Frequently Asked Questions

1. What the standard deviation is required for any population?

Answer:

Sample observations may have the same centre but look different because of the way the numbers spread out from the centre. The two distributions may have the same average but there may be a big difference in the way the measurements spread out or vary. Standard deviation is a vital measure for measuring the extent of deviation of the observations from the average. Without standard deviation, it is impossible to compute the standard error of the distribution which is very much important for ant statistical inference.

An estimate of population variance is usually needed before you can make inferences about population means. Sometimes however a population variance σ^2 is the primary objective in an experimental investigation. It may be more important for the experiment than the population mean.

2. Explain with examples why standard deviation is more important than the mean.

Answer:

Consider these examples:

- Scientific measuring instruments must provide unbiased readings with a very small error of measurement. An aircraft altimeter that measures the correct altitude on the average is fairly useless if the measurements are in error by as much as 1000 feet above or below the correct altitude
- Machined parts in a manufacturing process must be produced with minimum variability in order to reduce the out of size and hence defective parts
- Aptitude tests must be designed so that scores will exhibit a reasonable amount of variability. For example, an 800- point test is not very discriminatory if all students score between

3. Write a note on a sample standard variance.

Answer:

It is straightforward to calculate the variance from a sample of values. However, how accurate is that variance? Just by chance, we may have happened to obtain data that are closely bunched together, making the variance low. Or we may have randomly obtained values that are far more scattered than the overall population, making the variance high. The variance of our sample does not equal, and may be quite far from, the variance of the population.

$$s^2 = \frac{\sum_{i=1}^{n}(yi - \overline{y})^2}{n-1}$$ an unbiased estimator of the population variance σ^2

This means that in repeated sampling the average of all our sample estimates will equal the target parameter σ^2

However, how close or far from the target is our estimator s^2 likely to be? To answer this question we use the sampling distribution of $s^{2,}$ which describes it behaviour in repeated sampling.

4. What do you mean by Chi-square probability distribution?

Answer:

Consider the distribution of sample variance s^2 based on repeated random sampling from a Normal distribution with a specified mean and variance. We can show theoretically that the distribution begins at s^2=0, since the variance cannot be negative. With mean equal to σ^2

Its shape is non-symmetric and changes with each different sample size and each different value of σ^2. We can standardize the statistic as we did for Z the standardized statistic for variance is called chi-square = $(n-1)\ s^2 / \sigma^2$

Is called a Chi- square variable and has a sampling distribution called the Chi-square probability distribution with n-1degrees of freedom.

5. How wide should be the confidence interval of the standard deviation?

Answer:

The answer depends for the above question on sample size (n). With small samples, the interval is quite wide as shown in the table below.

n	**95% CI of SD**
2	0.45*SD to 31.9*SD
3	0.52*SD to 6.29*SD
5	0.60*SD to 2.87*SD
10	0.69*SD to 1.83*SD
25	0.78*SD to 1.39*SD

6. What do you mean by construction of 95% confidence Interval for variance?

Answer:

Practical problems very often lead to estimation of σ^2, the variance of the population. A confidence interval for a variance specifies a range of values within which the unknown population parameter, in this case the variance, may lie. The sample SD is just a value that we compute from a sample of data. It is not done often, but it is certainly possible to compute a CI for a SD. Interpreting the CI of the SD is straightforward 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 distribution, we can be 95% sure that the CI computed from the sample SD contains the true population SD.

7. Derive an interval estimate for the variance of the population when the population mean is known.

Answer:

Suppose $y_1, y_2, \ldots, y_n$ are the random samples drawn from a population of size N with a known mean μ and an unknown variance σ^2 we know that

We know $\mu = \dfrac{\sum_{i=1}^{N} Y_i}{N}$ and $\bar{y} = \dfrac{\sum_{i=1}^{n} y_i}{n}$

When yi ~ $N(\mu, \sigma^2)$

$$Z_i = \frac{(\text{yi} - \mu)}{\sigma} \sim N(0,1)$$ and we know that

$$\sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$$

$$\chi^2 = \frac{\sum_{\text{i}=1}^{\text{n}} (\text{yi} - \mu)^2}{\sigma^2} \sim \chi^2(n)$$

We can use a chi-square distribution to construct a confidence interval for the variance and standard deviation.

From the table of probabilities of Chi- square variables for n degrees of freedom we can always find two quantities A and B such that

$P[A \le \chi^2(n) \le B] = 1 - \alpha$

$$P\left[A \le \frac{\sum_{i=1}^{n} (yi-\mu)^2}{\sigma^2} \le B\right] = 1-\alpha$$

$$P\left[1/A > \frac{\sigma^2}{\sum_{i=1}^{n} (yi-\mu)^2} > 1/B \right)\left.\right] = 1-\alpha$$

$P[\sum(yi-\mu)^2/ A > \sigma^2 > \sum(yi-\mu)^2/ B] = 1-\alpha$

$P[\sum(yi-\mu)^2/ B < \sigma^2 < \sum(yi-\mu)^2/ A] = 1-\alpha$

Therefore 100 (1- α) % C.I for the population variance σ^2 when the mean is known as μ is given by

$[\sum(yi-\mu)^2/ B , \sum(yi-\mu)^2/ A]$

8. Derive an interval estimate for the unknown variance of the population when the mean is unknown.

Answer:

Suppose $y_1,y_2,\ldots,y_n$ are the random samples drawn from a population of size N with mean μ and variance σ^2 we know that

We know $\mu = \frac{\sum_{i=1}^{N} Y_i}{N}$ and $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ $s^2 = \frac{\sum_{i=1}^{n} (yi-\bar{y})^2}{n-1}$

$$Z_i = \frac{(yi-\mu)}{\sigma} \sim N(0,1)$$

Since mean μ is unknown we substitute its point estimate y bar then we know that $\sum_{i=1}^{n} Z_i^2 \sim \chi^2(n-1)$

$$\chi^2 = \frac{\sum_{i=1}^{n} (yi-\bar{y})^2}{\sigma^2} \sim \chi^2(n-1)$$

From the table of probabilities of Chi- square variables for (n -1)degrees of freedom we can always find two quantities A and B such that

$P[A \le \chi^2(n-1) \le B] = 1-\alpha$

$$P\left[A \le \frac{\sum_{i=1}^{n} (yi - \bar{y})^2}{\sigma^2} \le B\right] = 1- \alpha$$

$$P\left[1/A > \frac{\sigma^2}{\sum_{i=1}^{n} (yi - \bar{y})^2} > 1/B\ )\ \right] = 1- \alpha$$

$P[\sum(yi- \bar{y})^2 / A > \sigma^2 > \sum(yi- \bar{y})^2 / B] = 1- \alpha$

$P[\sum(yi- \bar{y})^2 / B < \sigma^2 < \sum(yi- \bar{y})^2 / A] = 1- \alpha$

Therefore, 100 (1- α) % C.I for the population variance σ^2 when the mean is unknown is given by

$[\sum(yi- \bar{y})^2 / B\ ,\ \sum(yi- \bar{y})^2 / A]$

9. How do you determine constants A and B while obtaining an interval estimate of the population variance with known mean?

Answer:

The constants A and B, which appears in the confidence limits of a CI for population variance is determined using Chi square critical values table

- There are two critical values for each level of confidence. The value of B= $\chi^2_{\alpha/2}(n)$ represents the right-tail critical value and A= $\chi^2_{(1-\alpha/2)}(n)$ represents the left-tail critical value. Each area in the table represents the region under the chi-square curve to the RIGHT of the critical value.
- One can use the critical values B= $\chi^2_{\alpha/2}(n)$ and A= $\chi^2_{(1-\alpha/2)}(n)$ to construct confidence intervals for a population variance and standard deviation with a known population mean where α is determined from the given values of the confidence coefficient and n is the number of observations

10. How do you determine constants A and B while obtaining an interval estimate of the population variance with unknown mean?

Answer:

Since the confidence Intervals are determined such that

$P[A \le \chi^2(n-1) \le B] = 1- \alpha$

The area between the constants A and B hi-square curve is 1- α. And hence $P[\chi^2<A]= \alpha/2$ and $P[\chi^2>B]= \alpha/2$

Therefore, the constants A and B which appears in the confidence limits of a CI for population variance is determined using Chi square critical values table for (n-1) degrees of freedom when mean is 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 chi-square curve to the RIGHT of the critical value. Therefore value of B is determined such that P[χ^2>B]= α/2 and the value of A is determined such that

P[χ^2>A]= 1-α/2

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

Hence, we use the critical values B= $\chi^2_{\alpha/2}(n-1)$ and A= $\chi^2_{(1-\alpha/2)}(n-1)$ to construct confidence intervals for a population variance and standard deviation with an unknown population mean where α is determined from the given values of the confidence coefficient and n is the number of observations

11. What are the steps to be followed for the construction of the C.I for the unknown population variance with known mean of the population?

Answer:

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

- First consider the value of population mean μ
- If *n* is large, then the Central Limit Theorem can be used and y bar is normally distributed with mean μ and standard deviation σ/√n
- Select a confidence level. The most common level is 95% or 99%
- Obtain the critical values of Chi square probabilities for α/2 and
- (1 - α/2) level of significance with n degrees of freedom if μ is known (n - 1)degrees of freedom if μ is unknown)
- Let B= $\chi^2_{\alpha/2}(n)$ and A= $\chi^2_{(1-\alpha/2)}(n)$
- Then the confidence interval for variance when mean is known is computed as [[∑(yi-μ)2 / B , ∑(yi-μ)2 / A]

- But when reporting a confidence interval, you 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 unknown population variance with an unknown mean of the population?

Answer:

Steps to obtain confidence interval for variance when the mean is unknown

- First obtain the point estimate of μ, that is, the sample mean y bar
- Select a confidence level. The most common level is 95% or 99%
- When the population mean is unknown Let B= $\chi^2_{\alpha/2}(n-1)$ and A= $\chi^2_{(1-\alpha/2)}(n-1)$.
- Then the confidence interval for variance when mean is unknown is computed as [∑(yi- $\bar{y}$)2 / B , ∑(yi- $\bar{y}$)2 / A] which is also equal to [$(n-1)s^2$ / B , $(n-1)s^2$ / A]

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

13. An experimenter is convinced that her measuring instrument had variability. During the experiment, she recorded the measurements 4.1, 5.2 1nd 10.2.Construct 90% confidence interval to estimate the true value of the population variance.

Answer:

Given n=3, and (1-α) = 90%=0.90 then α= 1-0.90 =0.10

Then α/2=0.05 and 1- α/2=0.95

From the table of Chi square distribution B= $\chi^2_{0.05}(2) =$ 5.99147

And A= $\chi^2_{0.95}(2) =$ 0.102587

s^2 =10.57

Hence 100 (1- α) % C.I for the population variance σ^2 when the mean is unknown as μ is given by

[$(n-1)s^2$ / B , $(n-1)s^2$ / A]
=[(3-1) 10.57/5-99147,(3-1)10.57/0.102587]

[3.53, 206.07]

Therefore, with 90% confidence, we can say that the population variance is between 3.53and 206.07. This very wide CI indicates how little information on the population variance is obtained from a sample of only three measurements.

14. A random sample of size 25 from the population gives the sample standard deviation to be 8.5. Obtain 95% Interval estimate for the population Standard deviation.

Answer:

Given, n=25, and s =8.5 , s^2 =72.25

Hence 100 (1- α) % C.I for the population variance σ^2 when the mean is unknown is given by

[$(n-1)s^2$ / B, $(n-1)s^2$ / A]

100 (1- α) %=95%

1 -α = 0.95 which implies α=0.05, α/2=0.025 and 1-α/2=.975

B= $\chi^2_{\alpha/2}(n-1) = \chi^2_{0.025}(24) =$ 39.3641

$$A = \chi^2_{(1-\alpha/2)}(n-1) = \chi^2_{(0.975)}(24) = 12.4011$$

By substituting the given values in the above we get

[(25-1) 72.25/39.3641, (25-1) 72.25/12.4011]

=[44.05,139.83]

The 95% confidence interval for the population standard deviation is between 0.54 and 11.83

15. A random sample of 10 students is selected from a P.U College and their weights are noted down. Set up 99% Confidence Interval estimate for the standard deviation of the weights of the students. 38,40,45,53,47,43,55,48,52,49

Answer:

Given, n=10, $\bar{y} = \frac{\sum yi}{n} = 47$

For the given data $\sum(yi- \bar{y})^2$ =280= (n-1) s^2

Hence, 100 (1- α)% C.I for the population variance σ^2 when the mean is unknown is given by

[$(n-1)s^2$ / B, $(n-1)s^2$ / A]

100 (1- α) %=99%

1 -α = 0.99 which implies α=0.01, α/2=0.005 and 1-α/2=.995

B= $\chi^2_{\alpha/2}(n-1) = \chi^2_{0.005}(9) =$ 23.5893

$$A = \chi^2_{(1-\alpha/2)}(n-1) = \chi^2_{(0.995)}(9) = 1.734926$$

By substituting the given values in the above, we get

[280/23.5893, 280/1.734926]

=[11.87,161.39]

The 99% confidence interval for the population standard deviation is between 3.45 and 12.70.