)158 = 63.358	401
$60,000 and over	(0.197)477 = 93.969	(0.197)365 = 71.905	(0.197)158 = 31.126	197
Total	477	365	158	1000

➤**(c)**.
Step 1: For the formula,

$$\chi^2 = \frac{(x_1 - e_1)^2}{e_1} + \frac{(x_2 - e_2)^2}{e_2} + \frac{(x_3 - e_3)^2}{e_3} + \ldots + \frac{(x_N - e_N)^2}{e_N}$$

Step 2: $x_1 = 111$, $x_2 = 27$, $x_3 = 2$, $x_4 = 208$, $x_5 = 41$, $x_6 = 13$, $x_7 = 107$ $x_8 = 197$,

$x_9 = 97$, $x_{10} = 51$, $x_{11} = 100$, $x_{12} = 46$
$e_1 = 66.78$, $e_2 = 51.1$, $e_3 = 22.12$, $e_4 = 124.974$, $e_5 = 95.63$, $e_6 = 41.396$,
$e_7 = 191.277$, $e_8 = 146.365$, $e_9 = 63.358$, $e_{10} = 93.969$, $e_{11} = 71.905$,
$e_{12} = 31.126$

Step 3: $$\chi^2 = \frac{(111 - 66.78)^2}{66.78} + \frac{(27 - 51.10)^2}{51.10} + \frac{(2 - 22.12)^2}{22.12} +$$

$$\frac{(208 - 124.974)^2}{124.974} + \frac{(41 - 95.63)^2}{95.64} + \frac{(13 - 41.396)^2}{41.396} +$$

$$\frac{(107 - 191.277)^2}{191.277} + \frac{(197 - 146.365)^2}{146.365} + \frac{(97 - 63.358)^2}{63.358} +$$

$$\frac{(51 - 93.969)^2}{93.969} + \frac{(100 - 71.905)^2}{71.905} + \frac{(46 - 31.969)^2}{31.969} \approx 275.04$$

Step 4: $d = (r - 1)(c - 1) = (4-1)(3-1) = 6$

➤**(d)**.
For $\alpha = 0.01$, and 6 degree of freedom, $\chi^2_{0.01} = 16.8$.

Since 275.04 > 16.8, reject H_o and accept H_a Therefore, we conclude that there is a statistical dependent relationship between income and education.

Unsolved Problems with Answers

43.5 - Problem.1: A large petroleum company recently tested three new gasoline additives in 100 automobiles of a certain model to determine if there is an association between additives and mileage. The following table is a summary of this test:

	0% - 4.99% Mileage increase	5% - 9.99% mileage increase	10% - 14.99% mileage increase	15% mileage increase or more	Total
Additive A	12	11	8	4	35
Additive B	11	9	12	9	41
Additive C	7	5	8	4	24
Total	27	25	28	20	100

(a). State H_o and H_a.

(b). Assume H_o is true, complete the expected frequencies for the following table:

	0% - 4.99% Mileage increase	5% - 9.99% mileage increase	10% - 14.99% mileage increase	15% mileage increase or more	Total
Additive A					35
Additive B					41
Additive C					24
Total	27	25	28	20	100

(c). Compute χ^2.

(d). For $\alpha = 0.05$, would you reject H_o?

Answers:

➤**(a).**
H_o: The association between these additives and gasoline mileage performance are statistically independent.
H_a: The association between these additives and gasoline mileage performance are statistically dependent.

➤**(b).**

	0% - 4.99% Mileage increase	5% - 9.99% mileage increase	10% - 14.99% mileage increase	15% mileage increase or more	Total
Additive A	9.45	8.75	9.80	7.0	35
Additive B	11.07	10.25	11.48	8.2	41
Additive C	6.48	6.0	6.72	4.8	24
Total	27	25	28	20	100

➤**(c)**. $\chi^2 \approx 3.72$

➤**(d)**. For $\alpha = 0.05$, and 6 degree of freedom, $\chi^2_{0.05} = 12.6$.

Since $3.72 < 12.6$, reject H_a and accept H_o. Therefore, we conclude that there is no statistical dependent relationship between these additives and improved gasoline mileage.

⇑ *Refer back to* **43.5 - Example 1 & 43.5 - Solved Problem 1**.

Supplementary Problems

Assume we have a normal population with a sample random distribution X where N is the sample size, s the standard deviation of the sample and σ the standard deviation of the population.

1. From the formula: $\chi^2 = \dfrac{Ns^2}{\sigma^2}$.

Find a formula for s^2 and s.

2. For a given sample size N and α, find a general confidence interval formula for s^2 and s.

3. The Sweet Water Bottling Company has a machine that fills bottles of water. Each morning the variance of the machine is set to $\sigma^2 = 0.1$ ounces. During the day, the vibrations in the operation of the machine can significantly change the variance of the machine. To check for significant changes in variation, a sample of 100 bottles is taken and the sample variance s^2 is recorded. Find a 95% confidence interval for s^2 and s.

4. Assume a sample of size N = 350. Find $\chi^2_{0.05}$.

5. Using the formula:

$$z = \sqrt{2\chi^2} - \sqrt{2d - 1}$$

where z is the standard normal distribution, complete the following table for N = 50:

z =					
χ^2_α =	10	20	30	40	60

A machine drills holes in metal plates. The diameter tolerance of each hole is σ = 0.001 millimeters. Each hour 50 plates are tested for drill accuracy by computing s.

6. State H_o and H_a.

7. Using the formula:

$$\chi^2 = \frac{Ns^2}{\sigma^2}$$

and significant levels $\alpha = 0.02$ and $1 - \alpha = 0.98$, establish a decision rule for stopping the machine.

8. According to the decision rule, If $s^2 = 0.0009$, and $\sigma = 0.001$ would the machine be functioning properly? Explain.

9. Assume the drilling precision changed to $\sigma = 0.0015$ and a sample of $N = 50$ resulted in a standard deviation of $s^2 = 0.0019$. Would the above decision rule shut the machine down.

In a small village in South Africa 250 people became infected with a certain disease. To test the effectiveness of a new drug, half the women and men infected were given the drug while the others infected were given a placebo. The following table gives the final results of this experiment:

Male patients	Recovered	Did not recover	Total
Drug	34	16	50
Placebo	23	27	50
Total	57	43	100

Female Patients	Recovered	Did not recover	Total
Drug	39	36	75
Placebo	38	37	75
Total	77	73	150

10. State H_o and H_a.

11. Compute χ^2.

12. For $\alpha = 0.05$, would you reject H_o?