CSL003P1M : Probability and Statistics Lecture 40 (Estimation-III (Interval Estimates-II))

Sumit Kumar Pandey

December 14, 2021

Confidence Intervals for the Variance of a Normal Distribution

If $X_1, \dots, X_n$ is a sample from a normal distribution having unknown parameters μ and σ^2, then we can construct a confidence interval for σ^2 by using the fact that

$$(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Hence,

$$P\left\{\chi^2_{-\alpha/2,n-1} \leq (n-1)\frac{S^2}{\sigma^2} \leq \chi^2_{\alpha/2,n-1}\right\} = 1-\alpha$$

or, equivalently,

$$P\left\{\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right\} = 1-\alpha$$

Confidence Intervals for the Variance of a Normal Distribution

Hence when $S^2 = s^2$, a $100(1-\alpha)$ percent confidence interval for σ^2 is

$$\left\{ \frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} \right\}$$

Exercise-1

A standardized procedure is expected to produce washers with very small deviation in their thickness. Suppose that 10 such washers were chosen and measured. If the thickness of these washers were, in inches,

0.123	0.133
0.124	0.125
0.126	0.128
0.120	0.124
0.130	0.126

what is a 90 percent confidence interval for the standard deviation of the thickness of a washer produced by this procedure?

Exercise-1

A computation gives that

$$S^2 = 1.366 \times 10^{-5}$$

Because $\chi^2_{0.05,9} = 16.917$ and $\chi^2_{0.95,9} = 3.334$ and because

$$\frac{9 \times 1.366 \times 10^{-5}}{16.917} = 7.267 \times 10^{-6}, \quad \frac{9 \times 1.366 \times 10^{-5}}{3.334} = 36.875 \times 10^{-6}$$

it follows that, with confidence 90 percent,

$$\sigma^2 \in (7.267 \times 10^{-6}, 36.875 \times 10^{-6})$$

Taking square root yields that, with confidence 90 percent,

$$\sigma \in (2.696 \times 10^{-3}, 6.072 \times 10^{-3})$$

Estimating the Difference in Means of Two Normal Populations

Let $X_1, X_2, \ldots, X_n$ be a sample of size n from a normal population having mean μ_1 and variance σ_1^2 and let $Y_1, Y_2, \ldots, Y_m$ be a sample of size m from a different normal population having mean μ_2 and σ_2^2 and suppose that the two samples are independent of each other. We are interested in estimating $\mu_1 - \mu_2$.

Since $\bar{X} = \sum_{i=1}^{n} X_i/n$ and $\bar{Y} = \sum_{i=1}^{m} Y_i/m$ are the maximum likelihood estimators of μ_1 and μ_2, it seems intuitive (and can be proven) that $\bar{X} - \bar{Y}$ is the maximum likelihood estimator of $\mu_1 - \mu_2$.

Estimating the Difference in Means of Two Normal Populations

To obtain a confidence interval estimator, we need the distribution of $\bar{X} - \bar{Y}$. Because

$$\bar{X} \sim \mathcal{N}(\mu_1, \sigma_1^2/n)$$
$$\bar{Y} \sim \mathcal{N}(\mu_2, \sigma_2^2/m)$$

it follows from the fact that the sum of independent normal random variables is also normal, that

$$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

Hence assuming σ_1^2 and σ_2^2 are known, we have that

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim \mathcal{N}(0, 1)$$

Estimating the Difference in Means of Two Normal Populations

Let us suppose now that we again desire an interval estimator of $\mu_1 - \mu_2$ but that the population variances σ_1^2 and σ_2^2 are unknown. In this case, it is natural to try to replace σ_1^2 and σ_2^2 by the sample variances

$$S_1^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$$

$$S_2^2 = \sum_{i=1}^{m} \frac{(Y_i - \bar{Y})^2}{m-1}$$

Thus, it is natural to base our interval estimate on something like

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n + S_2^2/m}}$$

Estimating the Difference in Means of Two Normal Populations

- However, to utilize the foregoing to obtain a confidence interval, we need its distribution and it must not depend on any of the unknown parameters σ_1^2 and σ_2^2.
- **Unfortunatley**, this distribution is both complicated and does indeed depend on the unknown parameters σ_1^2 and σ_2^2.
- In fact, it is only the special case when $\sigma_1^2 = \sigma_2^2$ that we will be able to obtain an interval estimator.
- So, let us suppose that the population variances, though unknown, are equal and let σ^2 denote their common value.

Estimating the Difference in Means of Two Normal Populations

Note that

$$(n-1)\frac{S_1^2}{\sigma^2} \sim \chi^2_{n-1}$$

and

$$(m-1)\frac{S_2^2}{\sigma^2} \sim \chi^2_{m-1}$$

- Also, because the samples are independent, it follows that these two chi-square random variables are independent.
- Hence, from from the additive property of chi-square random variables, which states that the sum of independent chi-square random variables is also chi-square with a degree of freedom equal to the sum of their degrees of freedom.

Estimating the Difference in Means of Two Normal Populations

$$(n-1)\frac{S_1^2}{\sigma^2} + (m-1)\frac{S_2^2}{\sigma^2} \sim \chi^2_{n+m-2}$$

Also, since

$$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}\right)$$

we see that

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} \sim \mathcal{N}(0,1)$$

Estimating the Difference in Means of Two Normal Populations

- Now it follows from the fundamental result in normal sampling $\bar{X}$ and S^2 are independent, that $\bar{X}_1$, S_1^2, $\bar{X}_2$, S_2^2 are independent random variables.
- Hence, use the definition of t-random variable (as the ratio of two independent random variables, the numerator being a standard normal and the denominator being the square root of a chi-square random variable divided by its degree of freedom parameter).

Estimating the Difference in Means of Two Normal Populations

If we let

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$$

then

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma^2(1/n + 1/m)}} \div \sqrt{S_p^2/\sigma^2} = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S_p^2(1/n + 1/m)}}$$

has a t-distribution with $n+m-2$ degrees of freedom. Consequently,

$$P\left\{-t_{\alpha/2,n+m-2} \le \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{1/n + 1/m}} \le t_{\alpha/2,n+m-2}\right\} = 1 - \alpha$$

Estimating the Difference in Means of Two Normal Populations

Therefore, when the data result in the values $\bar{X} = \bar{x}$, $\bar{Y} = \bar{y}$, $S_p = sp$, we obtain the following $100(1-\alpha)$ percent confidence interval for $\mu_1 - \mu_2$:

$$(\bar{x}-\bar{y}-t_{\alpha/2,n+m-2}s_p\sqrt{1/n+1/m}, \bar{x}-\bar{y}+t_{\alpha/2,n+m-2}s_p\sqrt{1/n+1/m})$$

One-sided confidence intervals are similarly obtained.

Exercise-2

Two different types of electrical cable insulation have recently been tested to determine the voltage level at which failures tend to occur. When specimens were subjected to an increasing voltage stress in a laboratory experiment, failures for the two types of cable insulation occurred at the following voltages:

Type A		Type B	
36	54	52	60
44	52	64	44
41	37	38	48
53	51	68	46
38	44	66	70
36	35	52	62
34	44		

Exercise-2

Suppose that it is known that the amount of voltage that cables having type A insulation can withstand is normally distributed with unknown mean μ_A and known variance $\sigma_A^2 = 40$, whereas the corresponding distribution for type B insulation is normal with unknown mean μ_B and known variance $\sigma_B^2 = 100$. Determine a 95 percent confidence interval for $\mu_A - \mu_B$. Determine a value that we can assert, with 95 percent confidence, exceeds $\mu_A - \mu_B$.

Exercise-3

There are two different techniques a given manufacturer can employ to produce batteries. A random selection of 12 batteries produced by technique I and of 14 produced by technique II resulted in the following capacities (in ampere hours):

Technique I		Technique II	
140	132	144	134
136	142	132	130
138	150	136	146
150	154	140	128
152	136	128	131
144	142	150	137
		130	135

Exercise-3

Determine a 90 percent level two-sided confidence interval for the difference in means, assuming a common variance. Also determine a 95 percent upper confidence interval for $\mu_I - \mu_{II}$.

Thank You