2. Confidence Intervals

Let $\alpha \in (0,1)$.

A 1-α confidence interval is an interval $C=(a,b)$, where $a=a(\underset{\sim}{X})$ and $b=b(\underset{\sim}{X})$ such that

$$P(\theta \in C) = 1-\alpha.$$

Note: $a(\underline{X})$, $b(\underline{X})$ are not allowed to depend on θ.

In words: (a,b) traps θ with probability $1-\alpha$

Warning: C is random and θ is fixed

Most common is $\alpha = 0.05$, i.e. 95% C.I

confidence interval

Interpretation: if we repeat an experiment many times, and construct a C.I. each time, then approx 95% of our intervals will contain the true value of θ (repeated sampling).

Example $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, 1)$

and $\left(\bar{X} \pm \frac{1.96}{\sqrt{n}}\right)$ is a 95% C.I. for θ.

We'll usually want a central (equal tail) interval as above.

[One-sided intervals of the form (a, ∞) or $(-\infty, b)$ are possible.]

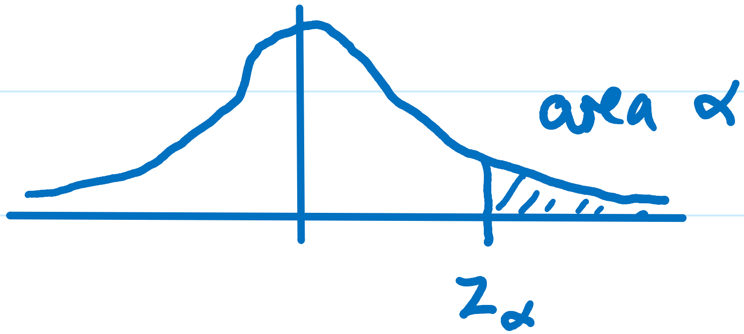
2.1 CIs using CLT

Plenty of examples in Prelims, and similar to the next section.

2·2 CIs using asymptotic distribution of MLE

We know $I(\theta)^{1/2} \cdot (\hat{\theta} - \theta) \xrightarrow{D} N(0,1)$

Let z_α be such that $P(N(0,1) > z_\alpha) = \alpha$



Then $1-\alpha \approx P\left(-z_{\frac{\alpha}{2}} < I(\theta)^{1/2} \cdot (\hat{\theta} - \theta) < z_{\alpha/2}\right)$ (1)

$$= P\left(\hat{\theta} - \frac{z_{\alpha/2}}{\sqrt{I(\theta)}} < \theta < \hat{\theta} + \frac{z_{\alpha/2}}{\sqrt{I(\theta)}}\right)$$

In general $I(\theta)$ depends on θ so (as in Prelims) replace $I(\theta)$ by $I(\hat{\theta})$ to get approx $1-\alpha$ C.I. of

$$\left(\hat{\theta} \pm \frac{z_{\alpha/2}}{\sqrt{I(\hat{\theta})}} \right). \qquad (2)$$

Why does replacing $I(\theta)$ by $I(\hat{\theta})$ work?

We are assuming $\hat{\theta} \xrightarrow{P} \theta$ and that $I(\theta)$ is continuous, hence $\left(\frac{I(\hat{\theta})}{I(\theta)}\right)^{1/2} \xrightarrow{P} 1$.

So $$I(\hat{\theta})^{1/2} \cdot (\hat{\theta} - \theta) = \underbrace{\left(\frac{I(\hat{\theta})}{I(\theta)}\right)^{1/2}}_{\xrightarrow{P} 1} \cdot \underbrace{I(\theta)^{1/2}(\hat{\theta} - \theta)}_{\xrightarrow{D} N(0,1).}$$

Hence by Slutsky (ii), $I(\hat{\theta})^{1/2} \cdot (\hat{\theta} - \theta) \xrightarrow{D} N(0,1)$.

So ① holds with $I(\theta)^{1/2}$ replaced by $I(\hat{\theta})^{1/2}$ and then the same rearrangement as above leads to C.I. ②.

Example $X_1, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$. Then $\hat{\theta} = \bar{X}$ and

$$I(\theta) = \frac{n}{\theta(1-\theta)}$$ and interval ② is

$$\left(\hat{\theta} \pm z_{\alpha/2} \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \right).$$

If $n = 30$, $\sum_1^n x_i = 5$, then this gives a 99% interval of: $(-0.008, 0.342)$.

But we know $\theta > 0$!

We can avoid negative values by reparametrising as follows.

Let $\psi = g(\theta) = \log \frac{\theta}{1-\theta}$ "log odds" (of success)

Since $\theta \in (0,1)$ we have $\psi \in (-\infty, \infty)$ so using a normal distribution can't produce impossible ψ values

Note $\hat{\theta} \approx N\left(\theta, \frac{\theta(1-\theta)}{n}\right)$ and delta method gives $\hat{\psi} \approx N\left(\psi, \frac{\theta(1-\theta)}{n} g'(\theta)^2\right)$. (3)

Use (3) to find approx $1-\alpha$ C.I. for ψ, say (ψ_1, ψ_2).

$$1-\alpha \approx P(\psi_1 < \psi < \psi_2)$$

$$= P\left(\frac{e^{\psi_1}}{1+e^{\psi_1}} < \theta < \frac{e^{\psi_2}}{1+e^{\psi_2}}\right) \quad \text{since } \theta = \frac{e^{\psi}}{1+e^{\psi}}.$$

This $1-\alpha$ C.I. for θ definitely won't contain negative values.

2.3 Distributions related to $N(0,1)$

Definition Let $Z_1, \dots, Z_r \overset{iid}{\sim} N(0,1)$. We say that $Y = Z_1^2 + \dots + Z_r^2$ has the chi-squared distribution with r degrees of freedom. Write $Y \sim \chi_r^2$.

In fact $\chi_r^2 \sim \text{Gamma}\left(\frac{r}{2}, \frac{1}{2}\right)$.

If $Y \sim \chi_r^2$ then $E(Y) = r$ and $\text{var}(Y) = 2r$.

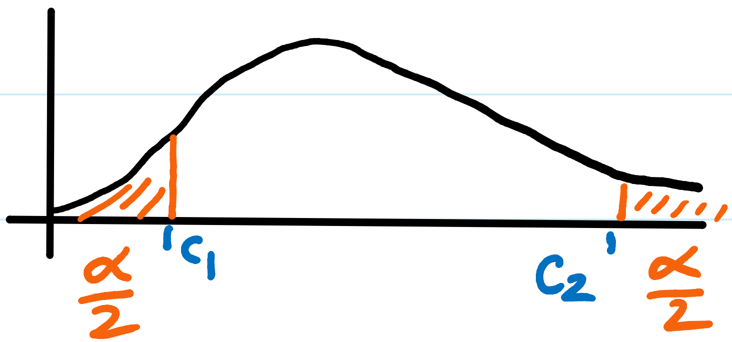
If $Y_1 \sim \chi_r^2$ and $Y_2 \sim \chi_s^2$ are independent, then $Y_1 + Y_2 \sim \chi_{r+s}^2$.

Example $X_1, \dots, X_n \overset{iid}{\sim} N(0, \sigma^2)$.

Then $\frac{X_i}{\sigma} \sim N(0,1)$, hence $\frac{\sum X_i^2}{\sigma^2} \sim \chi^2_n$.

Hence $P\left(c_1 < \frac{\sum X_i^2}{\sigma^2} < c_2\right) = 1 - \alpha$

where c_1, c_2 are such that $P(\chi^2_n < c_1) = P(\chi^2_n > c_2) = \frac{\alpha}{2}$



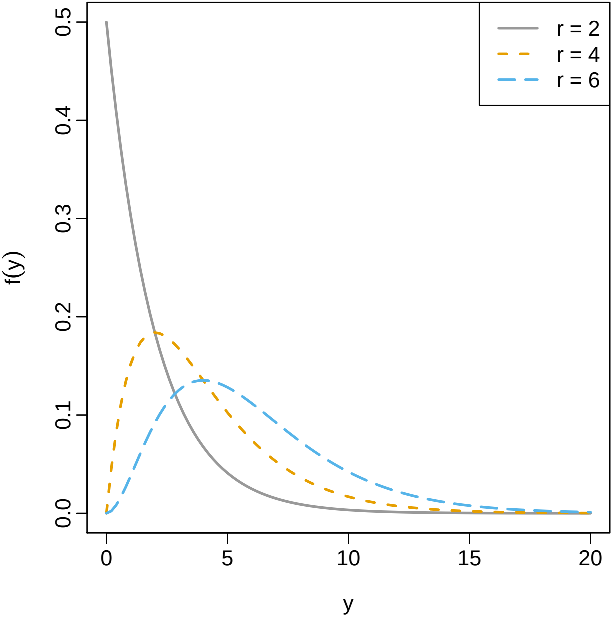
So $P\left(\frac{\sum X_i^2}{c_2} < \sigma^2 < \frac{\sum X_i^2}{c_1}\right) = 1 - \alpha$

and we've found a $1-\alpha$ CI for σ^2.

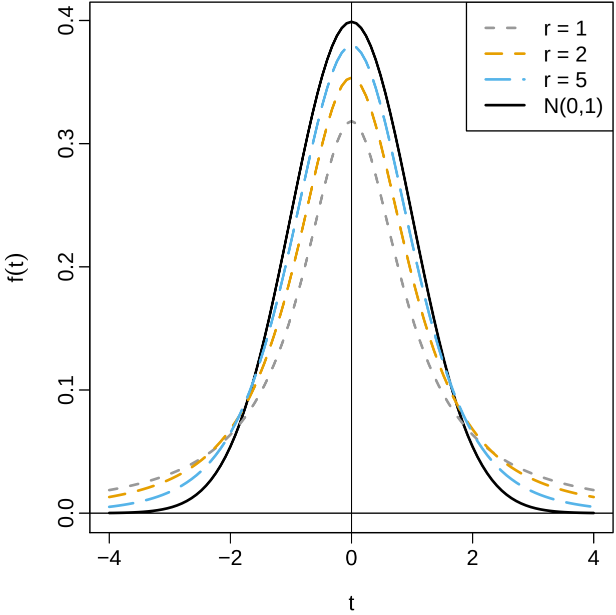
Definition Let $Z \sim N(0,1)$ and $Y \sim \chi^2_r$ be independent.
We say that
$$T = \frac{Z}{\sqrt{Y/r}}$$
has a **(Student) t-distribution** with **r degrees of freedom**.
Write $T \sim t_r$.

We have $t_r \xrightarrow{D} N(0,1)$ as $r \to \infty$.

Chi-squared pdfs



t distribution pdfs



2.4 Independence of $\bar{X}$ and S^2 for normal samples

Suppose $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

Consider $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ sample mean

$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ sample variance

<u>Theorem 2.1</u> $\bar{X}$ and S^2 are independent and their marginal distributions are

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

(ii) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.

Proof Let $Z_i = \dfrac{X_i - \mu}{\sigma}$. Then $Z_1 \dots Z_n \overset{iid}{\sim} N(0,1)$

and so have joint pdf

$$f(\underset{\sim}{z}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2} = (2\pi)^{-n/2} e^{-\frac{1}{2}\sum z_i^2} \quad ①$$

Now consider a transformation from $\underset{\sim}{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$ to $\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$.

Let $\underline{Y} = A \underline{Z}$ where A is an orthogonal $n \times n$ matrix with first row $\left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right)$.

Orthogonal: $A^T A = I$, so $(\det A)^2 = 1$.

If $\underset{\sim}{y} = A\underset{\sim}{z}$ then $\underline{\underline{z}} = A^T \underset{\sim}{y}$ so $z_i = \sum_k a_{ki} y_k$

and $\dfrac{\partial z_i}{\partial y_j} = a_{ji}$.

Hence the Jacobian $J = J(y_1, \dots, y_n) = \det(A^T)$, so $|J| = 1$ ②

Also $\sum_1^n y_i^2 = \underset{\sim}{y}^T \underset{\sim}{y} = \underset{\sim}{z}^T A^T A \underline{z} = \underline{z}^T \underline{z} = \sum_1^n z_i^2$ ③

Hence the pdf of $\underset{\sim}{Y}$ is $g(\underset{\sim}{y}) = f(\underset{\sim}{z}(\underset{\sim}{y})) . |J|$

$$= (2\pi)^{-n/2} e^{-\frac{1}{2}\sum y_i^2} . 1$$

Hence $Y_1 \dots Y_n \overset{iid}{\sim} N(0,1)$, using ①, ②, ③.

Now $Y_1 = (\text{first row of } A) \cdot \underset{\sim}{Z} = \frac{1}{\sqrt{n}} \sum_1^n Z_i = \sqrt{n}\, \bar{Z}$

and $$\sum_1^n (Z_i - \bar{Z})^2 = \sum_1^n Z_i^2 - 2\bar{Z} \sum Z_i + n\bar{Z}^2$$

$$= \sum_1^n Z_i^2 - n\bar{Z}^2$$

$$= \sum_1^n Y_i^2 - Y_1^2$$

$$= \sum_2^n Y_i^2$$

So $\bar{Z}$ is a function of Y_1 only

$\sum_1^n (Z_i - \bar{Z})^2$ - - - - - - $Y_2, \dots, Y_n$ only

and the Y_i are indep, hence $\bar{Z}$ and $\sum_1^n (Z_i - \bar{Z})^2$ are indep.

Then $\bar{X}$ and S^2 are indep because $\bar{X} = \sigma \bar{Z} + \mu$

and $S^2 = \frac{\sigma^2}{n-1} \sum_1^n (Z_i - \bar{Z})^2$.

Finally:

(i) $Y_1 \sim N(0,1)$ so $\bar{X} = \sigma \bar{Z} + \mu = \frac{\sigma}{\sqrt{n}} Y_1 + \mu \sim N(\mu, \frac{\sigma^2}{n})$

(of course!)

(ii) $\frac{(n-1)S^2}{\sigma^2} = \sum_1^n (Z_i - \bar{Z})^2 = \sum_2^n Y_i^2 \sim \chi^2_{n-1}$. $\square$

So we have $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ and $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

and these two random variables are <u>independent</u>

From the definition of a t_{n-1} distribution this gives

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

(the σ in numerator & denominator cancels).

The quantity $\dfrac{\bar{X}-\mu}{S/\sqrt{n}}$ is called a **pivotal quantity**

or **pivot**, meaning that it is a function of $\underset{\sim}{X}$ and $\theta=(\mu,\sigma^2)$ whose distribution does not depend on θ.

Similarly $\dfrac{(n-1)S^2}{\sigma^2}$ is another pivot.

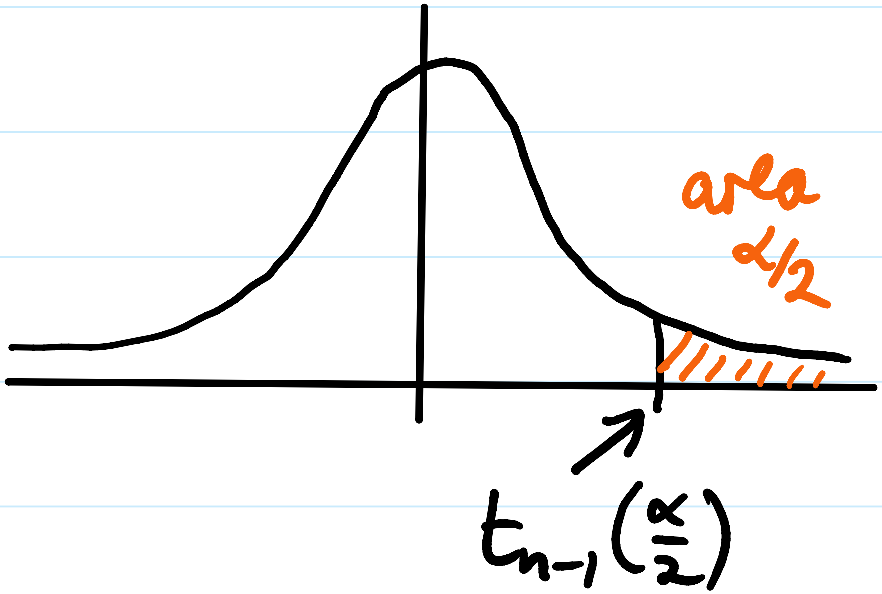
Example $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, μ, σ^2 unknown.

Let's find a C.I. for μ.

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

where $t_{n-1}\left(\frac{\alpha}{2}\right)$ is such that

$$P\left(t_{n-1} > t_{n-1}\left(\tfrac{\alpha}{2}\right)\right) = \frac{\alpha}{2}.$$



Hence

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

and we have a $1-\alpha$ C.I. for μ.

When $\sigma = \sigma_0$ is known, the corresponding interval from Prelims is

$$\left(\bar{X} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma_0}{\sqrt{n}}\right).$$

Student's Sleep data

"Student" = W.S. Gosset

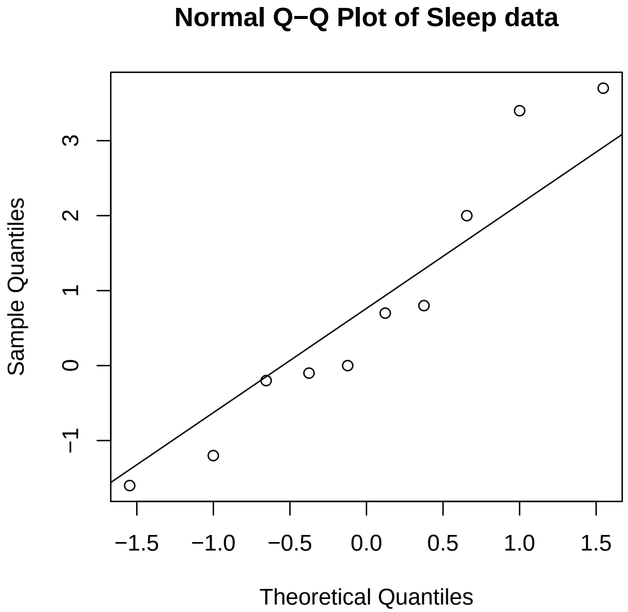
Below is half of Student's sleep data (1908):

$$0.7, \;\; -1.6, \;\; -0.2, \;\; -1.2, \;\; -0.1, \;\; 3.4, \;\; 3.7, \;\; 0.8, \;\; 0.0, \;\; 2.0.$$

The data give the number of hours of sleep gained, by 10 patients, following a low dose of a drug.

[The other half of the data give the sleep gained following a normal dose of the drug.]

A point estimate of the sleep gained is $\overline{x} = 0.75$ hours.


Normal Q−Q Plot of Sleep data
Sample Quantiles
Theoretical Quantiles
−1
0
1
2
3
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5

Treating the sample as iid $N(\mu, \sigma^2)$, with μ and σ^2 unknown, a 95% CI for μ is

$$\left(\overline{x} \pm t_{n-1}(\tfrac{\alpha}{2})\frac{s}{\sqrt{n}}\right) = (-0.53, 2.03)$$

using $\overline{x} = 0.75$, $s^2 = 3.2$, $n = 10$, $\alpha = 0.05$, $t_9(0.025) = 2.262$.

The value of $t_9(0.025)$ comes from statistical tables, or from R.

Here, it would be *incorrect* to use a $N(0,1)$ distribution instead of a t_9.

E.g. Suppose we "assume" $\sigma^2 = s^2 = 3.2$ (the sample variance) and calculate the interval

$$\left(\overline{x} \pm 1.96\sqrt{\frac{3.2}{10}}\right) = (-0.36, 1.86).$$

The interval $(-0.53, 2.03)$ obtained using the t_9 distribution is wider than the interval $(-0.36, 1.86)$.

The interval from the t_9 distribution is the correct one here. Since σ^2 is unknown, we need to estimate it (our estimate is s^2). Since we are estimating σ^2, there is more uncertainty than if σ^2 were known, and the t_9 distribution correctly takes this uncertainty into account.

Sleep data (low dose)

Number of hours of sleep gained, by 10 patients:

$$0.7,\ \ -1.6,\ \ -0.2,\ \ -1.2,\ \ -0.1,\ \ 3.4,\ \ 3.7,\ \ 0.8,\ \ 0.0,\ \ 2.0.$$

Do the data support the conclusion that a low dose of the drug makes people sleep more, or not?

- We will start from the default position that the drug has no effect,
- and we will only reject this default position if the data contain "sufficient evidence" for us to reject it.

So we would like to consider

(i) the "null hypothesis" that the drug has no effect, and

(ii) the "alternative hypothesis" that the drug makes people sleep more.

We will denote the "null hypothesis" by H_0, and the "alternative hypothesis" by H_1.

Sleep data (normal dose)

The other half of the sleep data is the number of hours of sleep gained, by the same 10 patients, following a normal dose of the drug:

$$1.9,\ 0.8,\ 1.1,\ 0.1,\ -0.1,\ 4.4,\ 5.5,\ 1.6,\ 4.6,\ 3.4.$$

Is there evidence that a normal dose of the drug makes people sleep more than not taking a drug at all, or not?

3. Hypothesis Testing

3.1 Introductory example

We denote the null hypothesis by H_0 and the alternative hypothesis by H_1.

Suppose $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ with μ and σ^2 unknown.

We'll consider

$H_0 : \mu = \mu_0$ (and σ^2 is unknown) "drug has no effect"

$H_1 : \mu > \mu_0$ (and σ^2 unknown) "drug makes people sleep more" (on average)

$\mu_0 = 0$ for sleep data, but sometimes $\mu_0 \neq 0$.

Let $t_{obs} = t(\underline{x}) = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ and let $t(\underline{X}) = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$.

The idea is:

- a small/moderate value of t_{obs} is consistent with H_0
  including negative
- whereas a very large value of t_{obs} is not consistent with H_0 and instead suggests H_1.

For sleep data (low dose) $t_{obs} = 1.326$.

Is this t_{obs} large?

If H_0 is true then $t(\underset{\sim}{X}) \sim t_{n-1}$ and the probability of observing a value $\geqslant t_{obs}$ is

$$p = P(t(\underset{\sim}{X}) \geq t_{obs}) = P(t_9 \geq 1.326) = 0.109.$$

This p is called the <u>p-value</u> or <u>significance level</u>.

$p = 0.109$ is not particularly small, we'd observe $t(\underset{\sim}{x}) \geqslant 1.326$ over 10% of the time (if H_0 true).

So we don't have much evidence to reject H_0, so we'll retain H_0 and say the data are consistent with H_0 being true.

What we have done here: look to see if data seem inconsistent with H_0

- if so, reject H_0
- if not, keep / retain H_0.

Usually we say "don't reject H_0" or "reject H_0"
(retain H_0)

or say data are "consistent with H_0" or "not consistent with H_0"

(rather than "accept H_0" or "accept H_1")

2nd example (sleep data, full dose)

$t_{obs} = 3.68$

This time p-value $= P(t(\underline{X}) \geq 3.68) = P(t_9 \geq 3.68)$

$= 0.0025.$

This is very small, we'd see $t(\underline{X}) \geq 3.68$ only 0.25% of the time if H_0 true, very rare.

We conclude that there is very strong evidence to reject H_0 in favor of H_1.

How small is small for a p-value?

We might say something like:

$p < 0.01$	very strong evidence against H_0
$0.01 < p < 0.05$	strong - - - - - - -
$0.05 < p < 0.1$	weak - - - - - - - - - -
$0.1 < p$	little or no - - - - - - - - - .

$X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ $\quad\quad$ μ, σ both unknown

null $H_0: \mu = \mu_0$, $\quad\quad$ alternative $H_1: \mu > \mu_0$

$$T = T(\underline{X}) = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

when H_0 true

The further $\bar{x}$ above μ_0, the more evidence to reject H_0.

the further $t(\underline{x})$ above zero

One-sided and two-sided alternative hypotheses

$H_1: \mu > \mu_0$ is a one-sided alternative. The larger t_{obs} the more evidence to reject H_0.

Similarly $H_1: \mu < \mu_0$ is also one-sided. The p-value would be $p = P(t_{n-1} \leq t_{obs})$.

A different type of alternative is $H_1: \mu \neq \mu_0$. This is a two-sided alternative.

$H_1: \mu \neq \mu_0$ two-sided

For this H_1:

if t_{obs} is very lage (i.e. very positive) then we have evidence to reject H_0

AND also

if t_{obs} is very small (i.e. very negative) then we have evidence to reject H_0

Let $t_0 = |t_{obs}| = \left| \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} \right|$

The p-value is the probability $t(\underset{\sim}{X})$ takes a value "at least as extreme" as t_{obs}:

$$p = P(|t(\underset{\sim}{X})| \geq t_0)$$
$$= P(t(\underset{\sim}{X}) \geq t_0) + P(t(\underset{\sim}{X}) \leq -t_0)$$
$$= 2P(t(\underset{\sim}{X}) \geq t_0).$$

This p-value, and all others, are calculated under the assumption that H_0 is true. From now on we write $p = P(t(\underset{\sim}{X}) \geq t_{obs} \mid H_0)$ or $p = P(|t(\underset{\sim}{X})| \geq t_0 \mid H_0)$ to indicate this.

3.2 Tests for normally distributed samples

Example (z-test) $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ with μ unknown and $\sigma^2 = \sigma_0^2$ known.

To test $H_0: \mu = \mu_0$ against $H_1: \mu > \mu_0$ we use the test statistic
$$Z = \frac{\bar{X} - \mu_0}{\sigma_0 / \sqrt{n}}.$$

Let
$$z_{obs} = \frac{\bar{x} - \mu_0}{\sigma_0 / \sqrt{n}}.$$

If H_0 is true then $Z \sim N(0,1)$.

p-value $\quad p = P(Z \geq z_{obs} \mid H_0) = P(N(0,1) \geq z_{obs})$

$$= 1 - \Phi(z_{obs})$$

For H_0 versus $H_1': \mu < \mu_0$

p-value $\quad p' = P(Z \leq z_{obs} \mid H_0) = \Phi(z_{obs})$

For H_0 versus $H_1'': \mu \neq \mu_0$

p-value $\quad p'' = P(|Z| \geq z_0 \mid H_0) = 2(1 - \Phi(z_0))$

where $z_0 = |z_{obs}|$.

Example (t-test) $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ with μ and σ^2 unknown.

There are similar expressions for p-values like p, p', p'' above, after replacing:

- Z by $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$
- Φ by cdf of t_{n-1}.

t-test (one sample)

[Example from Dalgaard (2008).] Data on the daily energy intake (in kJ) of 11 women:

$$5260, \ 5470, \ 5640, \ 6180, \ 6390, \ 6515,$$
$$6805, \ 7515, \ 7515, \ 8230, \ 8770.$$

Do these values deviate from a recommended value of 7725 kJ?

We consider testing $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, where $\mu_0 = 7725$, and we make the standard assumptions for a t-test.

We have $t_{\text{obs}} = \frac{\overline{x}-\mu_0}{s/\sqrt{n}} = -2.821$.

The p-value is $p = 2P(t_{10} \geqslant |t_{\text{obs}}|) = 0.018$. So we conclude that there is good evidence to reject the null hypothesis that the mean intake is 7725 kJ.

Testing $H_0 : \mu = 7725$ against $H_1^- : \mu < 7725$,

the p-value is $p^- = P(t_{10} \leqslant t_{\text{obs}}) = 0.009$.

Conclusion: there is good evidence to reject H_0 in favour of H_1^-.

Testing $H_0 : \mu = 7725$ against $H_1^+ : \mu > 7725$,

the p-value is $p^+ = P(t_{10} \geqslant t_{\text{obs}}) = 0.991$.

Conclusion: there is no evidence to reject H_0 in favour of H_1^+.

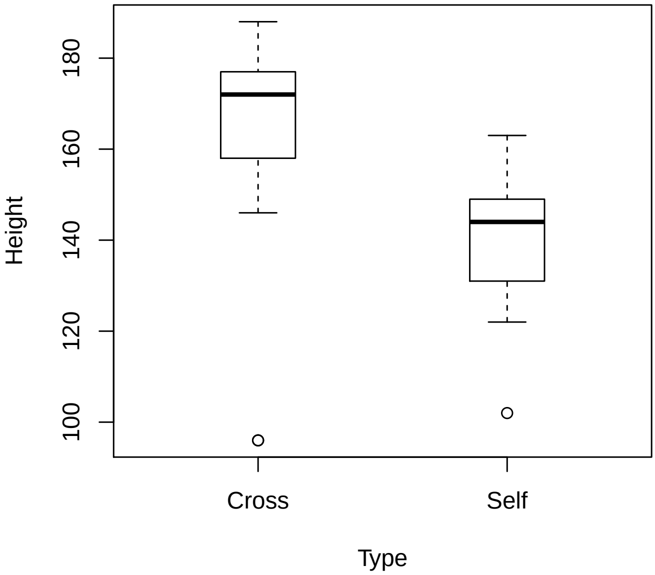
t-test (two sample)

Darwin's *Zea mays* data – heights of young maize plants.

Height (eighths of an inch)			
Crossed		Self-fertilized	
188	146	139	132
96	173	163	144
168	186	160	130
176	168	160	144
153	177	147	102
172	184	149	124
177	96	149	144
163		122	

Are the heights of the two types of plant the same?

[In fact, the plants were in pairs – one cross- and one self-fertilized in each pair – we ignore this pairing for now. We'll see how to deal with pairing later.]



Height
Type
Cross
Self
100
120
140
160
180

Assume we have two independent samples $X_1, \ldots, X_m \overset{\text{iid}}{\sim} N(\mu_X, \sigma^2)$, and $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} N(\mu_Y, \sigma^2)$, where σ^2 is unknown.

Suppose we would like to test $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X \neq \mu_Y$.

Let

$$T = \frac{\overline{X} - \overline{Y}}{S\sqrt{\frac{1}{m} + \frac{1}{n}}}$$

where $S^2 = \frac{1}{m+n-2}[\sum(X_i - \overline{X})^2 + \sum(Y_i - \overline{Y})^2]$.

Assuming H_0 is true, we have $T \sim t_{m+n-2}$.

For the maize data, the observed value of T is

$$t = \frac{\overline{x} - \overline{y}}{s\sqrt{\frac{1}{m} + \frac{1}{n}}} = 2.437.$$

The alternative hypothesis $(\mu_X \neq \mu_Y)$ is two-sided, so the p-value of this test is

$$p = 2P(t_{28} \geqslant 2.437) = 0.021.$$

Conclusion: there is good evidence to reject the null hypothesis $\mu_X = \mu_Y$.

t-test (paired)

Suppose we have pairs of RVs (X_i, Y_i), $i = 1 \ldots, n$. Let $D_i = X_i - Y_i$.

Suppose $D_1, \ldots, D_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, with σ^2 unknown, and that we want to test a hypothesis about μ. We can use the test statistic

$$\frac{\overline{D} - \mu_0}{S_D/\sqrt{n}}$$

which has a t_{n-1} distribution under $H_0 : \mu = \mu_0$. (Here, S_D^2 is the sample variance of the D_i.)

The kind of situation where a paired test is used is when there are two measurements on the same "experimental unit", e.g. in the sleep data, low and normal doses were given to the same 10 patients.

Two sample t and paired t

Is the amount of sleep gained with a low dose the same as the amount gained with a high dose?

low (X)	0.7	-1.6	-0.2	-1.2	-0.1	3.4	3.7	0.8	0.0	2.0
normal (Y)	1.9	0.8	1.1	0.1	-0.1	4.4	5.5	1.6	4.6	3.4
difference (D)	1.2	2.4	1.3	1.3	0.0	1.0	1.8	0.8	4.6	1.4

- Two sample t-test of $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X \neq \mu_Y$: the p-value is 0.079.
- Paired t-test (of $\mu_0 = 0$), based on the differences D_i: the p-value is 0.0028.

The paired test uses the information that the observations are paired: i.e. we have one low and one high dose observation per patient. The two sample test ignores this information. Prefer the paired test here.

Could consider one-sided alternatives here.