11.2.1: Facts About the Chi-Square Distribution

The notation for the chi-square distribution is:

$$\chi \sim \chi^2_{df} \tag{11.2.1.1}$$

where $df =$ degrees of freedom which depends on how chi-square is being used. (If you want to practice calculating chi-square probabilities then use $df = n - 1$. The degrees of freedom for the three major uses are each calculated differently.)

For the χ^2 distribution, the population mean is $\mu = df$ and the population standard deviation is

$$\sigma = \sqrt{2(df)}. \tag{11.2.1.2}$$

The random variable is shown as χ^2, but may be any upper case letter. The random variable for a chi-square distribution with k degrees of freedom is the sum of k independent, squared standard normal variables.

$$\chi^2 = (Z_1)^2 + \ldots + (Z_k)^2 \tag{11.2.1.3}$$

a. The curve is nonsymmetrical and skewed to the right.
b. There is a different chi-square curve for each df.

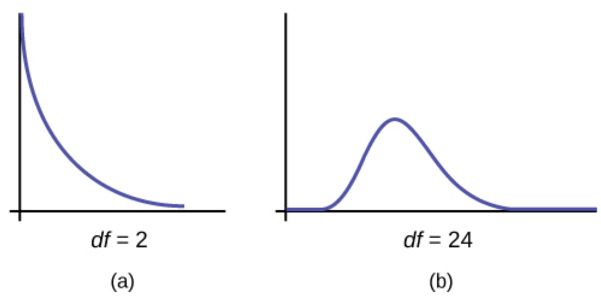

Figure 11.2.1.1

c. The test statistic for any test is always greater than or equal to zero.
d. When $df > 90$, the chi-square curve approximates the normal distribution. For $\chi \sim \chi^2_{1,000}$ the mean, $\mu = df = 1,000$ and the standard deviation, $\mu = \sqrt{2(1,000)}$. Therefore, $X \sim N(1,000, 44.7)$, approximately.
e. The mean, μ, is located just to the right of the peak.

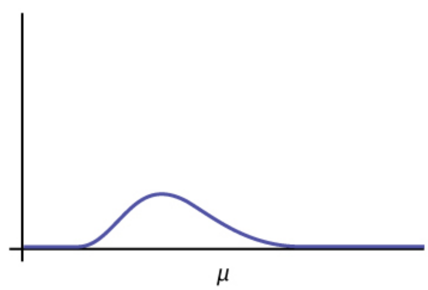

Figure 11.2.1.2

References

1. Data from *Parade Magazine*.
2. “HIV/AIDS Epidemiology Santa Clara County.”Santa Clara County Public Health Department, May 2011.

Review

The chi-square distribution is a useful tool for assessment in a series of problem categories. These problem categories include primarily (i) whether a data set fits a particular distribution, (ii) whether the distributions of two populations are the same, (iii) whether two events might be independent, and (iv) whether there is a different variability than expected within a population.

An important parameter in a chi-square distribution is the degrees of freedom df in a given problem. The random variable in the chi-square distribution is the sum of squares of *df* standard normal variables, which must be independent. The key characteristics of the chi-square distribution also depend directly on the degrees of freedom.

The chi-square distribution curve is skewed to the right, and its shape depends on the degrees of freedom df. For $df > 90$, the curve approximates the normal distribution. Test statistics based on the chi-square distribution are always greater than or equal to zero. Such application tests are almost always right-tailed tests.

Formula Review

$$\chi^2 = (Z_1)^2 + (Z_2)^2 + \ldots + (Z_{df})^2 \tag{11.2.1.4}$$

chi-square distribution random variable

$\mu_{\chi^2} = df$ chi-square distribution population mean

$\sigma_{\chi^2} = \sqrt{2(df)}$ Chi-Square distribution population standard deviation

Exercise 11.2.1.1

If the number of degrees of freedom for a chi-square distribution is 25, what is the population mean and standard deviation?

Answer

mean $= 25$ and standard deviation $= 7.0711$

Exercise 11.2.1.2

If $df > 90$, the distribution is _____________. If $df = 15$, the distribution is ________________.

Exercise 11.2.1.3

When does the chi-square curve approximate a normal distribution?

Answer

when the number of degrees of freedom is greater than 90

Exercise 11.2.1.4

Where is μ located on a chi-square curve?

Exercise 11.2.1.5

Is it more likely the *df* is 90, 20, or two in the graph?

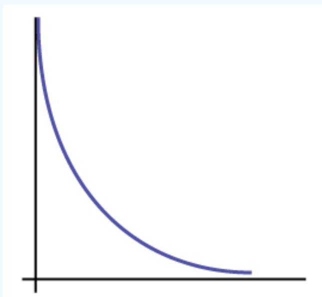

Figure 11.2.1.3.

Answer

$df = 2$