Statistical Inference

Interval Estimation – Hypothesis Testing

BIOSTATISTICS
Biomedical Engineering - Medical Radiation Physics

George Sakellaropoulos (presentation) - Sonia Malefaki (slides)

Department of Mechanical Engineering & Aeronautics
University of Patras, Greece.

☎: 2610 997673, ✉: smalefaki@upatras.gr

November 18, 2019

Outline

Introduction

Introduction

Goal: To draw conclusions about the population parameters from experimental data.

Introduction

Goal: To draw conclusions about the population parameters from experimental data.

Statistical Inference consists of those methods by which one makes inference or generalization about a population.

Introduction

Goal: To draw conclusions about the population parameters from experimental data.

Statistical Inference consists of those methods by which one makes inference or generalization about a population.

Statistical Inference

- Estimation
    - Point estimation
    - Interval estimation

Introduction

Goal: To draw conclusions about the population parameters from experimental data.

Statistical Inference consists of those methods by which one makes inference or generalization about a population.

Statistical Inference

- Estimation
  - Point estimation
  - Interval estimation
- Hypothesis Testing

Interval estimation

Confidence Intervals

Definition

Let $X_1, X_2, \ldots, X_n$ a random sample and θ an unknown parameter of the population and $1-a$ a probability. Then, the interval (l, u), where these endpoints are values of the corresponding random variables L and U, for which it holds

$$P(L < \theta < U) = 1 - a$$

is called $100(1-a)\%$ confidence interval for the parameter θ.

Confidence Intervals

Definition

Let $X_1, X_2, \ldots, X_n$ a random sample and θ an unknown parameter of the population and $1 - a$ a probability. Then, the interval (l, u), where these endpoints are values of the corresponding random variables L and U, for which it holds

$$P(L < \theta < U) = 1 - a$$

is called $100(1 - a)\%$ confidence interval for the parameter θ.

Confidence Intervals

Definition

Let $X_1, X_2, \ldots, X_n$ a random sample and θ an unknown parameter of the population and $1 - a$ a probability. Then, the interval (l, u), where these endpoints are values of the corresponding random variables L and U, for which it holds

$$P(L < \theta < U) = 1 - a$$

is called $100(1 - a)\%$ confidence interval for the parameter θ.

$1 - a$: confidence level

Confidence Intervals

Definition

Let $X_1, X_2, \ldots, X_n$ a random sample and θ an unknown parameter of the population and $1 - a$ a probability. Then, the interval (l, u), where these endpoints are values of the corresponding random variables L and U, for which it holds

$$P(L < \theta < U) = 1 - a$$

is called $100(1-a)\%$ confidence interval for the parameter θ.

$1 - a$: confidence level

- The $100(1-a)\%$ confidence interval provides an estimate of the accuracy of the point estimator.

Confidence Interval for the mean μ of a population

- with known variance

Let $\bar{X}$ be an estimator of μ.

Confidence Interval for the mean μ of a population

- with known variance

Let $\bar{X}$ be an estimator of μ.

Confidence Interval for the mean μ of a population

- with known variance

Let $\bar{X}$ be an estimator of μ. $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$,

Confidence Interval for the mean μ of a population

- with known variance

Let $\bar{X}$ be an estimator of μ. $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$

Confidence Interval for the mean μ of a population

- with known variance

Let $\bar{X}$ be an estimator of μ. $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$

$$P\left(A < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < B\right) = 1 - a$$

Confidence Interval for the mean μ of a population

- with known variance

Let $\bar{X}$ be an estimator of μ. $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$

$$P\left(A < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < B\right) = 1 - a$$

$$P\left(-z_{a/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{a/2}\right) = 1 - a$$

Confidence Interval for the mean μ of a population

- with known variance

Let $\bar{X}$ be an estimator of μ. $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$

$$P\left(A < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < B\right) = 1 - a$$

$$P\left(-z_{a/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{a/2}\right) = 1 - a$$

$$P\left(-\frac{z_{a/2}\sigma}{\sqrt{n}} < \bar{X}-\mu < \frac{z_{a/2}\sigma}{\sqrt{n}}\right) = 1 - a$$

Confidence Interval for the mean μ of a population

- with known variance

Let $\bar{X}$ be an estimator of μ. $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$

$$P\left(A < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < B\right) = 1 - a$$

$$P\left(-z_{a/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{a/2}\right) = 1 - a$$

$$P\left(-\frac{z_{a/2}\sigma}{\sqrt{n}} < \bar{X}-\mu < \frac{z_{a/2}\sigma}{\sqrt{n}}\right) = 1 - a$$

$$P\left(\bar{X} - \frac{z_{a/2}\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{z_{a/2}\sigma}{\sqrt{n}}\right) = 1 - a$$

Confidence Interval for the mean μ of a population

- with known variance

Let $\bar{X}$ be an estimator of μ. $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$

$$P\left(A < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < B\right) = 1 - a$$

$$P\left(-z_{a/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{a/2}\right) = 1 - a$$

$$P\left(-\frac{z_{a/2}\sigma}{\sqrt{n}} < \bar{X} - \mu < \frac{z_{a/2}\sigma}{\sqrt{n}}\right) = 1 - a$$

$$P\left(\bar{X} - \frac{z_{a/2}\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{z_{a/2}\sigma}{\sqrt{n}}\right) = 1 - a$$

$$\left(\bar{x} - \frac{z_{a/2}\sigma}{\sqrt{n}},\ \bar{x} + \frac{z_{a/2}\sigma}{\sqrt{n}}\right)$$

Confidence Interval for the mean μ of a population

- Error $|\bar{X} - \mu|$

If $\bar{x}$ is used as an estimator of μ, we can be $100(1 - a)\%$ confident that the error will not exceed

$$z_{a/2}\frac{\sigma}{\sqrt{n}}.$$

Confidence Interval for the mean μ of a population

- Error $|\bar{X} - \mu|$

If $\bar{x}$ is used as an estimator of μ, we can be $100(1-a)\%$ confident that the error will not exceed

$$z_{a/2}\frac{\sigma}{\sqrt{n}}.$$

- Sample size

Let $\bar{X}$ be an estimator of μ, we are $100(1-a)\%$ confident that the error will not exceed a specified amount ϵ when the sample size is

$$n \geq \left(\frac{z_{a/2}\sigma}{\epsilon}\right)^2$$

Confidence Interval for the mean μ of a population

- with unknown variance

$S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$,

Confidence Interval for the mean μ of a population

- with unknown variance

$S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$,

Confidence Interval for the mean μ of a population

- with unknown variance

$S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$, $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$

Confidence Interval for the mean μ of a population

- with unknown variance

$S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$, $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

$$P\left(-t_{n-1,a/2} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{n-1,a/2}\right) = 1 - a$$

Confidence Interval for the mean μ of a population

- with unknown variance

$S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$, $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

$$P\left(-t_{n-1,a/2} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{n-1,a/2}\right) = 1-a$$

$$P\left(-\frac{t_{n-1,a/2}S}{\sqrt{n}} < \bar{X}-\mu < \frac{t_{n-1,a/2}S}{\sqrt{n}}\right) = 1-a$$

Confidence Interval for the mean μ of a population

- with unknown variance

$S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$, $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

$$P\left(-t_{n-1,a/2} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{n-1,a/2}\right) = 1-a$$

$$P\left(-\frac{t_{n-1,a/2}S}{\sqrt{n}} < \bar{X}-\mu < \frac{t_{n-1,a/2}S}{\sqrt{n}}\right) = 1-a$$

$$P\left(\bar{X}-\frac{t_{n-1,a/2}S}{\sqrt{n}} < \mu < \bar{X}+\frac{t_{n-1,a/2}S}{\sqrt{n}}\right) = 1-a$$

Confidence Interval for the mean μ of a population

- with unknown variance

$S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$, $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

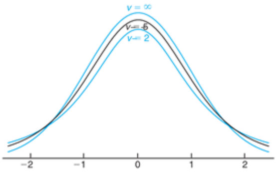
$$P\left(-t_{n-1,a/2} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{n-1,a/2}\right) = 1-a$$

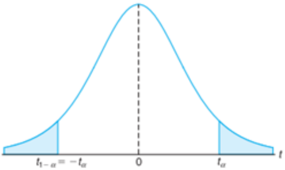
$$P\left(-\frac{t_{n-1,a/2}S}{\sqrt{n}} < \bar{X}-\mu < \frac{t_{n-1,a/2}S}{\sqrt{n}}\right) = 1-a$$

$$P\left(\bar{X} - \frac{t_{n-1,a/2}S}{\sqrt{n}} < \mu < \bar{X} + \frac{t_{n-1,a/2}S}{\sqrt{n}}\right) = 1-a$$

$$\left(\bar{x} - \frac{t_{n-1,a/2}s}{\sqrt{n}},\ \bar{x} + \frac{t_{n-1,a/2}s}{\sqrt{n}}\right)$$

t-Distribution





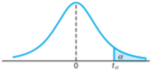


Table A.4 Critical Values of the *t*-Distribution

	α						
v	**0.40**	**0.30**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

Degrees of freedom

Although the sample size is n, the degrees of freedom of the random variable $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ is $n-1$. The degrees of freedom of t distribution are determined by the degrees of freedom of the sample variance.

Degrees of freedom

Although the sample size is n, the degrees of freedom of the random variable $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ is $n-1$. The degrees of freedom of t distribution are determined by the degrees of freedom of the sample variance.

▶Why does the sample variance has $n-1$ degrees of freedom?

Degrees of freedom

Although the sample size is n, the degrees of freedom of the random variable $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ is $n-1$. The degrees of freedom of t distribution are determined by the degrees of freedom of the sample variance.

▶Why does the sample variance has $n-1$ degrees of freedom?

It is known that $s^2 = \frac{\sum(x_i-\bar{x})^2}{n-1}$. In order to compute this quantity the squares of the differences $x_1 - \bar{x},\ x_2 - \bar{x}, \ldots, x_n - \bar{x}$ must be added. The sum of these differences is equal to zero

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + \ldots + (x_n - \bar{x}) = 0.$$

Thus, if $n-1$ values of these differences are known, the last one can be computed. So, from the n differences, only the $n-1$ can be changed freely. The last one is related to the rest of them. For this reason the sample variance has $n-1$ degrees of freedom.

Example

- Scholastic Aptitude Test (SAT) mathematics scores of a random sample of 25 high school seniors in the state of Texas are collected, and the sample mean and standard deviation are found to be 501 and 112, respectively. Find a 99% confidence interval on the mean SAT mathematics score for seniors in the state of Texas.

Prediction Intervals

Sometimes, we are also interested in predicting the possible value of a future observation X_0.

A prediction interval is an estimate of an interval
in which a future observation will fall, with a
certain probability, given what has already been observed

Prediction Intervals

Sometimes, we are also interested in predicting the possible value of a future observation X_0.

A natural point estimator of a new observation is $\bar{X}$.

Prediction Intervals

Sometimes, we are also interested in predicting the possible value of a future observation X_0.

A natural point estimator of a new observation is $\bar{X}$.

$X_0 - \bar{X} \sim \mathcal{N}(0, \sigma^2 \left(1 + \frac{1}{n}\right))$

Prediction Intervals

Sometimes, we are also interested in predicting the possible value of a future observation X_0.

A natural point estimator of a new observation is $\bar{X}$.

$X_0 - \bar{X} \sim \mathcal{N}(0, \sigma^2\left(1 + \frac{1}{n}\right))$

$$\left(\bar{x} - z_{a/2}\sigma\sqrt{1 + \frac{1}{n}},\ \bar{x} + z_{a/2}\sigma\sqrt{1 + \frac{1}{n}}\right)$$

$$\left(\bar{x} - t_{n-1,a/2}s\sqrt{1 + \frac{1}{n}},\ \bar{x} + t_{n-1,a/2}s\sqrt{1 + \frac{1}{n}}\right)$$

Prediction Intervals

Sometimes, we are also interested in predicting the possible value of a future observation X_0.

A natural point estimator of a new observation is $\bar{X}$.

$X_0 - \bar{X} \sim \mathcal{N}(0, \sigma^2 \left(1 + \frac{1}{n}\right))$

$$\left(\bar{x} - z_{a/2}\sigma\sqrt{1 + \frac{1}{n}},\ \bar{x} + z_{a/2}\sigma\sqrt{1 + \frac{1}{n}}\right)$$

$$\left(\bar{x} - t_{n-1,a/2}s\sqrt{1 + \frac{1}{n}},\ \bar{x} + t_{n-1,a/2}s\sqrt{1 + \frac{1}{n}}\right)$$

- The Prediction Intervals are used for Outlier detection

Tolerance Intervals

- One may be interested where the majority of the population falls.

Tolerance Intervals

- One may be interested where the majority of the population falls.

- In this case, one must attempt to determine bounds that, in some probability sense, "cover" values in a population.

Tolerance Intervals

- One may be interested where the majority of the population falls.

- In this case, one must attempt to determine bounds that, in some probability sense, "cover" values in a population.

- One method for establishing the desired bounds is to determine a confidence interval on a fix proportion of the measurements.

Tolerance Intervals

- One may be interested where the majority of the population falls.
- In this case, one must attempt to determine bounds that, in some probability sense, "cover" values in a population.
- One method for establishing the desired bounds is to determine a confidence interval on a fix proportion of the measurements.
- For example, $\mu \pm z_{a/2} \cdot \sigma$ covers exactly $(1 - a)100\%$ of the population of observations.

Tolerance Intervals

- One may be interested where the majority of the population falls.
- In this case, one must attempt to determine bounds that, in some probability sense, "cover" values in a population.
- One method for establishing the desired bounds is to determine a confidence interval on a fix proportion of the measurements.
- For example, $\mu \pm z_{a/2} \cdot \sigma$ covers exactly $(1 - a)100\%$ of the population of observations.
- When μ and σ are unknown, tolerance limits are given by $\bar{x} \pm k \cdot s$ where k is determined such that one can assert with $100(1 - \gamma)\%$ confidence that the given limits contain at least the proportion $1 - a$ of the measurements.

$$[\overline{X} - ks, \overline{X} + ks], \qquad k = \sqrt{\frac{(n^2 - 1)\, z^2_{1-\gamma/2}}{n\, \chi^2_{n-1,\alpha}}}$$

With a confidence of $1 - \alpha$, the proportion $1 - \gamma$ of population measurements will fall between the lower and upper bounds shown above.
This interval is called a $(1 - \gamma, 1 - \alpha)$-tolerance interval.

Example

- Machine Quality: A machine produces metal pieces that are cylindrical in shape. A sample of these pieces is taken and the diameters are found to be 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01, and 1.03 centimeters. For all computations, assume an approximately normal distribution.
(a) Find a 99% confidence interval on the mean diameter.
(b) Compute a 99% prediction interval on a measured diameter of a single metal piece taken from the machine.
(c) Find the 99% tolerance limits that will contain 95% of the metal pieces produced by this machine.

Example

- Machine Quality: A machine produces metal pieces that are cylindrical in shape. A sample of these pieces is taken and the diameters are found to be 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01, and 1.03 centimeters. For all computations, assume an approximately normal distribution.
(a) Find a 99% confidence interval on the mean diameter.
(b) Compute a 99% prediction interval on a measured diameter of a single metal piece taken from the machine.
(c) Find the 99% tolerance limits that will contain 95% of the metal pieces produced by this machine.

- Solution
The sample mean and standard deviation for the given data are $\bar{x} = 1.0056$ and $s = 0.0246$.

Example

- Machine Quality: A machine produces metal pieces that are cylindrical in shape. A sample of these pieces is taken and the diameters are found to be 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01, and 1.03 centimeters. For all computations, assume an approximately normal distribution.
(a) Find a 99% confidence interval on the mean diameter.
(b) Compute a 99% prediction interval on a measured diameter of a single metal piece taken from the machine.
(c) Find the 99% tolerance limits that will contain 95% of the metal pieces produced by this machine.

- Solution
The sample mean and standard deviation for the given data are $\bar{x} = 1.0056$ and $s = 0.0246$.
(a) $(0.9781,\ 1.0331)$

Example

- Machine Quality: A machine produces metal pieces that are cylindrical in shape. A sample of these pieces is taken and the diameters are found to be 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01, and 1.03 centimeters. For all computations, assume an approximately normal distribution.
(a) Find a 99% confidence interval on the mean diameter.
(b) Compute a 99% prediction interval on a measured diameter of a single metal piece taken from the machine.
(c) Find the 99% tolerance limits that will contain 95% of the metal pieces produced by this machine.

- Solution
The sample mean and standard deviation for the given data are $\bar{x} = 1.0056$ and $s = 0.0246$.
(a) $(0.9781,\ 1.0331)$
(b) $(0.9186,\ 1.0926)$

Example

- Machine Quality: A machine produces metal pieces that are cylindrical in shape. A sample of these pieces is taken and the diameters are found to be 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01, and 1.03 centimeters. For all computations, assume an approximately normal distribution.
(a) Find a 99% confidence interval on the mean diameter.
(b) Compute a 99% prediction interval on a measured diameter of a single metal piece taken from the machine.
(c) Find the 99% tolerance limits that will contain 95% of the metal pieces produced by this machine.

- Solution
The sample mean and standard deviation for the given data are $\bar{x} = 1.0056$ and $s = 0.0246$.
(a) $(0.9781, \; 1.0331)$
(b) $(0.9186, \; 1.0926)$
(c) $(0.8937, \; 1.1175)$

Solution: (a) The 99% confidence interval for the mean diameter is given by

$$\bar{x} \pm t_{0.005}s/\sqrt{n} = 1.0056 \pm (3.355)(0.0246/3) = 1.0056 \pm 0.0275.$$

Thus, the 99% confidence bounds are 0.9781 and 1.0331.

(b) The 99% prediction interval for a future observation is given by

$$\bar{x} \pm t_{0.005}s\sqrt{1 + 1/n} = 1.0056 \pm (3.355)(0.0246)\sqrt{1 + 1/9},$$

with the bounds being 0.9186 and 1.0926.

(c) From Table A.7, for $n = 9$, $1 - \gamma = 0.99$, and $1 - \alpha = 0.95$, we find $k = 4.550$ for two-sided limits. Hence, the 99% tolerance limits are given by

$$\bar{x} + ks = 1.0056 \pm (4.550)(0.0246),$$

with the bounds being 0.8937 and 1.1175. We are 99% confident that the tolerance interval from 0.8937 to 1.1175 will contain the central 95% of the distribution of diameters produced.

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ random samples form two independent populations with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2 respectively.

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ random samples form two independent populations with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2 respectively.

- with known variance

$\bar{X}$ an estimator of μ_X, $\bar{Y}$ an estimator of μ_Y $\bar{X} \sim \mathcal{N}(\mu_X, \frac{\sigma_X^2}{n})$, $\bar{Y} \sim \mathcal{N}(\mu_Y, \frac{\sigma_Y^2}{m})$,

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ random samples form two independent populations with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2 respectively.

- with known variance

$\bar{X}$ an estimator of μ_X, $\bar{Y}$ an estimator of μ_Y $\bar{X} \sim \mathcal{N}(\mu_X, \frac{\sigma_X^2}{n})$, $\bar{Y} \sim \mathcal{N}(\mu_Y, \frac{\sigma_Y^2}{m})$,

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ random samples form two independent populations with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2 respectively.

- with known variance

$\bar{X}$ an estimator of μ_X, $\bar{Y}$ an estimator of μ_Y $\bar{X} \sim \mathcal{N}(\mu_X, \frac{\sigma_X^2}{n})$, $\bar{Y} \sim \mathcal{N}(\mu_Y, \frac{\sigma_Y^2}{m})$,
$\bar{X} - \bar{Y} \sim \mathcal{N}(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m})$

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ random samples form two independent populations with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2 respectively.

▶ with known variance

$\bar{X}$ an estimator of μ_X, $\bar{Y}$ an estimator of μ_Y $\bar{X} \sim \mathcal{N}(\mu_X, \frac{\sigma_X^2}{n})$, $\bar{Y} \sim \mathcal{N}(\mu_Y, \frac{\sigma_Y^2}{m})$,
$\bar{X} - \bar{Y} \sim \mathcal{N}(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m})$

$$P\left(-z_{a/2} < \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}} < z_{a/2}\right) = 1 - a$$

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ random samples form two independent populations with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2 respectively.

- with known variance

$\bar{X}$ an estimator of μ_X, $\bar{Y}$ an estimator of μ_Y $\bar{X} \sim \mathcal{N}(\mu_X, \frac{\sigma_X^2}{n})$, $\bar{Y} \sim \mathcal{N}(\mu_Y, \frac{\sigma_Y^2}{m})$,
$\bar{X} - \bar{Y} \sim \mathcal{N}(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m})$

$$P\left(-z_{a/2} < \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}} < z_{a/2}\right) = 1 - a$$

$$P\left(-z_{a/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} < \bar{X} - \bar{Y} - (\mu_X - \mu_Y) < \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right) = 1 - a$$

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ random samples form two independent populations with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2 respectively.

- with known variance

$\bar{X}$ an estimator of μ_X, $\bar{Y}$ an estimator of μ_Y $\bar{X} \sim \mathcal{N}(\mu_X, \frac{\sigma_X^2}{n})$, $\bar{Y} \sim \mathcal{N}(\mu_Y, \frac{\sigma_Y^2}{m})$,
$\bar{X} - \bar{Y} \sim \mathcal{N}(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m})$

$$P\left(-z_{a/2} < \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}} < z_{a/2}\right) = 1 - a$$

$$P\left(-z_{a/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} < \bar{X} - \bar{Y} - (\mu_X - \mu_Y) < \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right) = 1 - a$$

$$P\left(\bar{X} - \bar{Y} - z_{a/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} < \mu < \bar{X} - \bar{Y} + z_{a/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right) = 1 - a$$

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ random samples form two independent populations with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2 respectively.

▶ with known variance

$\bar{X}$ an estimator of μ_X, $\bar{Y}$ an estimator of μ_Y $\bar{X} \sim \mathcal{N}(\mu_X, \frac{\sigma_X^2}{n})$, $\bar{Y} \sim \mathcal{N}(\mu_Y, \frac{\sigma_Y^2}{m})$,
$\bar{X} - \bar{Y} \sim \mathcal{N}(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m})$

$$P\left(-z_{a/2} < \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}} < z_{a/2}\right) = 1 - a$$

$$P\left(-z_{a/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} < \bar{X} - \bar{Y} - (\mu_X - \mu_Y) < \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right) = 1 - a$$

$$P\left(\bar{X} - \bar{Y} - z_{a/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} < \mu < \bar{X} - \bar{Y} + z_{a/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right) = 1 - a$$

$$\left(\bar{x} - \bar{y} - z_{a/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}},\ \bar{x} - \bar{y} + z_{a/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right)$$

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

- with unknown variances, but equal $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

- with unknown variances, but equal $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

- with unknown variances, but equal $\sigma_1^2 = \sigma_2^2 = \sigma^2$

The common variance σ^2 is estimated by

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$$

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

- with unknown variances, but equal $\sigma_1^2 = \sigma_2^2 = \sigma^2$

The common variance σ^2 is estimated by

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$$

$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{n+m-2}$

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

- with unknown variances, but equal $\sigma_1^2 = \sigma_2^2 = \sigma^2$

The common variance σ^2 is estimated by

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$$

$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{n+m-2}$

$$P\left(-t_{n+m-2,a/2} < \frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} < t_{n+m-2,a/2}\right) = 1-a$$

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

- with unknown variances, but equal $\sigma_1^2 = \sigma_2^2 = \sigma^2$

The common variance σ^2 is estimated by

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$$

$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{n+m-2}$

$$P\left(-t_{n+m-2,a/2} < \frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} < t_{n+m-2,a/2}\right) = 1-a$$

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

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

- with unknown variances and not equal $\sigma_1^2 \neq \sigma_2^2$
  - $m = n$ then

$$\left(\bar{x} - \bar{y} - t_{v,a/2} \sqrt{\frac{s_x^2 + s_y^2}{v}},\ \bar{x} - \bar{y} + t_{v,a/2} \sqrt{\frac{s_x^2 + s_y^2}{v}} \right)$$

where $v = 2(n-1)$

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

- with unknown variances and not equal $\sigma_1^2 \neq \sigma_2^2$
  - $m = n$ then

$$\left(\bar{x} - \bar{y} - t_{v,a/2}\sqrt{\frac{s_x^2 + s_y^2}{v}},\ \bar{x} - \bar{y} + t_{v,a/2}\sqrt{\frac{s_x^2 + s_y^2}{v}}\right)$$

where $v = 2(n-1)$

Confidence intervals for the difference $\mu_1 - \mu_2$ of two independent populations

- with unknown variances and not equal $\sigma_1^2 \neq \sigma_2^2$
  - $m = n$ then

$$\left(\bar{x} - \bar{y} - t_{v,a/2}\sqrt{\frac{s_x^2 + s_y^2}{v}},\ \bar{x} - \bar{y} + t_{v,a/2}\sqrt{\frac{s_x^2 + s_y^2}{v}}\right)$$

where $v = 2(n-1)$

  - $m \neq m$ then

$$\left(\bar{x} - \bar{y} - t_{v,a/2}\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}},\ \bar{x} - \bar{y} + t_{v,a/2}\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}\right)$$

$$v = \frac{\left(s_x^2/n + s_y^2/m\right)^2}{\frac{(s_X^2/n)^2}{n-1} + \frac{(s_Y^2/m)^2}{m-1}}$$

(the value of v is rounded to the nearest integer)

Example

- A researcher wants to compare the mean gain in weight of some animals using two different diets. The researcher gives diet A in 10 of the available animals and diet B in 6 animals and he gets the following results:

A	113	135	91	104	135	107	152	97	145	129
B	126	73	102	110	79	104				

Find a 95% confidence interval on the difference of the mean weight increase by the two diets, under the assumption that the two populations are normally distributed and have equal variances.

Confidence interval of $\mu_1 - \mu_2$ of paired samples

Paired observations $(x_i, y_i),\ i = 1, 2, \ldots n$

- We run a test on a new diet using n individuals, where x_i is the weight before and y_i after going on the diet forms the information of the two samples.

Confidence interval of $\mu_1 - \mu_2$ of paired samples

Paired observations $(x_i, y_i),\ i = 1, 2, \ldots n$

- We run a test on a new diet using n individuals, where x_i is the weight before and y_i after going on the diet forms the information of the two samples.
- The differences $d_i = x_i - y_i, i = 1, \ldots, n$ are considered.

Confidence interval of $\mu_1 - \mu_2$ of paired samples

Paired observations $(x_i, y_i),\ i = 1, 2, \ldots n$

- We run a test on a new diet using n individuals, where x_i is the weight before and y_i after going on the diet forms the information of the two samples.

- The differences $d_i = x_i - y_i, i = 1, \ldots, n$ are considered.

These differences are the values of a random sample $D_1, \ldots, D_n$ from a population of differences that we shall assume to be normally distributed with mean $\mu_d = \mu_1 - \mu_2$ and variance

$$Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y).$$

Confidence interval of $\mu_1 - \mu_2$ of paired samples

Paired observations (x_i, y_i), $i = 1, 2, \dots n$

- We run a test on a new diet using n individuals, where x_i is the weight before and y_i after going on the diet forms the information of the two samples.

- The differences $d_i = x_i - y_i, i = 1, \dots, n$ are considered.

These differences are the values of a random sample $D_1, \dots, D_n$ from a population of differences that we shall assume to be normally distributed with mean $\mu_d = \mu_1 - \mu_2$ and variance

$$Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y).$$

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \quad \text{and} \quad s_s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$$

Confidence interval of $\mu_1 - \mu_2$ of paired samples

Paired observations (x_i, y_i), $i = 1, 2, \ldots n$

- We run a test on a new diet using n individuals, where x_i is the weight before and y_i after going on the diet forms the information of the two samples.

- The differences $d_i = x_i - y_i, i = 1, \ldots, n$ are considered.

These differences are the values of a random sample $D_1, \ldots, D_n$ from a population of differences that we shall assume to be normally distributed with mean $\mu_d = \mu_1 - \mu_2$ and variance

$$Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y).$$

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \quad \text{and} \quad s_s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$$

$$\frac{\bar{D} - \mu_d}{S_d/\sqrt{n}} \sim t_{n-1}$$

Confidence interval of $\mu_1 - \mu_2$ of paired samples

Paired observations (x_i, y_i), $i = 1, 2, \ldots n$

- We run a test on a new diet using n individuals, where x_i is the weight before and y_i after going on the diet forms the information of the two samples.

- The differences $d_i = x_i - y_i, i = 1, \ldots, n$ are considered.

These differences are the values of a random sample $D_1, \ldots, D_n$ from a population of differences that we shall assume to be normally distributed with mean $\mu_d = \mu_1 - \mu_2$ and variance

$$Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y).$$

$$\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i \quad \text{and} \quad s_s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})^2$$

$$\frac{\bar{D} - \mu_d}{S_d / \sqrt{n}} \sim t_{n-1}$$

$$\left(\bar{d} - t_{n-1,a/2} \frac{s_d}{\sqrt{n}},\ \bar{d} + t_{n-1,a/2} \frac{s_d}{\sqrt{n}} \right)$$

Confidence interval of the proportion p of a population

Let x be the number of items with a specific characteristic in a sample of size n. Then the proportion p of the population with the specific characteristic is estimated by $\hat{p} = \frac{x}{n}$.

Confidence interval of the proportion p of a population

Let x be the number of items with a specific characteristic in a sample of size n. Then the proportion p of the population with the specific characteristic is estimated by $\hat{p} = \frac{x}{n}$.

From the CLT, for large n,

$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right) \quad \text{or} \quad \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0, 1)$$

Confidence interval of the proportion p of a population

Let x be the number of items with a specific characteristic in a sample of size n. Then the proportion p of the population with the specific characteristic is estimated by $\hat{p} = \frac{x}{n}$.

From the CLT, for large n,

$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right) \quad \text{or} \quad \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0, 1)$$

$$\left(\hat{p} - z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

Confidence interval of the proportion p of a population

Let x be the number of items with a specific characteristic in a sample of size n. Then the proportion p of the population with the specific characteristic is estimated by $\hat{p} = \frac{x}{n}$.

From the CLT, for large n,

$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right) \quad \text{or} \quad \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0, 1)$$

$$\left(\hat{p} - z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

`Wilson score interval (better performance when p near 0 or 1)`

$$\left(\frac{\hat{p} + \frac{z_{a/2}^2}{2n}}{1 + \frac{z_{a/2}^2}{n}} - \frac{z_{a/2}}{1 + \frac{z_{a/2}^2}{n}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{a/2}^2}{4n^2}},\ \frac{\hat{p} + \frac{z_{a/2}^2}{2n}}{1 + \frac{z_{a/2}^2}{n}} + \frac{z_{a/2}}{1 + \frac{z_{a/2}^2}{n}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{a/2}^2}{4n^2}}\right)$$

Confidence interval of the proportion p of a population

Let x be the number of items with a specific characteristic in a sample of size n. Then the proportion p of the population with the specific characteristic is estimated by $\hat{p} = \frac{x}{n}$.

From the CLT, for large n,

$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right) \quad \text{or} \quad \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0,1)$$

$$\left(\hat{p} - z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

$$\left(\frac{\hat{p}+\frac{z_{a/2}^2}{2n}}{1+\frac{z_{a/2}^2}{n}} - \frac{z_{a/2}}{1+\frac{z_{a/2}^2}{n}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z_{a/2}^2}{4n^2}},\ \frac{\hat{p}+\frac{z_{a/2}^2}{2n}}{1+\frac{z_{a/2}^2}{n}} + \frac{z_{a/2}}{1+\frac{z_{a/2}^2}{n}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z_{a/2}^2}{4n^2}}\right)$$

- A coin is tossed 100 times and 45 times are head. Find a 95% confidence interval for the true proportion of head.

Confidence interval for the difference $p_1 - p_2$ of two population

Let x_1 and x_2 be the number of the items with a specific characteristic in two independent samples with sample sizes n and m respectively. The proportions p_1 and p_2 with the specific characteristic are estimated with the quantities $\hat{p}_1 = \frac{x_1}{n}$ and $\hat{p}_2 = \frac{x_2}{m}$ respectively.

Confidence interval for the difference $p_1 - p_2$ of two population

Let x_1 and x_2 be the number of the items with a specific characteristic in two independent samples with sample sizes n and m respectively. The proportions p_1 and p_2 with the specific characteristic are estimated with the quantities $\hat{p}_1 = \frac{x_1}{n}$ and $\hat{p}_2 = \frac{x_2}{m}$ respectively.

From the CLT, for large n and m,

$$\hat{p}_1 \sim \mathcal{N}\left(p_1, \frac{p_1(1-p_1)}{n}\right) \quad \text{and} \quad \hat{p}_2 \sim \mathcal{N}\left(p_2, \frac{p_2(1-p_2)}{m}\right)$$

Confidence interval for the difference $p_1 - p_2$ of two population

Let x_1 and x_2 be the number of the items with a specific characteristic in two independent samples with sample sizes n and m respectively. The proportions p_1 and p_2 with the specific characteristic are estimated with the quantities $\hat{p}_1 = \frac{x_1}{n}$ and $\hat{p}_2 = \frac{x_2}{m}$ respectively.

From the CLT, for large n and m,

$$\hat{p}_1 \sim \mathcal{N}\left(p_1, \frac{p_1(1-p_1)}{n}\right) \quad \text{and} \quad \hat{p}_2 \sim \mathcal{N}\left(p_2, \frac{p_2(1-p_2)}{m}\right)$$

$$\hat{p}_1 - \hat{p}_2 \sim \mathcal{N}\left(p_1 - p_2, \frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}\right)$$

Confidence interval for the difference $p_1 - p_2$ of two population

Let x_1 and x_2 be the number of the items with a specific characteristic in two independent samples with sample sizes n and m respectively. The proportions p_1 and p_2 with the specific characteristic are estimated with the quantities $\hat{p}_1 = \frac{x_1}{n}$ and $\hat{p}_2 = \frac{x_2}{m}$ respectively.

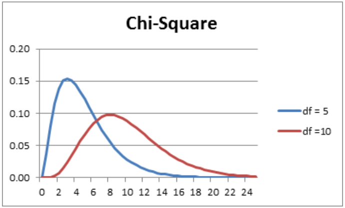
From the CLT, for large n and m,

$$\hat{p}_1 \sim \mathcal{N}\left(p_1, \frac{p_1(1-p_1)}{n}\right) \quad \text{and} \quad \hat{p}_2 \sim \mathcal{N}\left(p_2, \frac{p_2(1-p_2)}{m}\right)$$

$$\hat{p}_1 - \hat{p}_2 \sim \mathcal{N}\left(p_1 - p_2, \frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}\right)$$

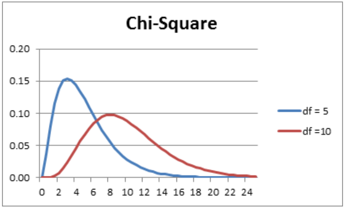
$$\left(\hat{p}_1 - \hat{p}_2 - z_{a/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}},\ \hat{p}_1 - \hat{p}_2 + z_{a/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}\right)$$

Confidence interval of the variance and standard deviation of a population



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

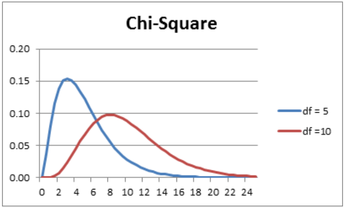
Confidence interval of the variance and standard deviation of a population



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

$$P\left(\mathcal{X}^2_{n-1,1-a/2} < \frac{(n-1)S^2}{\sigma} < \mathcal{X}^2_{n-1,a/2}\right) = 1 - a$$

Confidence interval of the variance and standard deviation of a population

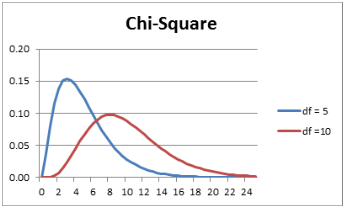


$$\frac{(n-1)S^2}{\sigma^2} \sim \mathcal{X}^2_{n-1}$$

$$P\left(\mathcal{X}^2_{n-1,1-a/2} < \frac{(n-1)S^2}{\sigma} < \mathcal{X}^2_{n-1,a/2}\right) = 1-a$$

$$\left(\frac{(n-1)s^2}{\mathcal{X}^2_{n-1,a/2}}, \frac{(n-1)s^2}{\mathcal{X}^2_{n-1,1-a/2}}\right)$$

Confidence interval of the variance and standard deviation of a population



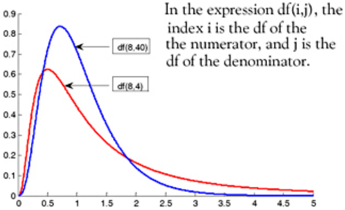
$$\frac{(n-1)S^2}{\sigma^2} \sim \mathcal{X}^2_{n-1}$$

$$P\left(\mathcal{X}^2_{n-1,1-a/2} < \frac{(n-1)S^2}{\sigma} < \mathcal{X}^2_{n-1,a/2}\right) = 1-a$$

$$\left(\frac{(n-1)s^2}{\mathcal{X}^2_{n-1,a/2}}, \frac{(n-1)s^2}{\mathcal{X}^2_{n-1,1-a/2}}\right)$$

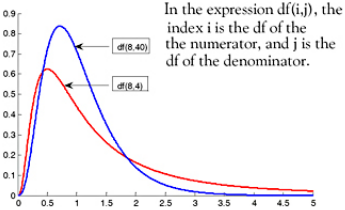
$$\left(\sqrt{\frac{(n-1)s^2}{\mathcal{X}^2_{n-1,a/2}}}, \sqrt{\frac{(n-1)s^2}{\mathcal{X}^2_{n-1,1-a/2}}}\right)$$

Confidence interval of the ration of σ_X^2/σ_Y^2 of two independent populations



$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F_{n-1,m-1}$$

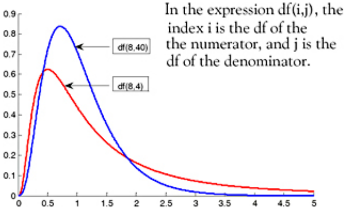
Confidence interval of the ration of σ_X^2/σ_Y^2 of two independent populations



$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F_{n-1,m-1}$$

$$P\left(F_{n-1,m-1,1-a/2} < \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} < F_{n-1,m-1,a/2}\right) = 1 - a$$

Confidence interval of the ration of σ_X^2/σ_Y^2 of two independent populations



$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F_{n-1,m-1}$$

$$P\left(F_{n-1,m-1,1-a/2} < \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} < F_{n-1,m-1,a/2}\right) = 1 - a$$

$$\left(\frac{1}{F_{n-1,m-1,a/2}} \cdot \frac{s_X^2}{s_Y^2}, \ \frac{1}{F_{n-1,m-1,1-a/2}} \cdot \frac{s_X^2}{s_Y^2}\right)$$