Classical hypothesis tests are close relatives of the classical confidence intervals.
Some general statements will be introduced after the first example.

Example 11.01

The lifetime X of a particular brand of filaments is known to be normally distributed. A random sample of six filaments is tested to destruction. Those six filaments are found to last for an average of 1,007 hours with a sample standard deviation of 6.2 hours.

Is there sufficient evidence to conclude, at a level of significance of 5%, that the true mean lifetime of this brand of filaments is not 1,000 hours?

Repeat this question with a level of significance of 1%.

Test the **null hypothesis** $\mathcal{H}_o : \mu = 1000$

against the **alternative hypothesis** $\mathcal{H}_a : \mu \neq 1000$.

Distribution: $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$

Data: $n = 6, \quad \bar{x} = 1007, \quad s = 6.2$

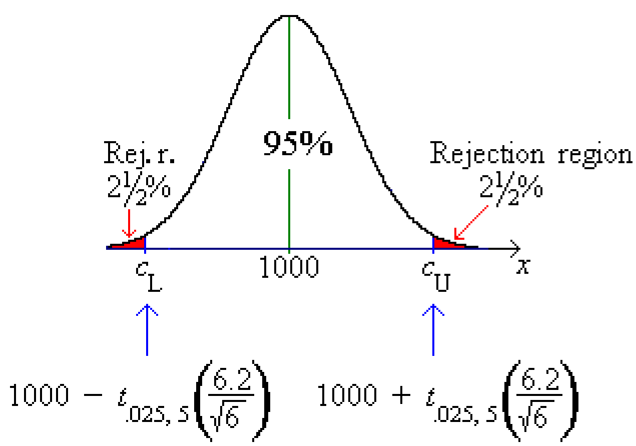
If $\mathcal{H}_o$ is true, then

$$\bar{X} \sim \mathrm{N}\left(1000, \frac{\sigma^2}{6}\right) \quad \Rightarrow \quad Z = \frac{\bar{X} - 1000}{\left(\sigma / \sqrt{6}\right)} \sim \mathrm{N}(0,1)$$

But σ is not known.

$$\Rightarrow \quad T = \frac{\bar{X} - 1000}{\left(S / \sqrt{6}\right)} \sim t_5$$

$t_{.025,\,5} \approx$ **2.57058**



[Note: "S" is upper case because it is a random quantity.
$\nu = n - 1 = 5$ is the number of degrees of freedom for the t distribution.]

Example 11.01 (continued)

Method 1

Reject $\mathcal{H}_o$ in favour of $\mathcal{H}_a$ iff $\bar{x} > c_U$ or $\bar{x} < c_L$

$\because \quad \mathrm{P}\left[\left(\bar{X} > c_U \quad \text{or} \quad \bar{X} < c_L\right) \middle| \mathcal{H}_\circ \text{ true}\right] = 5\%$

$c_L, c_U = 1000 \pm 2.57\ldots \times \frac{6.2}{\sqrt{6}} = 1000 \pm 6.51\ldots$

$\therefore \quad c_L, c_U = [993.5,\ 1006.5] \quad (\text{to 1 d.p.})$

$\bar{x} = 1007 > c_U$

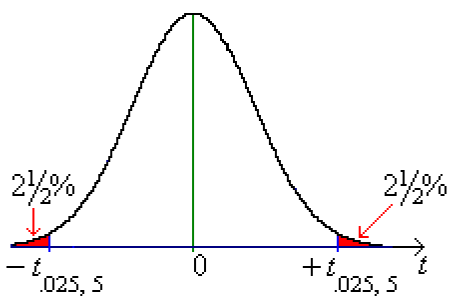
Therefore REJECT $\mathcal{H}_o$ at a level of significance of $\alpha = .05$.
[This result is equivalent to the classical two-sided confidence interval of example 10.04.]

OR

Method 2

$\bar{x} = 1007 \Rightarrow$

$$t_{\text{obs}} = \frac{\bar{x} - \mu_\circ}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{1007 - 1000}{\left(\frac{6.2}{\sqrt{6}}\right)} = \frac{7}{2.531\ldots} = 2.77 \quad (2 \text{ d.p.})$$



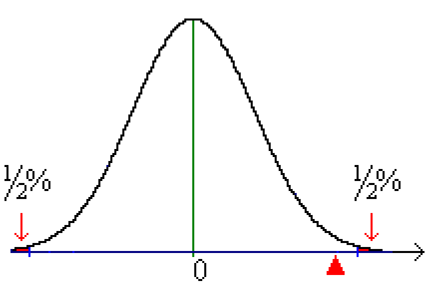
Reject $\mathcal{H}_o$ in favour of $\mathcal{H}_a$ iff

$$|t_{\text{obs}}| > t_{.025,5}$$

2.77... > 2.57...

Therefore REJECT $\mathcal{H}_o$ at a level of significance of $\alpha = .05$.

At $\alpha = 1\%$,



Method 1

$c_L, c_U = \mu_\circ \pm t_{.005,5}\left(\frac{s}{\sqrt{n}}\right)$

$= 1000 \pm 4.03\ldots \times \frac{6.2}{\sqrt{6}} = 1000 \pm 10.20\ldots$

$c_L, c_U = [989.8,\ 1010.2]$ but $\bar{x} = 1007$

$c_L < \bar{x} < c_U$. **Do NOT reject $\mathcal{H}_o$.**

Example 11.01 (continued)

α = 1%, Method 2

$\bar{x} = 1007 \quad \Rightarrow \quad t_{\text{obs}} \approx 2.77$

$t_{.005,\,5} \;=\; 4.03\ldots$

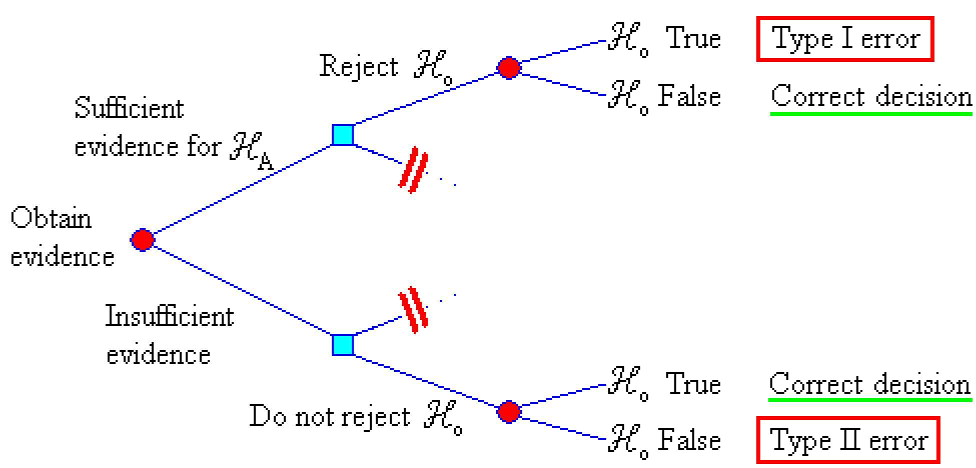
$|\, t_{\text{obs}} \,| \not> t_{.005,\,5}$

Therefore do NOT reject $\mathcal{H}_o$.

Interpretation:

If $\mathcal{H}_o$ is true, then the ***p*-value** (the probability that $\bar{X}$ is further away from $\mu = 1000$ than $\bar{x} = 1007$) is between 5% and 1%. The level of significance α is an upper bound to the probability of committing a type I error: P[reject $\mathcal{H}_o$ | $\mathcal{H}_o$ true] ≤ α .

Decision Tree: **[from page 9.19]**



P[type I error] = P[reject $\mathcal{H}_o$ | $\mu = \mu_o$ ($\mathcal{H}_o$ true)] ≤ α

P[type II error] = P[accept $\mathcal{H}_o$ | $\mu = \mu_1$ ($\mathcal{H}_o$ false)] = $\beta(\mu_1)$

$1 - \beta$ = **power** of the test.

General method for two-tailed tests:

State hypotheses:

$$\mathcal{H}_o : \quad \mu = \mu_o \quad \text{vs.} \quad \mathcal{H}_a : \quad \mu \neq \mu_o$$

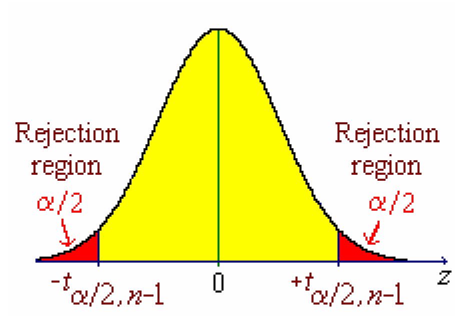
The burden of proof is on $\mathcal{H}_a$.

Choose the **level of significance** α .

State your **assumptions**
(for example, the random quantity X
is nearly normal).

Find $\bar{x}$ (the **test statistic**).

If σ is unknown, then estimate it using s .



Case 1: σ is unknown and n is small

	$\bar{x}$ space		t space
Find	$\mu_o \pm t_{\alpha/2,\, n-1}\left(\frac{s}{\sqrt{n}}\right)$	Find	$t_{\alpha/2,\, n-1}$
Iff	$\bar{x} < \mu_o - t_{\alpha/2,\, n-1}\left(\frac{s}{\sqrt{n}}\right)$	and	$t_{obs} = \dfrac{\bar{x}-\mu_o}{\left(\frac{s}{\sqrt{n}}\right)}$
or	$\bar{x} > \mu_o + t_{\alpha/2,\, n-1}\left(\frac{s}{\sqrt{n}}\right)$	Iff	$\lvert t_{obs} \rvert > t_{\alpha/2, n-1}$

then reject $\mathcal{H}_o$ in favour of $\mathcal{H}_a$.

Case 2: n is large (> 30) is the same as Case 1 except that $t_{\alpha/2,\, n-1}$ is replaced by $t_{\alpha/2,\, \infty} = z_{\alpha/2}$.

Common values: $z_{.025} = 1.95996$, $z_{.005} = 2.57583$.

Case 3: σ is known is the same as Case 2 except that s is replaced by σ .

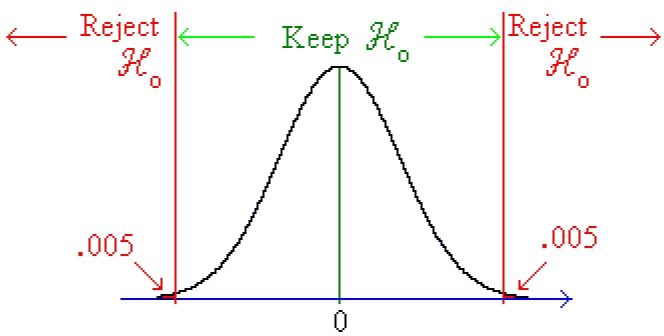
<u>Example 11.02</u>

A manufacturer claims that replacement machinery fills paper bags with exactly one kilogramme of sugar each, on average. A random sample of 400 bags of sugar is weighed, producing a sample mean mass of 996.5 grammes and a sample standard deviation of 25.1 grammes. At a level of significance of .01, is there sufficient evidence to doubt the manufacturer's claim?

$n = 400$, $\bar{x} = 996.5$, $s = 25.1$

Test $\mathcal{H}_o: \mu = 1000$ vs. $\mathcal{H}_a: \mu \neq 1000$ **at** $\alpha = .01$

[Reason for selecting a two-sided alternative hypothesis rather than one-sided: Before we have any data to examine, if the manufacturer's claim is false, then we have no pre-conceptions as to whether the true value of μ is greater than or less than 1000. We are seeking only evidence that μ is different from 1000. We are not seeking evidence, *a priori*, for a decrease.]



<u>Method 1</u>

$$\mu_\circ \pm t_{\alpha/2,\,n-1}\left(\frac{s}{\sqrt{n}}\right)$$

$$= 1000 \pm t_{.005,\,399} \times \frac{25.1}{\sqrt{400}}$$

$$\approx 1000 \pm 2.60 \times 1.255$$

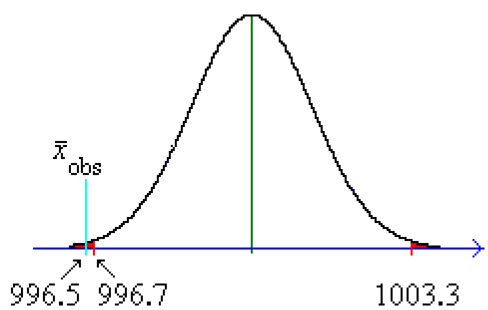
$$= 1000 \pm 3.263$$

$$= (996.7,\ 1003.3) \quad (1\text{ d.p.})$$

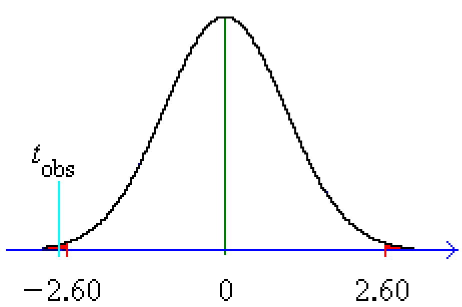
$\bar{x} = 996.5 < c_L$

Therefore reject $\mathcal{H}_o$ **.**

<u>YES</u>, $\mu \neq 1000$.



<u>Method 2</u>



$t_{.005,\,399} \approx t_{.005,\,200} = 2.60...$

$$t_{\text{obs}} = \frac{\bar{x} - \mu_\circ}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{996.5 - 1000}{\left(\frac{25.1}{\sqrt{400}}\right)} = -2.78\ldots$$

$$|t_{\text{obs}}| = 2.78\ldots > 2.60\ldots$$

Therefore reject $\mathcal{H}_o$ **.** **<u>YES</u>, $\mu \neq 1000$.**

***p*-value** (Method 3)**:**

Find $z_{\text{obs}} = \dfrac{\bar{x} - \mu_\circ}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}$ or $t_{\text{obs}} = \dfrac{\bar{x} - \mu_\circ}{\left(\dfrac{s}{\sqrt{n}}\right)}$

Find $p = \text{P}[\,|Z| > |z_{\text{obs}}|\,]$ or $p = \text{P}[\,|T| > |t_{\text{obs}}|\,]$

Compare p to α .

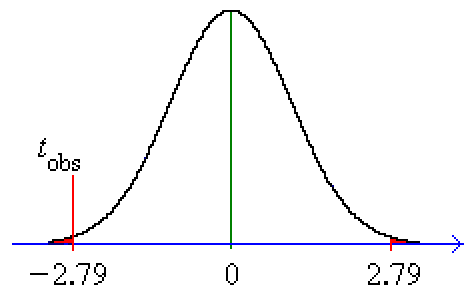
Example 11.02 (continued, using method 3):

$t_{\text{obs}} = -2.78... = -2.79$ (2 d.p.)

Using $t_{\alpha, 399} \approx z_\alpha$,

$\text{P}[\,|Z| > 2.79] = 2\,\Phi(-2.79)$

$= 2 \times .00264 = .00528 < .01000 = \alpha$.



Therefore reject $\mathcal{H}_o$. YES, $\mu \neq 1000$.

Note:

Tables are not usually provided for $\text{P}[T < t_{\text{obs}}]$,
but the values can be obtained from software, such as the Excel file at
`www.engr.mun.ca/~ggeorge/3423/demos/t1test.xls`.

$t_{.005, 399} = 2.588204... \rightarrow c_L = 996.7518...$, $c_U = 1003.248...$

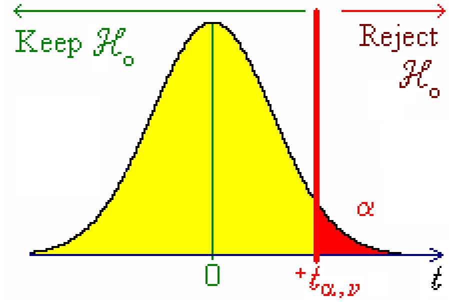
$t_{\text{obs}} = -2.78884... \rightarrow p = \text{P}[\,|T| > t_{\text{obs}}\,] = .005543$

The corresponding, more precise, confidence interval allows us to claim that "we are 99% sure that $993.25... < \mu \leq 999.74...$".

General Method (upper-tailed tests):

State hypotheses:

$\mathcal{H}_o : \quad \mu = \mu_o \quad$ vs. $\quad \mathcal{H}_a : \quad \mu > \mu_o$



The burden of proof is on $\mathcal{H}_a$.

Choose the **level of significance** α .

State your **assumptions**
(for example, the random quantity X is nearly normal).

Find $\bar{x}$ (the **test statistic**).

If σ is unknown, then estimate it using s .

Method 1:
Evaluate

$$c = \mu_\circ + t_{\alpha,(n-1)}\left(\frac{s}{\sqrt{n}}\right)$$

Reject $\mathcal{H}_o$ iff $\bar{x} > c$.

Method 2:
Reject $\mathcal{H}_o$ iff

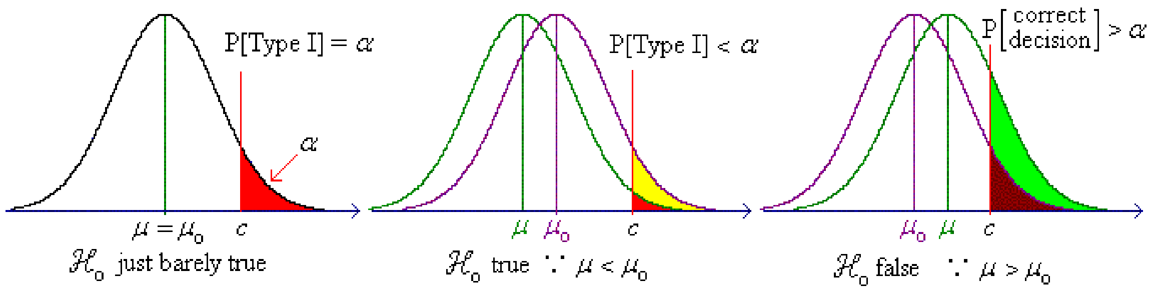
$$t_{\text{obs}} = \frac{\bar{x} - \mu_\circ}{\left(\frac{s}{\sqrt{n}}\right)} > t_{\alpha,(n-1)}$$

Method 3:

Evaluate $t_{\text{obs}} = \dfrac{\bar{x} - \mu_\circ}{\left(\frac{s}{\sqrt{n}}\right)}$ and $p = \text{P}[\,T > t_{\text{obs}}]$

Reject $\mathcal{H}_o$ iff $p < \alpha$.

Let us explore the meaning of α, the probability of committing a Type I error, in the case when the alternative hypothesis is one (upper) tailed, $\mathcal{H}_a : \quad \mu > \mu_o$:

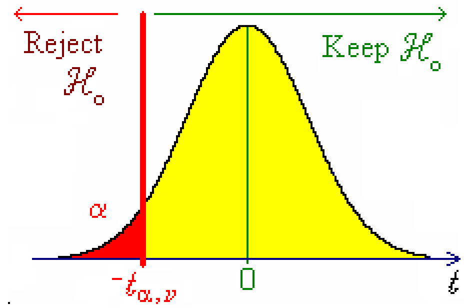


α is actually an upper bound to P[Type I error], the "worst case scenario", which occurs when the null hypothesis is just barely true.

General Method (lower-tailed tests):

State hypotheses:

$\mathcal{H}_o: \quad \mu = \mu_o \quad$ vs. $\quad \mathcal{H}_a: \quad \mu < \mu_o$



The burden of proof is on $\mathcal{H}_a$.

Choose the **level of significance** α .

State your **assumptions**
(for example, the random quantity X is nearly normal).

Find $\bar{x}$ (the **test statistic**).

If σ is unknown, then estimate it using s .

Method 1:
Evaluate

$$c = \mu_o - t_{\alpha,(n-1)} \left(\frac{s}{\sqrt{n}} \right)$$

Reject $\mathcal{H}_o$ iff $\bar{x} < c$.

Method 2:
Reject $\mathcal{H}_o$ iff

$$t_{\text{obs}} = \frac{\bar{x} - \mu_o}{\left(\frac{s}{\sqrt{n}} \right)} < -t_{\alpha,(n-1)}$$

Method 3:

Evaluate $t_{\text{obs}} = \dfrac{\bar{x} - \mu_o}{\left(\frac{s}{\sqrt{n}} \right)}$ and $p = \mathrm{P}[\, T < t_{\text{obs}}]$

Reject $\mathcal{H}_o$ iff $p < \alpha$.

Example 11.03

An opinion poll of 100 randomly selected customers produces 58 customers who state a preference for brand A. Does a majority of the population of customers prefer brand A?

From the random sample of 100 customers, how many must state a preference for brand A in order for the inference "a majority of the population of customers prefers brand A" to be valid?

$\mathcal{H}_o$: $p = .5$ (or less)

$\mathcal{H}_a$: $p > .5$

Choose $\alpha = .05$

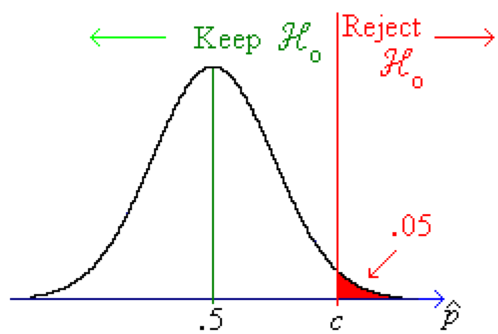
Assume that the sample is random, so that, to a good approximation,

$$\hat{P} \sim \mathrm{N}\left(p, \frac{pq}{n}\right)$$

$$\hat{p} = \frac{x}{n} = \frac{58}{100} = .58 \quad \Rightarrow \quad \hat{q} = 1 - \hat{p} = .42$$

If $\mathcal{H}_o$ is true, then $\frac{pq}{n} = \frac{.5 \times .5}{100} = .0025$

Use method 1 (because of the second part of the question).



$$c = \text{"}\mu_o + z_\alpha \frac{\sigma}{\sqrt{n}}\text{"}$$

$$= p_o + z_{.05}\sqrt{\frac{p_o q_o}{n}}$$

$$= .5 + 1.64\ldots \times \sqrt{.0025}$$

$$= .582\ldots$$

$\hat{p} = .58 < c$, $\therefore$

do NOT reject $\mathcal{H}_o$.

There is insufficient evidence for a majority.

$c = .582\ldots \Rightarrow x = 58.2$. Therefore

$$x_{\min} = \underline{\underline{59}}$$

Two sample z test

From the **central limit theorem**, we know that, for sufficiently large sample sizes from two independent populations of means μ_1, μ_2 and variances σ_1^2, σ_2^2, the sample means are distributed as

$$\bar{X}_1 \sim \mathrm{N}\left(\mu_1, \frac{\sigma_1^2}{n_1}\right), \qquad \bar{X}_2 \sim \mathrm{N}\left(\mu_2, \frac{\sigma_2^2}{n_2}\right), \text{ with } \quad \bar{X}_1 - \bar{X}_2 \sim \mathrm{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Example 11.04

A large corporation wishes to determine the effectiveness of a new training technique. A random sample of 64 employees is tested after undergoing the new training technique and obtains a mean test score of 62.1 with a standard deviation of 5.12 . Another random sample of 100 employees, serving as a control group, is tested after undergoing the old training methods. The control group has a sample mean test score of 58.3 with a standard deviation of 6.30 .

(a) Use a two-sided confidence interval to determine whether the new training technique has led to a significant *change* in test scores.

(b) Use an appropriate hypothesis test to determine whether the new training technique has led to a significant *increase* in test scores.

(a)

$n_1 = 64 \quad \bar{x}_1 = 62.1 \quad s_1 = 5.12$

$n_2 = 100 \quad \bar{x}_2 = 58.3 \quad s_2 = 6.30$

Two different groups of employees; may assume independence.
Both sample sizes are large (>> 30) $\Rightarrow$ normal. Choose $\alpha = 1\%$.

$\bar{x}_1 - \bar{x}_2 = 62.1 - 58.3 = 3.8$

$$\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} = \frac{5.12^2}{64} + \frac{6.30^2}{100} = 0.8065$$

The 99% CI for $\mu_1 - \mu_2$ has its boundaries at

$3.8 \pm 2.575\ldots \times \sqrt{0.8065} \approx 3.8 \pm 2.31 = [1.49, 6.11]$

The CI does not include 0.
Therefore YES, the new training technique has led to a significant change in test scores.

[Note that if $t_{.005,\,162} = 2.60\ldots$ is used instead of $z_{.005}$, then the CI would be $3.8 \pm 2.34\ldots$ instead of $3.8 \pm 2.31\ldots$, leading to *no change* to 1 d.p.!
It is usually valid to replace t by z when $\nu > 100$.]

Example 11.04 (continued)

(b)

$$\mathrm{V}\left[\overline{X}_1 - \overline{X}_2\right] = \left(\frac{5.12^2}{64} + \frac{6.30^2}{100}\right) \rightarrow \overline{X}_1 - \overline{X}_2 \sim \mathrm{N}(\mu, 0.8065)$$

Seeking evidence for an increase.

Therefore use an upper-tailed test. [Again choose $\alpha = 1\%$].

Test $\mathcal{H}_o : \mu_1 - \mu_2 = 0$ vs. $\mathcal{H}_a : \mu_1 - \mu_2 > 0$.

$$\mathcal{H}_o \text{ true} \quad \Rightarrow \quad \overline{X}_1 - \overline{X}_2 \sim \mathrm{N}(0, 0.8065)$$

Method 1:

$$c = "\mu_\circ + z_\alpha \sigma" = 0 + 2.32\ldots \times \sqrt{0.8065} = 2.089\ldots$$

$$\overline{x}_1 - \overline{x}_2 = 3.8 > c$$

Therefore reject $\mathcal{H}_o$ in favour of $\mathcal{H}_a : \mu_1 - \mu_2 > 0$.

[Expressed crudely, "we are 99% sure that the training process has increased test scores."]

Method 2:

$$t_{\text{obs}} = \frac{(\overline{x}_1 - \overline{x}_2) - \Delta_\circ}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{3.8 - 0}{\sqrt{0.8065}} = 4.23\ldots$$

$t_{\alpha,\infty} = z_\alpha = 2.32...$

$t_{\text{obs}} > z_\alpha$

Therefore reject $\mathcal{H}_o$ in favour of $\mathcal{H}_a : \mu_1 - \mu_2 > 0$.

Method 3:

$t_{\text{obs}} = 4.23...$

$\mathrm{P}[\,Z > t_{\text{obs}}] = \Phi(-4.23...) < .0003$ (from Table A.3)

OR, using `www.engr.mun.ca/~ggeorge/3423/demos/tCalculator.xls`

with 63+99 = 162 degrees of freedom,

$\mathrm{P}[\,T > t_{\text{obs}}] = .0000194... <$ any reasonable α .

Therefore reject $\mathcal{H}_o$ in favour of $\mathcal{H}_a : \mu_1 - \mu_2 > 0$.

General Method (Method 2 illustrated here):

Establish the null hypothesis $\mathcal{H}_o : \mu_1 - \mu_2 = \Delta_o$ (often $\Delta_o = 0$)

Select the appropriate alternative hypothesis $\mathcal{H}_a$.

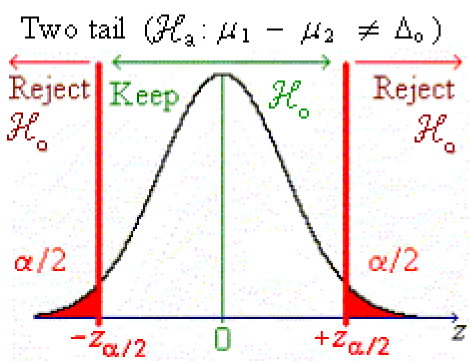
Select the level of significance α , which leads to the boundaries of the rejection region for z
(assuming either σ known or large n or both):

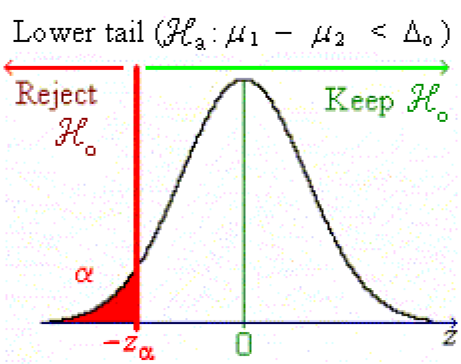
z_c	$\alpha = 5\%$	$\alpha = 1\%$
1 - tail	1.64485	2.32634
2 - tail	1.95996	2.57583

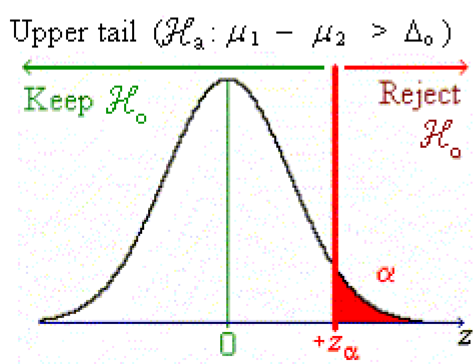
Find

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_o}{\sqrt{\dfrac{s_1^{\,2}}{n_1} + \dfrac{s_2^{\,2}}{n_2}}}$$

Compare z to z_c .







Two sample *t* test:

If n_1 and/or n_2 is/are small (< 30) **and** the population variances are both equal to an unknown number ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) **and** the random quantities X_1 and X_2 are independent and have normal (or nearly normal) distributions, then a *t* test may be used.

The separate sample variances s_1^2 and s_2^2 are both point estimates of the same unknown population parameter σ^2. A better point estimate of σ^2 is a weighted average of these two estimates, with the weights given by the numbers of degrees of freedom. Thus both sample variances are replaced by the **pooled sample variance**

$$\boxed{s_P^2 = \frac{\nu_1 s_1^2 + \nu_2 s_2^2}{\nu_1 + \nu_2}} \qquad \text{where} \quad \nu_1 = n_1 - 1 \ \text{ and } \ \nu_2 = n_2 - 1 .$$

In the hypothesis test, $z = \dfrac{(\bar{x}_1 - \bar{x}_2) - \Delta_o}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ is replaced by $\boxed{t = \dfrac{(\bar{x}_1 - \bar{x}_2) - \Delta_o}{\sqrt{s_P^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}}$,

which has $\nu = \nu_1 + \nu_2$ degrees of freedom.

<u>Example 11.05</u>

An investigator wants to know which of two electric toasters has the greater ability to resist the abnormally high electrical currents that occur during an unprotected power surge. Random samples of six toasters from factory *A* and five toasters from factory *B* were subjected to a destructive test, in which each toaster was subjected to increasing currents until it failed. The distribution of currents at failure (measured in amperes) is known to be approximately normal for both products, with a common (but unknown) population variance. The results are as follows:

Factory *A*:	20	28	24	26	23	26
Factory *B*:	21	18	19	17	22	

(a) State the hypotheses that are to be tested.
(b) State the assumptions that you are making.
(c) Conduct the appropriate hypothesis test.

(a) $\mathcal{H}_o$: $\boldsymbol{\mu_A - \mu_B = 0}$ (no difference between toasters)

$\mathcal{H}_a$: $\boldsymbol{\mu_A - \mu_B \neq 0}$ (significant difference between toasters)

[In advance of examining the data, we have no preconceptions of which toaster might be better.]

(b) Given in the question:

$X_A \sim \mathrm{N}\left(\mu_A, \sigma^2\right)$

$X_B \sim \mathrm{N}\left(\mu_B, \sigma^2\right)$

Assumption:

X_A, X_B are independent.

(c) The **summary statistics** are

$n_A = 6$ $\quad \bar{x}_A = 24.5$ $\quad s_A = 2.81 \ldots$

$n_B = 5$ $\quad \bar{x}_B = 19.4$ $\quad s_B = 2.07 \ldots$

$\nu_A = n_A - 1 = 5$ and $\nu_B = n_B - 1 = 4 \Rightarrow \nu = 5 + 4 = 9$

$$s_P^{\,2} = \frac{\nu_A s_A^{\,2} + \nu_B s_B^{\,2}}{\nu_A + \nu_B} = \frac{5 \times (2.81\ldots)^2 + 4 \times (2.07\ldots)^2}{5+4} \approx 6.300$$

$$\text{standard error} = s_P \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} = \sqrt{6.300\ldots \times \left(\frac{1}{6} + \frac{1}{5}\right)} = 1.519\ldots$$

$$\Rightarrow \quad t_{\text{obs}} = \frac{\left(\bar{x}_A - \bar{x}_B\right) - \Delta_o}{\sqrt{s_P^{\,2}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} = \frac{(24.5 - 19.4) - 0}{1.519\ldots} \approx \underline{\underline{\mathbf{3.356}}}$$

With $\alpha = .01$, $t_{\alpha/2,\,\nu} = t_{.005,\,9} = \mathbf{3.249\ldots}$

$|\, t_{\text{obs}} \,| > t_{\alpha/2,\,\nu}$, therefore reject $\mathcal{H}_o$ in favour of $\mathcal{H}_a$: $\mu_A - \mu_B \neq 0$.

From the data, we can conclude, with a high level of confidence, that toaster *A* is more robust.

Paired t test

Example 11.06

Nine volunteers are tested before and after a training programme. Based on the data below, can you conclude that the programme has improved test scores?

Volunteer:	1	2	3	4	5	6	7	8	9
After training:	75	66	69	45	54	85	58	91	62
Before training:	72	65	64	39	51	85	52	92	58

Let X_A = score **a**fter training and X_B = score **b**efore training.

$$\text{Test} \quad \mathcal{H}_o: \mu_A - \mu_B = 0 \quad \text{vs.} \quad \mathcal{H}_a: \mu_A - \mu_B > 0$$

Choose $\alpha = .01$.

INCORRECT METHOD:

$$n_A = n_B = 9 \quad \Rightarrow \quad \nu_A = \nu_B = 8 \quad \Rightarrow \quad \nu = 16$$

$\bar{x}_A = 67.222...$ $\qquad s_A = 14.695...$

$\bar{x}_B = 64.222...$ $\qquad s_B = 16.820...$

$$s_P^{\,2} = \frac{\nu_A s_A^{\,2} + \nu_B s_B^{\,2}}{\nu_A + \nu_B} = \frac{8 \times (14.695...)^2 + 8 \times (16.820...)^2}{8+8}$$

$$= \frac{215.9444... + 282.9444...}{2} = 249.444...$$

$$\Rightarrow \quad \text{s.e.} = \sqrt{s_P^{\,2}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)} = \sqrt{249.\dot{4}\left(\frac{1}{9} + \frac{1}{9}\right)} = 7.445\ldots$$

$$\Rightarrow \quad t = \frac{(\bar{x}_A - \bar{x}_B) - \Delta_o}{\text{s.e.}} = \frac{(67.\dot{2} - 64.\dot{2}) - 0}{7.445...} \approx \underline{\underline{\mathbf{0.403}}}$$

Compare with $t_{\alpha,\,\nu} = t_{.010,\,16} = 2.583...$

$0.403 \not> 2.583$ Therefore do **not** reject $\mathcal{H}_o$: no increase in test scores !

The error is that

the two test scores are NOT independent.

[They are highly correlated.]

The **correct method** is to take account of the fact that X_A and X_B are paired, by examining the **differences** $D = X_A - X_B$.

Volunteer:	1	2	3	4	5	6	7	8	9
After training x_A:	75	66	69	45	54	85	58	91	62
Before training x_B:	72	65	64	39	51	85	52	92	58
Difference d	**3**	**1**	**5**	**6**	**3**	**0**	**6**	**−1**	**4**

Test $\mathcal{H}_o : \mu_D = 0$ vs. $\mathcal{H}_a : \mu_D > 0$ with $\alpha = .01$.

Summary statistics:

$$n = 9 \quad \Rightarrow \quad \nu = 8 \quad , \quad \bar{d} = 3 \; , \quad s_D = 2.5495...$$

$$\Rightarrow \qquad t = \frac{\bar{d} - \mu_{Do}}{\left(\dfrac{s_D}{\sqrt{n}}\right)} = \frac{3 - 0}{\left(\dfrac{2.5495...}{\sqrt{9}}\right)} \approx \underline{\underline{\mathbf{3.530}}}$$

Compare with $t_{\alpha,\, \nu} = t_{.010,\, 8} = 2.896...$

Therefore **reject** $\mathcal{H}_o$.

At a 1% level of significance, we conclude that the training has, indeed, increased the test scores.

An Excel spreadsheet file for both methods is available at
`http://www.engr.mun.ca/~ggeorge/3423/demos/t2test.xls` .

When should we use a paired two sample *t* test?

When samples of equal size n are taken from two populations, the unpaired two sample t test will have $\nu = 2n - 2$ degrees of freedom, but the paired two sample t test will have only $\nu = n - 1$ degrees of freedom. The power of the unpaired test to distinguish between null and alternative hypotheses is greater, especially for small sample sizes.

The paired test is valid even if the two populations are strongly correlated, whereas the unpaired test is based on the assumption that the two populations are independent (or at least uncorrelated).

We should use the paired t test if there is reason to believe that the two populations from which the samples come may be correlated, or if the variance within the samples is high.

If the samples are pairs of observations of two different effects on the ***same set of individuals***, then independence between the populations is unlikely and one should use the **paired** t test.

Otherwise, (and especially if the sample size is very small), use the unpaired t test.

Note (<u>not examinable</u>):

The correlation ρ is a measure of the linear dependence of a pair of random quantities.

Independence $\Rightarrow$ $\rho = 0$

The relationship between the t statistics for the unpaired and paired two sample t tests is

$$T_{\text{pair}} = \frac{T_{\text{unpair}}}{\sqrt{1-\rho}}$$

The unpaired t test can therefore be used only if the random quantities are uncorrelated. And, upon replacing the unknown underlying true correlation ρ by the observed sample correlation coefficient r, the two observed values of t are related by

$$t_{\text{pair}} = \frac{t_{\text{unpair}}}{\sqrt{1 - \dfrac{2 r s_A s_B}{s_A^{\,2} + s_B^{\,2}}}}$$

where s_A and s_B are the two observed standard deviations from samples A and B respectively.

In Example 11.06, $r = .996$, leading to an error factor of 8.76... .
$t_{\text{unpair}} = 0.402...$, $t_{\text{pair}} = 3.53...$ and one can verify that
$3.53... = 0.402... \times 8.76...$

Inferences on Differences in Population Proportions
[not examinable (except for bonus)]

We have seen that the sample proportion $\hat{P}$ is distributed approximately as $\hat{P} \sim \mathrm{N}\left(p, \frac{pq}{n}\right)$,
where n is the sample size, p is the population proportion and $q = 1 - p$.
This approximation holds provided that np (the expected number of successes) and nq (the expected number of failures) are both sufficiently large (both numbers greater than 10 is usually sufficient).

We have also seen that for any two random quantities X, Y:
$\mathrm{E}[X - Y] = \mathrm{E}[X] - \mathrm{E}[Y]$ and
for any two *uncorrelated* random quantities X, Y: $\mathrm{V}[X - Y] = \mathrm{V}[X] + \mathrm{V}[Y]$.

For two independent large random samples, it then follows that

$$\hat{P}_1 - \hat{P}_2 \sim \mathrm{N}\left((p_1 - p_2), \left(\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} \right) \right)$$

$\Rightarrow$ a $(1-\alpha)\times 100\%$ confidence interval estimate for $p_1 - p_2$ is

$$\hat{p}_1 - \hat{p}_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

A **special case** arises in hypothesis tests whenever the null hypothesis is $\mathcal{H}_o : p_1 = p_2$.
In this case the two sample proportions are point estimates of the same unknown population proportion p.

The **pooled estimate of p** is

$$\hat{p} = \frac{\text{Total number of successes}}{\text{Total sample size}} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$$

and the standard error becomes

$$s = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} .$$

Compare $z = \dfrac{\hat{p}_1 - \hat{p}_2}{s}$ to $z_{\alpha/2}$ (two tailed test),
or $-z_\alpha$ (lower tailed test) or z_α (upper tailed test).

Example 11.07

A random sample of 100 customers produces 42 customers who like brand A (as opposed to not liking brand A). Another random sample of 225 customers produces 81 customers who like brand B.

(a) Find a standard 95% confidence interval for the difference in population proportions $p_A - p_B$.

(b) Is there sufficient evidence to conclude, at a level of significance of five per cent, that brand A is more popular than brand B?

(a) $x_A = 42 \qquad n_A = 100 \qquad \Rightarrow \qquad \hat{p}_A = .42$

$x_B = 81 \qquad n_B = 225 \qquad \Rightarrow \qquad \hat{p}_B = .36$

$$V\left[\hat{P}_A - \hat{P}_B\right] \approx s^2 = \frac{\hat{p}_A \hat{q}_A}{n_A} + \frac{\hat{p}_B \hat{q}_B}{n_B} = \frac{.42 \times .58}{100} + \frac{.36 \times .64}{225} = .002436 + .001024$$

$$= .003460$$

The 95% confidence interval estimate is

$$\hat{p}_A - \hat{p}_B \pm z_{.025} \bullet (\text{s.e.}) = .42 - .36 \pm 1.960\sqrt{.003460}$$
$$= .06 \pm .115...$$
$$= [\,-5.5\%\,, +17.5\%\,] \quad (1 \text{ d.p.})$$

(b) The 95% confidence interval estimate includes $p_A - p_B = 0$

$\Rightarrow$ insufficient evidence to conclude that $p_A \neq p_B$

But the effect for which evidence is being sought is $p_A - p_B > 0$, (not $p_A \neq p_B$).

Conduct an hypothesis test

$$\mathcal{H}_o: \ p_A - p_B = 0 \quad \text{vs.} \quad \mathcal{H}_a: p_A - p_B > 0$$

Pooled sample proportion $\hat{p} = \dfrac{x_A + x_B}{n_A + n_B} = \dfrac{42 + 81}{100 + 225} = \dfrac{123}{325} = .3784\ldots$

Standard error $s = \sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n_A} + \dfrac{1}{n_B}\right)} = \sqrt{.37\ldots \times .62\ldots \times \left(\dfrac{1}{100} + \dfrac{1}{225}\right)} = 0.05829\ldots$

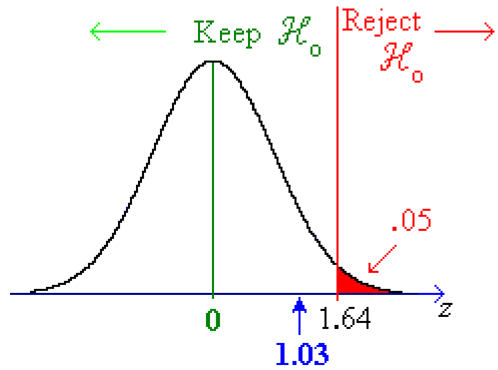
$$\Rightarrow \quad z = \frac{\hat{p}_A - \hat{p}_B}{s} = \frac{.42 - .36}{.058\ldots} = 1.029\ldots$$

$z_\alpha = z_{.050} = 1.644...$

$z < z_\alpha$

Therefore do **not** reject $\mathcal{H}_o : p_A = p_B$

There is insufficient evidence (at a level of significance of 5%) that brand A is more popular than brand B.



Example 11.08 (not examinable except for bonus)

A manager wishes to find a 95% confidence interval for the difference in the proportions of successful sales attempts between sales teams A and B. Random samples of n sales attempts are examined for each team. How large must the sample sizes n be in order to ensure that the confidence interval has a width of less than .10 ? [In other words, find the minimum sample size $n_{\min}$ to estimate $p_A - p_B$ to within five percentage points either way nineteen times out of twenty.]

The confidence interval estimate for $p_A - p_B$ is $\hat{p}_A - \hat{p}_B \pm \underbrace{z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_A \hat{q}_A}{n_A} + \frac{\hat{p}_B \hat{q}_B}{n_B}}}_{= w/2}$

Maximum width occurs when $\hat{p}_A = \hat{p}_B = \frac{1}{2} \Rightarrow \hat{q}_A = \hat{q}_B = \frac{1}{2}$

$n_A = n_B = n$

$$\Rightarrow \frac{w}{2} \geq z_{\frac{\alpha}{2}} \sqrt{\frac{\frac{1}{2} \times \frac{1}{2}}{n} + \frac{\frac{1}{2} \times \frac{1}{2}}{n}} = z_{\frac{\alpha}{2}} \sqrt{\frac{1}{2n}}$$

$$\Rightarrow \sqrt{2n} \geq \frac{2 z_{\alpha/2}}{w} \Rightarrow n \geq 2 \left(\frac{z_{\alpha/2}}{w} \right)^2$$

$\Rightarrow \quad n \geq 2\,(1.95... / 0.10)^2 = 768.3...$

Therefore

$$n_{\min} = \underline{\mathbf{769}}$$