Two-Sample T-tests

Decisions with Data | Inference for means

STAT5002

The University of Sydney

May 2025

Decisions with Data

Topics 8 and 9: Confidence intervals and the z-test

Topic 10: The t-test

Topic 11: The two-sample test

Topic 12: χ^2-test

Outline

Redbull data

Comparing two (sample) means

The Classical Two-Sample T-test

The Welch Test

Using simulation

Paired (two-sample) T-test

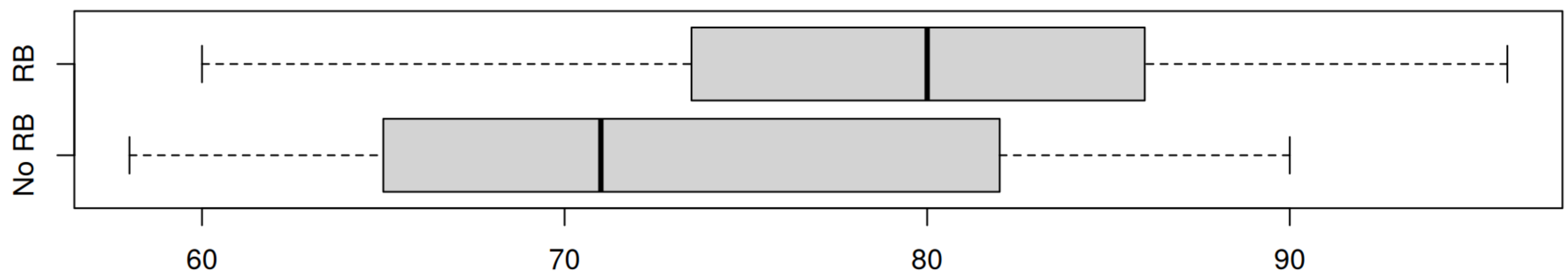
Red bull

Red Bull is an energy drink advertised to “give you wings”.

- We want to understand how much drinking Red Bull affects people medically (in terms of the heart rate).
- Consider the following data on heart rates (beats per minute), for 2 independent groups of Sydney students, collected 20 minutes after the ‘RedBull’ group had drunk a 250ml cold can of Red Bull.

No Red Bull	84	76	68	80	62	58	74	84	68	90	82	64	65	66
Red Bull	72	88	72	88	76	75	84	80	60	96	80	84	-	-

```r
No_RB <- c(84,76,68,80,62,58,74,84,68,90,82,64,65,66)
RB <- c(72,88,72,88,76,75,84,80,60,96,80,84)
boxplot(No_RB, RB, names=c("No RB", "RB"),  horizontal=T)
```



- The Red Bull group seems to have a higher heart rate. **Is the difference significant?**

Comparing two (sample) means

Two-box model

We can model the two groups as samples taken from two separate boxes (independently of each other).

That is, we model

- the "No Red Bull" group as a random sample $X_1, \ldots, X_m$ taken (with repl.) from a box with
  - mean μ_X and
  - SD σ_X;
- the "Red Bull" group as a random sample $Y_1, \ldots, Y_n$ taken (with repl.) from a box with
  - mean μ_Y and
  - SD σ_Y;

We *really* wish to make a statement about the **population** mean difference μ_X and μ_Y, based on the **sample** mean difference $\bar{X} - \bar{Y}$.

Expected values and SEs

- We know that
  - $E(\bar{X}) = \mu_X$;
  - $SE(\bar{X}) = \frac{\sigma_X}{\sqrt{m}}$;
  - $E(\bar{Y}) = \mu_Y$;
  - $SE(\bar{Y}) = \frac{\sigma_Y}{\sqrt{n}}$.
- But what about $\bar{X} - \bar{Y}$?

Revisit Topic 6

- Note that the difference $\bar{X} - \bar{Y} = \bar{X} + (-\bar{Y})$ where $-\bar{Y}$ has
  - $E(-\bar{Y}) = -E(\bar{Y})$;
  - $SE(-\bar{Y}) = SE(\bar{Y})$.
- Hence we can use results in Topic 6 (box models) to conclude
  - $E(\bar{X} - \bar{Y}) = E(\bar{X}) + E(-\bar{Y}) = E(\bar{X}) - E(\bar{Y}) = \mu_X - \mu_Y$;
  - Most importantly

$$SE(\bar{X} - \bar{Y})^2 = SE(\bar{X})^2 + SE(-\bar{Y})^2 = \frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}.$$

- That is

$$SE(\bar{X} - \bar{Y}) = \sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}.$$

Two-sample Test Statistics

- We wish to test the null hypothesis H_0: $\mu_X = \mu_Y$.
- If the two box SDs σ_X and σ_Y were known, we could test H_0 using the Z-statistic

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \sim N(0,1)$$

  assuming H_0 true.
- In general, σ_X and σ_Y are both unknown.
- In this case we have two options:
  - ➟ Assume $\sigma_X = \sigma_Y = \sigma$ is the same in both boxes: **Classical Two-Sample T-test**
  - ➟ Do **not** assume $\sigma_X = \sigma_Y = \sigma$ the same in both boxes: **Welch Test**.

The Classical Two-Sample T-test

“Equal variance” assumption

- In some cases it is reasonable to assume $\sigma_X = \sigma_Y = \sigma$
    - This is often called an **equal variances** assumption, i.e. $\sigma_X^2 = \sigma_Y^2$
        - for “normal populations” (idealised, infinite boxes whose histograms follow a normal curve exactly) it is more common to refer to variances, i.e. squared SDs.
- Then the SE may be written as

$$SE(\bar{X} - \bar{Y}) = \sigma\sqrt{\frac{1}{m} + \frac{1}{n}}\,.$$

Extra assumptions: Student's t-distribution

- In this case, *if* it is also assumed the boxes are **approximately normal-shaped**, a special **pooled** estimate $\widehat{\sigma}_p$ of the common σ is used

$$\widehat{\sigma}_p = \sqrt{\frac{\sum_{i=1}^{m}(X_i - \bar{X})^2 + \sum_{j=1}^{n}(Y_j - \bar{Y})^2}{m+n-2}} = \sqrt{\frac{(m-1)\widehat{\sigma}_X^2 + (n-1)\widehat{\sigma}_Y^2}{m+n-2}}.$$

- The estimated variance $\widehat{\sigma}_p^2$ is a **weighted average** of $\widehat{\sigma}_X^2$ and $\widehat{\sigma}_Y^2$.
    - ➟ The bigger sample gets more weight.
    - ➟ The estimate from the larger sample is somehow "more trustworthy".
- Then Student's theory can be applied to show the statistic

$$T = \frac{\bar{X} - \bar{Y}}{\widehat{\sigma}_p\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t_{m+n-2}$$

  i.e. has Student's-t distribution with $m+n-2$ degrees of freedom
- Why do we have $m+n-2$ (the derivation is not for assessment)?

Squared estimate $\widehat{\sigma}_p^2$ is "on target" for σ^2

- Recall that each sample variance (squared sample SD) estimates σ^2 "in expectation", in that

$$E(\widehat{\sigma}_X^2) = E\left(\frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2\right) = \sigma^2$$

and so

$$E\left((m-1)\widehat{\sigma}_X^2\right) = E\left(\sum_{i=1}^{m}(X_i - \bar{X})^2\right) = (m-1)\sigma^2$$

- Similarly we have

$$E\left((n-1)\widehat{\sigma}_Y^2\right) = E\left(\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right) = (n-1)\sigma^2$$

- Then the numerator inside the $\sqrt{\cdot}$ has

$$
\begin{aligned}
E\left(\sum_{i=1}^{m}(X_i - \bar{X})^2 + \sum_{j=1}^{n}(Y_j - \bar{Y})^2\right) &= E\left((m-1)\widehat{\sigma}_X^2\right) + E\left((n-1)\widehat{\sigma}_Y^2\right) \\
&= (m-1)\sigma^2 + (n-1)\sigma^2 \\
&= (m+n-2)\sigma^2 \,.
\end{aligned}
$$

- Dividing through by $m + n - 2$ we get

$$
E\left(\widehat{\sigma}_p^2\right) = \sigma^2 \,,
$$

so $\widehat{\sigma}_p^2$ shares the "on-target in expectation" property that $\hat{\sigma}_X^2$ and $\widehat{\sigma}_Y^2$ have.

- As in the one-sample T-test, the denominator in the estimate of σ^2 is also the degrees of freedom.
    - ➟ We "lose one degree of freedom" for each sample SD we estimate.
    - ➟ Degrees of freedom is then total sample size, minus 2.

Red Bull example

$\boxed{\text{H}}$ We want to test whether there is a significant difference between the two groups: those who consumed Red Bull (RB) and those who did not (No RB).

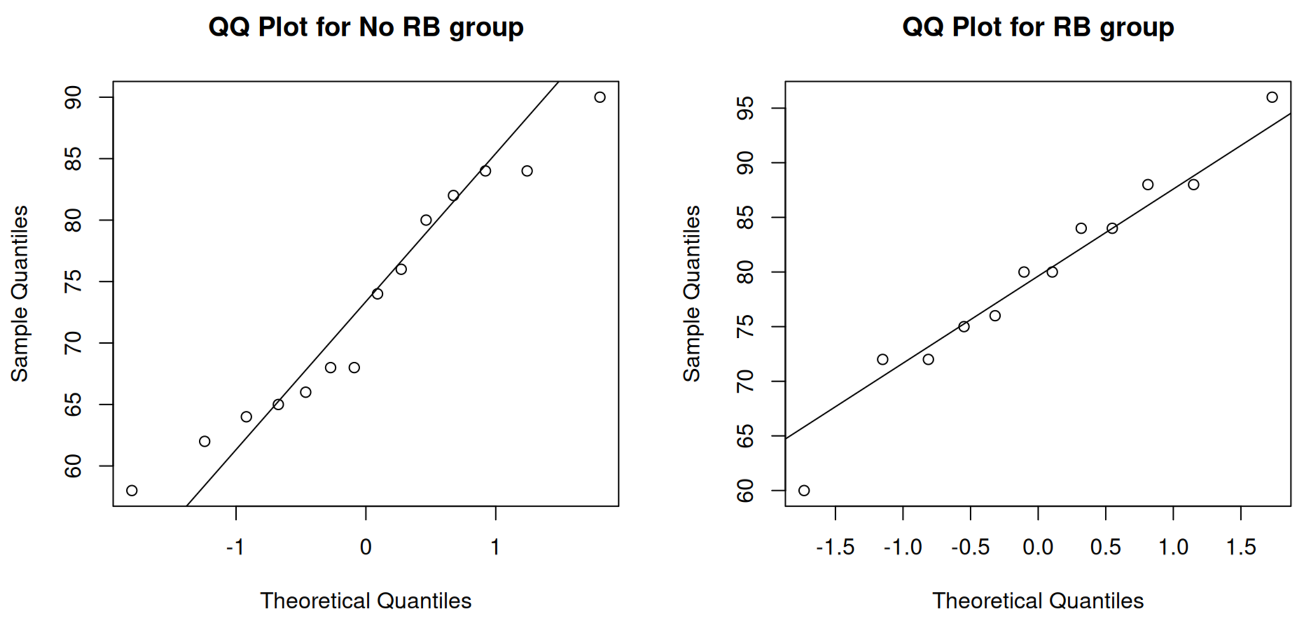
- **Null hypothesis** (H_0): There is no difference in the means of the two groups.
  - $H_0 : \mu_{\text{No RB}} = \mu_{\text{RB}}$
- **Alternative hypothesis** (H_1): There is a difference in the means of the two groups.
  - $H_1 : \mu_{\text{No RB}} \neq \mu_{\text{RB}}$

Since we are testing for the difference between the two groups, this is a two-sided test.

A Check whether the assumptions of a two sample t-test are reasonable:

- Approximately normal distributions?

```r
par(mfrow=c(1,2))
qqnorm(No_RB, main = "QQ Plot for No RB group")
qqline(No_RB)
qqnorm(RB, main = "QQ Plot for RB group")
qqline(RB)
```



- Each group looks roughly normal shaped, but QQ plots don't follow QQ lines closely, we may also need to simulate T-statistics and calculate the simulation-based P-value to validate.

- Similar spreads?

```
sd(No_RB)
```

```
[1] 9.848579
```

```
sd(RB)
```

```
[1] 9.452833
```

- The standard deviations of both groups are similar
- This suggests we can assume a common standard deviation across groups

$\boxed{\mathrm{T}}$ Two-sided test: small and large test statistic argues against H_0

- Compute the pooled standard deviation

```
m = length(No_RB)
n = length(RB)

numer = (m-1)*(sd(No_RB)^2) +  (n-1)*(sd(RB)^2)
dof = m+n-2
sig.hat.p = sqrt(numer/dof)
sig.hat.p
```

```
[1] 9.669206
```

- Compute the standard error

```
est.SE = sig.hat.p*sqrt((1/m)+(1/n))
est.SE
```

```
[1] 3.803845
```

- Compute the test statistic

```
mean.diff = mean(No_RB)-mean(RB)
stat = mean.diff/est.SE
stat
```

```
[1] -1.749483
```

P A two-sided P-value is

```r
2*pt(abs(stat), df=m+n-2, lower.tail=F)
```

```
[1] 0.09298616
```

Of course, the `t.test()` function can do all of this in one line;

- We must supply the `var.equal=T` parameter:

```r
t.test(No_RB, RB, var.equal=T)
```

```

	Two Sample t-test

data:  No_RB and RB
t = -1.7495, df = 24, p-value = 0.09299
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -14.505513   1.195989
sample estimates:
mean of x mean of y 
 72.92857  79.58333 
```

C The difference is not significant for any false alarm rate less than ~9.3%.

- If we use the default 5% false alarm rate, the data is consistent with H_0 (Red Bull does not significantly change the heart rate).

Confidence interval

- Note that the confidence interval (for the difference between means) given here is obtained in the familar way,

$$\frac{(\bar{X}-\bar{Y})-E(\bar{X}-\bar{Y})}{SE(\bar{X}-\bar{Y})}=\frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\widehat{\sigma}_p\sqrt{\frac{1}{m}+\frac{1}{n}}}\sim t_{m+n-2}$$

so we need to find multipliers l and u such that

$$P\left(l\leq\frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{SE(\bar{X}-\bar{Y})}\leq u\right)=0.95$$

which gives the 95% confidence interval for the **unknown difference between population means**:

$$P\left((\bar{X}-\bar{Y})-u\times SE(\bar{X}-\bar{Y})\leq\mu_X-\mu_Y\leq(\bar{X}-\bar{Y})-l\times SE(\bar{X}-\bar{Y})\right)=0.95$$

- By symmetry of Student's T, we have $-l = u$
- We can use `qt()` and to find the upper quantile $u =$ `qt(` $1 - \frac{\alpha}{2}$ `, df=m+n-2)` where $1 - \alpha$ gives the percentage of the confidence interval:

```
qt(0.975, df=m+n-2) # the multiplier
```

```
[1] 2.063899
```

```
mean.diff+c(-1,1)*qt(0.975, df=m+n-2)*est.SE
```

```
[1] -14.505513   1.195989
```

- It contains 0, so the "no difference" claim of H_0 is consistent with observed data at the 5% level of significance.

The Welch Test

Relaxing the equal variance assumption

- If we want to apply Student's theory directly, we need to assume $\sigma_X = \sigma_Y$.
- What if it is not reasonable to assume this?
  - ➠ E.g., the two boxplots may have very different spreads.
- An "obvious" approach would be to instead consider the statistic

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\hat{\sigma}_X^2}{m} + \frac{\hat{\sigma}_Y^2}{n}}},$$

  which just plugs in the two sample SD estimates in for σ_X and σ_Y.
- How to get a P-value?
  - ➠ what is the distribution of T if $H_0\colon \mu_X = \mu_Y$ is true?

Welch's paper

- In 1947 (some time after Student's paper) B. L. Welch found that the test statistic behaved **approximately** like a Students-t distribution whose degrees of freedom is a complicated function of m, n, σ_X and σ_Y.
- The **Welch Test** obtains the P-value using a Student's-t distribution with a **data-dependent degrees of freedom** (beyond the scope of this unit).

THE GENERALIZATION OF 'STUDENT'S' PROBLEM WHEN SEVERAL DIFFERENT POPULATION VARIANCES ARE INVOLVED

BY B. L. WELCH, B.A., PH.D.

1. *Introduction and summary.* Let η be a population parameter which is estimated by an observed quantity y, normally distributed with variance σ_y^2. Let $\sigma_y^2 = \sum_{i=1}^{k} \lambda_i \sigma_i^2$, where the λ_i are known positive numbers and the σ_i^2 are unknown variances. Suppose that the observed data provide estimates s_i^2 of these variances, based on f_i degrees of freedom, respectively, so that the sampling distribution of s_i^2 is

$$p(s_i^2)\,ds_i^2 = \frac{1}{\Gamma(\frac{1}{2}f_i)}\left(\frac{f_i s_i^2}{2\sigma_i^2}\right)^{\frac{1}{2}f_i-1}\exp\left[-\frac{1}{2}\frac{f_i s_i^2}{\sigma_i^2}\right]d\left(\frac{f_i s_i^2}{2\sigma_i^2}\right), \tag{1}$$

and that these estimates are distributed independently of each other and of y.

A very simple particular case of this set-up occurs when we have samples of n_1 and n_2, respectively, from two normal populations with true means α_1 and α_2 and standard deviations σ_1 and σ_2. If η is the true difference $(\alpha_1 - \alpha_2)$ between the means, the estimated difference is $y = (\bar{x}_1 - \bar{x}_2)$. The variance of the estimate is $\sigma_y^2 = (\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2)$, where $\lambda_1 = 1/n_1$ and $\lambda_2 = 1/n_2$. The estimated values of σ_1^2 and σ_2^2 are $s_1^2 = \Sigma_1/f_1$ and $s_2^2 = \Sigma_2/f_2$, where Σ_1 and Σ_2 are the respective sums of squares of observations from the individual sample means and $f_1 = (n_1 - 1)$ and $f_2 = (n_2 - 1)$. These s^2 are distributed in the form (1) and the postulated conditions of independence hold.

Default two-sample `t.test()`

- It turns out Welch's procedure works very well. R uses the Welch test as the default two-sample T-test:

```r
t.test(No_RB, RB) # note: data-dependent d.f. close to Classical (which was 24 d.f.)
```

```

	Welch Two Sample t-test

data:  No_RB and RB
t = -1.7552, df = 23.66, p-value = 0.09216
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -14.485720   1.176197
sample estimates:
mean of x mean of y 
 72.92857  79.58333 
```

- Note that in **the degrees of freedom**:
  - the Welch test uses a complicated formula of the standard deviations of samples.
  - the two-sample t-test (under the equal variance assumption)
    - takes the degrees of freedom as $df = m + n - 2$.
    - the degrees of freedom is independent of the sample SDs
- The test statistics are also different, as different estimated SEs are used.

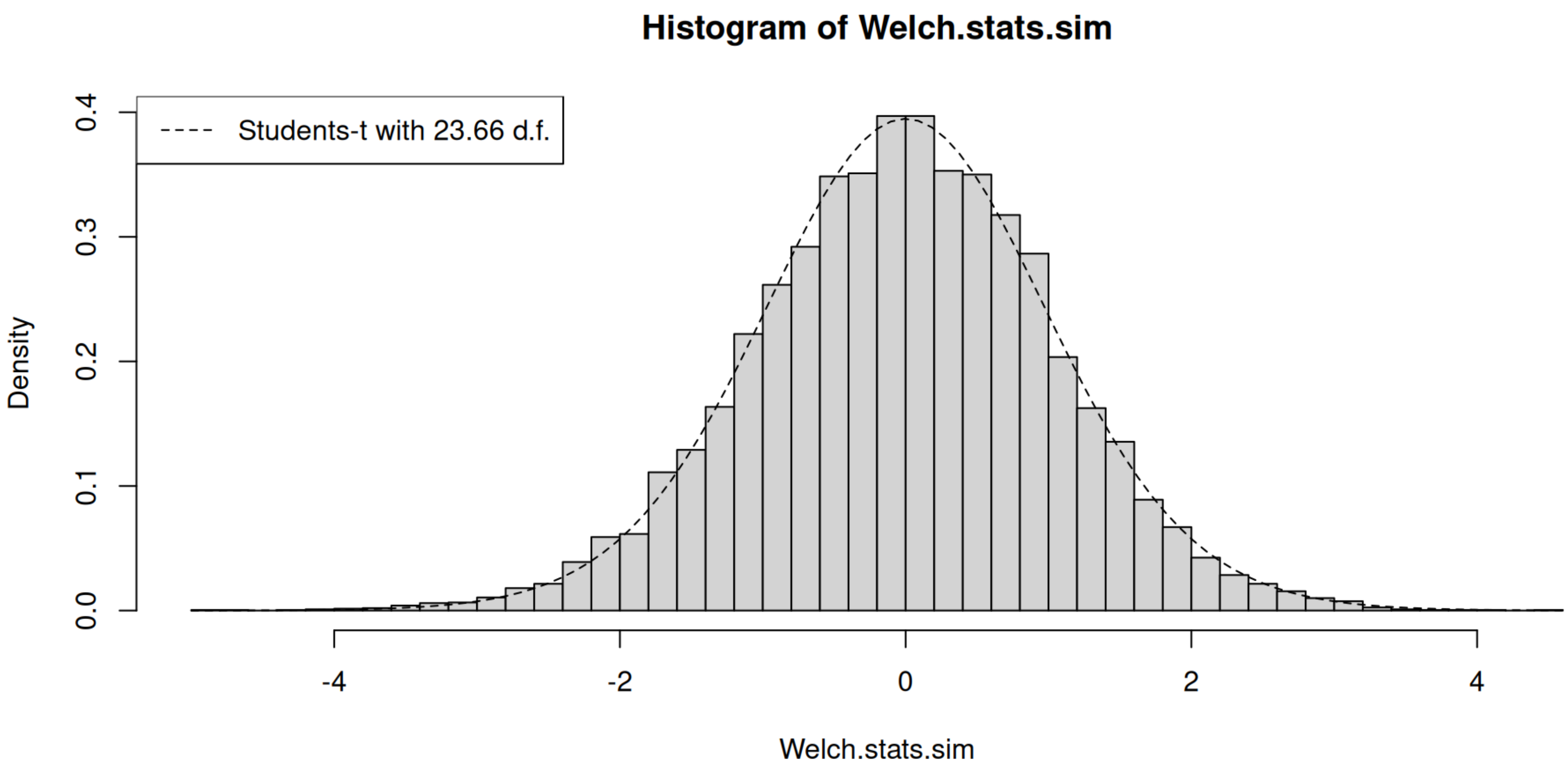
Using simulation

Using simulation

- The Welch test does not assume $\sigma_X = \sigma_Y$, but it still assumes the two boxes are "approximately normal".
- What if we are uncomfortable making this assumption?
- We can try simulating from two surrogate boxes which
  - have "similar shapes" to the "true" boxes that generated our data;
  - have **equal means** (H_0).
- We thus sample from each observed sample (surrogate population) with replacement; but we subtract the means so both surrogate populations have the same mean (i.e. zero).

```
No_RB.g = No_RB-mean(No_RB)  # both sorrogate boxes have
RB.g = RB-mean(RB)           # mean zero
Welch.stats.sim = 0
for(i in 1:10000) {
    samp.x = sample(No_RB.g, size = m, replace=T)
    samp.y = sample(RB.g, size = n, replace=T)
    est.SE = sqrt( (sd(samp.x)^2)/m + (sd(samp.y)^2)/n )
    Welch.stats.sim[i] = (mean(samp.x)-mean(samp.y))/est.SE
}
```

```r
hist(Welch.stats.sim, n=50, freq=F)
curve(dt(x, df=23.66), add=T, lty=2)      # data-dependent d.f. from original sample
legend("topleft", legend=c("Students-t with 23.66 d.f."), lty=2)
```



- The histogram is a little left-skewed.
  - This is maybe due to the slight departure from normal curve in the `No_RB` sample.

Two-sided P-value by simulation

- First calculate the Welch statistic

```
mean.diff = mean(No_RB)-mean(RB)
est.SE = sqrt( (sd(No_RB)^2)/m + (sd(RB)^2)/n )
stat = mean.diff/est.SE
stat
```

```
[1] -1.755237
```

- Then calculate the simulation-based P-value

```
mean(abs(Welch.stats.sim) >= abs(stat))
```

```
[1] 0.0936
```

This is very close to the earlier P-values (both Classical and Welch).

Confidence interval by simulation

- We use the simulated values in `Welch.stats.sim` to approximate the "true distribution" of the Welch statistic when $\mu_X = \mu_Y$ (under H_0):

```
u.l = quantile(Welch.stats.sim, prob=c(.975, .025))
u.l
```

```
   97.5%      2.5% 
2.017352 -2.160676 
```

- That these are not the same magnitude indicates the slight lack of symmetry.

```
mean.diff - u.l*est.SE
```

```
     97.5%       2.5% 
-14.303296   1.537169 
```

- The interval is quite close to those obtained by (both versions of) `t.test()`.

Paired (two-sample) T-test

Two paired samples

- A common scenario is where we have two samples of data (X, Y) **but**
  - ➠ are obtained from reading a **pair** of data (X_i, Y_i) from n individuals.
- In this case, the two samples are **not independent** and we cannot compare the two sample means using the methods we have already seen.

Example

The data from Student's original 1908 paper, which involve the treatment of 10 patients using two different treatment plans (intended to increase sleep time).

```r
dextro = c(.7, -1.6, -.2, -1.2, -.1, 3.4, 3.7, .8, 0, 2)
laevo = c(1.9, .8, 1.1, .1, -.1, 4.4, 5.5, 1.6, 4.6, 3.4)
diff = laevo-dextro
```

- The R comand `apply( ... , 2, mean)` applies the function `mean()` to each column; similarly `apply( ... , 2, sd)` gives column `sd()`s:

```r
sleep=data.frame(dextro, laevo, diff)
sleep
```

```
   dextro laevo diff
1     0.7   1.9  1.2
2    -1.6   0.8  2.4
3    -0.2   1.1  1.3
4    -1.2   0.1  1.3
5    -0.1  -0.1  0.0
6     3.4   4.4  1.0
7     3.7   5.5  1.8
8     0.8   1.6  0.8
9     0.0   4.6  4.6
10    2.0   3.4  1.4
```

```r
apply(sleep, 2, mean)
```

```
dextro  laevo   diff
  0.75   2.33   1.58
```

```r
apply(sleep, 2, sd)
```

```
  dextro    laevo     diff
1.789010 2.002249 1.229995
```

- Note here that `sd(diff)` is much smaller than it would be if the two samples were independent.

Paired T-test

- $\boxed{\text{H}}$ $H_0 : \mu_{\text{L}} = \mu_{\text{D}}$ or $\mu_{\text{diff}} = 0$ the population means are the same. To assess whether the sample mean difference is significantly
  - ➠ different to zero $H_1 : \mu_{\text{L}} \neq \mu_{\text{D}}$ or $\mu_{\text{diff}} \neq 0$;
  - ➠ greater than zero $H_1 : \mu_{\text{L}} > \mu_{\text{D}}$ or $\mu_{\text{diff}} > 0$;
  - ➠ less than zero $H_1 : \mu_{\text{L}} < \mu_{\text{D}}$ or $\mu_{\text{diff}} < 0$;

  we simply perform the appropriate **one-sample T-test** on the **sample differences**.
- $\boxed{\text{A}}$ Is the difference approximately normal?
  - ➠ If each data set is approximately normal, then the difference is approximately normal as well.

Two ways for a paired T-test using `t.test()`:

- Manually taking differences for $H_1 : \mu_{\text{diff}} \neq 0$

```r
t.test(laevo-dextro)
```

```

	One Sample t-test

data:  laevo - dextro
t = 4.0621, df = 9, p-value = 0.002833
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.7001142 2.4598858
sample estimates:
mean of x 
     1.58 
```

- Using `t.test( ... , paired=T)`:

```r
t.test(laevo, dextro, paired=T)
```

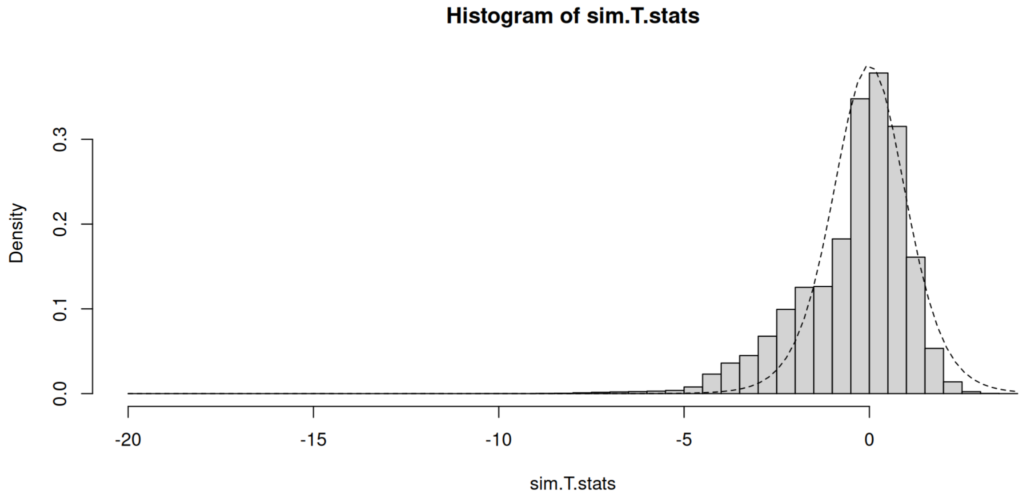
```

	Paired t-test

data:  laevo and dextro
t = 4.0621, df = 9, p-value = 0.002833
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 0.7001142 2.4598858
sample estimates:
mean difference 
           1.58 
```

Simulation-based P-value

- Use the difference as the surrogate box

```r
diff = laevo-dextro
sim.T.stats=0
n = length(diff)
box.guess = diff
for(i in 1:100000) {
    samp = sample(box.guess, size=n, replace=T)
    sim.T.stats[i] = sqrt(n)*(mean(samp)-mean(diff))/sd(samp)
}
```

```r
hist(sim.T.stats, pr=T, n=50)
curve(dt(x, df=9), add=T, lty=2)
```



- Observed T statistic

```r
t.stat = sqrt(10)*(mean(diff))/sd(diff)
t.stat
```

```
[1] 4.062128
```

- A two-sided P-value – the proportion of simulated T-statistics exceeding the observed statisitc in absolute value:

```r
mean(abs(sim.T.stats) >= abs(t.stat))
```

```
[1] 0.02109
```

One-sided tests

One-sided tests can be done using `alternative="greater"` or `alternative="less"` as usual;

- but be careful of the order;
- the following three tests are the same for the alternative $H_1 : \mu_{\mathrm{L}} > \mu_{\mathrm{D}}$ or $\mu_{\mathrm{diff}} = \mu_{\mathrm{L}} - \mu_{\mathrm{D}} > 0$

```r
# test 1
t.test(laevo, dextro, paired=T, alternative="greater")
```

```

	Paired t-test

data:  laevo and dextro
t = 4.0621, df = 9, p-value = 0.001416
alternative hypothesis: true mean difference is greater than 0
95 percent confidence interval:
 0.8669947       Inf
sample estimates:
mean difference 
           1.58 
```

- Equivalently, $H_1 : \mu_{\mathrm{L}} - \mu_{\mathrm{D}} > 0$

```r
# test 2
t.test(laevo-dextro, alternative="greater")
```

```

	One Sample t-test

data:  laevo - dextro
t = 4.0621, df = 9, p-value = 0.001416
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 0.8669947       Inf
sample estimates:
mean of x
     1.58
```

- Equivalently, $H_1 : \mu_{\mathrm{D}} < \mu_{\mathrm{L}}$ or $\mu_{\mathrm{D}} - \mu_{\mathrm{L}} < 0$

```r
# test 3
t.test(dextro, laevo, paired=T, alternative="less")
```

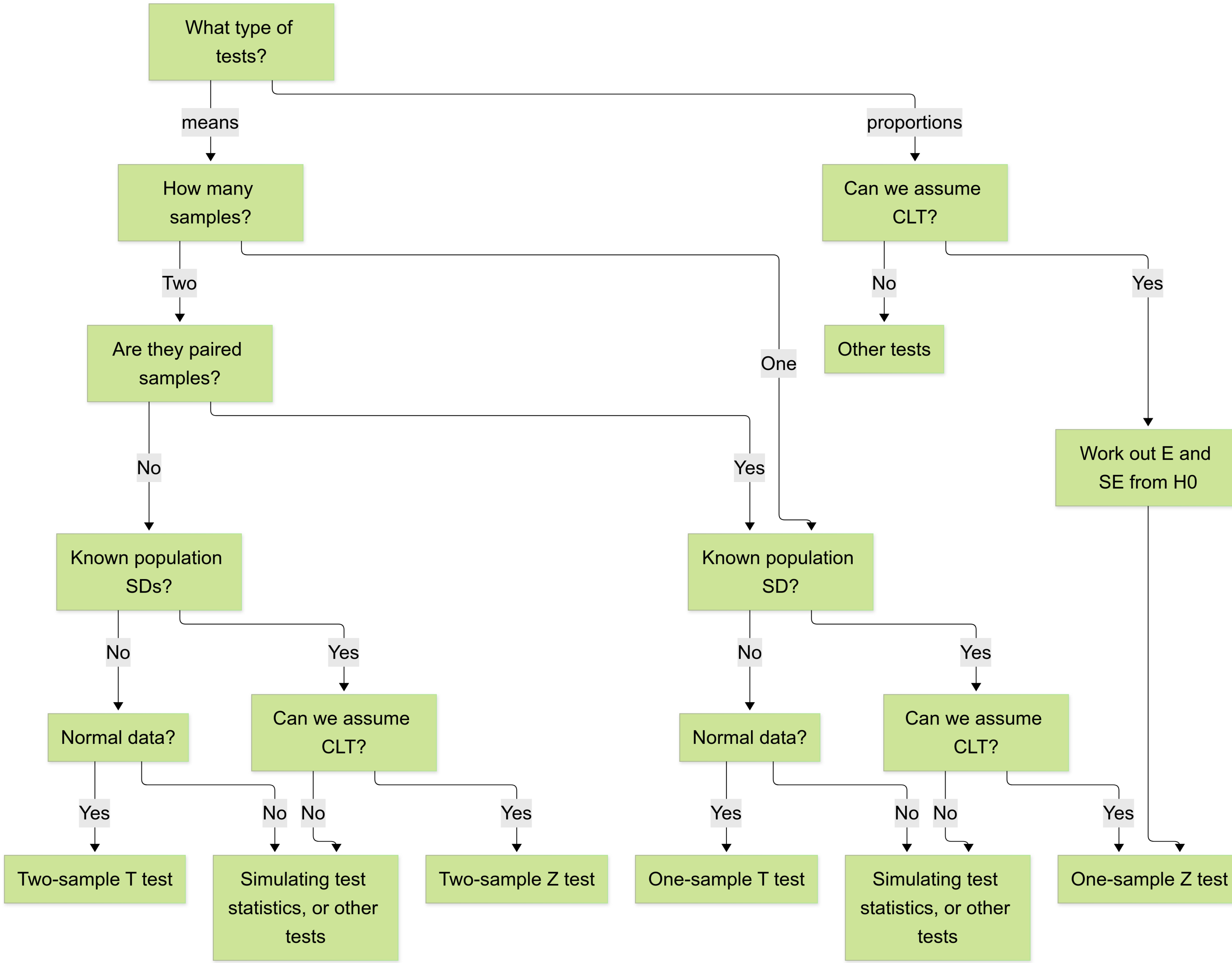
```

	Paired t-test

data:  dextro and laevo
t = -4.0621, df = 9, p-value = 0.001416
alternative hypothesis: true mean difference is less than 0
95 percent confidence interval:
       -Inf -0.8669947
sample estimates:
mean difference
          -1.58
```

- Using simulations, the one-sided P-value for $H_1 : \mu_{\mathrm{L}} > \mu_{\mathrm{D}}$ (or $\mu_{\mathrm{diff}} = \mu_{\mathrm{L}} - \mu_{\mathrm{D}} > 0$) is given by the proportion of simulated T-statistics exceeding the observed statisitc:

```r
mean(sim.T.stats >= t.stat)
```

```
[1] 0
```

Summary of Z and T tests


What type of tests?
means
proportions
How many samples?
Can we assume CLT?
Two
One
No
Yes
Are they paired samples?
Other tests
No
Yes
Work out E and SE from H0
Known population SDs?
Known population SD?
No
Yes
No
Yes
Normal data?
Can we assume CLT?
Normal data?
Can we assume CLT?
Yes
No
No
Yes
Yes
No
No
Yes
Two-sample T test
Simulating test statistics, or other tests
Two-sample Z test
One-sample T test
Simulating test statistics, or other tests
One-sample Z test

- One-sample Z test

$$Z = \frac{\bar{x} - E_0(\bar{X})}{SE_0(\bar{X})} \quad \text{where} \quad SE_0(\bar{X}) = \underbrace{\frac{\sigma}{\sqrt{n}}}_{\text{mean, known popu. SD}} \quad \text{or} \quad SE_0(\bar{X}) = \underbrace{\sqrt{\frac{p_0(1-p_0)}{n}}}_{\text{proportion}}$$

- One sample T test

$$T = \frac{\bar{x} - E_0(\bar{X})}{\frac{\hat{\sigma}}{\sqrt{n}}} \sim t_{n-1}$$

- Two-sample Z test

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}}$$

- Two-sample T test (classic)

$$T = \frac{\bar{X} - \bar{Y}}{\widehat{\sigma}_p \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t_{m+n-2}, \quad \widehat{\sigma}_p = \sqrt{\frac{(m-1)\widehat{\sigma}_X^2 + (n-1)\widehat{\sigma}_Y^2}{m+n-2}}$$

- Two-sample T test (Welch)

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\widehat{\sigma}_X^2}{m} + \frac{\widehat{\sigma}_Y^2}{n}}} \sim t_{\mathrm{dof}}$$

where the degrees of freesom (`dof`) is a complicated function of sample sizes and SDs (so we use R).