Introductory Statistics

Tatiana Tyukina
tt51@leicester.ac.uk

Topic 2 - Estimators
2.3 - Interval Estimation

- **Topic 0: Introduction**
- **Topic 1: Descriptive Statistics**
- **Topic 2: Estimators: Point estimation, Interval estimation**
  - 2.1 **Point estimation**,
  - 2.2 **Properties of estimators**,
  - 2.3 Interval estimation,
  - 2.4 Inference based on Normal Distribution
- Topic 3: Hypothesis Testing: Test for single population mean, test for two population mean
- Topic 4: Bayesian Estimation
- Topic 5: Goodness of Fit: The χ^2 test.

RECAP

Properties of the estimators:

1. unbiasedness - $E(\hat{\theta}) = \theta$
2. efficiency - minimum variance; Cramér-Rao lower bound;
3. sufficiency - $P_\theta(X_1, ..., X_n|\hat{\theta}) = P_\theta(X_1, ..., X_n)$;
4. consistency - $\lim_{n\to\infty} P\left(|\hat{\theta}_n - \theta| < \varepsilon\right) = 1$

INTERVAL ESTIMATION

Point estimators do not give us any information about their precision!

EXAMPLE

Let $X_i \sim Pois(\lambda)$, point estimator - $\hat{\lambda} = \bar{X}$.
For a sample of size 5, the estimate is $\hat{\lambda}(x) = 6.4$.
What can we say about the real value? Is this estimate is close enough?
Which one is the correct value: $\lambda = 6$, $\lambda = 1$ or $\lambda = 12$?

INTERVAL ESTIMATION

For an interval estimator of a single parameter θ, we use the random sample to find two quantities L and U, $L \leq \theta \leq U$ with some probability.
This interval $[L, U]$ should have two properties:

- $P(L \leq \theta \leq U)$ is high,

INTERVAL ESTIMATION

For an interval estimator of a single parameter θ, we use the random sample to find two quantities L and U, $L \leq \theta \leq U$ with some probability.
This interval $[L, U]$ should have two properties:

- $P(L \leq \theta \leq U)$ is high,
- the length of the interval $[L, U]$ should be relatively narrow on the average.

INTERVAL ESTIMATION

For an interval estimator of a single parameter θ, we use the random sample to find two quantities L and U, $L \leq \theta \leq U$ with some probability.
This interval $[L, U]$ should have two properties:

- $P(L \leq \theta \leq U)$ is high,
- the length of the interval $[L, U]$ should be relatively narrow on the average.

DEFINITION

*The **problem of confidence estimation** is that of finding a family of random sets $S(X)$ for a parameter θ such that for a given α, $0 < \alpha < 1$,*

$$P_\theta(S(X) \ni \theta) > 1 - \alpha, \text{ for all } \theta \in \Theta.$$

*Interval estimators are called **confidence intervals** (**CI**). The limits L and U are called the **lower** and the **upper confidence limits**, respectively.*

INTERVAL ESTIMATION

For an interval estimator of a single parameter θ, we use the random sample to find two quantities L and U, $L \leq \theta \leq U$ with some probability.
This interval $[L, U]$ should have two properties:

- $P(L \leq \theta \leq U)$ is high,
- the length of the interval $[L, U]$ should be relatively narrow on the average.

DEFINITION

*The **problem of confidence estimation** is that of finding a family of random sets $S(X)$ for a parameter θ such that for a given α, $0 < \alpha < 1$,*

$$P_\theta(S(X) \ni \theta) > 1 - \alpha, \text{ for all } \theta \in \Theta.$$

*Interval estimators are called **confidence intervals** (**CI**). The limits L and U are called the **lower** and the **upper confidence limits**, respectively.*

DEFINITION

*The probability $1 - \alpha$ that a confidence interval contains the true parameter θ is called **the confidence coefficient**.*

INTERVAL ESTIMATION:INTERPRETATION

EXAMPLE

Suppose that 6.5, 9.2, 9.9, 12.4 are the realisations of a random variable X from $N(\mu, 0.8^2)$.
Construct a 95% confidence interval for μ. $P(L \le \mu \le U) = 1-\alpha$, $\alpha = 0{,}05$

INTERVAL ESTIMATION:INTERPRETATION

EXAMPLE

Suppose that 6.5, 9.2, 9.9, 12.4 are the realisations of a random variable X from $N(\mu, 0.8^2)$.
Construct a 95% confidence interval for μ.
Solution.
The sample of 4 random variables $X_1, X_2, X_3, X_4 \sim N(\mu, \sigma^2)$, $\sigma = 0.8$.
MLE is $\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

INTERVAL ESTIMATION:INTERPRETATION

EXAMPLE

Suppose that 6.5, 9.2, 9.9, 12.4 are the realisations of a random variable X from $N(\mu, 0.8^2)$.
Construct a 95% confidence interval for μ.
Solution.
The sample of 4 random variables $X_1, X_2, X_3, X_4 \sim N(\mu, \sigma^2)$, $\sigma = 0.8$.
MLE is $\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Proposition
Let $X_1, X_2, ...X_n$ be independent normally distributed random variables such that $X_i \sim N(\mu_i, \sigma_i^2)$. Let $Y = c_1X_1 + c_2X_2 + ... + c_nX_n$.
Then $Y \sim N(\sum_{i=1}^{n} c_i\mu_i, \sum_{i=1}^{n} c_i^2\sigma_i^2)$.

INTERVAL ESTIMATION:INTERPRETATION

EXAMPLE

Suppose that 6.5, 9.2, 9.9, 12.4 are the realisations of a random variable X from $N(\mu, 0.8^2)$.
Construct a 95% confidence interval for μ.
Solution.
The sample of 4 random variables $X_1, X_2, X_3, X_4 \sim N(\mu, \sigma^2)$, $\sigma = 0.8$.
MLE is $\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Proposition
Let $X_1, X_2, ...X_n$ be independent normally distributed random variables such that $X_i \sim N(\mu_i, \sigma_i^2)$. Let $Y = c_1X_1 + c_2X_2 + ... + c_nX_n$.
Then $Y \sim N(\sum_{i=1}^{n} c_i\mu_i, \sum_{i=1}^{n} c_i^2\sigma_i^2)$.

Proposition
If $X \sim N(\mu, \sigma^2)$ then $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$, where $a,\ b \in R\backslash\{0\}$ are constants.
If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

INTERVAL ESTIMATION:INTERPRETATION

EXAMPLE

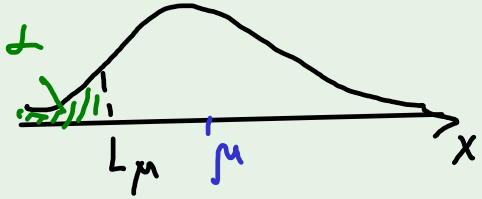
MLE is $\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Hence, $\bar{X} \sim N(\mu, \sigma^2/n)$ and $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

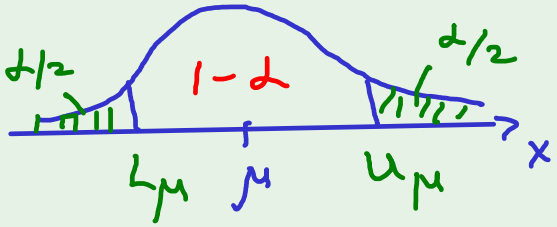
$P(L \le \mu \le U) = 0{,}95$ $\quad P(L < \bar{X} - \mu < U) = 0{,}95 = 1 - \alpha$, $\alpha = 0{,}05$

$P(L \le \mu) = 0{,}95$

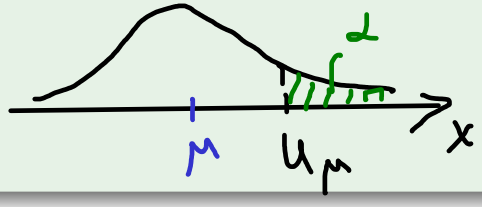
$P(\mu \le U) = 0{,}95$

$P(\underbrace{\bar{X} - U}_{L_\mu} < \mu < \underbrace{\bar{X} - L}_{U_\mu}) = 0{,}95$ $\quad \alpha/2 = 0{,}25$



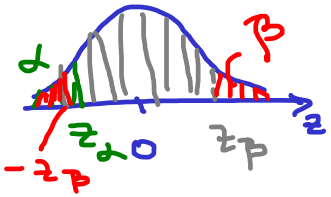


$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$



$P(L_z < z < U_z) = ?$

INTERVAL ESTIMATION:INTERPRETATION



$\alpha = 0{,}01 = 1\%$

$P(z < z_{\alpha}) = 0{,}01$

$z_{\alpha} = -2.31$

$z_{\frac{\alpha}{2}} = -2{,}93$

$P(z < -2{,}93) = 0{,}0017$

$z = 0$

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

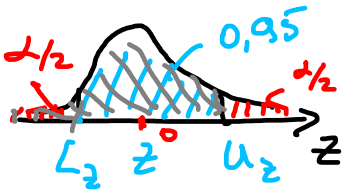
Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592

$\alpha/2 = 0{,}05$

$P(z < z_{\beta}) =$

$1 - P(z > z_{\beta}) =$

$= 1 - P(z < -z_{\beta})$



$P(L_z \le z \le u_z) = 0{,}95$

$L_z = -1{,}96$

$u_z = 1{,}96$

MATLAB: >> x = norminv([0.025 0.975])

R: > x=qnorm(0.025)
> x=qnorm(0.975)

EXCEL: =NORMINV(0.025,0,1)
=NORMINV(0.975,0,1)

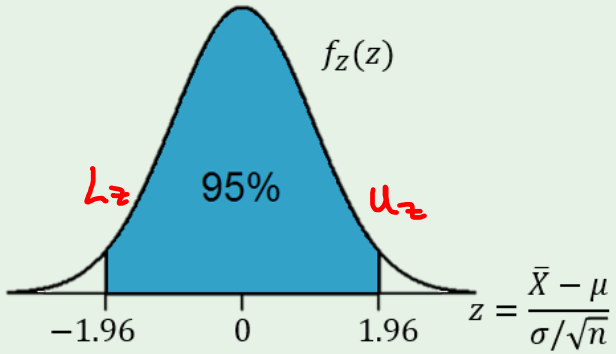
INTERVAL ESTIMATION:INTERPRETATION

EXAMPLE

MLE is $\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Hence, $\bar{X} \sim N(\mu, \sigma^2/n)$ and $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$$P(L_z \le z \le U_z) = 0{,}95$$

$$P\left(L_z \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le U_z\right) = P\left(\bar{X} - U_z \cdot \sigma/\sqrt{n} \le \mu \le \bar{X} - L_z \cdot \sigma/\sqrt{n}\right) = 0{,}95$$



https://en.wikipedia.org/wiki/1.96

$\sigma = 0{,}8$

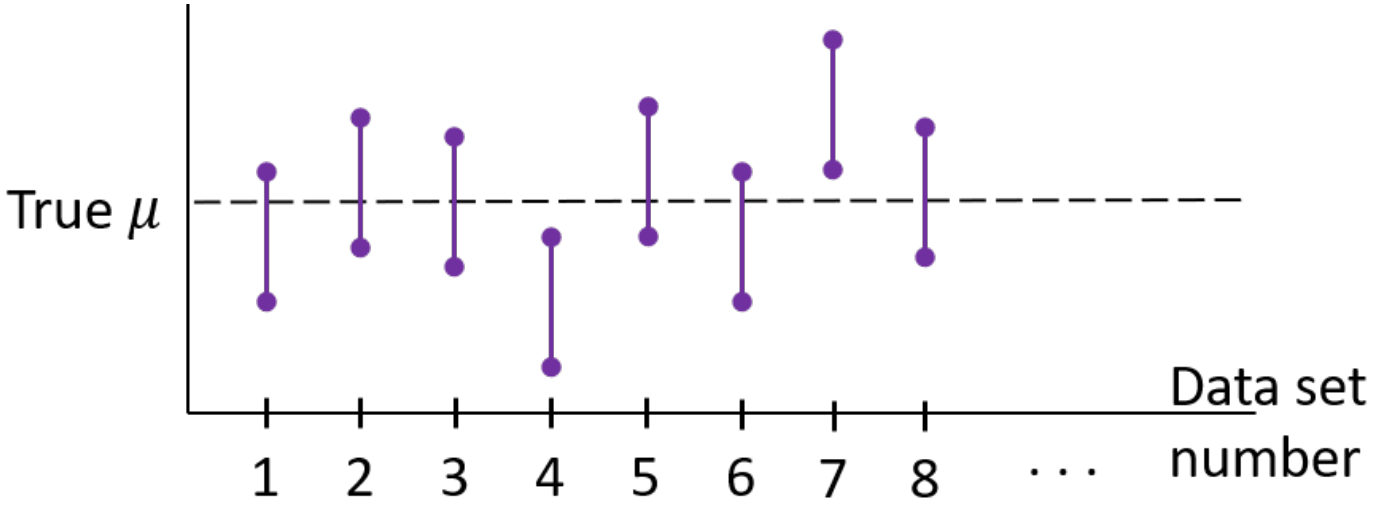
$n = 4$

$\bar{x} = 9{,}5$

$$P(8{,}75 \le \mu \le 10{,}28) = 0{,}95$$

95% CI for μ is $(8{,}75, 10{,}28)$

INTERVAL ESTIMATION:INTERPRETATION



The correct interpretation of confidence interval for the population mean is that if samples of the same size, n, are drawn repeatedly from a population, and a confidence interval is calculated from each sample, then 95% of these intervals should contain the population mean.

INTERVAL ESTIMATION USING PIVOTS

DEFINITION

Let $X \sim P_\theta$. A random variable $T(X, \theta)$ is known as a ***pivot*** *if the distribution of $T(X, \theta)$ does not depend on θ.*

INTERVAL ESTIMATION USING PIVOTS

DEFINITION

Let $X \sim P_\theta$. A random variable $T(X, \theta)$ is known as a ***pivot*** *if the distribution of $T(X, \theta)$ does not depend on θ.*

The *pivotal method* relies on our knowledge of sampling distributions.
The pivotal quantity should have the following two characteristics:

INTERVAL ESTIMATION USING PIVOTS

DEFINITION

Let $X \sim P_\theta$. A random variable $T(X, \theta)$ is known as a ***pivot*** *if the distribution of $T(X, \theta)$ does not depend on θ.*

The *pivotal method* relies on our knowledge of sampling distributions.
The pivotal quantity should have the following two characteristics:

- It is a function of the random sample (a statistic or an estimator $\hat{\theta}$) and the unknown parameter θ, where θ is the only unknown quantity, and

INTERVAL ESTIMATION USING PIVOTS

DEFINITION

Let $X \sim P_\theta$. A random variable $T(X, \theta)$ is known as a ***pivot*** *if the distribution of $T(X, \theta)$ does not depend on θ.*

The *pivotal method* relies on our knowledge of sampling distributions.
The pivotal quantity should have the following two characteristics:

- It is a function of the random sample (a statistic or an estimator $\hat{\theta}$) and the unknown parameter θ, where θ is the only unknown quantity, and
- It has a probability distribution that does not depend on the parameter θ.

INTERVAL ESTIMATION USING PIVOTS

DEFINITION

*Let $X \sim P_\theta$. A random variable $T(X, \theta)$ is known as a **pivot** if the distribution of $T(X, \theta)$ does not depend on θ.*

The *pivotal method* relies on our knowledge of sampling distributions.
The pivotal quantity should have the following two characteristics:

- It is a function of the random sample (a statistic or an estimator $\hat{\theta}$) and the unknown parameter θ, where θ is the only unknown quantity, and
- It has a probability distribution that does not depend on the parameter θ.

Suppose that $\hat{\theta} = g(X)$ is a point estimate of θ, and let $T(\hat{\theta}, \theta)$ be the pivotal quantity. Let a and b be constants with $(a < b)$, such that

$$P(a \leq T(\hat{\theta}, \theta) \leq b) = 1 - \alpha$$

for a given value of α, $(0 < \alpha < 1)$.
Hence, if we resolve this inequality in terms of θ, the result will be a desired confidence interval.

INTERVAL ESTIMATION USING PIVOTS

$$T(\bar{X}, \mu) = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}, \quad T \sim N(0, 1)$$

INTERVAL ESTIMATION USING PIVOTS

THEOREM

Let $T(X,\theta)$ be a pivot such that for each θ, $T(X,\theta)$ is a statistic, and as a function of θ, T is either strictly increasing or decreasing at each $x \in R$.
Let $\Lambda \subseteq R$ be the range of T, and for every $\lambda \in \Lambda$ and $x \in R$, let the equation $\lambda = T(x,\theta)$ be solvable. Then one can construct a confidence interval for θ at any level.

INTERVAL ESTIMATION USING PIVOTS

THEOREM

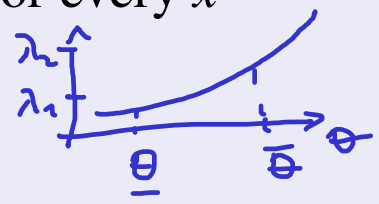
Let $T(X,\theta)$ be a pivot such that for each θ, $T(X,\theta)$ is a statistic, and as a function of θ, T is either strictly increasing or decreasing at each $x \in R$.
Let $\Lambda \subseteq R$ be the range of T, and for every $\lambda \in \Lambda$ and $x \in R$, let the equation $\lambda = T(x,\theta)$ be solvable. Then one can construct a confidence interval for θ at any level.

PROOF.

Let $0 < \alpha < 1$, choose a pair of numbers $\lambda_1(\alpha), \lambda_2(\alpha) \in \Lambda$:

$$P_\theta(\lambda_1(\alpha) < T(X,\theta) < \lambda_2(\alpha)) \geq 1 - \alpha$$

T is monotone in θ, hence, $T(x,\theta) = \lambda_1(\alpha)$ and $T(x,\theta) = \lambda_2(\alpha)$ for every x uniquely for θ.



$$P_\theta(\underline{\theta}(X) < \theta < \bar{\theta}(X)) \geq 1 - \alpha \quad \text{for all } \theta,$$

$\underline{\theta}(X) < \bar{\theta}(X)$ are random variables. □

INTERVAL ESTIMATION USING PIVOTS

The condition that $\lambda = T(x, \theta)$ *be solvable* will be satisfied if, for example, T is continuous and strictly increasing or decreasing as a function of $\theta \in \Theta$.

Relation $P_\theta(\lambda_1(\alpha) < T(X, \theta) < \lambda_2(\alpha)) \geq 1 - \alpha$ is valid even when the assumption of monotonicity of T in the theorem is dropped. In that case, inversion of the inequalities may yield a set of intervals (random set) $S(X)$ in θ instead of a confidence interval.

INTERVAL ESTIMATION USING PIVOTS

Procedure for finding CI for θ using pivot

1. Find an estimator $\hat{\theta} = g(X)$ of θ (usually MLE of θ works).

INTERVAL ESTIMATION USING PIVOTS

Procedure for finding CI for θ using pivot

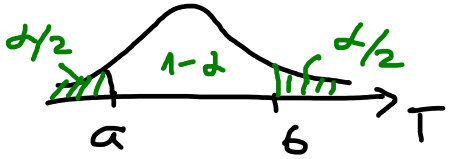
1. Find an estimator $\hat{\theta} = g(X)$ of θ (usually MLE of θ works).
2. Find a function of $\hat{\theta}$ and θ, $T(\hat{\theta}, \theta)$ (pivot), such that the probability distribution of $T(\hat{\theta}, \theta)$ does not depend on θ.

INTERVAL ESTIMATION USING PIVOTS

$P(a < T) = 1-\alpha$

Procedure for finding CI for θ using pivot

1. Find an estimator $\hat{\theta} = g(X)$ of θ (usually MLE of θ works).
2. Find a function of $\hat{\theta}$ and θ, $T(\hat{\theta}, \theta)$ (pivot), such that the probability distribution of $T(\hat{\theta}, \theta)$ does not depend on θ.
3. Find $a < b$ such that $P(a < T(\hat{\theta}, \theta) < b) = 1 - \alpha$.
   Choose a and b such that $P(a \geq T(\hat{\theta}, \theta)) = \frac{\alpha}{2} = P(T(\hat{\theta}, \theta) \geq b)$

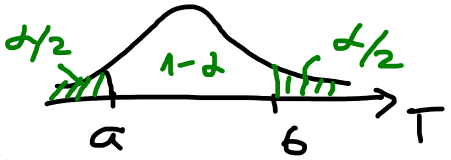


INTERVAL ESTIMATION USING PIVOTS

$P(a < T) = 1-\alpha$

Procedure for finding CI for θ using pivot

1. Find an estimator $\hat{\theta} = g(X)$ of θ (usually MLE of θ works).
2. Find a function of $\hat{\theta}$ and θ, $T(\hat{\theta}, \theta)$ (pivot), such that the probability distribution of $T(\hat{\theta}, \theta)$ does not depend on θ.



3. Find $a < b$ such that $P(a < T(\hat{\theta}, \theta) < b) = 1 - \alpha$.
   Choose a and b such that $P(a \geq T(\hat{\theta}, \theta)) = \frac{\alpha}{2} = P(T(\hat{\theta}, \theta) \geq b)$
4. Transform the pivot confidence interval to a confidence interval for the parameter θ: $P(L < \theta < U) = 1 - \alpha$, where L is the lower confidence limit and U is the upper confidence limit.

T increasing

$L = T^{-1}(a)$

$U = T^{-1}(b)$

INTERVAL ESTIMATION USING PIVOTS

EXAMPLE

Hemoglobin levels in 11-year old boys have a normal distribution with unknown mean μ and $\sigma = 1.209\ g/dl$. Suppose that a random sample of size 10 has the sample mean 12 g/dl.

Find the 90% confidence interval for μ.

Solution.

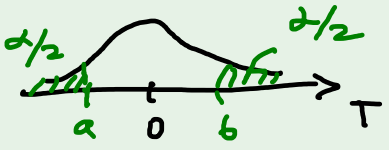
X - level of hemoglobin, $X \sim N(\mu, 1.209^2)$

1) $n = 10$, $\bar{x} = 12$, $\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) = N(\mu, \frac{1.209^2}{10})$

2) $T(\bar{X}, \mu) = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ - does not depend on μ

decreasing function wrt μ

3) $P(a < T(\bar{X}, \mu) < b) = 1 - \alpha = 0{,}9$, $\alpha = 0.1$



$\alpha/2 = 0{,}05$ $a = -1.65$, $b = 1{,}65$

$P(-1{,}65 < T < 1{,}65) = 0{,}9 \Rightarrow L = \bar{X} - 1{,}65 \cdot \frac{\sigma}{\sqrt{n}}$, $U = \bar{X} + 1{,}65 \cdot \frac{\sigma}{\sqrt{n}}$

4) $P(11.37 < \mu < 12.63) = 0{,}9 \Rightarrow$ 90% CI for μ is $(11.37, 12.63)$

INTERVAL ESTIMATION USING PIVOTS

EXAMPLE

Suppose the random sample $X_1, ..., X_n$ has $U(0, \theta)$ distribution. Construct a 90% confidence interval for θ. Identify the upper and lower confidence limits.

INTERVAL ESTIMATION USING PIVOTS

EXAMPLE

Suppose the random sample $X_1, ..., X_n$ has $U(0, \theta)$ distribution. Construct a 90% confidence interval for θ. Identify the upper and lower confidence limits.

Solution.

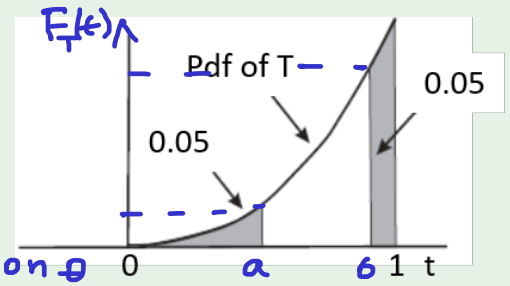
1) The MLE estimator for θ is $Y = \max_{1 \le i \le n} X_i$. $\quad y = \hat{\theta} = X_{(n)}$

The random variable Y has the pdf $f_Y(y) = ny^{n-1}/\theta^n$, is not independent of the parameter θ.

2) Consider $T = Y/\theta$

the pdf of T is $\quad f_T(t) = nt^{n-1},\ 0 \le t \le 1$

and the cdf is $\quad F_T(t) = t^n,\ 0 \le t \le 1$ $\quad$ do not depend on θ



3) $P(a<T<b) =$ $P(a \le \frac{Y}{\theta} \le b) = 0.90 \quad \Rightarrow F_T(a) = 0.05 \quad$ and $\quad F_T(b) = 0.95$

$$a = \sqrt[n]{0.05} \quad \text{and} \quad b = \sqrt[n]{0.95}$$

4) So the 90% confidence interval is:

$$P\left(\frac{Y}{\sqrt[n]{0.95}} \le \theta \le \frac{Y}{\sqrt[n]{0.05}}\right) = 0.90 \quad \Leftrightarrow \quad \left(\frac{Y}{\sqrt[n]{0.95}}, \frac{Y}{\sqrt[n]{0.05}}\right) \text{ for } \theta$$

INTERVAL ESTIMATION: FOR LARGE SAMPLE

EXAMPLE

A random sample of 50 mathematics grades out of a total of 200 showed a sample mean of 75 and a standard deviation of 10.
(a) What are the 95% confidence limits for estimates of the mean of the 200 grades?
(b) With what degree of confidence could we say that the mean of all 200 grades is 75 ± 1?

$(74, 76)$ $P(74 < \mu < 76) = 1 - \alpha$ – ?

INTERVAL ESTIMATION: FOR LARGE SAMPLE

EXAMPLE

A random sample of 50 mathematics grades out of a total of 200 showed a sample mean of 75 and a standard deviation of 10.
(a) What are the 95% confidence limits for estimates of the mean of the 200 grades?
(b) With what degree of confidence could we say that the mean of all 200 grades is 75 ± 1?

Solution:

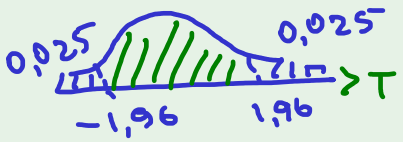
(a) Since the population size is not very large compared with the sample size, we must adjust for it:

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

$$T(\bar{X}, \mu) = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} \sim N(0, 1)$$

Then the 95% confidence limits are

$$P\left(L < \underbrace{\frac{\bar{X} - \mu}{\sigma_{\bar{X}}}}_{T} < U\right) = P\left(L < z < U\right) = 0.95$$



$L = z_{0.05} = -1.96$, $U = z_{0.95} = 1.96$, then

INTERVAL ESTIMATION: FOR LARGE SAMPLE

EXAMPLE

$$P\left(75 - 1.96 \cdot \frac{10}{\sqrt{50}}\sqrt{\frac{150}{199}} < \mu < 75 + 1.96 \cdot \frac{10}{\sqrt{50}}\sqrt{\frac{150}{199}}\right) = 0.95$$

$$P(75 - 2.4 < \mu < 75 + 2.4) = P(72.6 < \mu < 77.4) = 0.95$$

Hence, the 95% confidence interval for population mean is $(72.6, 77.4)$.

INTERVAL ESTIMATION: FOR LARGE SAMPLE

EXAMPLE

$$P\left(75 - 1.96 \cdot \frac{10}{\sqrt{50}}\sqrt{\frac{150}{199}} < \mu < 75 + 1.96 \cdot \frac{10}{\sqrt{50}}\sqrt{\frac{150}{199}}\right) = 0.95$$

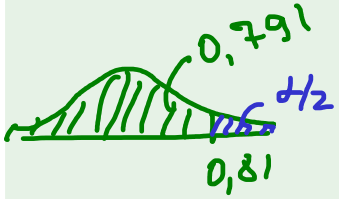
$$P(75 - 2.4 < \mu < 75 + 2.4) = P(72.6 < \mu < 77.4) = 0.95$$

Hence, the 95% confidence interval for population mean is $(72.6, 77.4)$.
(b)

$$P\left(75 - 1 < \mu < 75 + 1\right) = 1 - \alpha$$

$P(z_L < T < z_u) = 1 - \alpha$

$$-1 = z_L \cdot \frac{10}{\sqrt{50}}\sqrt{\frac{150}{199}}, \quad 1 = z_U \cdot \frac{10}{\sqrt{50}}\sqrt{\frac{150}{199}}$$



$$z_L = \frac{-1}{\frac{10}{\sqrt{50}}\sqrt{\frac{150}{199}}} = -0.81, \quad z_U = \frac{1}{\frac{10}{\sqrt{50}}\sqrt{\frac{150}{199}}} = 0.81$$

The area under the normal curve from $z = -\infty$ to $z = 0.81$ is 0.791, hence $\alpha/2 = 1 - 0.791 = 0.209$, and the required degree of confidence is $1 - \alpha = 1 - 2 \cdot 0.209 = 0.582$, or 58.2%. confidence level.

INTERVAL ESTIMATION: EXERCISES

Exercises for practice:

1. Let X be the mean of a random sample of size n from a distribution that is $N(\mu, 9)$. Find n such that $P(X - 1 < \mu < X + 1) = 0.90$, approximately.

2. Let a random sample of size 17 from the normal distribution $N(\mu, \sigma^2)$ yield $\bar{x} = 4.7$ and $\sigma^2 = 5.76$. Determine a 90% confidence interval for μ.

INTERVAL ESTIMATION USING PIVOTS

Theorem

Let X be a binomial variable defined on n independent trials for which $p = P(\text{success})$. For any a and b

$$\lim_{n\to\infty} P\left(a \leq \frac{X - np}{\sqrt{np(1-p)}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-z^2/2} dz$$

INTERVAL ESTIMATION USING PIVOTS

Theorem

Let X be a binomial variable defined on n independent trials for which $p = P(\text{success})$. For any a and b

$$\lim_{n\to\infty} P\left(a \leq \frac{X - np}{\sqrt{np(1-p)}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-z^2/2} dz$$

The random variable $\frac{X-np}{\sqrt{np(1-p)}}$ can be approximated by the standard normal distribution.

INTERVAL ESTIMATION USING PIVOTS

Theorem

Let X be a binomial variable defined on n independent trials for which $p = P(\text{success})$. For any a and b

$$\lim_{n\to\infty} P\left(a \le \frac{X - np}{\sqrt{np(1-p)}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-z^2/2} dz$$

The random variable $\frac{X-np}{\sqrt{np(1-p)}}$ can be approximated by the standard normal distribution.

Recall that X/n is a MLE for p, $E(X/n) = p$ and $Var(X/n) = p(1-p)/n$, then

$$\frac{X - np}{\sqrt{np(1-p)}} = \frac{X/n - p}{\sqrt{(X/n)(1-(X/n))/n}} \sim N(0,1),$$

INTERVAL ESTIMATION USING PIVOTS

Theorem

Let X be a binomial variable defined on n independent trials for which $p = P(\text{success})$. For any a and b

$$\lim_{n\to\infty} P\left(a \le \frac{X - np}{\sqrt{np(1-p)}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-z^2/2} dz$$

The random variable $\frac{X-np}{\sqrt{np(1-p)}}$ can be approximated by the standard normal distribution.

Recall that X/n is a MLE for p, $E(X/n) = p$ and $Var(X/n) = p(1-p)/n$, then

$$\frac{X - np}{\sqrt{np(1-p)}} = \frac{X/n - p}{\sqrt{(X/n)(1-(X/n))/n}} \sim N(0,1),$$

and a confidence interval:

$$P\left(-z_{\alpha/2} \le \frac{X/n - p}{\sqrt{(X/n)(1-(X/n))/n}} \le z_{\alpha/2}\right) = 1 - \alpha$$

INTERVAL ESTIMATION USING PIVOTS

$$T\left(\frac{X}{n}, p\right) = \frac{X/n - p}{\sqrt{(X/n)(1-(X/n))/n}}, \quad T\left(\frac{X}{n}, p\right) \sim N(0,1)$$

A *population proportion* is the proportion (percentage) of a population that has a specified attribute.

Exercise

Suppose 250 randomly selected people are surveyed to determine if they own a tablet. Of the 250 surveyed, 98 reported owning a tablet. Using a 95% confidence level, compute a confidence interval estimate for the true proportion of people who own tablets.

Exercise

A student polls her school to see if students in the school district are for or against the new legislation regarding school uniforms.

She surveys 600 students and finds that 480 are against the new legislation.

a. Compute a 90% confidence interval for the true percent of students who are against the new legislation, and interpret the confidence interval.

b. In a sample of 300 students, 68% said they own an iPad and an iPhone. Compute a 97% confidence interval for the true percent of students who own an iPad and an iPhone.

INTERVAL ESTIMATION USING PIVOTS

For X (the number of successes in n independent trials) from Binomial distribution, where n is large and $p = P(\text{success})$ is unknown.
Let $\hat{p} = X/n$ be an estimate for p where X is the number of successes in n trials, then $Var(X/n) = p(1-p)/n$, depends on p.
This means that, in general case, we cannot get exact confidence interval

$$\left[\frac{X}{n} - z_{\alpha/2}\sqrt{\frac{(X/n)(1-(X/n))}{n}}, \frac{X}{n} + z_{\alpha/2}\sqrt{\frac{(X/n)(1-(X/n))}{n}}\right],$$

but we can obtain an approximate $100(1-\alpha)\%$ confidence interval for p, if we note that $Var(X/n)$ has maximum value that at the point $p = 1/2$, then $Var(X/n) \leq 1/(4n)$ for any $0 < p < 1$.

INTERVAL ESTIMATION: MARGIN OF ERROR

Symmetric confidence intervals can be represented in the form

point estimate $\pm$ margin of error.

DEFINITION

The margin of error E for the estimate of μ is

$$E = z_{\alpha/2} \cdot \sigma / \sqrt{n}.$$

The margin of error for the estimate of a population mean indicates the accuracy with which a sample mean estimates the unknown population mean.

INTERVAL ESTIMATION: MARGIN OF ERROR

Let d be the width of a $(1-\alpha)\%$ confidence interval for the true proportion, p. Then

$$d = \frac{X}{n} + z_{\alpha/2}\sqrt{\frac{(X/n)(1-(X/n))}{n}} - \left(\frac{X}{n} - z_{\alpha/2}\sqrt{\frac{(X/n)(1-(X/n))}{n}}\right)$$

$$= 2z_{\alpha/2}\sqrt{\frac{(X/n)(1-(X/n))}{n}} \leq 2z_{\alpha/2}\sqrt{\frac{1}{4n}} = \frac{z_{\alpha/2}}{\sqrt{n}}$$

DEFINITION

*The **margin of error** for confidence level $100(1-\alpha)\%$ associated with an estimate $\frac{X}{n}$, where X is a number of success in n independent trials, and p is unknown, is*

$$E = \frac{z_{\alpha/2}}{2\sqrt{n}}.$$

INTERVAL ESTIMATION: MARGIN OF ERROR

Let d be the width of a $(1-\alpha)\%$ confidence interval for the true proportion, p. Then

$$d = \frac{X}{n} + z_{\alpha/2}\sqrt{\frac{(X/n)(1-(X/n))}{n}} - \left(\frac{X}{n} - z_{\alpha/2}\sqrt{\frac{(X/n)(1-(X/n))}{n}}\right)$$

$$= 2z_{\alpha/2}\sqrt{\frac{(X/n)(1-(X/n))}{n}} \leq 2z_{\alpha/2}\sqrt{\frac{1}{4n}} = \frac{z_{\alpha/2}}{\sqrt{n}}$$

DEFINITION

*The **margin of error** for confidence level $100(1-\alpha)\%$ associated with an estimate $\frac{X}{n}$, where X is a number of success in n independent trials, and p is unknown, is*

$$E = \frac{z_{\alpha/2}}{2\sqrt{n}}.$$

DEFINITION

*If we can estimate that true value of p is greater than $\frac{1}{2}$ (or less than $\frac{1}{2}$) the **margin of error** for confidence level $100(1-\alpha)\%$ associated with an estimate $\frac{X}{n}$ is*

$$E = \frac{z_{\alpha/2}\sqrt{p_g(1-p_g)}}{\sqrt{n}}.$$

INTERVAL ESTIMATION: SAMPLE SIZE

The minimal sample size required for estimation of the population mean μ at level $(1-\alpha)$ with margin of error E is given by

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{E^2}$$

and, if this is not an integer, rounding up to the next integer.

Similarly, a $(1-\alpha)$-level confidence interval for a population proportion p, if p is unknown, that has a margin of error of at most E can be obtained by choosing

$$n = \frac{z_{\alpha/2}^2}{4E^2}$$

rounded up to the next integer.
Or, if we can make "educated guess" about the value of p (p_g), then the smallest sample sample required is

$$n = \frac{z_{\alpha/2}^2}{E^2} p_g(1-p_g)$$

(Interval Estimation: Sample size)

EXAMPLE

Hemoglobin levels in 11-year old boys have a normal distribution with unknown mean μ and $\sigma = 1.209\ g/dl$. How large a sample is needed to estimate μ with 95% confidence and a margin of error of 0.5?

Solution. <for Exercise>

INTERVAL ESTIMATION: SAMPLE SIZE

EXAMPLE

The *Bureau of Labor Statistics* collects information on the ages of people in the civilian labor force and publishes the results in *Current Population Survey*.
i) Determine the sample size needed to be collected in order to be 95% confident that μ (mean age of all people in the civilian labor force) is within 0.5 year of the point estimate, $\bar{X}$. Assuming that $\sigma = 12.1$ years.
ii) Find a 95% confidence interval for μ if a sample of the size determined in part i) has a mean age of 43.8 years.

INTERVAL ESTIMATION: SAMPLE SIZE

EXAMPLE

The *Bureau of Labor Statistics* collects information on the ages of people in the civilian labor force and publishes the results in *Current Population Survey*.
i) Determine the sample size needed to be collected in order to be 95% confident that μ (mean age of all people in the civilian labor force) is within 0.5 year of the point estimate, $\bar{X}$. Assuming that $\sigma = 12.1$ years.
ii) Find a 95% confidence interval for μ if a sample of the size determined in part i) has a mean age of 43.8 years.

Solution.
i) $E = 0.5$, for 95% $\alpha = 0.05$ and $z_{\alpha/2} = 1.96$

$$n \geq \frac{z_{\alpha/2}^2 \sigma^2}{E^2} = \frac{1.96^2 \cdot 12.1^2}{0.05^2} \approx 2249.8$$

Hence, $n = 2250$.

INTERVAL ESTIMATION: SAMPLE SIZE

EXAMPLE

The *Bureau of Labor Statistics* collects information on the ages of people in the civilian labor force and publishes the results in *Current Population Survey*.
i) Determine the sample size needed to be collected in order to be 95% confident that μ (mean age of all people in the civilian labor force) is within 0.5 year of the point estimate, $\bar{X}$. Assuming that $\sigma = 12.1$ years.
ii) Find a 95% confidence interval for μ if a sample of the size determined in part i) has a mean age of 43.8 years.

Solution.

i) $E = 0.5$, for 95% $\alpha = 0.05$ and $z_{\alpha/2} = 1.96$

$$n \geq \frac{z_{\alpha/2}^2 \sigma^2}{E^2} = \frac{1.96^2 \cdot 12.1^2}{0.05^2} \approx 2249.8$$

Hence, $n = 2250$.

ii) $\bar{X} = 43.8$, hence the 95% CI is

$$\left[43.8 - 1.96\frac{12.1}{\sqrt{2250}},\ 43.8 + 1.96\frac{12.1}{\sqrt{2250}}\right] = [43.3,\ 44.3]$$

INTERVAL ESTIMATION: FOR LARGE SAMPLE

If θ is an unknown parameter (such as μ, p, $(\mu 1 - \mu 2)$, $(p1 - p2)$), then for large samples, by the **Central Limit Theorem**, the z-transform

$$z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}$$

possesses an approximately standard normal distribution, where $\hat{\theta}$ is the MLE of θ and $\sigma_{\hat{\theta}}$ is its standard deviation.

INTERVAL ESTIMATION: FOR LARGE SAMPLE

Procedure to calculate large sample confidence intervals for θ

- Find an estimator (such as the MLE) of θ say $\hat{\theta}$;

INTERVAL ESTIMATION: FOR LARGE SAMPLE

Procedure to calculate large sample confidence intervals for θ

- Find an estimator (such as the MLE) of θ say $\hat{\theta}$;
- Obtain the standard error, $\sigma_{\hat{\theta}}$ of $\hat{\theta}$;

INTERVAL ESTIMATION: FOR LARGE SAMPLE

Procedure to calculate large sample confidence intervals for θ

- Find an estimator (such as the MLE) of θ say $\hat{\theta}$;
- Obtain the standard error, $\sigma_{\hat{\theta}}$ of $\hat{\theta}$;
- Find the z-transform: $z = (\hat{\theta} - \theta)/\sigma_{\hat{\theta}}$. Then z has an approximately standard normal distribution;

INTERVAL ESTIMATION: FOR LARGE SAMPLE

Procedure to calculate large sample confidence intervals for θ

- Find an estimator (such as the MLE) of θ say $\hat{\theta}$;
- Obtain the standard error, $\sigma_{\hat{\theta}}$ of $\hat{\theta}$;
- Find the z-transform: $z = (\hat{\theta} - \theta)/\sigma_{\hat{\theta}}$. Then z has an approximately standard normal distribution;
- Using the normal table, find two tail values $-z_{\alpha/2}$ and $z_{\alpha/2}$;

INTERVAL ESTIMATION: FOR LARGE SAMPLE

Procedure to calculate large sample confidence intervals for θ

- Find an estimator (such as the MLE) of θ say $\hat{\theta}$;
- Obtain the standard error, $\sigma_{\hat{\theta}}$ of $\hat{\theta}$;
- Find the z-transform: $z = (\hat{\theta} - \theta)/\sigma_{\hat{\theta}}$. Then z has an approximately standard normal distribution;
- Using the normal table, find two tail values $-z_{\alpha/2}$ and $z_{\alpha/2}$;
- An approximate $(1-\alpha)100\%$ confidence interval for θ is

$$P\left(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right) = 1 - \alpha$$

INTERVAL ESTIMATION: FOR LARGE SAMPLE

Procedure to calculate large sample confidence intervals for θ

- Find an estimator (such as the MLE) of θ say $\hat{\theta}$;
- Obtain the standard error, $\sigma_{\hat{\theta}}$ of $\hat{\theta}$;
- Find the z-transform: $z = (\hat{\theta} - \theta)/\sigma_{\hat{\theta}}$. Then z has an approximately standard normal distribution;
- Using the normal table, find two tail values $-z_{\alpha/2}$ and $z_{\alpha/2}$;
- An approximate $(1-\alpha)100\%$ confidence interval for θ is

$$P\left(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right) = 1 - \alpha$$

- *Conclusion*: We are $(1-\alpha)100\%$ confident that the true parameter θ lies in the interval

$$\left(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}, \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right)$$

INTERVAL ESTIMATION: FOR LARGE SAMPLE

EXAMPLE

Two statistics professors want to estimate average scores for an elementary statistics course that has two sections. Each professor teaches one section and each section has a large number of students.

A random sample of 50 scores from each section produced the following results:

(a) Section I: $\bar{x}_1 = 77.01$, $s_1 = 10.32$ (sample standard deviation).

(b) Section II: $\bar{x}_2 = 72.22$, $s_2 = 11.02$.

Calculate 95% confidence intervals for each of these two samples.

INTERVAL ESTIMATION: FOR LARGE SAMPLE

EXAMPLE

Two statistics professors want to estimate average scores for an elementary statistics course that has two sections. Each professor teaches one section and each section has a large number of students.

A random sample of 50 scores from each section produced the following results:

(a) Section I: $\bar{x}_1 = 77.01$, $s_1 = 10.32$ (sample standard deviation).

(b) Section II: $\bar{x}_2 = 72.22$, $s_2 = 11.02$.

Calculate 95% confidence intervals for each of these two samples.

Solution.

For $\alpha = 0.05$, $z_{\alpha/2} = 1.96$, hence,

INTERVAL ESTIMATION: FOR LARGE SAMPLE

EXAMPLE

Two statistics professors want to estimate average scores for an elementary statistics course that has two sections. Each professor teaches one section and each section has a large number of students.

A random sample of 50 scores from each section produced the following results:

(a) Section I: $\bar{x}_1 = 77.01$, $s_1 = 10.32$ (sample standard deviation).

(b) Section II: $\bar{x}_2 = 72.22$, $s_2 = 11.02$.

Calculate 95% confidence intervals for each of these two samples.

Solution.

For $\alpha = 0.05$, $z_{\alpha/2} = 1.96$, hence,

$$\bar{x}_1 \pm z_{\alpha/2} \cdot \frac{s_1}{\sqrt{n}} = 77.01 \pm 1.96 \cdot \left(\frac{10.32}{\sqrt{50}}\right)$$

which gives a 95% confidence interval $(74.149,\ 79.871)$ $(to\ 3\ d.p.)$.

And

$$\bar{x}_2 \pm z_{\alpha/2} \cdot \frac{s_2}{\sqrt{n}} = 72.22 \pm 1.96 \cdot \left(\frac{11.02}{\sqrt{50}}\right)$$

which gives the interval $(69.165,\ 75.275)$ $(to\ 3\ d.p.)$.

INTERVAL ESTIMATION: FOR LARGE SAMPLE

EXAMPLE

Fifteen vehicles were observed at random for their speeds (in mph) on a highway with speed limit posted as 70 mph, and it was found that their average speed was 72.5 mph. Suppose that from past experience we can assume that vehicle speeds are normally distributed with $\sigma = 3.5$.
Construct a 90% confidence interval for the true mean speed μ, of the vehicles on this highway. Interpret the result.

INTERVAL ESTIMATION: FOR LARGE SAMPLE

EXAMPLE

Fifteen vehicles were observed at random for their speeds (in mph) on a highway with speed limit posted as 70 mph, and it was found that their average speed was 72.5 mph. Suppose that from past experience we can assume that vehicle speeds are normally distributed with $\sigma = 3.5$.
Construct a 90% confidence interval for the true mean speed μ, of the vehicles on this highway. Interpret the result.
Solution.
The population is given to be Normal with $\sigma = 3.5$. Hence, the distribution of the sample mean $\bar{x} = 72.5$ is also Normal.

INTERVAL ESTIMATION: FOR LARGE SAMPLE

EXAMPLE

Fifteen vehicles were observed at random for their speeds (in mph) on a highway with speed limit posted as 70 mph, and it was found that their average speed was 72.5 mph. Suppose that from past experience we can assume that vehicle speeds are normally distributed with $\sigma = 3.5$.

Construct a 90% confidence interval for the true mean speed μ, of the vehicles on this highway. Interpret the result.

Solution.

The population is given to be Normal with $\sigma = 3.5$. Hence, the distribution of the sample mean $\bar{x} = 72.5$ is also Normal.

For $\alpha = 0.1$, $z_{\alpha/2} = 1.645$, hence, the confidence interval for μ with the sample size $n = 15$:

$$72.5 - 1.645 \cdot \frac{3.5}{\sqrt{15}} < \mu < 72.5 + 1.645 \cdot \frac{3.5}{\sqrt{15}}$$

INTERVAL ESTIMATION: FOR LARGE SAMPLE

EXAMPLE

Fifteen vehicles were observed at random for their speeds (in mph) on a highway with speed limit posted as 70 mph, and it was found that their average speed was 72.5 mph. Suppose that from past experience we can assume that vehicle speeds are normally distributed with $\sigma = 3.5$.
Construct a 90% confidence interval for the true mean speed μ, of the vehicles on this highway. Interpret the result.

Solution.

The population is given to be Normal with $\sigma = 3.5$. Hence, the distribution of the sample mean $\bar{x} = 72.5$ is also Normal.
For $\alpha = 0.1$, $z_{\alpha/2} = 1.645$, hence, the confidence interval for μ with the sample size $n = 15$:

$$72.5 - 1.645 \cdot \frac{3.5}{\sqrt{15}} < \mu < 72.5 + 1.645 \cdot \frac{3.5}{\sqrt{15}}$$

$$71.013 < \mu < 73.987 \ \ (to\ 3\ d.p.)$$

INTERVAL ESTIMATION: FOR LARGE SAMPLE

EXAMPLE

Fifteen vehicles were observed at random for their speeds (in mph) on a highway with speed limit posted as 70 mph, and it was found that their average speed was 72.5 mph. Suppose that from past experience we can assume that vehicle speeds are normally distributed with $\sigma = 3.5$.
Construct a 90% confidence interval for the true mean speed μ, of the vehicles on this highway. Interpret the result.

Solution.

The population is given to be Normal with $\sigma = 3.5$. Hence, the distribution of the sample mean $\bar{x} = 72.5$ is also Normal.
For $\alpha = 0.1$, $z_{\alpha/2} = 1.645$, hence, the confidence interval for μ with the sample size $n = 15$:

$$72.5 - 1.645 \cdot \frac{3.5}{\sqrt{15}} < \mu < 72.5 + 1.645 \cdot \frac{3.5}{\sqrt{15}}$$

$$71.013 < \mu < 73.987 \ \ (to\ 3\ d.p.)$$

Interpretation: We are 90% confident that the true mean speed μ of the vehicles on this highway is between $(71.013,\ 73.987)$

KEY FACTS ABOUT CONFIDENCE INTERVALS

When to Use the One-Mean z-Interval Procedure:

- For small samples - of size less than 15 - the z-interval procedure should be used only when the variable under consideration is normally distributed or very close to being so.

KEY FACTS ABOUT CONFIDENCE INTERVALS

When to Use the One-Mean z-Interval Procedure:

- For small samples - of size less than 15 - the z-interval procedure should be used only when the variable under consideration is normally distributed or very close to being so.
- For samples of moderate size - between 15 and 30 - the z-interval procedure can be used unless the data contain outliers or the variable under consideration is far from being normally distributed.

KEY FACTS ABOUT CONFIDENCE INTERVALS

When to Use the One-Mean z-Interval Procedure:

- For small samples - of size less than 15 - the z-interval procedure should be used only when the variable under consideration is normally distributed or very close to being so.
- For samples of moderate size - between 15 and 30 - the z-interval procedure can be used unless the data contain outliers or the variable under consideration is far from being normally distributed.
- For large samples - of size 30 or more - the z -interval procedure can be used essentially without restriction. However, if outliers are present and their removal is not justified, you should compare the confidence intervals obtained with and without the outliers to see what effect the outliers have. If the effect is substantial, use a different procedure or take another sample, if possible.

KEY FACTS ABOUT CONFIDENCE INTERVALS

When to Use the One-Mean z-Interval Procedure:

- For small samples - of size less than 15 - the z-interval procedure should be used only when the variable under consideration is normally distributed or very close to being so.
- For samples of moderate size - between 15 and 30 - the z-interval procedure can be used unless the data contain outliers or the variable under consideration is far from being normally distributed.
- For large samples - of size 30 or more - the z -interval procedure can be used essentially without restriction. However, if outliers are present and their removal is not justified, you should compare the confidence intervals obtained with and without the outliers to see what effect the outliers have. If the effect is substantial, use a different procedure or take another sample, if possible.
- If outliers are present but their removal is justified and results in a data set for which the z-interval procedure is appropriate (as previously stated), the procedure can be used.

QUESTIONS TO TAKE HOME

1. Define the notion of confidence interval;

QUESTIONS TO TAKE HOME

1. Define the notion of confidence interval;
2. Define the notion of the confidence coefficient;

QUESTIONS TO TAKE HOME

1. Define the notion of confidence interval;
2. Define the notion of the confidence coefficient;
3. Define the notion of pivot random variable;

QUESTIONS TO TAKE HOME

1. Define the notion of confidence interval;
2. Define the notion of the confidence coefficient;
3. Define the notion of pivot random variable;
4. Define the main properties of the pivot RV;

QUESTIONS TO TAKE HOME

1. Define the notion of confidence interval;
2. Define the notion of the confidence coefficient;
3. Define the notion of pivot random variable;
4. Define the main properties of the pivot RV;
5. State the pivot for mean values of the population;

QUESTIONS TO TAKE HOME

1. Define the notion of confidence interval;
2. Define the notion of the confidence coefficient;
3. Define the notion of pivot random variable;
4. Define the main properties of the pivot RV;
5. State the pivot for mean values of the population;
6. When this pivot can be used;

QUESTIONS TO TAKE HOME

1. Define the notion of confidence interval;
2. Define the notion of the confidence coefficient;
3. Define the notion of pivot random variable;
4. Define the main properties of the pivot RV;
5. State the pivot for mean values of the population;
6. When this pivot can be used;
7. Which statements (theorems) allow us to construct CI for population proportion or for population mean;

QUESTIONS TO TAKE HOME

1. Define the notion of confidence interval;
2. Define the notion of the confidence coefficient;
3. Define the notion of pivot random variable;
4. Define the main properties of the pivot RV;
5. State the pivot for mean values of the population;
6. When this pivot can be used;
7. Which statements (theorems) allow us to construct CI for population proportion or for population mean;
8. State the pivot for proportion of a population;

QUESTIONS TO TAKE HOME

1. Define the notion of confidence interval;
2. Define the notion of the confidence coefficient;
3. Define the notion of pivot random variable;
4. Define the main properties of the pivot RV;
5. State the pivot for mean values of the population;
6. When this pivot can be used;
7. Which statements (theorems) allow us to construct CI for population proportion or for population mean;
8. State the pivot for proportion of a population;
9. When this pivot can be used;

QUESTIONS TO TAKE HOME

1. Define the notion of confidence interval;
2. Define the notion of the confidence coefficient;
3. Define the notion of pivot random variable;
4. Define the main properties of the pivot RV;
5. State the pivot for mean values of the population;
6. When this pivot can be used;
7. Which statements (theorems) allow us to construct CI for population proportion or for population mean;
8. State the pivot for proportion of a population;
9. When this pivot can be used;
10. Define the notion of margin error;

QUESTIONS TO TAKE HOME

1. Define the notion of confidence interval;
2. Define the notion of the confidence coefficient;
3. Define the notion of pivot random variable;
4. Define the main properties of the pivot RV;
5. State the pivot for mean values of the population;
6. When this pivot can be used;
7. Which statements (theorems) allow us to construct CI for population proportion or for population mean;
8. State the pivot for proportion of a population;
9. When this pivot can be used;
10. Define the notion of margin error;
11. How to define the minimal sample size required to construct a confidence interval at desired confidence level and a given margin of error;