

7.5: Claims on Population Variances - Optional Material

Learning Objectives

- Conduct hypothesis testing on claims regarding population variance using the p -value method on one-tailed tests
- Introduce the critical value method
- Conduct hypothesis testing on claims regarding population variance using the critical value method on two-tailed tests

▮ [Section 7.5 Excel File](#): (contains all of the data sets for this section)

Review and Preview

Having developed hypothesis testing for claims on population means, paired variables, and proportions, we are aware that the process is supported by our understanding of the sampling distributions of particular sample statistics. This remains the case, when considering claims on population variance and standard deviation. Recall that the sample standard deviation is not an unbiased estimator of the population standard deviation but that the sample variance is an unbiased estimator of the population variance. Therefore, to test any claims on a population's standard deviation, we must first translate them into equivalent claims regarding the population's variance, test these new claims, and then translate the results back into the realm of standard deviation.

Once we formulate our hypotheses and collect our evidence, we assess the significance of the evidence using the p -value for one-tailed tests. Difficulties arise in determining the p -value when conducting a two-tailed test; they stem from determining what is equally extreme in the opposite direction when the distribution is not symmetric. We will address this difficulty in more detail later in the section and subsequently develop a different, yet common, approach to hypothesis testing. For the remainder of this introductory section, we will focus on the process for one-tailed tests because of its similarity to all that we have developed.

Recall that when the parent distribution is normal, we transformed the sampling distribution of sample variances into a χ^2 -distribution with $n - 1$ degrees of freedom to compute probabilities. We will need to utilize test statistics when testing claims on population variance. The test statistic is the value produced by mapping the evidence from a particular sample into the common distribution under the assumption that the null hypothesis is true. In assuming the null hypothesis is true, we will have some hypothesized value of the population variance, σ_0^2 , leading to the following test statistic.

$$\chi_{n-1}^2 = \frac{(n-1)}{\sigma_0^2} \cdot s^2$$

With the test statistic in hand, we compute the p -value and make a conclusion based on the comparison between the α value and the p -value. Let us begin testing claims on population variance and standard deviation.

Claims on Population Variance: One-Tailed Tests

In order to conduct hypothesis testing on claims regarding population variance, we will need to have a random sample taken from a normally distributed parent population. As with all hypothesis tests, checking that the requirements of the test are met is important! Let us consider an example situation together.

Many farmers spray their fields to prevent weeds and pests from negatively affecting their harvests. When spraying a field, it is important to get sufficient and even coverage. We need the average ratio of volume to area high enough to meet our needs and the standard deviation to be low enough to imply consistent application.

A company that manufactures sprayers conducted a test on a recently developed prototype to see if it met company standards regarding consistent, even coverage. The company will not produce a sprayer unless the standard deviation is less than a quarter of a gallon per acre. To test the consistency of the sprayer, the prototype sprayed three fields each containing 100 collection devices scattered sporadically throughout the field. When all was said and done, the 300 measurements averaged out to 15.3 gallons per acre with a standard deviation of 0.235 gallons per acre. They formulated the hypothesis test choosing a significance level of 0.10 and the following hypotheses.

$$\begin{aligned}H_0 &: \sigma \geq 0.25 \text{ gallons per acre} \\H_1 &: \sigma < 0.25 \text{ gallons per acre}\end{aligned}$$

This formulation of the hypotheses, however, is not the formulation that the company used in testing because the hypothesis testing needs to be done in the realm of variance, which yields the following set of hypotheses.

$$H_0 : \sigma^2 \geq 0.0625 \text{ gallons per acre}^2$$

$$H_1 : \sigma^2 < 0.0625 \text{ gallons per acre}^2$$

To conduct the hypothesis test, we need that the sample was randomly selected from a parent distribution that is normally distributed. Given the random placement of the 300 collection devices, the sample was randomly chosen. The company felt confident that the distribution was normally distributed based on past history, but they conducted a test on the sample data to see if it was reasonable based on the observed data (recall that such tests exist but are outside of the scope of this course). The test affirmed the reasonableness of the assumption that the parent distribution was normally distributed. So, the hypothesis test could be conducted. Note that the sample variance is 0.055225 square gallons per square acre. We compute our test statistic and produce our visualization.

$$\chi^2_{299} = \frac{(300 - 1)}{0.0625} \cdot 0.055225 \approx 264.1964$$

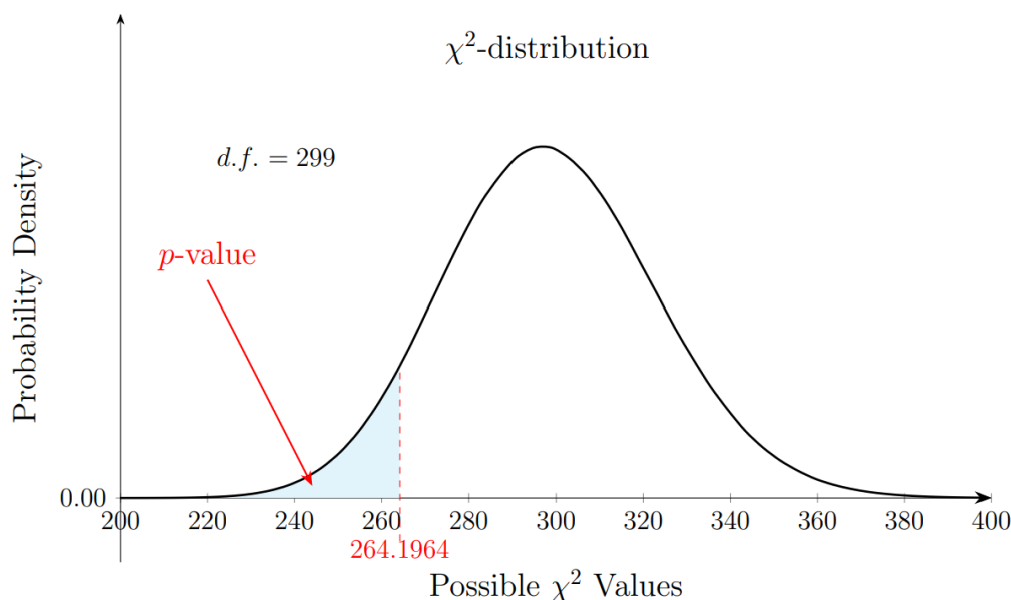


Figure 7.5.1: χ^2 -distribution

We note that the χ^2 -distribution appears to be symmetric as opposed to the asymmetrical appearance we have come to recognize. This is because the sample size is so large. The amount of skew present in χ^2 -distributions decreases as the degrees of freedom increase. From our visualization we compute our p -value in order to conclude the hypothesis test.

$$p\text{-value} \approx \text{CHISQ.DIST}(264.1964, 299, 1) \approx 0.0728$$

Given that the level of significance for this test was 0.10, there is sufficient evidence to reject the null hypothesis. The company can begin to produce the first generation of this prototype sprayer.

? Text Exercise 7.5.1

An amateur game developer is designing a game with AI generated open worlds in hopes of building a game that is essentially endless. The developer does not, however, want the game to become monotonous and has tried to incorporate a great variability between worlds. One of the metrics the developer decided to use to test if the AI is producing enough variability is the distance the first significant encounter occurs from the starting position. The developer does not want the distance to be too long and does not want it to be too consistent. The developer designed the AI to produce worlds so that the average distance is about 550 game paces with a standard deviation of more than 170 game paces.

To make sure the AI was working properly, the developer randomly chose 50 game backers to play randomly chosen AI generated worlds in order to find the distances to the first significant encounter. The sample data was analyzed and was found to have an average distance of 497 game paces and standard deviation of 200 game paces. Test the hypothesis at the 0.05

significance level under the assumption that the distribution of the number of game paces to the first significant encounter is normally distributed.

Answer

We can conduct the hypothesis test because the sample was randomly selected and we were told to assume the parent distribution is normally distributed. The problem is framed within the context of standard deviation; so, we must translate the problem to variance. If the standard deviation is supposed to be more than 170 game paces, the variance would need to be more than $170^2 = 28,900$ square game paces. Since this game is just being developed and tested to see if it is working correctly, we do not want to assume that the population variance is greater than 28,900 square game paces. This helps us to set our hypotheses as follows.

$$H_0 : \sigma \leq 28,900 \text{ game paces}^2$$

$$H_1 : \sigma > 28,900 \text{ game paces}^2$$

We have a right-tailed test. We compute our test statistic using the sample variance and then produce our visualization to help compute the p -value.

$$\chi_{49}^2 = \frac{(50 - 1)}{40,000} \cdot 28,900 \approx 67.8201$$

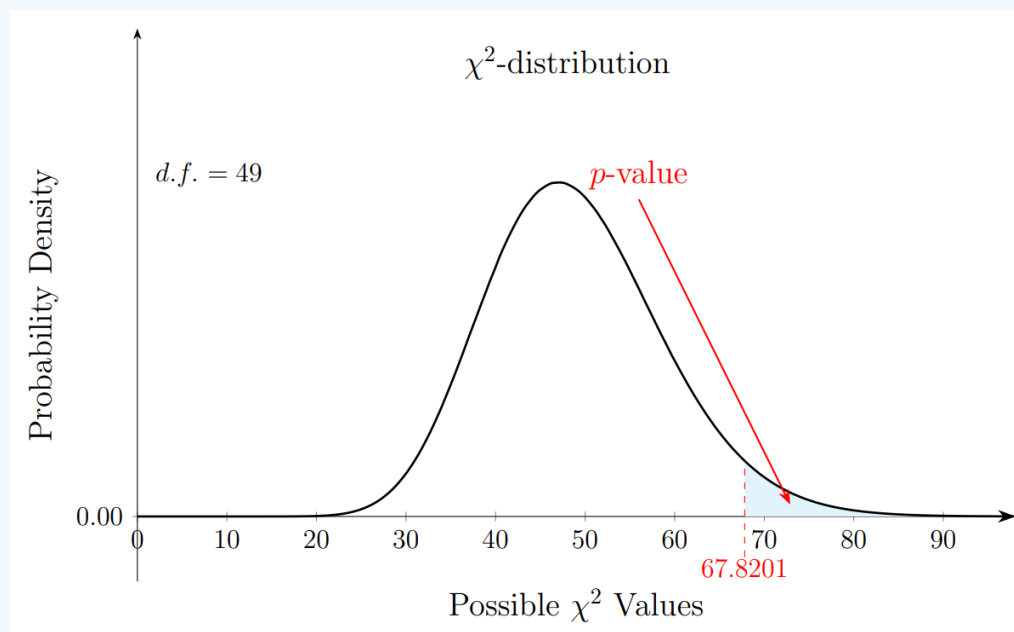


Figure 7.5.2 χ^2 -distribution

$$p\text{-value} \approx 1 - \text{CHISQ.DIST}(67.8201, 49, 1) \approx 0.0387$$

The p -value is smaller than the level of significance; therefore we reject the null hypothesis. This provides sufficient evidence for the developer to assert that the AI is working for the variation in the game. It looks like it may not be meeting specifications regarding the average distance though. That would require a test on means. An interested reader is encouraged to consider how to conduct such a test.

Claims on Population Variance: Two-Tailed Tests

As hinted in the Review and Preview section, we will take a separate approach to conducting two-tailed tests on population variances. This second technique can be applied to the hypothesis tests in general, but we leave such application to the reader. Let us examine why an issue arises with two-tailed tests on population variances. Recall two main ideas: the standard normal distribution and the t -distribution are symmetric about 0 (the expected value of each distribution) and the p -value is the probability of obtaining something at least as extreme as what was observed under the assumption that the null hypothesis is true. In the two-

tailed case, we needed to consider the value of a test statistic equally extreme as the test statistic computed from the observed sample statistic but in the opposite direction. We chose the value that was the same distance away from the mean just with the opposite sign. Given the symmetry of the previous distributions, three facts about the two values coincide: equidistant from the mean, equal probability in the tails, and the heights of the density function at those values match. As it turns out, these serve as three different possibilities for determining the value that would be equally extreme just in the opposite direction. Since the χ^2 -distribution is not symmetric, these three facts do not coincide in the χ^2 -distribution. Arguments can be made for the legitimacy of each possible definition; we leave such discussion for more advanced studies, and instead introduce a method common to many textbooks that can be applied just as easily in this context as in the other contexts considered thus far in the book.

Critical Value Method

The p -value and critical value methods share much in common: the requirements to conduct the hypothesis test, the designation of an α value, and the computation of a test statistic under the assumption that the null hypothesis is true to name a few. The primary difference lies in how to assess the significance of the collected evidence. In the p -value method, we compare the probability of getting something at least as extreme as what was observed to the α value. If the p -value is less than the α value, we have sufficient evidence to reject the null hypothesis. In the critical value method, we determine, based on the α value, what values of the test statistic constitute significant evidence. We segment the distribution of the test statistics into regions based on whether or not we will reject or fail to reject the null hypothesis if the computed test statistic falls in them or not. The regions where we would reject the null hypothesis are called rejection regions. The boundary points of these regions are called critical values, hence the name of the method. We must address how to identify these regions and set their boundaries.

If we are conducting a right-tailed test, we are looking for evidence against the null hypothesis by looking for test statistics far to the right of the expected value. If we are conducting a left-tailed test, we are looking for evidence against the null hypothesis by looking for test statistics far to the left of the expected value. If we are conducting two-tailed tests, we are looking for evidence against the null hypothesis by looking for a test statistic differing from the expected value in either direction. From these thoughts, we identify our rejection regions. If we have a right-tailed test, our rejection region lies in the right tail. If we have a left-tailed test, our rejection region lies in the left tail. And, similarly, if we have a two-tailed test, our rejection region has two components: both the left and right tails.

But how far along the tails must the computed test statistic be in order for us to fall in the rejection region? It depends on the α value. The smaller the α is the farther along the tail we need the computed test statistic to fall. Recall that we can understand the α value as the probability of making a type I error given that the null hypothesis is actually true. Once we have selected a particular α value for a test, that value represents the expected rate of making a type I error if the null hypothesis is true and we repeatedly collect random samples to test the hypothesis. We thus determine the size of our rejection region by setting the probability of a test statistic falling in the region to be the α value. In one-tailed tests, the entire area naturally falls in one tail, but with two-tailed tests, the area must be split between the two tails; each having an area of $\frac{\alpha}{2}$.

We have the following formulation of the critical value method.

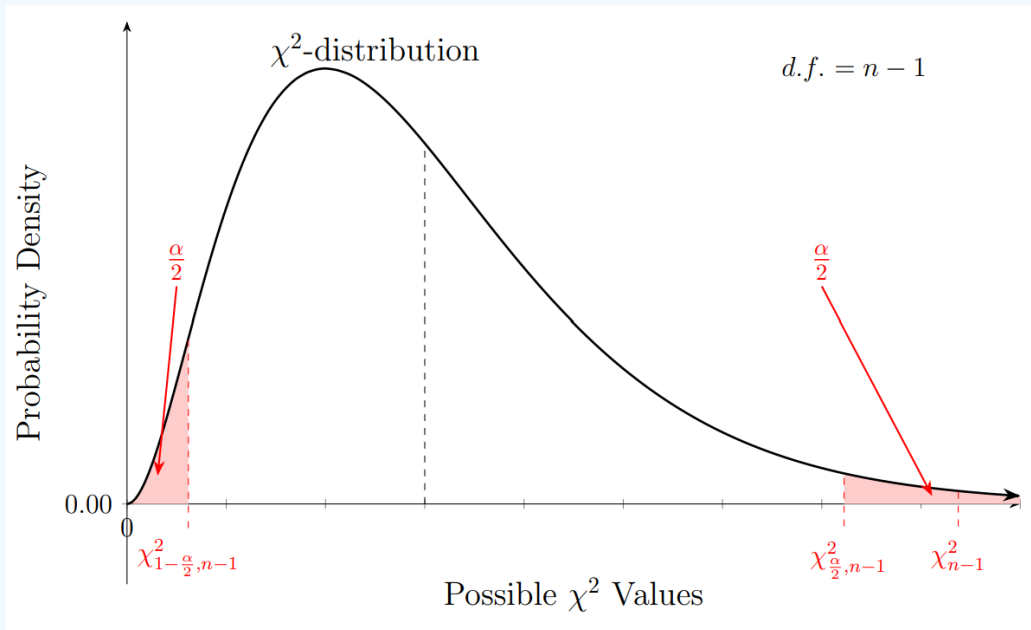
1. Use natural observation, previous experimental results, or the claims of others to formulate a hypothesis that warrants testing.
2. Identify a competing hypothesis. Set the null and alternative hypotheses.
3. Set the α value for this particular hypothesis test
4. Determine the methodology of collecting evidence against the null hypothesis and determine what constitutes sufficient evidence by setting the level of significance. Make sure the design meets the requirements of the tests intended to be conducted.
5. Conduct the experiment and collect the evidence.
6. Compute the test statistic.
7. Use the hypotheses to determine whether a test is a left-tailed, right-tailed, or two-tailed test. Note that the directions match with sign in the alternative hypothesis.
8. Determine the rejection region of the appropriate distribution of the test statistics based on the hypothesis.
9. Determine if the test statistic falls within the rejection region. If so, reject the null hypothesis. If the test statistic falls on the boundary or outside of the rejection region, fail to reject the null hypothesis.

When using the same α value, the critical value method will produce the same conclusions to hypothesis tests as the p -value method when conducting one-tailed tests on claims regarding means, proportions, and variances and when conducted two-tailed tests on claims regarding means and proportions. The p -value method is the most prevalent method in part because simply relaying

the p -value allows the readers to assess the strength of the evidence and to apply their personal thresholds without additional work which is not realistic with the critical value method.

? Text Exercise 7.5.2

When conducting hypothesis tests regarding claims on population variance, there are three forms that the hypotheses can have and there are two conclusions that can be drawn from each form. This yields six total possibilities. The following pictures visualize the implementation of the critical value method for claims on population variance with a random sample of size n taken from a normally distributed parent population. Note that the critical values are denoted using similar notation as the critical values used in constructing confidence intervals for population variance, the rejection region is colored a light red, and the test statistic is denoted χ^2_{n-1} . For each picture, deduce the formulation of the hypotheses and determine the conclusion of the test.



1.

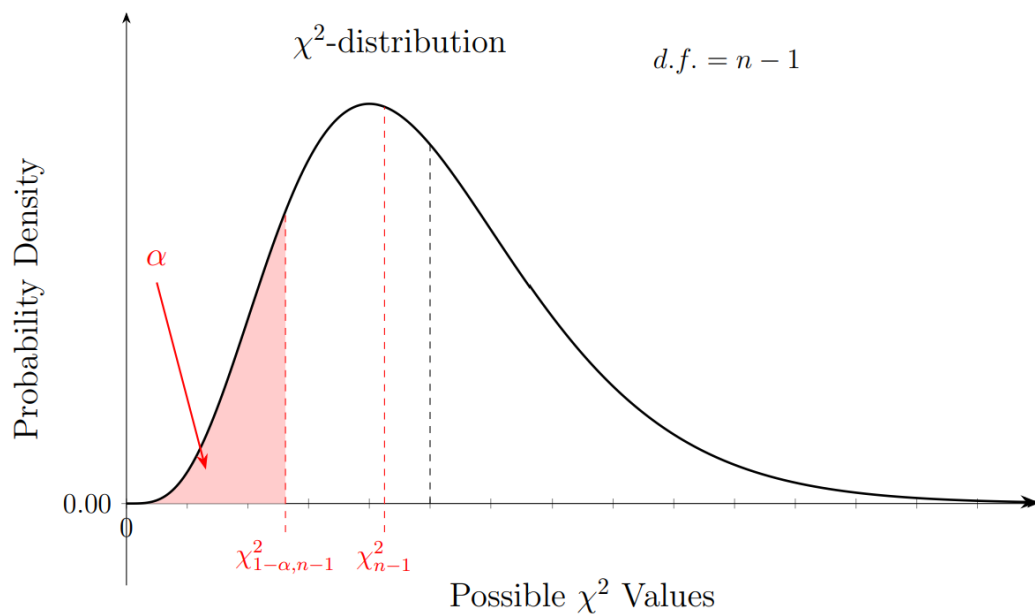
Figure 7.5.3: χ^2 -distribution

Answer

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

$$H_1 : \sigma^2 \neq \sigma_0^2$$

Since the critical value is in the shaded rejection region, we have sufficient evidence to reject the null hypothesis.



2.

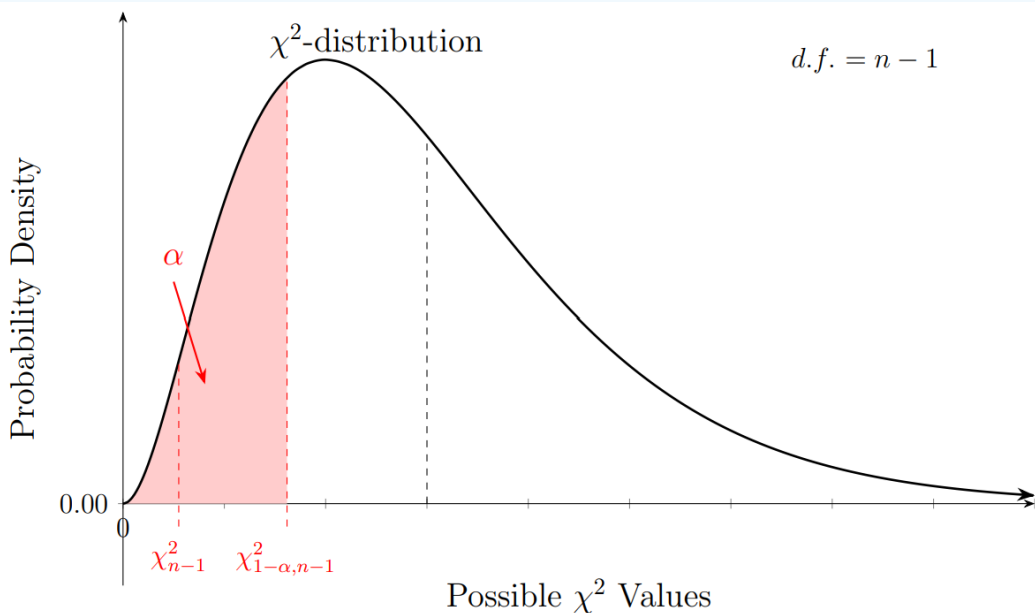
Figure 7.5.4: χ^2 -distribution

Answer

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

$$H_1 : \sigma^2 < \sigma_0^2$$

Since the critical value is not in the shaded rejection region, we do not have sufficient evidence to reject the null hypothesis.



3.

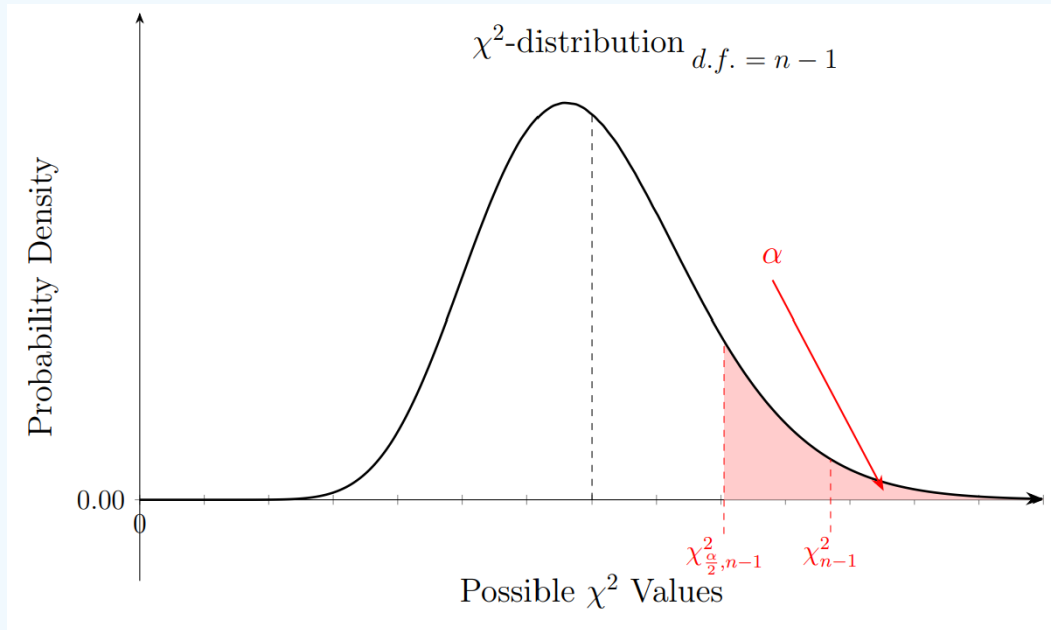
Figure 7.5.5: χ^2 -distribution

Answer

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

$$H_1 : \sigma^2 < \sigma_0^2$$

Since the critical value is in the shaded rejection region, we have sufficient evidence to reject the null hypothesis.



4.

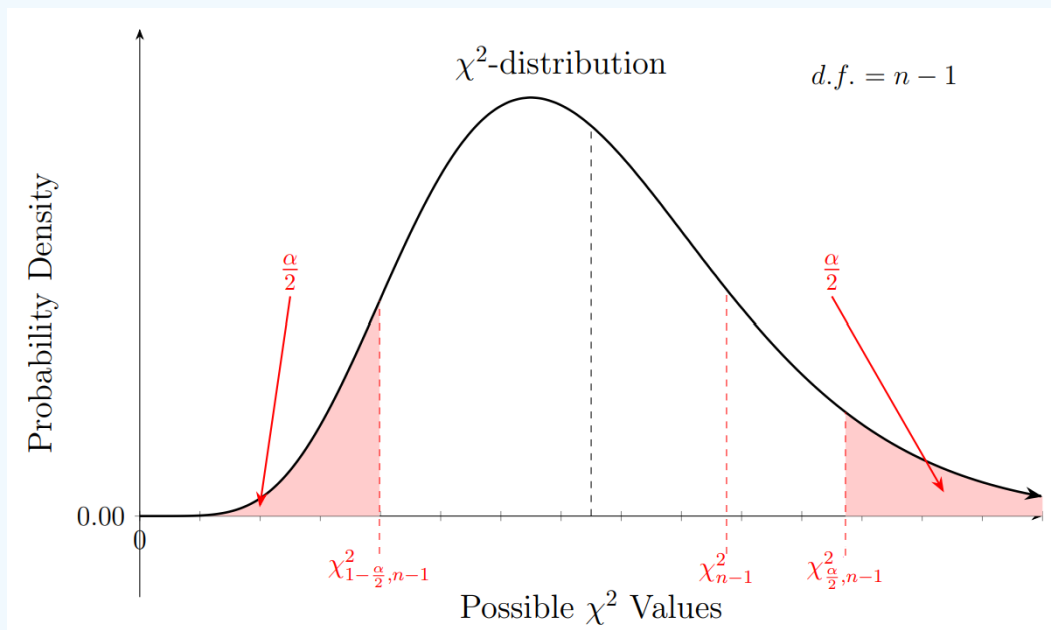
Figure 7.5.6: χ^2 -distribution

Answer

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

$$H_1 : \sigma^2 > \sigma_0^2$$

Since the critical value is in the shaded rejection region, we have sufficient evidence to reject the null hypothesis.



5.

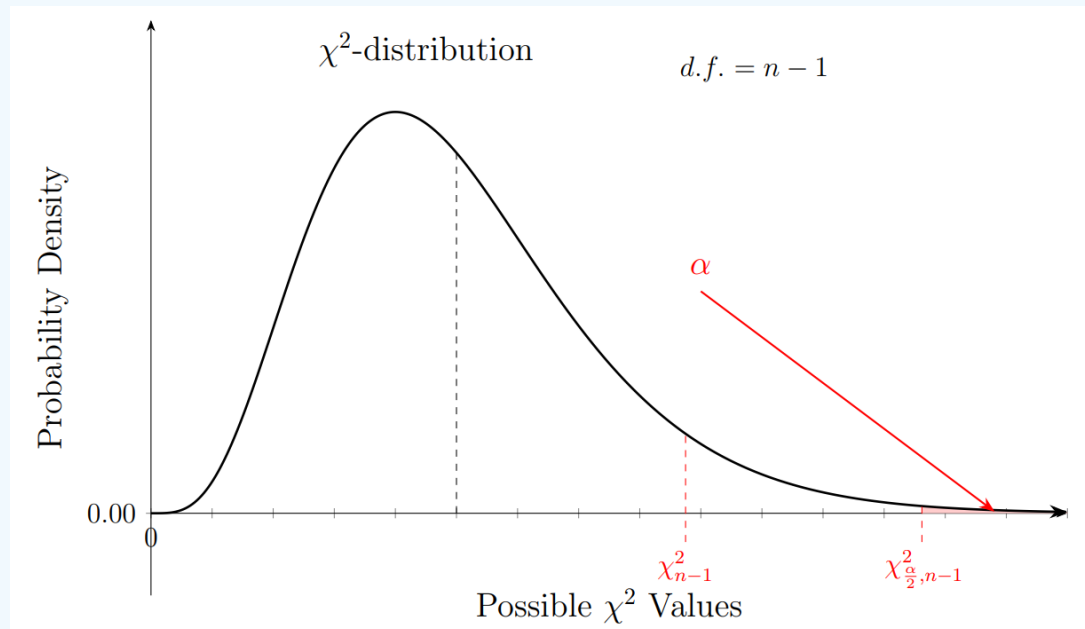
Figure 7.5.7: χ^2 -distribution

Answer

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

$$H_1 : \sigma^2 \neq \sigma_0^2$$

Since the critical value is not in the shaded rejection region, we do not have sufficient evidence to reject the null hypothesis.



6.

Figure 7.5.8: χ^2 -distribution

Answer

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

$$H_1 : \sigma^2 > \sigma_0^2$$

Since the critical value is not in the shaded rejection region, we do not have sufficient evidence to reject the null hypothesis.

? Text Exercise 7.5.3

In [Text Exercise 5.2.3](#), we claimed that the heights of adult females followed a normal distribution with an average height of 64 inches and a standard deviation of 2.5 inches. A researcher thinks that the variation of adult female heights changes with time due to a combination of genetics, nutrition, and lifestyle. The researcher decides to test this claim at a level of significance of 0.01 by randomly sampling 15 adult females. Their heights are reported below. Conduct the test using the critical value method.

59, 59, 61, 62, 63, 63, 64, 64, 65, 66, 68, 69, 69, 69, 70

Answer

The heights of adult females are known to be normally distributed and the sample was randomly selected. We can, therefore, conduct the hypothesis test. Since the researcher is interested in any difference in the variability, we will have a two-tailed test. We do not want to assume that the researcher is correct without evidence. We settle on the following hypotheses.

$$H_0 : \sigma^2 = 6.25 \text{ inches}^2$$

$$H_1 : \sigma^2 \neq 6.25 \text{ inches}^2$$

We now compute the sample variance from the collected data and arrive at $s^2 \approx 13.4952$ square inches. We compute our test statistic.

$$\chi_{14}^2 \approx \frac{15-1}{6.25} \cdot 13.4962 \approx 30.2293$$

In order to compute the critical values, we recall that the degrees of freedom are $n - 1$ and that we must split the α equally between the two tails. This means that only $\frac{0.01}{2} = 0.005$ will be in each tail. We compute the critical values using technology.

$$\chi_{0.005,14}^2 = \text{CHISQ.INV}(0.005, 14) \approx 4.0747$$

$$\chi_{0.995,14}^2 = \text{CHISQ.INV}(0.995, 14) \approx 31.3194$$

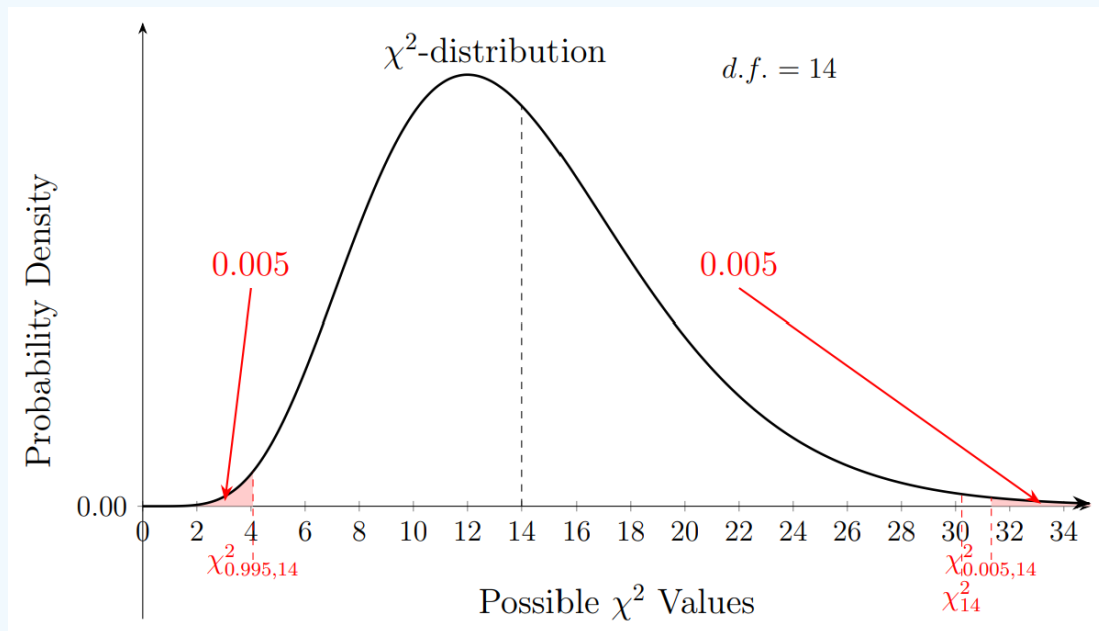


Figure 7.5.9 χ^2 -distribution

The test statistic is greater than the smaller critical value while being smaller than the larger critical value. The test statistic, therefore, falls in the fail to reject region of the distribution of test statistics. We conclude that there is not sufficient evidence to reject the null hypothesis. We cannot affirm the researchers' claims that the variability present in adult female heights is different than it once was with a standard deviation of 2.5 inches.

7.5: Claims on Population Variances - Optional Material is shared under a [Public Domain](#) license and was authored, remixed, and/or curated by The Math Department at Fort Hays State University.

- **3.3: Measures of Central Tendency** by [David Lane](#) is licensed [Public Domain](#). Original source: <https://onlinestatbook.com>.