# 5.4: Sampling Distribution of Sample Variances - Optional Material

**Learning Objectives**

- State the expected value and variance of the sampling distribution of sample variances from a normally distributed parent population
- Discuss transforming the sampling distribution of sample variances to a $\chi^2$-distribution
- Calculate probabilities of sample variances from normally distributed parent populations using $\chi^2$-distributions

## Review and Preview

At this stage, we have a relatively robust understanding of a sampling distribution, but we reiterate it once more within the context of sample variances. For a particular population, the sampling distribution of sample variances for a given sample size $n$ is constructed by considering all possible samples of size $n$ and computing the sample variances for each one. The values of the sample variances are the values that our random variable takes on. We then build the probability distribution with the understanding that the sampling method is simple random sampling. As such, we understand the sample variances as a random variable, which we typically treat as a continuous random variable.

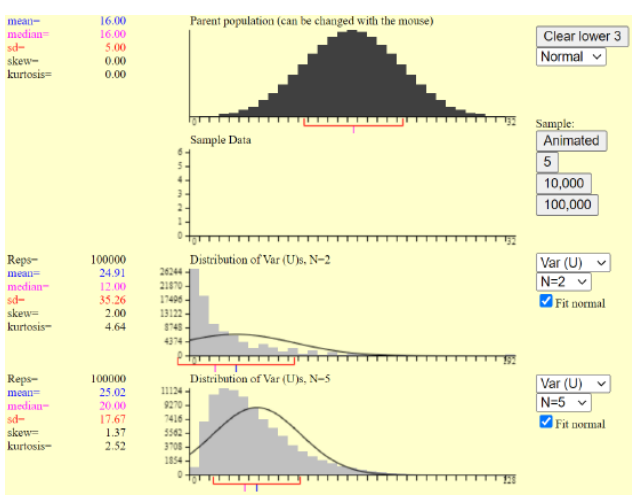
The construction of a sampling distribution is always the same. We have used this process for three sections. When considering sampling distributions of sample means, the Central Limit Theorem asserts that the sampling distribution becomes approximately normal as the sample size increases. This is true for any parent population. The smallest sample size at which the sampling distribution is approximately normal depends on the parent population.

We had something similar with the sampling distributions of sample proportions in the last section; if the sample was large enough to expect at least 10 observations with the given characteristic and at least 10 without the given characteristic, the sampling distribution was approximately normal. This was true for any parent population. Again, the smallest sample size at which the sampling distribution is approximately normal depends on the population proportion, but there is, theoretically, a sample size where it is approximately normal.

One may hypothesize that this is true for any sample statistic and population. Unfortunately, this is not the case when considering the sampling distribution of sample variances.

## Sampling Distribution of Sample Variances

We have seen approximations to the sampling distribution of sample variances in the first section of this chapter. We are now using the same program to thoroughly explore the sampling distribution of sample variances. Using a normal parent population, we simulate the sampling distribution of sample variances through the same progression of sample sizes used in our previous development: $n = 2,\ 5,\ 10,\ 16,\ 20,$ and $25$. We have fit a normal curve to each distribution to emphasize that these sampling distributions are not approximately normal.



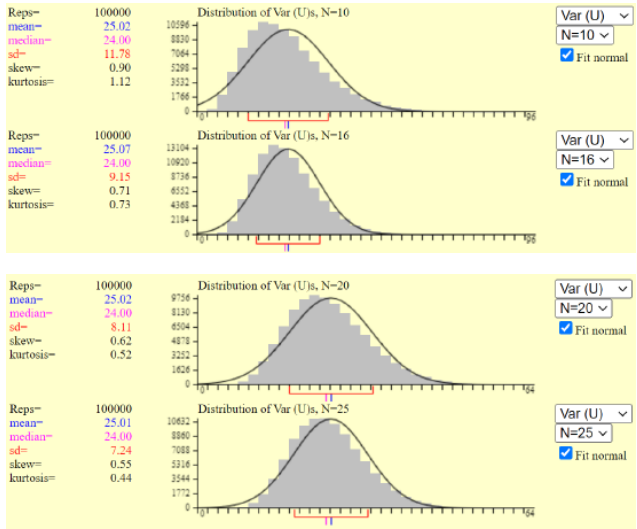



Figure 5.4.1: Sampling Distributions of Sample Variances for various sample sizes

Note that each sampling distribution of sample variances is centered about $25$, the population variance. This happens because sample variance is an unbiased estimator of the population variance. Note that we have only estimated the sampling distribution of sample variances with a single example where the parent population is normal. We previously considered various parent populations. This reduction in scope is no accident; the method we describe works only for normally distributed parent populations. Some methods work for all parent populations, but those are beyond the scope of this course.

We provide, without proof, the expected value and standard deviation of the sampling distribution of sample variances in the case of a parent population that is normal with population standard deviation $\sigma$.

$$\mu_{s^2} = \mathrm{E}(s^2) = \sigma^2 \qquad \sigma_{s^2}^2 = \mathrm{Var}(s^2) = \frac{2\sigma^4}{n-1}$$

Again, we emphasize that the distributions above are not represented well by normal curves. Recall that in chapter 4 we discussed a family of distributions that were positively skewed and followed a similar progression in shape as the degrees of freedom increased. We introduced the $\chi^2$-distributions because they are at play in the sampling distributions of sample variances when the parent population is normally distributed. Consider the progression below (see that the figures above are frequency distributions with various scales, while the figures below are probability density functions all on the same scale).

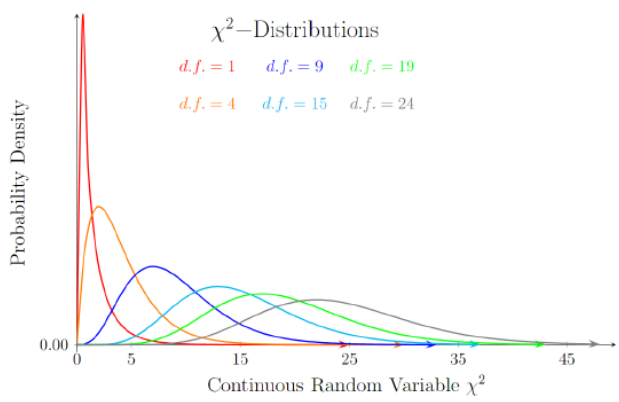



Figure 5.4.2: Chi-Square Distributions with Various Degrees of Freedoms

A rigorous development of the relationship between the sampling distribution of sample variances and the $\chi^2$-distribution requires mathematical tools beyond the scope of this text. We provide only a brief exposition. Recall that the $z$-score can transform any normal distribution with mean $\mu$ and standard deviation $\sigma$ into the standard normal distribution with mean $0$ and standard deviation $1$. This means we can study any normal distribution using the standard normal distribution. A similar situation is at play with the

sampling distribution of sample variances from a normal parent population. We must standardize the distribution and use technology to find the area. We must use a different transformation and probability distribution.

We will introduce the transformation within the familiar context of adult female heights. Recall that adult female heights are normally distributed with a mean of $64$ inches and a standard deviation of $2.5$ inches. We know that the population variance of adult female heights is $2.5^2 = 6.25$ square inches. We consider the sampling distribution of sample variances with a sample size of $10$ and assess the probability of randomly selecting a sample of size $10$ and getting a sample variance between $3$ square inches and $9.5$ square inches, $P(3 < s^2 < 9.5)$. Consider the following figures that illustrate the conversion.

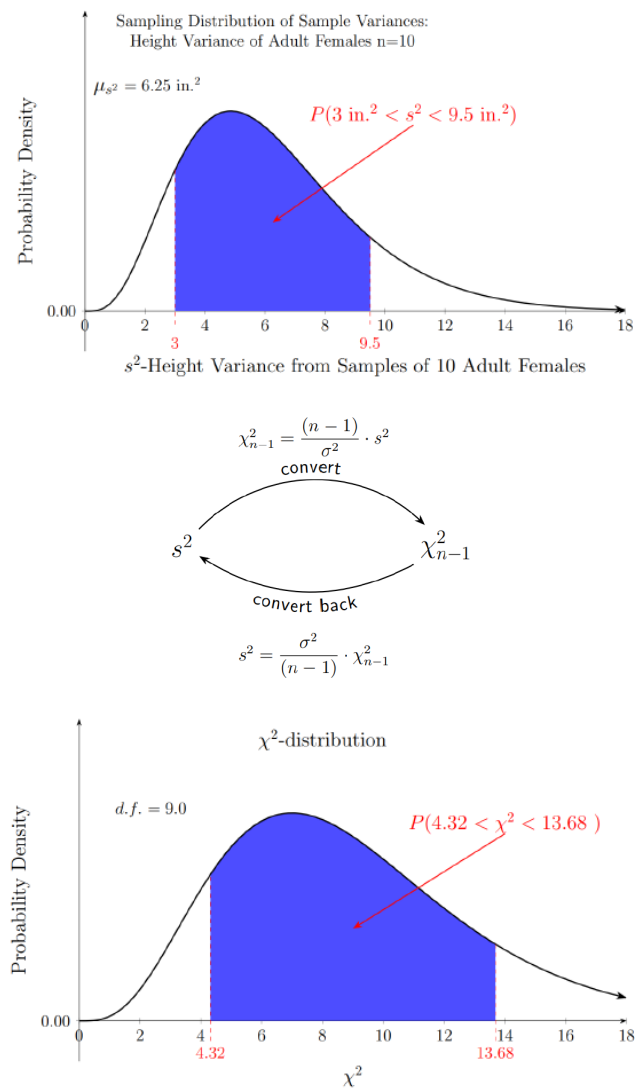



Figure $5.4.3$: Transformation of a sampling distribution of sample variances to $\chi^2$-distribution

Using the transformation $\chi^2_{n-1} = \frac{(n-1)}{\sigma^2} \cdot s^2$ , with $n = 10$ and $\sigma^2 = 6.25$, we transform the sampling distribution of sample variances to the $\chi^2$-distribution with $9$ degrees of freedom, which we sometimes denote using $\chi^2_{10-1} = \chi^2_9$ to emphasize or provide the degrees of freedom. In this context, the degrees of freedom will always be $n-1$, one less than the sample size. We can compute $P(3 < s^2 < 9.5)$ by transforming the interval in terms of $s^2$ to an interval in terms of the $= \chi^2_9$ variable and computing the area using technology, as we did in chapter $4$. Note that $\frac{(10-1)}{6.25} \cdot 3 = 4.32$ and $\frac{(10-1)}{6.25} \cdot 9.5 = 13.68$. So the interval, $3 < s^2 < 9.5$ gets transformed to $4.32 < \chi^2 < 13.68$. It is difficult to tell that these two areas are equal simply from the figure above, but it is indeed the case. We have plotted both distributions using the exact coordinates below. Compare the red and the blue areas to help convince yourself the claim of equal areas is possible/reasonable (for an interested reader, further exploration can be done using this Desmos comparison).

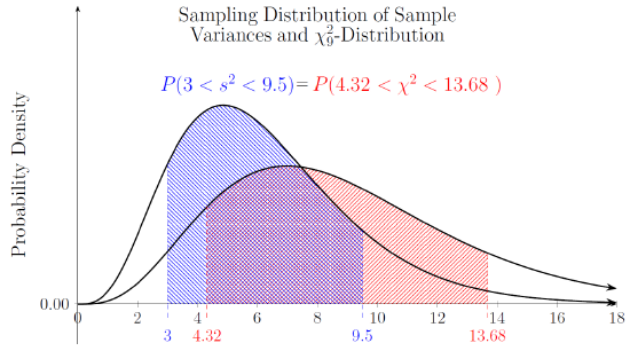



Figure 5.4.4: Sampling distribution of sample variances and $\chi^2$-distribution plotted together to illistrate the preservation of area

We must introduce an accumulation function to calculate the area beneath $\chi^2$-distributions. The function name and syntax may vary depending on the technology. We present a left-tailed accumulation function from Excel: $\text{CHISQ.DIST}$. The syntax in Excel is $= \text{CHISQ.DIST}(x \text{ value, degrees of freedom, cumulative})$. Since we are trying to find areas, we want $\text{cumulative}$ to be marked true using $\text{TRUE}$ or $1$.

$$P(\chi^2_{n-1} < x) = \text{CHISQ.DIST}(x, n-1, 1)$$

With these tools, we may now compute $P(3 < s^2 < 9.5)$.

$$P(3 < s^2 < 9.5) = P(4.32 < \chi^2 < 13.68) = \text{CHISQ.DIST}(13.68, 9, 1) - \text{CHISQ.DIST}(4.32, 9, 1) \approx 75.4945\%$$

**? Text Exercise 5.4.1**

Adult IQ scores are thought to be normally distributed with a mean of $100$ and a standard deviation of $15$. Determine the probability that a random sample of $16$ adults has a sample standard deviation less than $10$.

**Answer**

We will not develop a sampling distribution of sample standard deviations since sample standard deviation is <u>not</u> an unbiased estimator of population standard deviation, even though sample variance is an unbiased estimator of population variance. If we are to proceed, we must translate our prompt into one that considers variance rather than standard deviation. If $s < 10$, then $s^2 < 100$ since standard deviation is non-negative. We are interested in computing $P(s^2 < 100)$ when the population is normally distributed, $n = 16$, and $\sigma^2 = 15^2 = 225$.

Let us transform our probability statement into one about a $\chi^2$-distribution.

$$P(s^2 < 100) = P\left(\frac{n-1}{\sigma^2} \cdot s^2 < \frac{16-1}{225} \cdot 100\right) = P\left(\chi^2_{15} < \frac{20}{3}\right) = \text{CHISQ.DIST}(\frac{20}{3}, 15, 1) \approx 3.3752\%$$

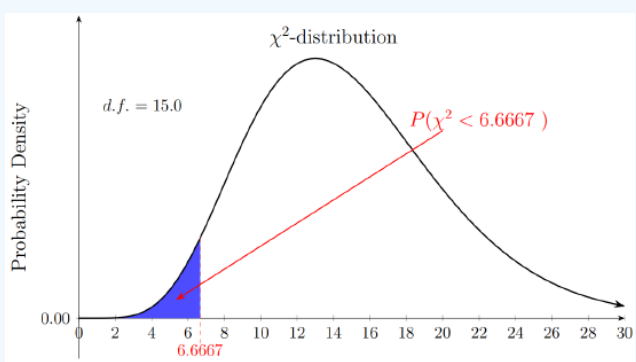



Figure 5.4.5: Sampling distribution of sample variances

5.4: Sampling Distribution of Sample 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.