Two-sample testing for difference in mean (檢定二組樣本是否有相同 population mean).

- 問題: 有二種降血壓藥物A和B. 想知道二藥效果是否有差. 假設有二組病人, 一組病人試用A藥而另一組病人試用B藥. 假設A藥組有 n_1 人, B藥組有 n_2人. 試用結果, A藥組血壓降低量爲 $X_1, \ldots, X_{n_1}$, 而B藥組血壓降低量爲 $Y_1, \ldots, Y_{n_2}$. 血壓測量單位爲mm-Hg.

  如何根據 $(X_1, \cdots, X_{n_1})$ 和 $(Y_1, \cdots, Y_{n_2})$二組樣本來決定二藥效果是否有差?

- Data assumptions for two independent samples
  - (i) $(X_1, \cdots, X_{n_1})$ and $(Y_1, \cdots, Y_{n_2})$ are independent.
  - (ii) $(X_1, \cdots, X_{n_1})$ and $(Y_1, \cdots, Y_{n_2})$ are two random samples.
  - (iii) $X_i \sim N(\mu_1, \sigma_1^2)$ for $i = 1, \ldots, n_1$; $Y_j \sim N(\mu_2, \sigma_2^2)$ for $j = 1, \ldots, n_2$.

- Notation:
  - $\bar{X}$ and S_X: sample mean and sample standard deviation for $(X_1, \cdots, X_{n_1})$;
  - $\bar{Y}$ and S_Y: sample mean and sample standard deviation for $(Y_1, \cdots, Y_{n_2})$.

- 若二組樣本 population variances σ_1, σ_2爲已知，可使用Z檢定
  - 資料: $(X_1, \cdots, X_{n_1})$ 和 $(Y_1, \cdots, Y_{n_2})$二組樣本
  - 假設(i)(ii)(iii)成立，即二組樣本爲獨立，每組樣本爲 random sample, $X_i \sim N(\mu_1, \sigma_1^2)$, $Y_i \sim N(\mu_2, \sigma_2^2)$.
  - 檢定問題:
  $$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 \neq \mu_2.$$
  - $\bar{X}$ and $\bar{Y}$: sample means for $(X_1, \cdots, X_{n_1})$ and $(Y_1, \cdots, Y_{n_2})$.
  - 檢定統計量
  $$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$
    - 令
    $$Z_0 = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}},$$
    則 $Z_0 \sim N(0, 1)$. 因此當$H_0 : \mu_1 = \mu_2$成立時， 檢定統計量$Z \sim N(0, 1)$.
  - 檢定規則：當顯著水準爲a時，若$|Z| > z_{a/2}$，則Z檢定拒絕$H_0 : \mu_1 = \mu_2$. Z檢定之p-value 爲
  $$2P\left(N(0, 1) > \text{ observed } |Z|\right).$$

- 假設(i)(ii)(iii)成立而$\sigma_1 = \sigma_2$爲未知時，可使用 pooled t test.

- Pooled t test.
  - Data assumptions
    (i) $(X_1, \cdots, X_{n_1})$ and $(Y_1, \cdots, Y_{n_2})$ are independent.
    (ii) $(X_1, \cdots, X_{n_1})$ and $(Y_1, \cdots, Y_{n_2})$ are two random samples.
    (iv) $X_i \sim N(\mu_1, \sigma^2)$ for $i = 1, \ldots, n_1$; $Y_j \sim N(\mu_2, \sigma^2)$ for $j = 1, \ldots, n_2$.
  - Testing problem (two-sided case): $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$.
  - Test statistic
$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\hat{\sigma}^2 \left(1/n_1 + 1/n_2\right)}},$$
    where
$$\hat{\sigma} = \sqrt{\frac{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2}{n_1 + n_2 - 2}}.$$
    * Let
$$T_0 = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\hat{\sigma}^2 \left(1/n_1 + 1/n_2\right)}},$$
      then $T_0 \sim t(n_1 + n_2 - 2)$, so $T \sim t(n_1 + n_2 - 2)$ when $\mu_1 = \mu_2$.
  - The pooled t test rejects H_0: $\mu_1 = \mu_2$ at level a when
$$|T| > t_{a/2, n_1 + n_2 - 2}.$$
    The test is of size a. The p-value of the test is
$$2P(t(n_1 + n_2 - 2) > \text{observed } |T|).$$
  - For the testing problem $H_0 : \mu_1 \leq \mu_2$ versus $H_1 : \mu_1 > \mu_2$, the pooled t test rejects $H_0 : \mu_1 \leq \mu_2$ at level a when
$$T > t_{a, n_1 + n_2 - 2}.$$
    The p-value of the test is
$$P(t(n_1 + n_2 - 2) > \text{observed } T).$$
  - For the testing problem $H_0 : \mu_1 \geq \mu_2$ versus $H_1 : \mu_1 < \mu_2$, the pooled t test rejects $H_0 : \mu_1 \geq \mu_2$ at level a when
$$T < -t_{a, n_1 + n_2 - 2}.$$
    The p-value of the test is
$$P(t(n_1 + n_2 - 2) < \text{observed } T).$$

- Example 1. 降血壓藥物問題. 有二種降血壓藥物A和B. 想知道二藥效果是否有差. 假設有二組病人, 一組病人試用A藥而另一組病人試用B藥.
  - A 組人數: 10; 血壓降低量(收縮壓) sample mean: 15 (mm-Hg); sample standard deviation: 2.8 (mm-Hg).

- B 組人數: 20; 血壓降低量 sample mean: 21 (mm-Hg); sample standard deviation: 3.2 (mm-Hg).

假設使用A或B藥物後, 血壓降低量分布約爲 normal, 且二 normal 分布有相同的 variance. 在 0.05 顯著水準下, 是否可推論二藥效果有差?

Sol. $t_{0.05/2,10+20-2} = t_{0.025,28} = 2.048$.

$$\text{observed } |T| = \frac{|15-21|}{\sqrt{(9(2.8^2)+19(3.2^2))(1/10+1/20)/28}} = 5.034582 > 2.048.$$

在 0.05 顯著水準下, 可推論二藥效果有差.

- Example 2. Example 1 中, 若顯著水準改爲0.04, 0.03, 或0.02, 是否可推論二藥效果有差?

  p-value $= 2P(t(28) > 5.034582)$. 用課程網站上提供的表 "Quantiles for t distributions", 可知 $2P(t(28) > 3.674) = 2(0.0005) = 0.001$, 所以$p$-value <0.001. 因此無論顯著水準改爲0.04, 0.03, 或0.02, 皆可推論二藥效果有差.

  用R指令 `2*(1-pt(5.034582, df=28))` 也可算出 p-value $= 2.523123 \times 10^{-5} < 0.001$.

- Example 3. Example 1 中, 如果 A, B 二組血壓降低量 population variances 分別爲 2.8^2 和 3.2^2, 則可使用 Z 統計量檢定 population means 是否相等. 此時

$$|Z| = \frac{|15-21|}{\sqrt{2.8^2/10+3.2^2/20}} = 5.270463$$

  而 p-value $= 2P(N(0,1) > 5.270463)$. 由表"Normal Probabilities" 查表得知,

$$P(N(0,1) > 5.270463) < P(N(0,1) > 3.09) = 0.5 - 0.499 = 0.001. \quad (1)$$

  比(1)中更精確的$P(N(0,1) > 5.270463)$的範圍可由表"Quantiles for t distributions" 得知:

$$P(N(0,1) > 5.270463) < P(N(0,1) > 3.29) = 0.0005. \quad (2)$$

  由(2)可知p-value $= 2P(N(0,1) > 5.270463) < 2(0.0005) = 0.001$. 有極端強烈(extremely strong)的證據顯示A, B 二組血壓降低量 population means 不同.

- 檢定二組樣本是否有相同 population mean 時, 如果二組樣本不獨立而樣本數相同時, 可考慮將二組樣本$(X_1, \cdots, X_n)$ and $(Y_1, \cdots, Y_n)$相減得到$(X_1 - Y_1, \ldots, X_n - Y_n)$, 再使用 one-sample t test 檢定相減得到的樣本 population mean 是否爲0. 這種檢定方式稱爲 pairwise t test 或 paired t test.

- Pairwise t test (paired t test).
  - 資料: $(X_1, \cdots, X_n)$ 和 $(Y_1, \cdots, Y_n)$二組樣本. $E(X_i) = \mu_1$ and $E(Y_i) = \mu_2$.

- 假設$(X_1 - Y_1, \ldots, X_n - Y_n)$爲 random sample 且 $X_i - Y_i \sim$ normal.
- 檢定問題 (two-sided case):

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 \neq \mu_2.$$

- $\bar{D}$ and S_D: sample mean and sample standard deviation for $(X_1 - Y_1, \cdots, X_n - Y_n)$.
- 檢定統計量:

$$T = \frac{\sqrt{n}\bar{D}}{S_D}.$$

  The pairwise t test rejects $H_0 : \mu_1 = \mu_2$ at level a when

$$|T| > t_{a/2,n-1}.$$

  The p-value for the test is

$$2P\left(t(n-1) > \text{ observed } |T|\right).$$

- For the testing problem $H_0 : \mu_1 \leq \mu_2$ versus $H_1 : \mu_1 > \mu_2$, the paired t test rejects $H_0 : \mu_1 \leq \mu_2$ at level a when

$$T > t_{a,n-1}.$$

  The p-value of the test is

$$P(t(n-1) > \text{observed } T).$$

- For the testing problem $H_0 : \mu_1 \geq \mu_2$ versus $H_1 : \mu_1 < \mu_2$, the paired t test rejects $H_0 : \mu_1 \geq \mu_2$ at level a when

$$T < -t_{a,n-1}.$$

  The p-value of the test is

$$P(t(n-1) < \text{observed } T).$$

- Example 4. Suppose that we want to test whether a drug has the side-effect of changing systolic blood pressure (收縮壓). The blood pressures for 6 patients before/after using the drug are given in the following table.

Patient ID	Blood pressure before treatment	Blood pressure after treatment
1	116.6	120.4
2	117.6	119.3
3	118.3	119.9
4	117.4	118.6
5	117.4	121.3
6	114.3	119.4

Can we conclude that the drug has the side-effect of changing blood pressure at the 0.05 significance level?

Sol. $t_{0.05/2,6-1} = t_{0.025,5} = 2.571$. The differences between the before-treatment blood pressures and the after-treatment blood pressures are $(120.4 - 116.6) = 3.8$,$(119.3 - 117.6) = 1.7$,$(119.9 - 118.3) = 1.6$,$(118.6 - 117.4) = 1.2$,$(121.3 - 117.4) = 3.9$,$(119.4 - 114.3) = 5.1$. Compute the sample mean and sample variance for the difference data and we have the sample mean is

$$\frac{3.8 + 1.7 + 1.6 + 1.2 + 3.9 + 5.1}{6} = \frac{17.3}{6},$$

the sample variance is

$$\frac{1}{6-1}\left(3.8^2 + 1.7^2 + 1.6^2 + 1.2^2 + 3.9^2 + 5.1^2 - 6\left(\frac{17.3}{6}\right)^2\right) = \frac{76.01}{30},$$

and the $|T|$ statistic for the pairwise t test is

$$\frac{\sqrt{6}(17.3/6)}{\sqrt{76.01/30}} = 4.437064.$$

$4.437064 > 2.571$, so we can conclude that the drug has the effect of changing blood pressure at the 0.05 significance level.

- 用 R 指令 t.test 可計算 one-sample t test 的 p-value 和 T statistic. 以下爲計算 Example 4 的 p-value 和 T statistic 所用的 R 指令. 執行

```
x <- c(116.6, 117.6, 118.3, 117.4, 117.4, 114.3)
y <- c(120.4, 119.3, 119.9, 118.6, 121.3, 119.4)
t.test(y-x)$p.value #or t.test(y, x, paired=TRUE)$p.value
```

結果爲 0.006783352, 即p-value. 繼續執行 `t.test(y-x)$statistic` 可得

```
       t
4.437064
```

所以根據 `y-x` 作 one-sample t test 的T statistic爲4.437064.

- 執行R 指令 `t.test(y, x, paired=TRUE)`和`t.test(y-x)`得到的檢定統計量和p-value都是一樣的。預設H_1爲$\mu_1 \neq \mu_2$.

- 用 R 指令 `t.test` 也可計算 pooled t test 的 p-value 和 T statistic. 執行

```
x <- c(116.6, 117.6, 118.3, 117.4, 117.4, 114.3)
y <- c(120.4, 119.3, 119.9, 118.6, 121.3, 119.4)
t.test(x,y, paired=FALSE, var.equal=TRUE)$p.value
t.test(x,y, paired=FALSE, var.equal=TRUE)$statistic
```

結果爲 0.001882114 (p-value)和-4.181868 (T statistic). 預設H_1爲$\mu_1 \neq \mu_2$. 爲了與使用公式算出的 T statistic進行比較， 以下程式中定義了`t_test`函數，用公式計算pooled t test 的 T statistic. 函數輸入變數包括二組樣本的sample mean/standard deviation/size. 執行

```
t_test <- function(mx, my, sx, sy, nx, ny){
 ss <- (nx-1)*sx^2 + (ny-1)*sy^2
 sigma2 <- ss/(nx+ny-2)
 T <- (mx-my)/sqrt(sigma2*(1/nx+1/ny))
 return(T)
}
```

即可定義函數。接著執行

```
x <- c(116.6, 117.6, 118.3, 117.4, 117.4, 114.3)
y <- c(120.4, 119.3, 119.9, 118.6, 121.3, 119.4)
t_test(mean(x), mean(y), sd(x), sd(y), 6, 6)
```

可得pooled t test 的 T statistic爲-4.181868.

- R commands for computing sample mean/standard deviation. Suppose that `x` is a data vector in R.
  - `mean(x)` computes the sample mean of `x`.
  - `sd(x)` computes the sample standard deviation of `x`.
- 以下3頁介紹 Welch two-sample t test, 供參考用，閱讀時可略過。

- Welch two-sample t test. Suppose that $(X_1, \cdots, X_{n_1})$ and $(Y_1, \cdots, Y_{n_2})$ are two random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. Let $\bar{X}$, $\bar{Y}$, S_X, S_Y be the sample means and sample standard deviations for $(X_1, \cdots, X_{n_1})$ and $(Y_1, \cdots, Y_{n_2})$ respectively. Consider the problem of testing

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 \neq \mu_2.$$

A reasonable testing statistic is

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}}}.$$

  - Under H_0, the exact distribution of T depends on σ_1/σ_2, n_1 and n_2.
  - An approximation of the distribution of T under H_0 is a t distribution with degree of freedom estimated by

$$d = \frac{\left(\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}\right)^2}{\frac{S_X^4}{n_1^2(n_1-1)} + \frac{S_Y^4}{n_2^2(n_2-1)}}. \tag{3}$$

Approximately, we can reject H_0 at level a if $|T| > t_{a/2,d}$, and the p-value for the test is $2P(t(d) >$ observed $|T|)$. This test is called Welch two-sample t test hereafter since the degree of freedom in (3) is derived using an approximation result due to Welch in [1].

  - A χ^2 distribution with fractional degree of freedom can be defined. Since

$$t(k) \sim \frac{N(0,1)}{\sqrt{\chi^2(k)/k}},$$

    a t distribution with fractional degree of freedom can also be defined.
  - The R commands `qt(a/2,df=d)` and `1-pt(C, df=d)` can still be used to find the $t_{a/2,d}$ and the probability $P(t(d) > C)$ even when d is not an integer.
  - The R command `t.test(x,y)` can be used to carry out the Welch two-sample t test, where `x` and `y` are the two samples $(X_1, \cdots, X_{n_1})$ and $(Y_1, \cdots, Y_{n_2})$.

- Example 5. Suppose that we have two independent random samples from normal distributions and we want to test whether the population means are the same. The observed values for the two samples are

Sample 1	52	67	56	45	70	54	64	
Sample 2	59	60	61	51	56	63	57	65

Is there a significant difference in the two population means at the 0.05 level?

Ans. The p-value is 0.8528643, so there is no significant difference in the two population means at the 0.05 level. 這題 p-value 用 R 計算, 算法有以下二種.

算法一: 用 R 計算T統計量和自由度, 再用`pt` 算p-value. 執行

```
x <- c(52,67,56,45,70,54,64)
y <- c(59,60,61,51,56,63,57,65)
sx <- sd(x); sy <- sd(y); n1 <- length(x); n2 <- length(y)
S2 <- sx^2/n1 + sy^2/n2; S3 <- sx^4/(n1^2*(n1-1)) + sy^4/(n2^2*(n2-1))
T <- (mean(x)-mean(y))/sqrt(S2); d <- S2^2/S3
```

繼續執行 T;d結果爲

```
[1] -0.1912380
[1] 8.431799
```

所以T統計量 $= -0.1912380$, 自由度爲8.431799. 繼續執行

```
2*(1-pt(abs(-0.1912380), df=8.431799))
```

結果爲 0.8528643, 即p-value.

算法二: 用 t.test. 執行

```
x <- c(52,67,56,45,70,54,64)
y <- c(59,60,61,51,56,63,57,65)
t.test(x,y)$p.value
```

結果爲 0.8528643, 所以 Welch two sample t-test 的 p-value $= 0.8528643$.

- Welch's approximation. Consider a linear combination $U = aK_1 + bK_2$, where a and b are constants, and $K_1 \sim \chi^2(d_1)$ and $K_2 \sim \chi^2(d_2)$ are independent. Welch suggested in [1] to approximate the distribution of U by the distribution of $c\chi^2(k)$, where

$$k = \frac{(ad_1 + bd_2)^2}{a^2d_1 + b^2d_2}$$

and

$$c = \frac{ad_1 + bd_2}{k} = \frac{a^2d_1 + b^2d_2}{ad_1 + bd_2}.$$

The above c and k are obtained by solving

$$E(U) = E(c\chi^2(k)) = kc$$

and

$$Var(U) = Var(c\chi^2(k)) = 2kc^2.$$

- Using Welch's approximation,

$$\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2} = a\left(\frac{(n_1-1)S_X^2}{\sigma_1^2}\right) + b\left(\frac{(n_2-1)S_Y^2}{\sigma_2^2}\right) \approx c\chi^2(k),$$

where $a = \sigma_1^2/(n_1(n_1-1))$ and $b = \sigma_2^2/(n_2(n_2-1))$,

$$k = \frac{(a(n_1-1) + b(n_2-1))^2}{a^2(n_1-1) + b^2(n_2-1)} = \frac{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^2}{\frac{\sigma_1^4}{n_1^2(n_1-1)} + \frac{\sigma_2^4}{n_2^2(n_2-1)}}$$

and

$$c = \frac{a(n_1 - 1) + b(n_2 - 1)}{k} = \frac{1}{k}\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right).$$

Therefore,

$$T \approx \frac{\bar{X} - \bar{Y}}{\sqrt{c\chi^2(k)}} \sim \frac{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} N(0,1)}{\sqrt{c\chi^2(k)}} \sim \frac{N(0,1)}{\sqrt{\chi^2(k)/k}} \sim t(k),$$

where

$$k = \frac{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^2}{\frac{\sigma_1^4}{n_1^2(n_1-1)} + \frac{\sigma_2^4}{n_2^2(n_2-1)}},$$

which can be estimated by the degree of freedom given in (3).

References

[1] B. L. Welch. The generalization of 'student's' problem when several different population variances are involved. *Biometrika*, 34:28–35, 1947.