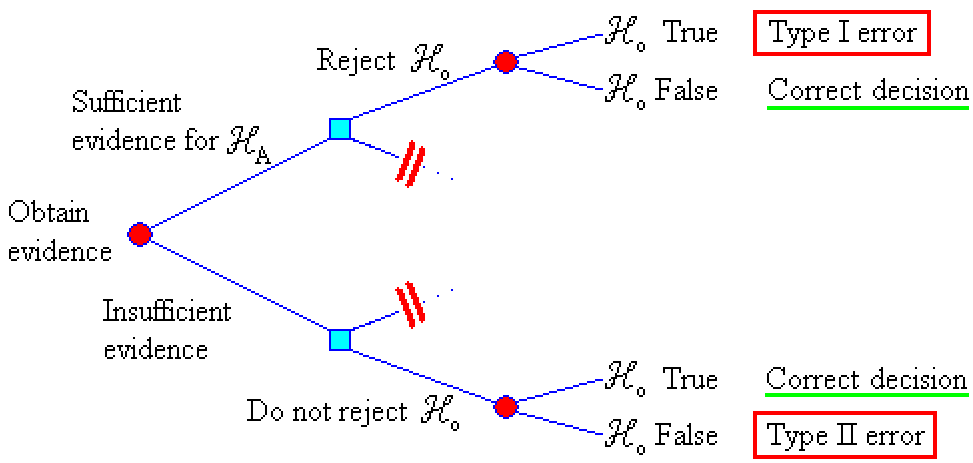
Hypothesis Tests

[Navidi sections 6.1-6.8 and 6.12-6.13; Devore chapter 8 and sections 9.1-9.4]
We begin with a reminder of page 7-14:

Is a **null hypothesis** $\mathcal{H}_o$ true (our "default belief"), or do we have sufficient evidence to reject $\mathcal{H}_o$ in favour of the **alternative hypothesis** $\mathcal{H}_A$?

$\mathcal{H}_o$ could be "defendant is not guilty" or "$\mu = \mu_o$", etc.
The corresponding $\mathcal{H}_A$ could be "defendant is guilty" or "$\mu \neq \mu_o$", etc.
The **burden of proof is on** $\mathcal{H}_A$ **.**



Of the two types of error, Type I is usually more serious.

Bayesian analysis: μ is treated as a random quantity. Data are used to modify prior belief about μ. Conclusions are drawn using both old and new information. It may be possible to construct valid statements involving $\mathrm{P}[\mathcal{H}_o \text{ true} \mid \text{data}]$.

Classical analysis: Data are used to draw conclusions about the unknown constant μ, without using any prior information. The results of a classical hypothesis test are often reported using *p*-values, where $p = \mathrm{P}[A \mid \mathcal{H}_o \text{ true}]$. The event A represents the event of obtaining data at least as extreme as the observed data. The probability that we usually want is $\mathrm{P}[\mathcal{H}_o \text{ true} \mid A]$.
But beware: it is usually the case that $\mathrm{P}[A \mid \mathcal{H}_o \text{ true}] \neq \mathrm{P}[\mathcal{H}_o \text{ true} \mid A]$.
In the criminal justice example, (as explored in a problem set question), this becomes
$\mathrm{P}[\text{incriminating evidence} \mid \text{defendent innocent}] \neq \mathrm{P}[\text{defendent innocent} \mid \text{incriminating evidence}]$
Indeed, the latter probability is sometimes orders of magnitude greater than the former.

Classical hypothesis tests are close relatives of classical confidence intervals.
We shall therefore revisit some examples from the previous two chapters.

Example 13.01 (Modification of Example 11.03)

The rate of energy loss X (watt) in a motor is known to be a normally distributed random quantity with standard deviation $\sigma = 3.0$ W. A random sample of 100 such motors produces a sample mean rate of energy loss of 58.3 W. Is this sample consistent with $\mu = 60$ W at a 1% level of significance?

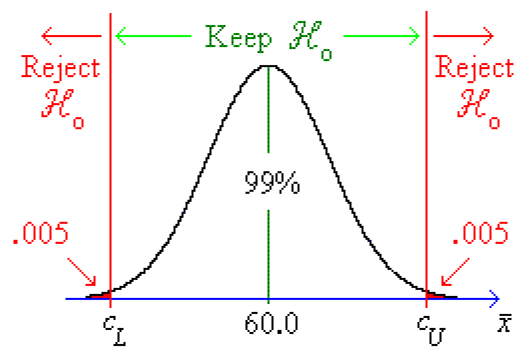
The null hypothesis is

The alternative hypothesis is

A "1% level of significance" is equivalent to a "99% level of confidence".

Construct a probability distribution on the [default] assumption that the null hypothesis is true:

$\mathcal{H}_o$ true $\Rightarrow$



If $\mathcal{H}_o$ is true, then values outside the interval $[c_L, c_U]$ occur in only 1% of all random samples drawn from this population. Therefore, if and only if the random sample returns a value of $\bar{x}$ outside this interval, we will have sufficient evidence to reject the null hypothesis.

$\bar{x} = 58.3$

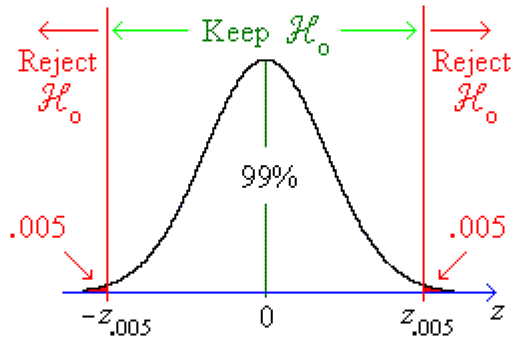
Example 13.01 (continued)

The method above we shall refer to as "Method 1".

Two alternative methods involve a transformation from values of $\bar{x}$ to values of z or t (depending on whether or not the population variance σ^2 is known).

Method 2:

Determine how many standard errors the observed sample mean $\bar{x}$ is away from the population mean μ if $\mathcal{H}_o$ is true. In other words, transform $\bar{x}$ into the equivalent value of z:



$z_{\text{obs}} =$

Any value $\bar{x}$ that is observed to be more than $z_{\alpha/2}$ standard errors away from μ_o is considered, at a level of significance α, to be too unlikely to have occurred by chance, thereby providing sufficient evidence to reject the null hypothesis in favour of the alternative hypothesis.

Method 3:

Again convert $\bar{x}$ to z_{obs}.

Evaluate $p = \text{P}\left[\,|Z| > z_{\text{obs}}\right]$.

Iff $p < \alpha$ then reject $\mathcal{H}_o$.

General method for two-tailed tests:

State hypotheses:

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

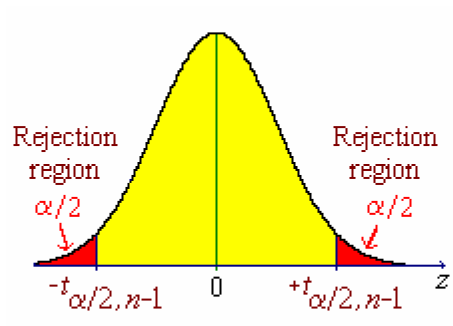
The burden of proof is on $\mathcal{H}_A$.

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 <u>and</u> n is small

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

Example 13.02 (Modification of Example 11.04)

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 and they are found to last for an average of 1,008 hours with a sample standard deviation of 6.2 hours.

Is there sufficient evidence, at a level of significance of 5%, to conclude that the population mean lifetime μ is greater than 1000 hours?

The null hypothesis is

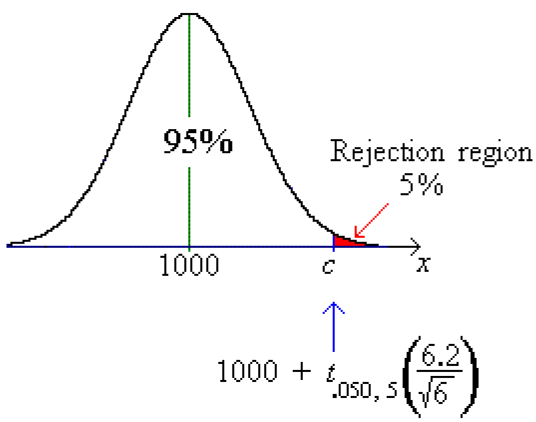
The alternative hypothesis is

A "5% level of significance" is equivalent to a "95% level of confidence".

Construct a probability distribution on the assumption that the null hypothesis is just barely true, that is $\mu = 1000$:

$\mathcal{H}_o$ true $\Rightarrow$

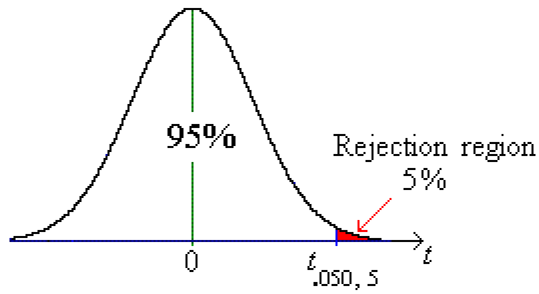
$\overline{X} \sim$



Example 13.02 (continued)

Method 2:

$t_{\text{obs}} =$



Method 3:

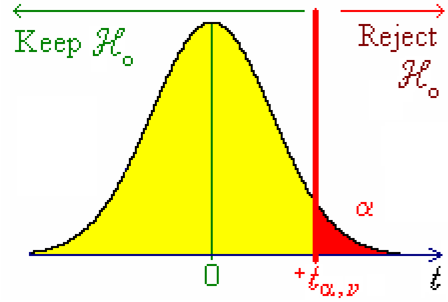
When the population variance is unknown and the sample size is too small for z_α to be an acceptable approximation to $t_{\alpha,\,n-1}$, Method 3 becomes difficult to use with standard printed statistical tables. However, when statistical software is applied to these hypothesis test problems, the *p*-value of Method 3 is usually reported. If this example were a question on a test or the final examination, then Method 3 should not be attempted.

$t_{\text{obs}} =$

General Method (upper-tailed tests):

State hypotheses:

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



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 .

Case 1: σ is unknown <u>and</u> n is small

<u>Method 1</u>:
Evaluate

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

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

<u>Method 2</u>:
Reject $\mathcal{H}_o$ iff

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

<u>Method 3</u>:

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

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

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

Common values: $z_{.050} = 1.64485$, $z_{.010} = 2.32635$

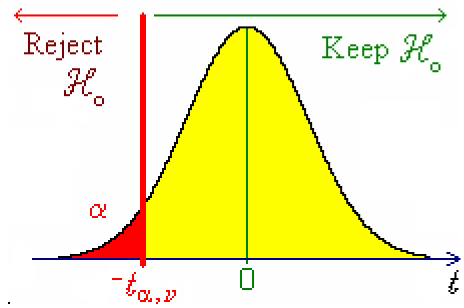
Case 3: σ is known is the same as Case 2 except that s is replaced by σ .

Lower-tailed tests are mirror-images of upper-tailed tests.

General Method (lower-tailed tests):

State hypotheses:

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



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 .

Case 1: σ is unknown <u>and</u> n is small

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

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

Method 3:

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

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

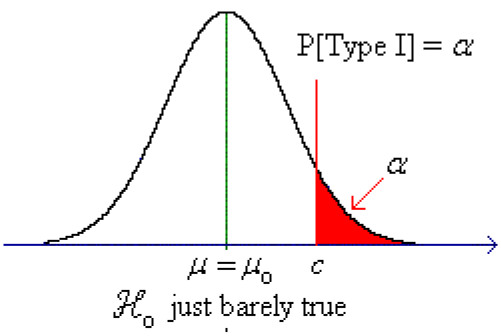
Case 2: n is large (> 30) is the same as Case 1 except that $t_{\alpha,\, n-1}$ is replaced by $t_{\alpha,\infty} = z_\alpha$.

Common values: $z_{.050} = 1.64485$, $z_{.010} = 2.32635$

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

α and the Probability of Committing a Type I Error:

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

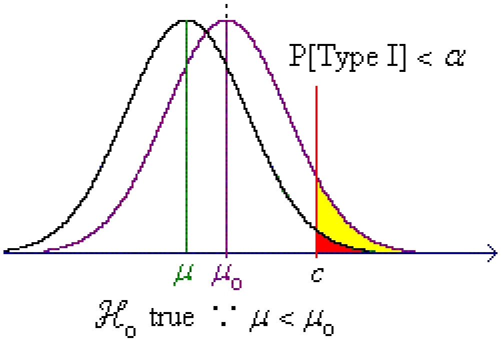


The boundary of the rejection region is calculated on the basis of the null hypothesis being just barely true; that is, the true population mean has a value right on the boundary between the two hypotheses.

$\mathcal{H}_o$ will be rejected iff the sample mean $\bar{x} > c$.

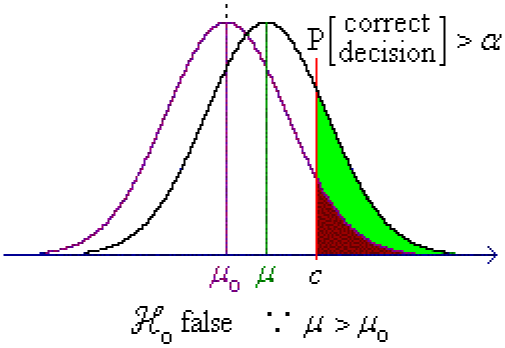
Rejection of $\mathcal{H}_o$ when it is true is a Type I error.

In this case, P[type I error] = α.



If the true value of the population mean is less than μ_o, then the true probability curve is to the left of the one used to calculate the boundary c of the rejection region. The area of the rejection region decreases.

The null hypothesis is still true, so rejecting $\mathcal{H}_o$ is still a type I error, but now P[type I error] < α.



If the true value of the population mean is greater than μ_o, then the true probability curve is to the right of the one used to calculate the boundary c of the rejection region. The area of the rejection region increases beyond α, but P[type I error] does ***not*** increase.

The null hypothesis is now false, so rejecting $\mathcal{H}_o$ is

Example 13.03 (Modification of Example 12.01)

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 .

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

Test $\mathcal{H}_o : \mu_{\text{new}} \le \mu_{\text{old}}$ vs. $\mathcal{H}_A : \mu_{\text{new}} > \mu_{\text{old}}$

or, equivalently, $\mathcal{H}_o : \mu_{\text{new}} - \mu_{\text{old}} = 0$ vs. $\mathcal{H}_A : \mu_{\text{new}} - \mu_{\text{old}} > 0$.

It is reasonable to assume that the samples drawn from the two populations are independent.

The sample sizes are large enough for us to estimate the unknown population variances σ^2 by the sample variances s^2 and to replace $t_{\alpha,\nu}$ by z_α.

<u>Example 13.04</u> (Modification of Example 12.04)

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. The results are as follows:

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

At a 5% level of significance, can one conclude that there is any difference between the mean failure currents of the two types of toaster?

Given in the question:

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

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

An essential assumption is that $X_{\mathrm{A}}, X_{\mathrm{B}}$ are independent.

The hypotheses being tested are $\mathcal{H}_{\mathrm{o}} : \mu_{\mathrm{A}} - \mu_{\mathrm{B}} = 0$ vs. $\mathcal{H}_{\mathrm{A}} : \mu_{\mathrm{A}} - \mu_{\mathrm{B}} \neq 0$ with $\alpha = 5\%$

The **summary statistics** are

$$n_A = 6 \qquad \bar{x}_A = 24.5 \qquad s_A^2 = 7.9 \qquad \Rightarrow \quad \frac{s_A^2}{n_A} = 1.31\dot{6}$$

$$n_B = 5 \qquad \bar{x}_B = 19.4 \qquad s_B^2 = 4.3 \qquad \Rightarrow \quad \frac{s_B^2}{n_B} = 0.86$$

$$\Rightarrow \quad (s.e.)^2 = \frac{s_A^2}{n_A} + \frac{s_B^2}{n_B} = 2.17\dot{6} \qquad \Rightarrow \quad s.e. = 1.475353\ldots$$

$$\nu = \mathrm{INT}\left(\frac{\left(\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B} \right)^2}{\frac{1}{n_A - 1}\left(\frac{s_A^2}{n_A} \right)^2 + \frac{1}{n_B - 1}\left(\frac{s_B^2}{n_B} \right)^2} \right) = \mathrm{INT}\left(\frac{\left(2.17\dot{6}\right)^2}{\frac{\left(1.31\dot{6}\right)^2}{5} + \frac{(0.86)^2}{4}} \right) = \mathrm{INT}(8.912\ldots) = 8$$

$$t_{\alpha/2,\nu} = t_{.025,8} = 2.30600$$

Example 13.04 (continued)

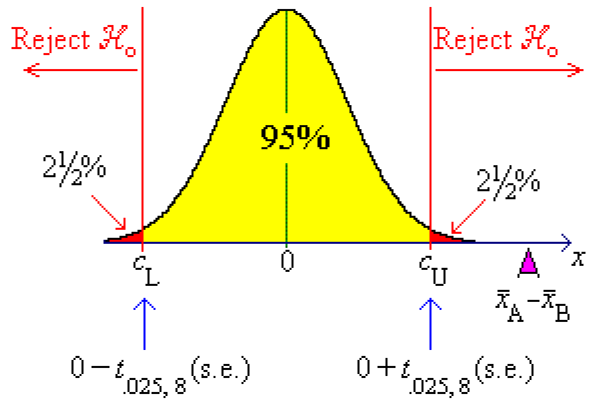
Method 1:

The boundaries of the rejection region are at

$$c_U, c_L = \Delta_o \pm t_{\alpha/2,\nu}(s.e.) =$$
$$0 \pm 2.30600 \times 1.475353\ldots = \pm 3.402\ldots$$

$$\bar{x}_A - \bar{x}_B = 24.5 - 19.4 = 5.1 > c_U$$

Therefore reject $\mathcal{H}_o$ in favour of $\mathcal{H}_A$.



OR

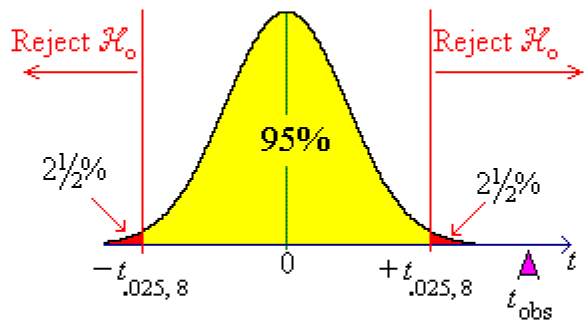
Method 2:

$$\bar{x}_A - \bar{x}_B = 24.5 - 19.4 = 5.1$$

$$\Rightarrow \quad t_{\text{obs}} = \frac{(\bar{x}_A - \bar{x}_B) - \Delta_o}{(s.e.)}$$

$$= \frac{5.1 - 0}{1.475353\ldots} = 3.456\ldots > t_{\alpha/2,\nu}$$

Therefore reject $\mathcal{H}_o$ in favour of $\mathcal{H}_A$.



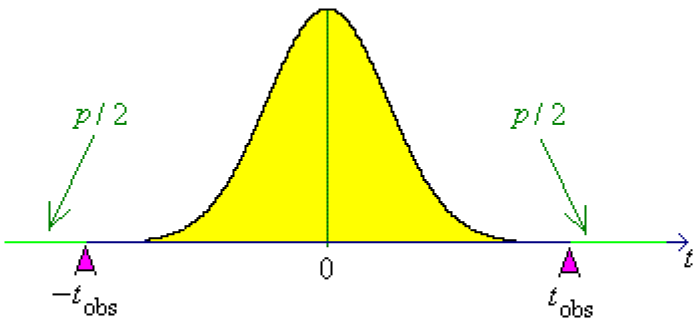
OR

Method 3 (requires access to software):

$$t_{\text{obs}} = \ldots = 3.456\ldots$$

$$\text{P}\left[|T_8| > |t_{\text{obs}}|\right] = .0086\ldots < \alpha$$

Therefore reject $\mathcal{H}_o$ in favour of $\mathcal{H}_A$.



YES, there is a significant difference between the mean failure currents of the two types of toaster.

Classical Hypothesis Tests for Paired Data

Example 13.05 (Modification of Example 12.05)

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

Test $\mathcal{H}_o : \mu_D = 0$ vs. $\mathcal{H}_A : \mu_D > 0$ and choose $\alpha = 1\%$

As discussed before in Example 12.05, the same individuals are present in both samples and there is very strong correlation between the samples. Therefore a paired test, based on the sample differences, is required.

Volunteer:	1	2	3	4	5	6	7	8	9
Differences:	3	1	5	6	3	0	6	–1	4

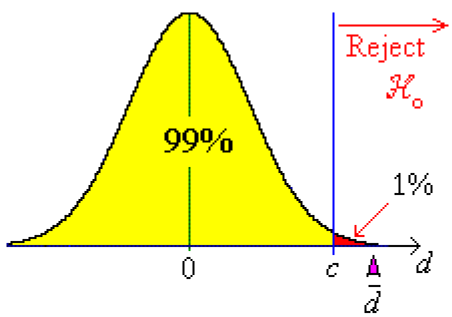
As before, $n = 9$, $\bar{d} = 3.0$, $s_D^2 = \dfrac{9 \times 133 - 27^2}{9 \times 8} = \dfrac{13}{2} = 6.5 \quad \Rightarrow \quad s_D = 2.549\ldots$

$s.e. = \dfrac{s_D}{\sqrt{n}} = 0.849836\ldots \qquad t_{\alpha,\nu} = t_{.010,8} = 2.89646$

Method 1:

$$c = \mu_o + t_{\alpha,\nu} \frac{s_D}{\sqrt{n}} = 0 + 2.89\ldots \times 0.84\ldots = 2.461\ldots$$

$\bar{d} = 3.0 > c \quad \Rightarrow$ reject $\mathcal{H}_o$

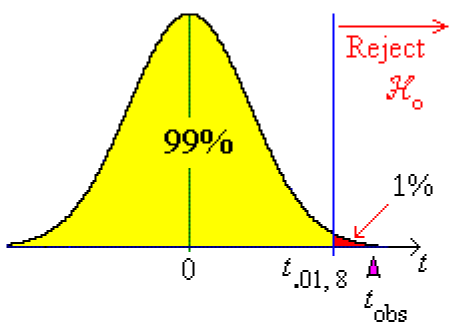


OR

Method 2:

$$t_{obs} = \frac{\bar{d} - \mu_o}{(s.e.)} = \frac{3.0 - 0}{0.849\ldots} = 3.530\ldots$$

$t_{obs} > t_{\alpha,\nu} \quad \Rightarrow$ reject $\mathcal{H}_o$



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

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

Independence $\Rightarrow \quad \rho = 0$

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

Classical Hypothesis Tests on Differences in Population Proportions
[Examinable in bonus questions only]
Example 13.06 (Modification of Example 12.02)

A random sample of 25 components (produced by one machine) yields 15 components that are longer than 10.0 cm. Another random sample of 30 components (produced by another machine) yields 12 components that are longer than 10.0 cm. Can one conclude, at a level of significance of 5%, that the two population proportions are different?

The hypotheses to be tested are: $\mathcal{H}_o : p_1 = p_2$ vs. $\mathcal{H}_A : p_1 \neq p_2$, or, equivalently,
$\mathcal{H}_o : p_1 - p_2 = 0$ vs. $\mathcal{H}_A : p_1 - p_2 \neq 0$, with $\alpha = 5\%$.

If the null hypothesis is true, then the two sample proportions $\hat{p}_1 = \frac{15}{25} = 0.6$ and $\hat{p}_2 = \frac{12}{30} = 0.4$ are two estimates of the same population proportion p. A better estimate of p can be obtained by pooling the two sample proportions together:

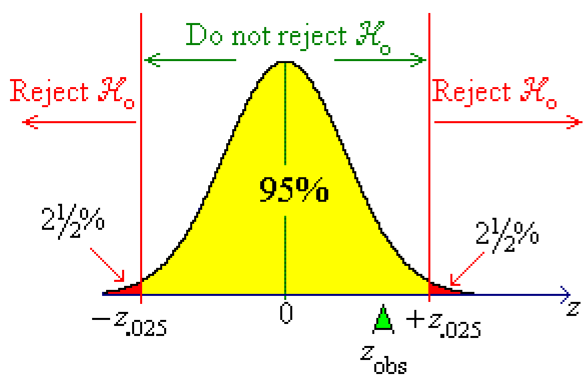
$$\hat{p} = \frac{(\text{total \# successes})}{(\text{total \# trials})} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{15+12}{25+30} = \frac{27}{55}$$

Using Method 2, the sample standard error is

$$s = \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{\frac{27}{55} \times \frac{28}{55} \times \left(\frac{1}{25} + \frac{1}{30}\right)} = 0.135378\ldots$$

$$\Rightarrow \quad z_{\text{obs}} = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{s} = \frac{.6 - .4}{0.135\ldots} = 1.477\ldots$$

$z_{.025} = 1.95\ldots \qquad |z_{\text{obs}}| \not> z_{.025}$



The Agresti-Caffo estimate is

$$p^* = \frac{x_1 + 1 + x_2 + 1}{n_1 + 2 + n_2 + 2} = \frac{16+13}{27+32} = \frac{29}{59}$$

$\Rightarrow \quad s = 0.135\ldots \quad \Rightarrow \quad z_{\text{obs}} = 1.38\ldots$
with the same conclusion.

Therefore there is insufficient evidence to reject $\mathcal{H}_o$.

Power of a Test [Examinable in bonus questions only]

Example 13.07

The completion time T for the assembly of a product is known to follow a normal distribution to an excellent approximation, with a known population standard deviation of 2.4 minutes. The existing process has a mean completion time of 40.1 minutes. A new process may decrease the mean completion time but won't affect σ. T will continue to follow a normal distribution.

A customer will invest in the new process only if there is clear evidence (at a level of significance of 5%) that its mean completion time is less than 38.0 minutes.
A random sample of six product assemblies is drawn.

a) Set up the appropriate hypotheses to be tested.
b) Find the critical value c below which an observed sample mean $\bar{t}$ will cause the null hypothesis to be rejected at a level of significance of 5%.
c) Find the probability that the customer will not invest in the new process when the true mean completion time is 37.0 minutes.

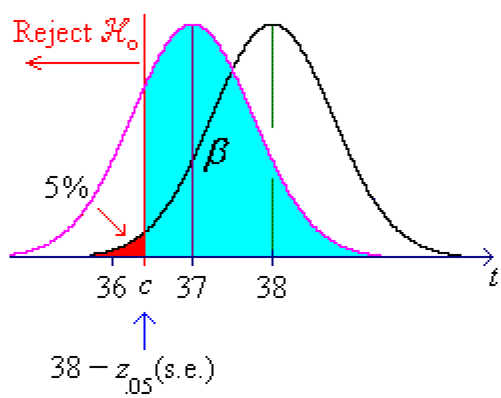
It is given that, under the new process, $T \sim \mathrm{N}\left(\mu, (2.4)^2\right)$

(a) $\mathcal{H}_o : \mu \geq 38.0$ vs. $\mathcal{H}_A : \mu < 38.0$ with $\alpha = 5\%$

(b) $c = \mu_o - z_\alpha \dfrac{\sigma}{\sqrt{n}} =$

$$38.0 - 1.64\ldots \times \frac{2.4}{\sqrt{6}} = 38.0 - 1.611\ldots$$

Therefore $c \approx 36.39$ minutes



(c) $\beta(37.0) = \mathrm{P}\left[\bar{T} > c \mid \mu = 37\right]$

$$= \mathrm{P}\left[Z > \frac{c-37}{\left(\frac{2.4}{\sqrt{6}}\right)}\right] = \mathrm{P}\left[Z > \frac{38 - 1.64\ldots\left(\frac{2.4}{\sqrt{6}}\right) - 37}{\left(\frac{2.4}{\sqrt{6}}\right)}\right]$$

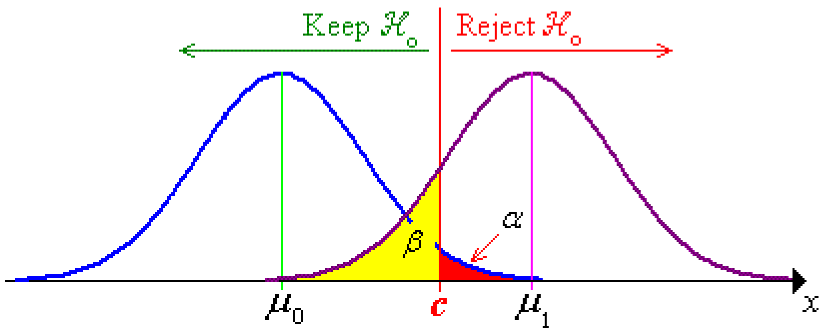
$$= \mathrm{P}\left[Z < 1.64\ldots - \frac{38-37}{\left(\frac{2.4}{\sqrt{6}}\right)}\right] = \Phi(1.64\ldots - 1.02\ldots) = \Phi(0.624\ldots)$$

Therefore $\beta(37.0) \approx 73\%$

Type II Errors, A More General Case

$\mathcal{H}_o : \mu \le \mu_o$ vs. $\mathcal{H}_A : \mu > \mu_o$, with σ known
(or n large enough for $t_{\alpha, n-1} \approx z_\alpha$ and $s^2 \approx \sigma^2$).

The probability β of committing a type II error depends on the true value of the parameter (μ).



$$\beta(\mu_1) = \mathrm{P}\left[\text{not rejecting } \mathcal{H}_o \mid \mathcal{H}_o \text{ false } \because \ \mu = \mu_1\right]$$

$$\Rightarrow \quad \beta(\mu_1) = \mathrm{P}\left[\overline{X} < c \mid \mu = \mu_1\right] \quad \text{where } c = \mu_o + z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$\beta(\mu_1) = \mathrm{P}\left[Z < \frac{c - \mu_1}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right] = \Phi\left(\frac{\left(\mu_o + z_\alpha \frac{\sigma}{\sqrt{n}}\right) - \mu_1}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right)$$

Therefore

$$\boxed{\beta(\mu_1) = \Phi\left(z_\alpha - \frac{\mu_1 - \mu_o}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right)}$$

The formula for a lower tail test is similar:

$$\boxed{\beta(\mu_1) = \Phi\left(z_\alpha - \frac{\mu_o - \mu_1}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right)}$$

A two tail test is more complicated, as is the case of unknown variance and a small sample.

The farther apart μ_1 and μ_o are, the less likely a type II error is and the greater the likelihood that the test can distinguish the two hypotheses correctly. The power of the test is defined to be $1 - \beta$.

[End of Chapter 13]

[Space for Additional Notes]