Math 4501 - Probability and Statistics II

7.2 - Confidence intervals for the difference of two means

Overview

We will see how to determine confidence intervals for the difference of means of:

- two normal distributions with known variance
- two normal distributions with unknown but equal variance

We will construct approximate confidence intervals for the difference of means of:

- two unknown distributions, using a large sample
- two normal distributions with unknown and eventually unequal variances

Normal distributions with known variances

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be, respectively, two independent random samples of sizes n and m from normal distributions $N\left(\mu_X, \sigma_X^2\right)$ and $N\left(\mu_Y, \sigma_Y^2\right)$.

σ_x^2 and σ_y^2 both known!

Suppose that:

- σ_X^2 and σ_Y^2 are known.
- the random samples are independent;

Then:

- the respective sample means $\bar{X}$ and $\bar{Y}$ are independent
- $\bar{X}$ is $N\left(\mu_X, \sigma_X^2/n\right)$ ← because $X_1, \ldots, X_n \sim N(\mu_X, \sigma_X^2)$
- $\bar{Y}$ is $N\left(\mu_Y, \sigma_Y^2/m\right)$. ← because $Y_1, \ldots, Y_m \sim N(\mu_Y, \sigma_Y^2)$

$\boxed{\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)}$

$E[\bar{X} - \bar{Y}] = E[\bar{X}] - E[\bar{Y}] = \mu_X - \mu_Y$

$Var(\bar{X} - \bar{Y}) = Var(\bar{X}) + Var(\bar{Y})$

$= \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$

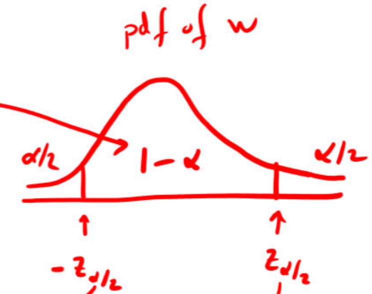
Consequently:

$$W = \frac{(\bar{X} - \bar{Y}) - \overbrace{(\mu_X - \mu_Y)}^{E[\bar{X} - \bar{Y}]}}{\sqrt{\underbrace{\sigma_X^2/n + \sigma_Y^2/m}_{Var(\bar{X} - \bar{Y})}}} \quad \text{is} \quad N(0,1) \ .$$

standardize $\bar{X} - \bar{Y}$

Note that

$$P\left(-z_{\alpha/2} \le \overbrace{\frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{\sigma_X^2/n+\sigma_Y^2/m}}}^{W \sim N(0,1)} \le z_{\alpha/2}\right) = 1-\alpha$$



solve inequalities for $\mu_X - \mu_Y$

can be rewritten as

$$P\left[(\bar{X}-\bar{Y})-z_{\alpha/2}\sigma_W \le \mu_X-\mu_Y \le (\bar{X}-\bar{Y})+z_{\alpha/2}\sigma_W\right] = 1-\alpha$$

where

$$\sigma_W = \sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}$$

is the standard deviation of $\bar{X}-\bar{Y}$.

Once the experiments have been performed and the means $\bar{x}$ and $\bar{y}$ computed, the interval

$$\left[\bar{x}-\bar{y}-z_{\alpha/2}\sigma_W, \bar{x}-\bar{y}+z_{\alpha/2}\sigma_W\right]$$

provides a $100(1-\alpha)\%$ confidence interval for $\mu_X-\mu_Y$.

Example

Let $X_1, X_2, \ldots, X_{15}$ be a random sample from a normal distribution with variance $\sigma_X^2 = 60$ and let $Y_1, Y_2, \ldots, Y_8$ be a random sample from a normal distribution with variance $\sigma_Y^2 = 40$.

Knowing that the observed sample means were $\bar{x} = 70.1$ and $\bar{y} = 75.3$, determine a 90% confidence interval for $\mu_X - \mu_Y$.

$1-\alpha = 0.9 \Rightarrow \alpha = 0.1 \Rightarrow \frac{\alpha}{2} = 0.05$

Note that:

- samples are both taken from normal distributions
- both variances are known

$X_1, \ldots, X_m \sim N(\mu_x, 60)$ $m = 15$ (σ_x^2)
$Y_1, \ldots, Y_m \sim N(\mu_y, 40)$, $m = 8$

Since $\bar{x} - \bar{y} = -5.2$, $z_{0.05} = 1.645$ and $z_{0.05}\sigma_W = 1.645\sqrt{\frac{60}{15} + \frac{40}{8}} = 4.935$:

$\bar{x} \sim N\left(\mu_x, \frac{60}{15}\right)$
$\bar{y} \sim N\left(\mu_y, \frac{40}{8}\right)$

$\Downarrow$

$W = \frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{60}{15}+\frac{40}{8}}} \sim N(0,1)$

$\mu_x - \mu_y \in$

$$
\begin{aligned}
& [\bar{x} - \bar{y} - z_{0.05}\sigma_W, \bar{x} - \bar{y} + z_{0.05}\sigma_W] = \\
= \quad & [-5.2 - 4.935, -5.2 + 4.935] = [-10.135, -0.265] .
\end{aligned}
$$

Because the confidence interval does not include zero, we suspect that μ_Y is greater than μ_X.

Normal distributions with unknown equal variances

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be, respectively, two independent random samples of sizes n and m from normal distributions $N\left(\mu_X, \sigma_X^2\right)$ and $N\left(\mu_Y, \sigma_Y^2\right)$.

Suppose that:

- σ_X^2 and σ_Y^2 are unknown.
- $\sigma_X^2 = \sigma_Y^2 = \sigma^2$
- the random samples are independent;

since $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ by assumption

instead of using S_X^2 to estimate σ_X^2 and S_Y^2 to estimate σ_Y^2, we use both samples to estimate common variance

Then

Just like previous case

$$\boxed{Z = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}} \quad \text{is} \quad N(0,1)}$$

σ^2 σ^2

and

$$U = \frac{(n-1)S_X^2}{\sigma^2} + \frac{(m-1)S_Y^2}{\sigma^2} \quad \text{is} \quad \chi^2(n+m-2)$$

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

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

$\Downarrow$ independence

$$U = \frac{(n-1)S_X^2}{\sigma^2} + \frac{(m-1)S_Y^2}{\sigma^2} \sim \chi^2(n+m-2)$$

and so

like 7.1 in the case of normal dist. with unknown variance

$$\boxed{T = \frac{Z}{\sqrt{U/(n+m-2)}} \quad \text{is} \quad t(n+m-2)\ .}$$

Note that

$N(0,1)$

$$T = \frac{Z}{\sqrt{U/(n+m-2)}} = \frac{\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\sigma^2/n+\sigma^2/m}}}{\sqrt{\left[\frac{(n-1)S_X^2}{\sigma^2}+\frac{(m-1)S_Y^2}{\sigma^2}\right]/(n+m-2)}} \sim t(n+m-2)$$

$\chi^2(n+m-2)$

can be rewritten as

$$T = \frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\underbrace{\sqrt{\left[\frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}\right]}}\left[\frac{1}{n}+\frac{1}{m}\right]}\ .$$

contains no σ^2!!!

If we let

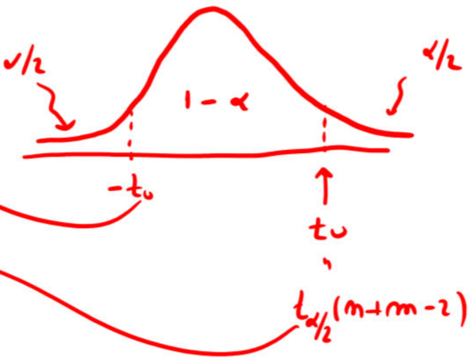
$$S_P = \sqrt{\frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}}$$

denote the pooled estimator of the common standard deviation, we can write

$$T = \frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{S_P\sqrt{\left[\frac{1}{n}+\frac{1}{m}\right]}} \quad \text{is} \quad t(n+m-2)\ .$$

Set $t_0 = t_{\alpha/2}(n+m-2)$ and note that

$$P\left(-t_0 \le \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_P\sqrt{\left[\frac{1}{n} + \frac{1}{m}\right]}} \le t_0\right) = 1 - \alpha$$



solve for $\mu_X - \mu_Y$

can be rewritten as

$$P\left(\bar{X} - \bar{Y} - t_0 S_P\sqrt{\frac{1}{n} + \frac{1}{m}} \le \mu_X - \mu_Y \le \bar{X} - \bar{Y} + t_0 S_P\sqrt{\frac{1}{n} + \frac{1}{m}}\right)$$

If $\bar{x}, \bar{y}$, and s_p are the observed values of $\bar{X}$, $\bar{Y}$, and S_P, then

$$\left[\bar{x} - \bar{y} - t_0 s_p\sqrt{\frac{1}{n} + \frac{1}{m}}, \bar{x} - \bar{y} + t_0 s_p\sqrt{\frac{1}{n} + \frac{1}{m}}\right]$$

is a $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$.

The assumption $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ is reasonable for some limited cases:

e.g.: sampling from the exact same distribution but at different instants of time

SIMPLIFYING ASSUMPTION:

- A point estimator for σ_X^2 is $S_X^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})^2$
- " " " " σ_Y^2 is $S_Y^2 = \frac{1}{(m-1)} \sum_{i=1}^{m} (Y_i - \bar{Y})^2$

If we assume that $\sigma_X^2 = \sigma_Y^2$, we can define a new estimator

$$\sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2} + \sum_{i=1}^{m} \frac{(Y_i - \bar{Y})^2}{\sigma^2} = \underbrace{\frac{(n-1)S_X^2}{\sigma^2}}_{\chi^2(n-1)} + \underbrace{\frac{(m-1)S_Y^2}{\sigma^2}}_{\chi^2(m-1)}$$

$$\sim \chi^2(n+m-2)$$

$$S_p^2 = \frac{1}{m+n-2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2 + \sum_{i=1}^{m}(y_i-\bar{y})^2\right] = \frac{1}{n+m-2}\cdot\left[(n-1)S_x^2 + (m-1)S_y^2\right]$$

so that

$$\frac{(m+n-2)\,S_p}{\sigma^2} \sim \chi^2(n+m-2)$$

Example

Let $X_1, X_2, \ldots, X_9$ and $Y_1, Y_2, \ldots, Y_{15}$ be two independent random samples from normal distributions with the same unknown variance σ^2.

$n = 9$ $m = 15$

Knowing that the observed sample means were $\bar{x} = 81.31$ and $\bar{y} = 78.61$ and the corresponding sample variances were $s_X^2 = 60.76$ and $s_Y^2 = 48.24$, determine a 95% confidence interval for $\mu_X - \mu_Y$.

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

Note that:

- samples are both taken from normal distributions
- variances are unknown but equal

$$\frac{(n+m-2)S_p^2}{\sigma^2} = \frac{(n-1)S_x^2 + (m-1)S_y^2}{\sigma^2} \sim \chi^2(n+m-2)$$

Set $n = 9$, $m = 15$ and note that $n + m - 2 = 22$ so that $t_0 = t_{0.025}(\overset{n+m-2}{22}) = 2.074$.
The endpoints of the confidence interval are

$$\bar{x} - \bar{y} \pm t_0 s_p \sqrt{\frac{1}{n} + \frac{1}{m}} = 81.31 - 78.61 \pm 2.074 \sqrt{\frac{8(60.76) + 14(48.24)}{22}} \sqrt{\frac{1}{9} + \frac{1}{15}}$$

$\sqrt{\frac{1}{n} + \frac{1}{m}}$

yielding the 95% confidence interval is $[-3.65, 9.05]$.

$$S_p = \sqrt{\frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}}$$

Large samples

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be random samples from distributions with unknown variances σ_X^2 and σ_Y^2, respectively.

For n and m large enough, we can use that:

NOT CLT Because σ_X^2 and σ_Y^2 are unknown

- $\dfrac{\bar{X}-\mu_X}{S_X/\sqrt{n}}$ is approximately $N(0,1)$
- $\dfrac{\bar{Y}-\mu_Y}{S_Y/\sqrt{m}}$ is approximately $N(0,1)$

$\Rightarrow W = \dfrac{\bar{X}-\bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}$ is approx. $N(0,1)$

where S_X^2 and S_Y^2 are the unbiased estimators of σ_X^2 and σ_Y^2, respectively.

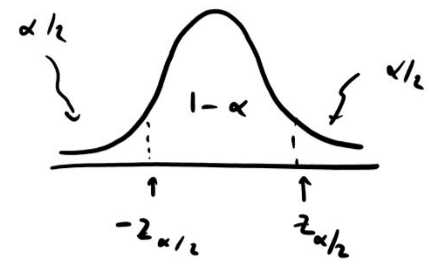
Consequently:

$$\boxed{W = \frac{(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{S_X^2/n + S_Y^2/m}}} \text{ is approximately } N(0,1) .$$

$\hookrightarrow \operatorname{Var}(\bar{X}-\bar{Y}) \underset{\text{ind.}}{=} \operatorname{Var}(\bar{X}) + \operatorname{Var}(\bar{Y}) \approx \dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}$

Note that

$$P\left(-z_{\alpha/2} \leq \overbrace{\frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{S_X^2/n+S_Y^2/m}}}^{W} \leq z_{\alpha/2}\right) \approx 1-\alpha$$



can be rewritten as

solve $\mu_X - \mu_Y$

$$\boxed{P\left[(\bar{X}-\bar{Y})-z_{\alpha/2}S_W \leq \mu_X-\mu_Y \leq (\bar{X}-\bar{Y})+z_{\alpha/2}S_W\right] \approx 1-\alpha}$$

where

$$\boxed{S_W = \sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}\ .}$$

If $\bar{x}$, $\bar{y}$, and s_W are the observed values of $\bar{X}$, $\bar{Y}$, and S_W, the interval

$$\left[\bar{x}-\bar{y}-z_{\alpha/2}s_W, \bar{x}-\bar{y}+z_{\alpha/2}s_W\right] \ ,$$

provides an approximate $100(1-\alpha)\%$ confidence interval for $\mu_X-\mu_Y$.

Note: if one or both of the variances happen to be known, we can use these known values instead of their estimates (sample variances).

Smaller samples and Welch's T

n, m not necessarily large

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be random samples from normal distributions with unknown variances σ_X^2 and σ_Y^2, respectively.

not necessarily the same !!!

Welch's T: We can resort to the fact that

$$\boxed{T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{S_X^2/n + S_Y^2/m}}} \text{ is approximately } t(r),$$

$t(n)$ has larger tails than $N(0,1)$

BUT $t(n) \to N(0,1)$ as $n \to \infty$

where r is the integer part of

$$\frac{\left(\frac{s_x^2}{n} + \frac{s_y^2}{m}\right)^2}{\frac{1}{n-1}\left(\frac{s_x^2}{n}\right)^2 + \frac{1}{m-1}\left(\frac{s_y^2}{m}\right)^2}.$$

$\frac{1}{n^2}, \frac{1}{m^2}, \frac{1}{nm}$

$\frac{1}{n^3}$ $\quad$ $1/m^3$

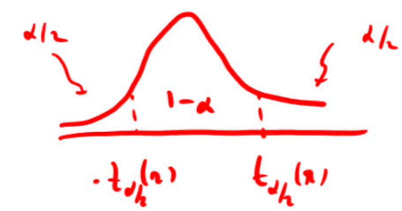
$\lfloor 1.5 \rfloor = 1$

$\lfloor \pi \rfloor = 3$

$\lfloor 23.75 \rfloor = 23$

NOTE: if n and m are large, then r increases and $t(r) \to N(0,1)$

Note that



$$P\left(-t_{\alpha/2}(r) \le \underbrace{\frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{S_X^2/n+S_Y^2/m}}}_{T} \le t_{\alpha/2}(r)\right) \approx 1-\alpha$$

solve for $\mu_X - \mu_Y$

can be rewritten as

$$P\left[(\bar{X}-\bar{Y})-t_{\alpha/2}(r)S_W \le \mu_X-\mu_Y \le (\bar{X}-\bar{Y})+t_{\alpha/2}(r)S_W\right] \approx 1-\alpha$$

where

$$S_W = \sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}.$$

If $\bar{x}$, $\bar{y}$, and s_W are the observed values of $\bar{X}$, $\bar{Y}$, and S_W, the interval

$$\left[\bar{x}-\bar{y}-t_{\alpha/2}(r)s_W, \bar{x}-\bar{y}+t_{\alpha/2}(r)s_W\right] ,$$

provides an approximate $100(1-\alpha)\%$ confidence interval for $\mu_X-\mu_Y$.

Dependent samples

In some applications, two measurements are taken on the same subject, usually at different instants of time, resulting in dependent random variables.

Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be n pairs of dependent measurements and set

$$D_i = X_i - Y_i \ , \quad i = 1, 2, \ldots, n.$$

Suppose that $D_1, D_2, \ldots, D_n$ can be thought of as a random sample from a

ASSUMPTION $\rightarrow$ $N\left(\mu_D, \sigma_D^2\right)$ distribution, where μ_D and σ_D are the sample mean and sample standard deviation of the n differences.

$E[D_i] = E[X_i - Y_i] = \underbrace{\mu_X - \mu_Y}_{\mu_D}$

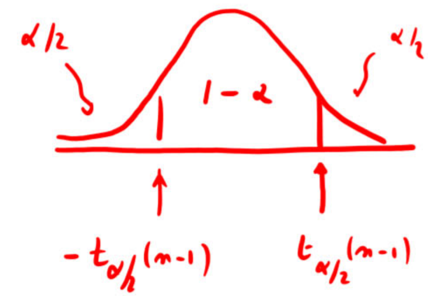
To form a confidence interval for $\mu_X - \mu_Y$, we use the fact that

Sec. 7.1
Confidence interval for the mean of a normal distr. with unknown variance $\rightarrow$

$$\underbrace{T = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}} \quad \text{is} \quad t(n-1)}\ ,$$

Note that

$$P\left(-t_{\alpha/2}(n-1) \leq \overbrace{\frac{\bar{D}-\mu_D}{S_D/\sqrt{n}}}^{T \sim t(n-1)} \leq t_{\alpha/2}(n-1)\right) = 1-\alpha$$



can be rewritten as

solve for $\mu_D = \mu_X - \mu_Y$

$$P\left[\bar{D} - t_{\alpha/2}(n-1)\frac{S_D}{\sqrt{n}} \leq \underset{\uparrow\ \mu_X - \mu_Y}{\mu_D} \leq \bar{D} + t_{\alpha/2}(n-1)\frac{S_D}{\sqrt{n}}\right] = 1-\alpha\ .$$

If $\bar{d}$ and s_d are the observed mean and standard deviation of the sample of the D values, the interval

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

provides a $100(1-\alpha)\%$ confidence interval for $\mu_D = \mu_X - \mu_Y$.

Example

An experiment was conducted to compare people's reaction times to a red light versus a green light. When signaled with either the red or the green light, the subject was asked to hit a switch to turn off the light. When the switch was hit, a clock was turned off and the reaction time in seconds was recorded. The following results give the reaction times for eight subjects:

$D_i = X_i - Y_i$

$\mu_D = \mu_X - \mu_Y$

Subject	Red (x) X_i	Green (y) Y_i	$d = x - y$ D_i
1	0.30	0.43	–0.13
2	0.23	0.32	–0.09
3	0.41	0.58	–0.17
4	0.53	0.46	0.07
5	0.24	0.27	–0.03
6	0.36	0.41	–0.05
7	0.38	0.38	0.00
8	0.51	0.61	–0.10

sample size $m = 8$

$\bar{D} = \frac{1}{m} \sum_{i=1}^{m} D_i$

$S_D^2 = \frac{1}{m-1} \sum_{i=1}^{m} (D_i - \bar{D})^2$

$D_i \sim N(\mu_D, \sigma_D^2)$

$\Rightarrow \frac{\bar{D} - \mu_D}{S_D/\sqrt{m}} \sim t(m-1)$

$1 - \alpha = 0.95$

$\Rightarrow \alpha = 0.05$

Determine a 95% confidence interval for $\boxed{\mu_D = \mu_X - \mu_Y}$.

Assume both samples are normally distributed. $\alpha/2$ table

Set $n = 8$ and note that $n - 1 = 7$ so that $t_{0.025}(7) = 2.365$. $n-1 = 8-1 = 7$

The differences mean and standard deviation are $\bar{d} = -0.0625$ and $s_d = 0.0765$.

The endpoints of the confidence interval are

$$\boxed{\bar{d} \pm t_{0.025}(7)\frac{s_d}{\sqrt{8}} = -0.0625 \pm 2.365\frac{0.0765}{\sqrt{8}}}$$

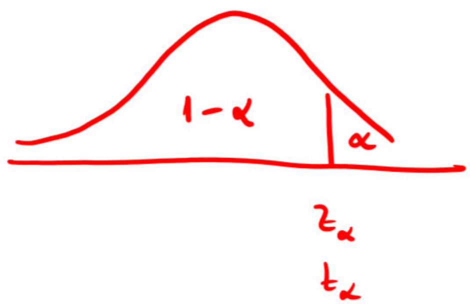
yielding the 95% confidence interval is $[-0.1265, 0.0015]$.

Interpretation:

- zero is included in the confidence interval but is close to the endpoint 0.0015.
- if more data were taken and zero happened to no longer be included in the confidence interval, this would suggest that people react faster to a red light.

One-sided confidence intervals

The techniques developed above can be used to obtain one-sided confidence intervals, in a similar way to what was done when discussing confidence intervals for the mean of a single sample.



Summary 1

Difference of means $\mu_X - \mu_Y$ of normal distributions $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$ with known variances σ_X^2 and σ_Y^2

- $X_1, \ldots, X_n$ random sample of size n
- $Y_1, \ldots, Y_m$ random sample of size m
- confidence level $1 - \alpha$
- sample means $\bar{x}$ and $\bar{y}$
- standard deviation of difference $\bar{X} - \bar{Y}$ is $\sigma_W = \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$
- two-sided confidence interval: $\bar{x} - \bar{y} \pm z_{\alpha/2}\sigma_W$
- upper bound: $\bar{x} - \bar{y} + z_{\alpha}\sigma_W$
- lower-bound: $\bar{x} - \bar{y} - z_{\alpha}\sigma_W$

Summary 2

Difference of means $\mu_X - \mu_Y$ of normal distributions $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$ with unknown but equal variances $\sigma_X^2 = \sigma_Y^2$

- $X_1, \ldots, X_n$ random sample of size n
- $Y_1, \ldots, Y_m$ random sample of size m
- confidence level $1 - \alpha$
- sample means $\bar{x}$ and $\bar{y}$
- sample variances s_X^2 and s_Y^2
- pooled estimate for common standard deviation $s_P = \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}}$
- two-sided confidence interval: $\bar{x} - \bar{y} \pm t_{\alpha/2}(n+m-2)s_p\sqrt{\frac{1}{n} + \frac{1}{m}}$
- upper bound: $\bar{x} - \bar{y} + t_{\alpha}(n+m-2)s_p\sqrt{\frac{1}{n} + \frac{1}{m}}$
- lower-bound: $\bar{x} - \bar{y} - t_{\alpha}(n+m-2)s_p\sqrt{\frac{1}{n} + \frac{1}{m}}$

Summary 3

Difference of means $\mu_X - \mu_Y$ of distributions with unknown variances σ_X^2 and σ_Y^2

- Large sample $X_1, \dots, X_n$ of size $n \geq 30$
- Large sample $Y_1, \dots, Y_m$ of size $m \geq 30$
- confidence level $1 - \alpha$
- sample means $\bar{x}$ and $\bar{y}$
- unpooled estimate for standard deviation of $\bar{X} - \bar{Y}$ is $s_W = \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}$
- two-sided confidence interval: $\bar{x} - \bar{y} \pm z_{\alpha/2} s_W$
- upper bound: $\bar{x} - \bar{y} + z_\alpha s_W$
- lower-bound: $\bar{x} - \bar{y} - z_\alpha s_W$

Note: In case σ_X^2 and σ_Y^2 are known, use $\sigma_W = \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$ instead of s_W.

Summary 4

Difference of means $\mu_X - \mu_Y$ of normal distributions $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$ with unknown (eventually unequal) variances σ_X^2 and σ_Y^2

- $X_1, \ldots, X_n$ random sample of size n
- $Y_1, \ldots, Y_m$ random sample of size m
- confidence level $1 - \alpha$
- sample means $\bar{x}$ and $\bar{y}$
- sample variances s_X^2 and s_Y^2
- unpooled estimate for standard deviation of $\bar{X} - \bar{Y}$ is $s_W = \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}$
- r is the integer part of $\left(\frac{s_x^2}{n} + \frac{s_y^2}{m}\right)^2 \Big/ \left[\frac{1}{n-1}\left(\frac{s_x^2}{n}\right)^2 + \frac{1}{m-1}\left(\frac{s_y^2}{m}\right)^2\right]$
- two-sided confidence interval: $\bar{x} - \bar{y} \pm t_{\alpha/2}(r) s_W$
- upper bound: $\bar{x} - \bar{y} + t_{\alpha}(r) s_W$
- lower-bound: $\bar{x} - \bar{y} - t_{\alpha}(r) s_W$

Summary 5

Difference of means $\mu_X - \mu_Y$ for distributions with (eventually unknown) variances σ_X^2 and σ_Y^2 using (eventually) dependent random samples.

- $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ pairs of dependent measurements
- set $D_i = X_i - Y_i$, for $i = 1, 2, \ldots, n$
- **Assumption**: $D_1, D_2, \ldots, D_n$ normally distributed
- confidence level $1 - \alpha$
- sample mean of differences $\bar{d}$
- sample variance of differences s_D^2
- two-sided confidence interval: $\bar{d} \pm t_{\alpha/2}(n-1)s_D/\sqrt{n}$
- upper bound: $\bar{d} + t_{\alpha}(n-1)s_D/\sqrt{n}$
- lower-bound: $\bar{d} - t_{\alpha}(n-1)s_D/\sqrt{n}$