For 31 trials, first has variance of 6 what is the probability to have a variance lower than 4 ?

Suponemos que si la primera prueba dio 6, que esa sera la varianza de referencia.

Cual es la probabilidad que los siguientes pruebas tengan una varianza menor a 4

A picture containing diagram

Description automatically generated

Como se calcula que de 20 es el 90%

from scipy.stats.distributions import chi2

chi2.sf(20,31-1)

output

0.91

A process has a variance of 23.6 what is the probability in 10 samples the variance exceed 50?

X2 = (n -1 )S2  = (10-1) 50 = 19.06 **= > 0.025 2.5 %**

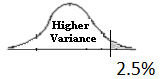
21 23.6

from scipy.stats.distributions import chi2

chi2.sf(19.06,10-1)

output

0.02468753415583865



**A process has a variance of 25, what would be the sample variance that will exceed 99% of the times?**

We want to be alerted when the sample variance exceed a value that should occur only once in 10 times.

X2 (0.01 , 9 ) = 21.7 (Area higher than 99%, to the left )

X2 = (n -1 )S2  => 21.7 = (10-1) S2 => S2 = 60.27 => 99% of the times the variance will be less of 60.27 for samples of 10

21 25

from scipy.stats.distributions import chi2

chi2.sf(60.27,10-1)

Output: 0.0000000011

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Samples (DF1) if DF2 < 3** |  | **F critic**  **(Like DF1)** | **Prob to reject Ho** | **Fail to reject Ho** |
| Decrease |  | Decrease | Increase | Decrease |
| Increase |  | Increase | Decrease | Increase |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Samples (DF1) if DF2 > 2** |  | **F critic**  **(Oposite of DF)** | **Prob to reject Ho** | **Fail to reject Ho** |
| Increase |  | Decrease | Increase | Decrease |
| Deacrease |  | Increase | Decrease | Increase |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Samples (DF2)** |  | **F critic**  **(Oposite of DF2)** | **Prob to reject Ho** | **Fail to reject Ho** |
| Increase |  | Decrease | Increase | Decrease |
| Deacrease |  | Increase | Decrease | Increase |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Alpha error** | **Confidence interval** | **F critic**  **(Oposite of Alpha)** | **Prob to reject Ho** | **Fail to reject Ho** |
| Increase | Decrease | Decrease | Increase | Decrease |
| Decrease | Increase | Increase | Decrease | Increase |

Ho: Both variances are equal / 1 is lower than 1 is numerator, the biggest varianace )

**1 TAIL (TEST WHICH VARIANCE IS LOWER)**

Determine if there is sufficient evidence to that route B is more consistent (1 tail) with 95%

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Route A | Route B |  |  | Bigger |
| Tests | 6 | 8 | Ho  b>= a | F critic ( 0.95, 5, 7) = 3.97 |  |
| Standard deviation | S1 =10.5 | S2 = 5.2 | H1 b < a | F = S1 / S2 = 10.52 / 5.22 = 4 | X: Reject Ho |

There is sufficient evidence to indicate Route B is more consistent, has less deviation

We want to know if there is an improvement after 1 year, (improve if variance is lower) with 95% confidence (1 tail)

We suppose that in the sample of the 1 year the variance is lower.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | At start | One year later |  |  | Bigger |
| Tests | 9 | 7 | Ho 1 <= 2 | F critic ( 0.95, 8, 6) = 4.15 |  |
| Standard deviation | S1 =900 | S2 = 400 | H1 1 > 2 | F = S1 / S2 = 9002 / 4002 = 5 | X: Reject Ho |

There is sufficient evidence to indicate a reduce variance and more consistence after 1 year. (1 year has lower variance)

For a confidence interval of 97.5 % F critic ( 0.975, 8, 6) = 5.6 We fail to reject Ho

As we increase confidence interval, we reduce alpha error, we increase posibility to accept Ho

We suppose that variance in the sample 2 is lower than sample 1 with confidence of 95% (1 tail)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Sample 1** | **Sample 2** |  |  | Bigger |
| Tests | 10 | 9 | Ho 1 <= 2 | F critic ( 0.95, 9, 8) = 3.39 | X: accept Ho |
| Variance | S12 =9 | S22 = 4 | H1 1 > 2 | F = S1 / S2 = 9/4 = 2.25 |  |

Fail to reject Ho. There is sufficient evidence to indicate that in sample 2 the variance is lower

**2 TAILS (TEST IF VARIANCES ARE DIFFERENT)**

Determine whether the following 2 types of rockets have significantly different variances (2 tails) at the 5% level.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Rocket A  **( Bigger Variance )** | Rocket B |  |  | Bigger |
| Tests | 31 | 61 | Ho  a = b | F critic ( **0.975** , 30 , 60) = 1.82 | X: Accept Ho |
| Variance | S1 =2237 | S2 = 1347 | H1 a <> b | F calc = 2,237 / 1347 = 1.66 |  |

Fail to reject Ho. There is NOT sufficient evidence to indicate that variances are different between rockets

We suppose that these variances are different at **90% (2 tails)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Sample 1 | Sample 2 |  |  | Bigger |
| Tests | 10 | 8 | Ho 1 = 2 | F critic ( 0.95,9, 7) = 3.68 | X: accept Ho |
| Variance | S12 =7.14 | S22 = 3.21 | H1 1 <> 2 | F = S1 / S2 = 7.14 / 3.21 = 2.22 |  |

Fail to reject Ho. There is sufficient evidence to indicate the variances are EQUAL in 90%

Calculate the F value and determine the 90% confidence limits for S12 = 50 n1=9 S22=29 n2 = 12



* More DF2 (more samples) lower F value
* Higher confidence interval, lower alpha, higher F value
* Populations must be continuous NOT DISCRETE

Making inferences about a population variance based on a single sample from that population what distribution is used: Chi square

Variation confidence interval is most likely to be non-symmetrical

Chi square test is the inference test that does not require some knowledge of a test of population variation

When evaluation data for goodness of fit to suspected distributions

* The X2 test can be used for hypothesis testing
* In all cases, the data is divided into cells
* It not true that most distributions have the same degrees of freedom,

Which distribution losses the most degrees of freedom NORMAL

* + Normal cells -3
  + Poisson cells -2
  + Binomial cells -2
  + Uniform cells -1

Process A produced 10 defective and 30 good units (40 units)

Process B produced 25 defective out of 60 units

What is the probability that the observed value could under the hypothesis that both processes have same quality

Text

Description automatically generated with low confidence

The probability that both processes are operating at the same quality is between 5% and 10%

Coefficient of contingency



N represent the grand frequency total, not the number of treatment columns

Maximun coeficient of contingency = **√** (( k-1)/k) k min of rows and columns = **√** ((2-1)/2) = **√** (0.5) = 0.7

Correlation coefficient



The expected (theoretical) value for a cell in a contingency table is calculated as **row total \* column total / grand total**

Prob ( good AND B ) = 35 / 100

Prob ( good / B ) = P (good AND B) / P (B) = ( 35/ 100 ) / ( 60/ 100 ) = 35 / 60

Observing for a period of 30 consecutive days, the error rate collected by hour of occurence has X2 =12

* First check if the distribution has the same X2 distribution (skew in one side)
* Kolmogorov-Smirnov analysis to assess the data because the data are ORDINAL (and should be nominal)
* Could it be uniform distribution, not X2 distribution