Defn: Let $\hat{\theta}_1$ & $\hat{\theta}_2$ be two unbiased estimators of θ.
The estimator $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$ if $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$.

(P1). Let X_1, X_2, X_3 be a random sample of size 3 from a population with mean μ & variance $\sigma^2 > 0$. If the statistics $\bar{X}$ & Y given by $Y = \dfrac{X_1 + 2X_2 + 3X_3}{1+2+3}$ are two unbiased estimators (verify) of the population mean μ, then which one is more efficient?

Confidence Intervals:

An interval estimate for a population parameter is called a confidence interval.

Defn:
An interval $[a, b]$ is a $100(1-\alpha)\%$ confidence interval for the parameter θ if it contains the parameter with probability $(1-\alpha)$, $\boxed{P(a \le \theta \le b) = 1 - \alpha.}$

degree of confidence
↑(or)

The coverage probability $(1-\alpha)$ is also called a confidence level. Note that the coverage probability refers to the chance that our interval covers a constant parameter θ.

Confidence interval on the Mean of a Normal Distribution With Variance Known:-

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with unknown mean μ & known Variance σ^2.

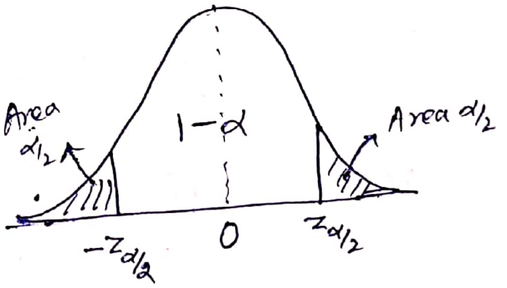
We know that the sample mean $\bar{X}$ is normally distributed with mean μ & variance $\frac{\sigma^2}{n}$

i.e $\quad \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

$\Rightarrow \quad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ has a standard normal distribution.

Now writing $z_{\alpha/2}$ for the z-value above which we find an area of $\alpha/2$. under the normal curve.

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha$$



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

Multiplying each term in the inequality by $\frac{\sigma}{\sqrt{n}}$ and then subtracting $\bar{x}$ from each term and multiplying by -1 (reversing the sense of inequalities), we obtain.

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

$$\Rightarrow P\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

Confidence interval on μ, σ^2 known.

If $\bar{x}$ is the value of the sample mean of a random sample of size 'n' from a normal population with known variance σ^2, a $100(1-\alpha)\%$ confidence interval for μ is given by

$$\boxed{\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

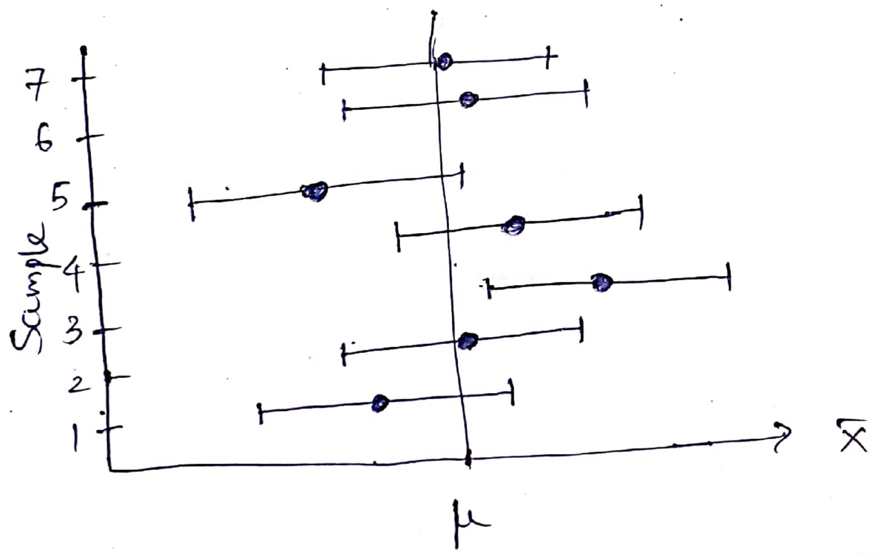
Where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

Interpreting a confidence interval.

Different samples will yield different values of $\bar{x}$ and therefore produce different interval estimates of the parameter μ, as shown in below figure. The dot at the center of each interval indicates the position of the point estimate $\bar{x}$ for that random sample. Note that all of these intervals are of the same width since their widths depend only on the choice of $z_{\alpha/2}$ once $\bar{x}$ is determined. The larger the value we choose for $z_{\alpha/2}$, the wider we make all the intervals and more confident we can be that the particular sample selected will produce an interval that contains the unknown parameter μ. In general, for a selection of $z_{\alpha/2}$, $100(1-\alpha)\%$

of the intervals will cover μ and 100α% of the intervals to miss it.

It is therefore wrong to say, "I computed a 90% confidence interval, it is [3,6]. Parameter belongs to this interval with Probability 90%. The parameter is constant, it either belongs to the interval [3,6] (with probability 1) or does not.



Interval estimates of μ for different samples.

(P1) The average zinc concentration recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 grams per ml. Find the 95% and 99% confidence intervals for the mean zinc concentration in the river. Assume that the population standard deviation is 0.3 gram per milliliter.

Solution:

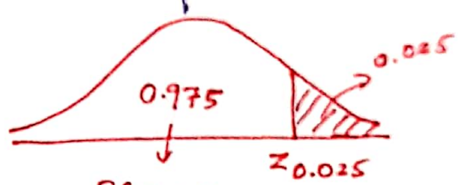
Given $\bar{x} = 2.6$, $n = 36$, $\sigma = 0.3$

(i) 95% confidence interval for μ:

$100(1-\alpha)\%$ Confidence interval for μ is

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Interval Estimate!



Here $100(1-\alpha)\% = 95\%$

$\Rightarrow \quad 1-\alpha = 0.95$

$\Rightarrow \quad \alpha = 0.05 \quad \Rightarrow \frac{\alpha}{2} = 0.025$

$P(Z \le z_{0.025}) = 0.975$

$\Rightarrow \Phi(z_{0.025}) = 0.975.$

The z-value leaving an area of 0.025 to the right, and therefore an area of 0.975 to the left, is $\boxed{z_{0.025} = 1.96}$ (from Standard Normal table).

Hence 95% confidence interval for μ is

$$2.6 - (1.96)\frac{(0.3)}{\sqrt{36}} < \mu < 2.6 + (1.96)\frac{(0.3)}{\sqrt{36}}$$

$$\Rightarrow \quad \boxed{2.50 < \mu < 2.70.}$$

(ii) To find a 99% confidence interval for μ:

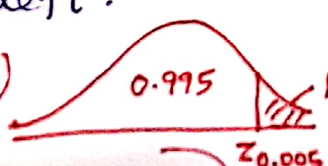
$100(1-\alpha)\% = 99\% \quad \Rightarrow 1-\alpha = 0.99 \quad \Rightarrow \alpha = 0.01 \Rightarrow \frac{\alpha}{2} = 0.005$

We find the z-value ($z_{\alpha/2}$) leaving an area of 0.005 to the right and 0.995 to the left. $\Rightarrow z_{0.005} = 2.58$ (from table)

$P(Z \le z_{\alpha/2}) = P(Z \le z_{0.005})$

$= \Phi(z_{0.005})$

$= 0.995 \Rightarrow z_{0.005} = 2.58$



The 99% Confidence interval is

$$2.6 - (2.58)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (2.58)\left(\frac{0.3}{\sqrt{36}}\right)$$

$$\Rightarrow \boxed{2.47 < \mu < 2.73}$$ ∎.

We now see that a longer interval is required to estimate μ with a higher degree of Confidence.

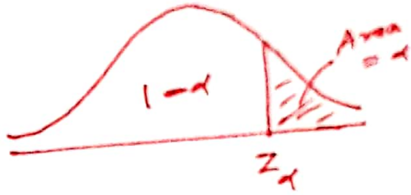
(P2). Let $X_1, X_2, \dots, X_{11}$ be a random sample of size 11 from a normal distribution with unknown mean μ & Variance $\sigma^2 = 9.9$. If $\sum_{i=1}^{11} x_i = 132$, then what is the 95% Confidence interval for μ?

(P3) Let us consider the data from the above Problem (P2), for what value of the constant 'k' is $(12 - k\sqrt{0.9}, 12 + k\sqrt{0.9})$ a 90% confidence interval for μ?

P4) Construct a 95% confidence intervals for the population mean based on a sample of measurements 2.5, 7.4, 8.0, 4.5, 7.4, 9.2. if measurement errors have Normal distribution, and the measurement device guarantees a standard deviation of $\sigma = 2.2$.

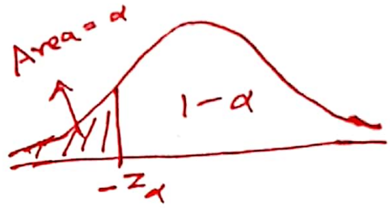
One Sided Confidence Bounds:

So far we discussed the confidence intervals are two sided (i.e., both upper & lower bounds are given). There are many applications in which only one bound is sought.



By central limit theorem.

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



$$\Rightarrow \quad P\left(\mu > \bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Similar manipulation of $P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > -z_\alpha\right) = 1-\alpha$

$$\Rightarrow P\left(\mu < \bar{X} + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

One sided confidence Bounds on μ, when σ^2 is known.

If $\bar{x}$ is the mean of a random sample of size n from a population with variance σ^2, the one-sided $100(1-\alpha)\%$ confidence bounds for μ are given by.

upper one-sided bound: $\bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$

Lower one-sided bound: $\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}$

Scanned with CamScanner

(P1). In a psychological testing experiment, 25 subjects are selected randomly and their reaction time, in seconds, to a particular Stimulus is measured. Past experience suggests that the variance in reaction times to these types of stimuli is 4 sec^2 and that the distribution of reaction times is approximately normal. The average time for the subjects is 6.2 seconds. Give an upper 95% bound for the mean reaction time.

Soln:

The upper 95% bound is given by $\bar{x} + Z_\alpha \frac{\sigma}{\sqrt{n}}$.

Here $n = 25$, $\sigma^2 = 4 \Rightarrow \sigma = 2$ & $\bar{x} = 6.2$

$1 - \alpha = 0.95$

$\alpha = 0.05 \Rightarrow Z_{0.05} = 1.645$

$$\therefore \bar{x} + Z_\alpha \frac{\sigma}{\sqrt{n}} = 6.2 + (1.645)\frac{(2)}{\sqrt{25}} = 6.2 + 0.658$$

$= 6.858$ seconds.

Hence, we are 95% confident that the mean reaction time is less than 6.858 seconds.

Confidence interval for μ when σ^2 is unknown

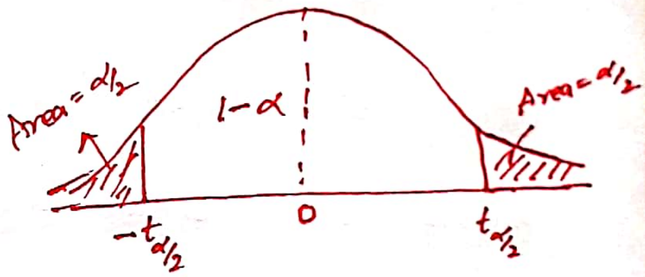
W.K.T if we have a random sample from a normal distribution, then the T-statistic $T = \dfrac{\bar{X}-\mu}{S/\sqrt{n}}$ has a Student t-distribution with n-1 degrees of freedom. Here S is sample standard deviation.

i.e, $$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}.$$

In this situation, with σ unknown, T can be used to construct a confidence interval on μ.

Referring to the Figure, we can assert that

$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1-\alpha.$$



Where $t_{\alpha/2}$ is the t-value with n-1 degrees of freedom, above which we find an area of $\alpha/2$. Because of symmetry, an equal area of $\alpha/2$ will fall to the left of $-t_{\alpha/2}$.

Now substituting for T, we write

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

Multiplying each term in the inequality by $\frac{S}{\sqrt{n}}$, and then subtracting $\bar{X}$ from each term and multiplying by -1, we get

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

Confidence Interval on μ, When σ^2 is Unknown

If $\bar{x}$ & s are the mean & standard deviation of a random sample from a normal population with unknown variance σ^2, a $100(1-\alpha)\%$ confidence interval for μ is

$$\boxed{\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}}$$ interval estimate of μ.

Where $t_{\alpha/2}$ is the t-value with $n-1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

One sided Confidence Bounds on μ, When σ^2 is Unknown.

The One-sided $100(1-\alpha)\%$ Confidence bounds for μ are

Upper one-sided bound : $\bar{x} + t_\alpha \frac{s}{\sqrt{n}}$

Lower one-sided bound : $\bar{x} - t_\alpha \frac{s}{\sqrt{n}}$

Here t_α is the t-value having an area of α to the right. With $n-1$ degrees of freedom.

(P1). The contents of seven similar containers of sulfuric acid are 9.8, 10.2, 10.4, 9.8, 10, 10.2, and 9.6 liters. Find a 95% confidence interval for the mean contents of all such containers, assuming an approximately normal distribution.

Soln: Sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, Sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$$= \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

Scanned with CamScanner

Here $n = 7$

$$\boxed{s^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}}$$

Sample mean $\bar{x} = 10.0$ & Sample standard deviation $s = 0.283$

$95\% = 100(1-\alpha)\%$

$\Rightarrow 1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$

$$t_{\alpha/2, n-1} = t_{0.025, 6} = 2.447$$

$100(1-\alpha)\%$ Confidence interval for μ when σ^2 is unknown

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

95% confidence interval for μ is

$$10 - (2.447)\frac{(0.283)}{\sqrt{7}} < \mu < 10 + (2.447)\left(\frac{0.283}{\sqrt{7}}\right)$$

$$\Rightarrow \boxed{9.74 < \mu < 10.26}.$$

Concept of a Large-Sample Confidence interval.

Often Statisticians recommend that even when normality cannot be assumed, σ is unknown, and $n \geq 30$, s can replace σ and the confidence interval

$$\boxed{\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}}$$ may be used.

This is often referred to as a large-Sample Confidence interval. It should be Omphasized that this is only an approximation and the quality of the result becomes better as the sample size grows larger.

(P1). Scholastic Aptitude Test (SAT) mathematics scores of a random sample of 500 high school seniors in the State of Texas are collected, and the sample mean and standard deviation are found to be 501 & 112, respectively. Find a 99% confidence interval on the mean SAT mathematics score for seniors in the State of Texas.

Soln:

Here $n = 500$, $\bar{x} = 501$, $s = 112$

Since the sample size is large, it is reasonable to use normal approximation.

$\Rightarrow$ $100(1-\alpha)\%$ confidence interval for μ when σ is unknown

$$\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}.$$

$\Rightarrow$ 99% confidence interval for μ is

$$\bar{x} - z_{0.005}\frac{s}{\sqrt{n}} < \mu < \bar{x} + z_{0.005}\frac{s}{\sqrt{n}}$$

Using std normal table, we find $z_{0.005} = 2.58$

$\therefore$ a 99% confidence interval for μ is

$$501 - (2.58)\frac{(112)}{\sqrt{500}} < \mu < 501 + (2.58)\frac{(112)}{\sqrt{500}}$$

$$\Rightarrow \quad 488.1 < \mu < 513.9.$$