Frequently Asked Questions

1. Explain with examples why comparison of two variances or standard deviations is more important than the means.

Answer:

Just as a single population variance, the difference of two population variances is sometimes important to an experimenter than means. One might also need to compare the two population variances. We may have to compare the precision of one measuring device with that of another, the stability of one manufacturing process with that of another or even the variability in the grading procedure of one college professor with that of another.

However when we want to compare the variances of two populations rather than looking at the difference of the two sample variances , it turns out to be more convenient if we look at their ratio which in turn gives the measure of efficiency.

For example, a prominent sociologist at a large Midwestern University wants to estimate ratio of variability in the earnings of college graduates to the earnings of those who did not attend college.

2. Write a note on a sampling distribution of ratio of two population variances.

Answer:

To get the sampling distribution of ratio of two population variances the sampling distribution of $(s_1^2 / \sigma_1^2)/(s_2^2 / \sigma_2^2)$ is used. Since the population variances are usually not known, the sample variances are used. The assumptions are that s_1^2 and s_2^2 are computed from independent samples of size m and n, respectively, drawn from two normally distributed populations. If the assumptions are met, $(s_1^2 / \sigma_1^2)/(s_2^2 / \sigma_2^2)$ follows a distribution known as the *F distribution* with *two* values used for degrees of freedom.

3. Briefly explain the characteristics of F-distribution.

Answer:

The confidence interval for the ratio of two variances requires the use of the probability distribution known as the *F*-distribution.

Characteristics of the *F*-Distribution

F-distributions are generally skewed. The shape of an *F*-distribution depends on the values of m and n, the numerator and denominator degrees of freedom, respectively. With the assumption that two independent parent populations are normally distributed samples of sizes m and n are drawn from the two populations respectively. Then, the ratio of two Chi square variates divided by the respective degrees of freedom gives an F variate and the distribution of F variate is known as a F distribution

4. Write a note on F table.

Answer:

One of the primary ways that we will need to interact with an *F*-distribution is by needing to know either (1) an *F*-value, or (2) the probabilities associated with an *F*-random variable, in order to complete a statistical analysis. Let us now explore how to use a typical *F*-table to look up *F*-values and/or *F*-probabilities. In order to refer F table we should know first the degrees of freedom and a level of significance.

Degrees of freedom

The F distribution uses two values for degrees of freedom. The numerator degree of freedom is the value of (m -1) which is used in calculating s_1^2. The denominator degree of freedom is the value of (n -1) which is used in calculating s_2^2.

<u>Level of significance</u>

Level of significance depends on the confidence coefficient

F tables come in denominations based on $F_{1-(\alpha/2)}$ which are $F_{0.995}$, $F_{0.99}$,

F 0.975, F0.95 and F0.90 with one tail. For two tail intervals, the lower boundary, $F_{(\alpha/2)}$, must be calculated to give values of F0.05 ,F0.025 and F0.005 .

5. What are the steps you should follow in using the F-table to find an F-value?

Answer:

Here are the steps one should take in using the *F*-table to find an *F*-value**:**

1. Find the column that corresponds to the relevant numerator degrees of freedom, m
2. Find the row that correspond to the relevant denominator degrees of freedom, m
3. Based on the probability of interest... whether it's **0.01**, **0.025**, **0.05**

Determine the *F*-value where the m column and the nth probability row identified in (2) intersect.

Now, at least theoretically, you could also use the *F*-table to find the probability associated with a particular *F*-value. However, as one can observe, the table is (very!) limited in that direction. Sometimes we may have to interpolate to get the F value for the required degrees of freedom

6. What do you mean by construction of 95% confidence Interval for the ratio of population variances?

Answer:

A confidence interval for the ratio of variances specifies a range of values within which the unknown population parameters, in this case the ratio of variances, may lie. The sample variances are just the values that we compute from a samples of data which gives point estimates of the variances of two populations . It's not done often, but it is certainly possible to compute a confidence interval for the ratio of two independent population variances. The idea of a confidence interval is very general, and we can express the precision of any computed value as a 95% confidence interval (CI).

If you assume that your data were randomly and independently sampled from a Gaussian (Normal) distribution, you can be 95% sure that the confidence interval computed from the sample variances contains the ratio of two true population variances.

7. Derive an interval estimate for the ratio of population variances when the population means are known.

Answer:

Assumptions:

- Populations means are known
- Population is normally distributed
- If population is not normal, use large sample

Suppose $x_1, x_2, \ldots, x_m$ and $y_1, y_2, \ldots, y_n$ are the random samples of size m and n drawn from Normal populations of size N_1 and N_2 with means μ_1 and μ_2 and variance σ_1^2 and σ_2^2 we know that

$X_i \sim N(\mu_1, \sigma_1^2)$ and $Y_i \sim N(\mu_2, \sigma_2^2)$

We know that $\bar{x} = \frac{\sum_{i=1}^{m} x_i}{m}$ and $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$

$\bar{x} \sim N(\mu_1, \sigma_1^2/m)$ and $\bar{y} \sim N(\mu_2, \sigma_2^2/n)$

$$\frac{\sum_{i=1}^{m}(x_i - \mu_1)^2}{\sigma_1^2} \sim \chi^2(m) \quad \text{and} \quad \frac{\sum_{i=1}^{n}(y_i - \mu_2)^2}{\sigma_2^2} \sim \chi^2(n)$$

$$F(m,n)=\frac{\frac{\chi^2(m)}{m}}{\frac{\chi^2(n)}{n}}=\frac{\frac{\sum_{i=1}^{m}(x_i-\mu_1)^2}{\sigma_1{}^2(m)}}{\frac{\sum_{i=1}^{n}(y_i-\mu_2)^2}{\sigma_2^2(n)}}=\frac{n\sum_{i=1}^{m}(x_i-\mu_1)^2}{\sigma_1{}^2}\frac{\sigma_2^2}{m\sum_{i=1}^{n}(y_i-\mu_2)^2}=\frac{\sigma_2^2}{\sigma_1{}^2}\frac{n\sum_{i=1}^{m}(x_i-\mu_1)^2}{m\sum_{i=1}^{n}(y_i-\mu_2)^2}$$

From the table of probabilities of Snedekor's F distribution we can always find two quantities A and B such that

$$P[A\le F(m,n)\le B]=1-\alpha$$

$$P[A\le \frac{\sigma_2^2}{\sigma_1{}^2}\frac{n\sum_{i=1}^{m}(x_i-\mu_1)^2}{m\sum_{i=1}^{n}(y_i-\mu_2)^2}\le B]=1-\alpha$$

$$P[\ A\ \frac{m\sum_{i=1}^{n}(y_i-\mu_2)^2}{n\sum_{i=1}^{m}(x_i-\mu_1)^2}\ \le\ \frac{\sigma_2^2}{\sigma_1{}^2}\ \le\ B\ \frac{m\sum_{i=1}^{n}(y_i-\mu_2)^2}{n\sum_{i=1}^{m}(x_i-\mu_1)^2}\ ] = 1-\alpha$$

$$P[\ \frac{n\sum_{i=1}^{m}(x_i-\mu_1)^2}{Am\sum_{i=1}^{n}(y_i-\mu_2)^2}\ >\ \frac{\sigma_1^2}{\sigma_2{}^2} >\ \frac{n\sum_{i=1}^{m}(x_i-\mu_1)^2}{Bm\sum_{i=1}^{n}(y_i-\mu_2)^2}\ ] = 1-\alpha$$

$$P[\ \frac{n\sum_{i=1}^{m}(x_i-\mu_1)^2}{Bm\sum_{i=1}^{n}(y_i-\mu_2)^2}\ \le\ \frac{\sigma_1^2}{\sigma_2{}^2}\ \le\ \frac{n\sum_{i=1}^{m}(x_i-\mu_1)^2}{Am\sum_{i=1}^{n}(y_i-\mu_2)^2}\ ] = 1-\alpha$$

Therefore, 100 (1- α) % C.I for the ratio of variances of two populations with known means as μ_1 and μ_2 is given by

$$[\ \frac{n\sum_{i=1}^{m}(x_i-\mu_1)^2}{Bm\sum_{i=1}^{n}(y_i-\mu_2)^2}\ ,\ \ \frac{n\sum_{i=1}^{m}(x_i-\mu_1)^2}{Am\sum_{i=1}^{n}(y_i-\mu_2)^2}\ ]$$

8. Derive an interval estimate for the ratio of two unknown population variances when the respective means are unknown.

Answer:

Assumptions:

- Populations means are unknown
- Population is normally distributed
- If population is not normal, use large sample

Suppose $x_1, x_2, \ldots, x_m$ and $y_1, y_2, \ldots, y_n$ are the random samples of size m and n drawn from Normal populations of size N_1 and N_2 with means μ_1 and μ_2 and variance σ_1^2 and σ_2^2 we know that

$X_i \sim N(\mu_1, \sigma_1^2)$ and $Y_i \sim N(\mu_2, \sigma_2^2)$

$$\bar{x} = \frac{\sum_{i=1}^{m} x_i}{m} \qquad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

We know that and

$\bar{x} \sim N(\mu_1, \sigma_1^2/m)$ and $\bar{y} \sim N(\mu_2, \sigma_2^2/n)$

Therefore, $\bar{x} - \bar{y} \sim N(\mu_1 - \mu_2, \sigma_1^2/m + \sigma_2^2/n)$

$$\frac{(m-1)s_1^2}{\sigma_1^2} \sim \chi^2(m-1) \quad \text{and} \quad \frac{(n-1)s_2^2}{\sigma_2^2} \sim \chi^2(n-1)$$

$$F(m-1, n-1) = \frac{\frac{\chi^2(m-1)}{m-1}}{\frac{\chi^2(n-1)}{n-1}} = \frac{\frac{(m-1)s_1^2}{\sigma_1^2(m-1)}}{\frac{(n-1)s_2^2}{\sigma_2^2(n-1)}} = \frac{s_1^2}{\sigma_1^2}\frac{\sigma_2^2}{s_2^2} = \frac{\sigma_2^2}{\sigma_1^2}\frac{s_1^2}{s_2^2}$$

From the table of probabilities of Snedekor's F distribution we can always find two quantities A and B such that

$$P[A \le F(m-1, n-1) \le B] = 1 - \alpha$$

$$P[A \le \frac{\sigma_2^2}{\sigma_1^2}\frac{s_1^2}{s_2^2} \le B] = 1 - \alpha$$

$$P\left[\ A\,\frac{s_2^2}{s_1^2} \le \frac{\sigma_2^2}{\sigma_1^{\,2}} \le B\,\frac{s_2^2}{s_1^2}\ \right] = 1-\alpha$$

$$P\left[\ \frac{s_1^2}{As_2^2} > \frac{\sigma_1^2}{\sigma_2^{\,2}} > \frac{s_1^2}{Bs_2^2}\ \right] = 1-\alpha$$

$$P\left[\ \frac{s_1^2}{Bs_2^2} \le \frac{\sigma_1^2}{\sigma_2^{\,2}} \le \frac{s_1^2}{As_2^2}\ \right] = 1-\alpha$$

Therefore, 100 (1- α) % C.I for the ratio of variances of two populations with unknown means is given by

$$\left[\ \frac{s_1^2}{Bs_2^2}\ ,\ \frac{s_1^2}{As_2^2}\ \right]$$

9. How do you determine constants A and B while obtaining an interval estimate of the ratio of two population variances with known means?

Answer:

The constants A and B which appears in the confidence limits of a CI for the ratio of population variances is determined using F distribution critical values table

- There are two critical values for each level of confidence. The value of $B = F_{\alpha/2}(m,n)$ represents the right-tail critical value and $A = F_{(1-\alpha/2)}(m,n)$ represents the left-tail critical value. Each area in the table represents the region under F curve to the RIGHT of the critical value.
- One can use the critical values $B = F_{\alpha/2}(m,n)$ and $A = F_{(1-\alpha/2)}(m,n)$ to construct confidence intervals for the ratio of population variances or standard deviations when the population means are known where α is determined from the given values of the confidence coefficient and m and n are the number of observations in the sample 1 and sample 2 respectively.

10. How do you determine constants A and B while obtaining an interval estimate of the ratio of two population variances when the population means are unknown?

Answer:

Since the confidence Intervals are determined such that

$P[A \le F(m-1,n-1) \le B\ ] = 1-\alpha$

The area between the constants A and B in a F curve is 1- α. And hence

$P[F < A] = \alpha/2$ and $P[F > B] = \alpha/2$

Therefore, the constants A and B which appears in the confidence limits of a CI for the ratio of two population variances is determined using F -Distribution critical values table for (m-1,n-1) degrees of freedom when the means are unknown

In this case, also there are two critical values for each level of confidence. Each area in the table represents the region under the F- curve to the RIGHT of the critical value. Therefore, value of B is determined such that

P[F>B]= α/2 and the value of A is determined such that

P[F>A]= 1-α/2

Therefore, the value of B= $F_{\alpha/2}(m-1,n-1)$ represents the right-tail critical value and A= $F_{(1-\alpha/2)}(m-1,n-1)$ represents the left-tail critical value.

Hence, we use the critical values B= $F_{\alpha/2}(m-1,n-1)$ and A= $F_{(1-\alpha/2)}(m-1,n-1)$ to construct confidence intervals for the ratio of population variances and standard deviations with unknown population means where α is determined from the given values of the confidence coefficient and m and n are the number of observations in the sample 1 and sample 2 respectively.

11. What are the steps to be followed for the construction of the C.I for the ratio of two population's unknown variances with known means of the populations?

Answer:

Steps to obtain confidence interval for the variance when the mean is known:

- First consider the values of two population means μ_1 and μ_2
- Let m be the sample size drawn from the first population and n be the sample size drawn from the second population
- Select a confidence level. The most common level is 95% or 99%
- Obtain the critical values of F- probabilities for α/2 and (1 - α/2) level of significance with (m ,n) degrees of freedom if μ_1 and μ_2 are known [(m-1, n - 1)degrees of freedom if μ_1 and μ_2 are unknown]

  Let B= $F_{\alpha/2}(m,n)$ and A= $F_{(1-\alpha/2)}(m,n)$

Then, the confidence interval for the ratio of variances when the means are known is computed as

$$\left[\frac{n\sum_{i=1}^{m}(x_i-\mu_1)^2}{Bm\sum_{i=1}^{n}(y_i-\mu_2)^2} \quad , \quad \frac{n\sum_{i=1}^{m}(x_i-\mu_1)^2}{Am\sum_{i=1}^{n}(y_i-\mu_2)^2} \right]$$

- However, when reporting a confidence interval, we have to report both the interval and the confidence level.

12. State the procedure to be followed for the construction of the C.I for the ratio of two population variances when the means of the populations are unknown?

Answer:

Steps to be followed for the construction of the C.I for the ratio of two population variances when the means of the populations are unknown

- First obtain the point estimate of μ_1 and μ_2 , that is, the sample means $\bar{x}$ and $\bar{y}$
- Let m be the sample size drawn from the first population and n be the sample size drawn from the second population
- Select a confidence level. The most common level is 95% or 99%
- When the population mean is unknown Let B= $F_{\alpha/2}(m-1,n-1)$ and A= $F_{(1-\alpha/2)}(m-1,n-1)$.
- Then the confidence interval for the ratio of variances when the population means are unknown is computed as

$$\left[\frac{s_1^2}{Bs_2^2} \quad , \quad \frac{s_1^2}{As_2^2} \right]$$

- When reporting a confidence interval, make sure you report both the interval and the confidence level.

13. An experimenter is concerned that the variability of responses using 2 different experimental procedures may not be the same. Before conducting his research he conducts a pre study with a random sample of 10 and 8 responses and gets s12 =7.14 and s22 =3.21 respectively. Construct 90% CI for σ12 / σ22

Answer:

Therefore, 100 (1- α) % C.I for the ratio of variances of two populations with unknown means is given by

$\left[\frac{s_1^2}{Bs_2^2} \quad , \quad \frac{s_1^2}{As_2^2} \right]$ where s_1^2 =7.14 and s_2^2 =3.21

Given 100(1- α) %=90%. Then α=0.1 and α/2= 0.05

From the table of F – distribution

B= $F_{\alpha/2}$(m-1,n-1) = $F_{0.05}$ (9,7)=3.68 and A = 1/B = 0.2717

Substituting these values into the formula for CI, we get

$$\left[\frac{7.14}{3.68(3.21)}, \frac{7.14}{0.2717(3.21)} \right] = [0.6044, 8.18]$$

14. The feeding habits of two-species of net-casting spiders are studied. The species, the deinopis and menneus, coexist in eastern Australia. The following summary statistics were obtained on the size, in millimeters, of the prey of the two species:

Adult DINOPIS Adult MENNEUS

m=10 n=10

$\bar{x} = 10.26$ mm $\bar{y} = 9.02$ mm

$S_x^2 = (2.51)^2$ $s_y^2 = (1.90)^2$

Estimate, with 95% confidence, the ratio of the two population variances.

Answer:

In order to estimate the ratio of the two population variances, we need to obtain two *F*-values from the *F*-table, namely:

B= $F_{\alpha/2}(m-1,n-1) = F_{0.025}(9,9)=4.03$ and $A = F_{0.975}(9,9)=1/B = 0.2481$

Therefore, 100 (1- α) % C.I for the ratio of variances of two populations with unknown means is given by

$$\left[\frac{s_x^2}{Bs_y^2}, \frac{s_x^2}{As_y^2} \right]$$ where $S_x^2 = (2.51)^2$ and $s_y^2 = (1.90)^2$

Then, the 95% confidence interval for the ratio of the two population variances is:

$$\left[\frac{(2.51)^2}{4.03(1.90)^2}, \frac{(2.51)^2}{0.2481(1.90)^2} \right]$$

Simplifying, we get [0.433, 7.033]

That is, we can be 95% confident that the ratio of the two population variances is between 0.433 and 7.033. (Because the interval contains the value 1, we cannot conclude that the population variances differ.)

15. Among 11 patients in a certain study, the standard deviation of the property of interest was 5.8. In another group of 4 patients, the standard deviation was 3.4. Construct a 95 percent confidence interval for the ratio of the standard deviations of these two.

Answer:

Given: $s_1 = 5.8$, $s_2 = 3.4$, m=11, n=4

Therefore, 100 (1- α) % C.I for the ratio of variances of two populations with unknown means is given by

$$\left[\frac{s_1^2}{Bs_2^2}, \frac{s_1^2}{As_2^2} \right]$$ where $s_1^2 = 33.64$ and $s_2^2 = 11.56$

Given 100(1- α) %=95%. Then α=0.05 and α/2= 0.025

From the table of F – distribution

$B = F_{\alpha/2}(m-1,n-1) = F_{0.025}(10,3) = 14.42$ and $A = F_{1-\alpha/2}(m-1,n-1) = F_{0.975}(10,3) = 1/B = 0.0694$

Substituting these values into the formula for CI, we get

$$\left[\ \frac{33.64}{14.42(11.56)}\ ,\ \frac{33.64}{0.0694(11.56)}\ \right] = [0.2018,\ 41.93]$$

Therefore, 95 % C.I for the ratio of standard deviations of two populations with unknown means is given by [0.4492, 6.4754]