

Introduction to application of hypothesis

Concept of hypothesis:
當你想做某些操作或研究時，你想要知道操作前後或者是兩個組別是否有差別時，以假設搭配統計理論來客觀描述結果的方法。

Hypothesis

H_0 (Null hypothesis): 你想爭辯的主題，通常是你想要反駁的假設，若 H_0 為 True，那麼這個測試結果就是 Negative。
 H_1 (Alternative Hypothesis): H_0 的對立面，通常是你想得到的結論，若 H_1 為 True，那麼這個測試結果就是 Positive。。

Example:

假設你拿到一些數據，你也得到了男生和女生的平均身高 (μ_{boy} and μ_{girl})，你想要知道男生和女生在身高上是否有差。但實驗必有誤差，所以從樣本算出來的平均數很大機會是不相同的，那這個不相同，到底是

1. 真相是男女生身高平均是沒差，但是實驗誤差造成你算出來不同還是
2. 兩者真的有差？。

假說的設定：

H_0 : 假設男生和女生的身高沒差， $\mu_{boy} = \mu_{girl}$

H_1 : 假設男生和女生的身高有差， $\mu_{boy} \neq \mu_{girl}$

做完hypothesis的設定以後，藉由統計的計算，會得出一個推論，但是你沒有辦法100%的有信心你說的肯定對，畢竟可能其實誤差造成的。至於怎麼個錯法就有以下兩種錯的可能：

Type I error(α):

真相是 H_0 是對的，但從樣本推論 H_0 是錯的，這種情況叫做False Positive，其機率定為 α 。

Type II error(β):

真相是 H_0 是錯的，但從樣本推論 H_0 是對的，這種情況叫做False Negative的，其機率定為 β 。

但要怎麼描述你認為發生這些錯誤的信心是多少？

針對Type I error，要用p-value來描述

P-value:

P-value:在 H_0 為真的情況下，你從實驗樣本得到的統計量(例如t或z值)，出現的機率是多少。

針對Type II error，會用power(解釋力)去描述，但他是以Type II error的另一面去描述。

Power:

若從樣本得到的推論是拒絕 H_0 ，而這也是真相(H_0 是錯的)的機率。

$$Power = 1 - \beta$$

將真相和假說檢定的關係製成表格就會變成如下關係：

實際上的計算

由以上介紹可以發現，進行假設檢定的時候有以下參數要設定，而這些參數如何決定或計算？

設定：

-
1. Type I error(α):要看你要有多少的信心水準，根據經驗設定，通常設為0.05，

2. Type II error(β):根據經驗獲得，若你要求解釋力(power)要0.8以上，也就是你是正確拒絕 H_0 的機率要80%以上，那你 β 就要設定0.2，表示錯誤接受 H_0 的機會低於20%。

計算：

-
- p-value:按照檢定是分為單尾還是雙尾，計算會有所不同

1. 假設為單尾，且是左尾的單邊檢定，也就是 $H_1:\mu_1 < \mu_0$ 這種的檢定，那麼在計算出統計量t或是z後，將其放入機率模型的累計機率分佈所計算出的函數值，計算出來就是p-value。

$$p-value = cdf(z, 0, 1)$$

1. 假設為單尾，且是右尾的單邊檢定，也就是 $H_1:\mu_1 > \mu_0$ 這種的檢定，那麼在計算出統計量t或是z後，將其放入機率模型的累計機率分佈所計算出的函數值，再用1減去函數值，就是p-value。

$$p-value = 1 - cdf(z, 0, 1)$$

1. 假設為雙尾檢定，也就是 $H_1:\mu_1 \neq \mu_0$ 這種的檢定，那麼在計算出統計量t或是z後，看是偏左邊還是右邊分別計算。

偏左邊(t or $z < 0$):

$$p-value = 2 \times cdf(z, 0, 1)$$

偏右邊(t or $z > 0$):

$$p-value = 2 \times (1 - cdf(z, 0, 1))$$

Remark:單尾或雙尾，會有係數2的差別，而雙尾還要小心計算出來的統計量是偏左還是偏右邊。

將以上關係以圖型表示如下(假設常態分佈):

圖片來源:[stackexchange](#)

如何決定實驗需要多少的樣本？

- 1. 前面提到實驗做完後，如何進行假設檢定的操作，但是在實驗前需要多少的樣本，是另一個重點，若樣本太少，顯著性會不夠，樣本數定太大，在搜集資料又曠日費時。
- 2. 實際上，根據你要求的 α, β , power, effect size(效應值)，會得出最低要求的n，才夠顯著的體現你的要求，也就是這幾個參數其實是有關係是連結的，沒辦法隨便訂出一個n。

而effect size(效應值)又是什麼？

Effect size(效應值)d:

干預組和控制組之間的標準化後的差異量，通常是用Cohen's d來表示。

計算方法根據不同檢定有不同的計算方法:

- 1. 兩獨立樣本：

$$Cohen's\ d = \frac{\mu_2 - \mu_1}{s}$$

- 1. 單樣本或是兩配對樣本(two paired or related sample):

$$Cohen's\ d = \frac{\mu_D}{s_D}$$

其中:

μ_D = 實驗組和基準值差異的平均值，或是兩配對樣本間差異的平均值，D表示Difference。

s_D = 實驗組和基準值差異的標準差，或是兩配對樣本間差異的標準差，D表示Difference。

根據Cohen定義的量表中，對應關係如下表。

效應值(d)	意義
0 - 0.2	可忽視的
0.2 - 0.5	小
0.5 - 0.8	中
0.80 +	大

而effect size和sample size, α 又是彼此相關的，若d很小的情況下要求顯著的差異，那就會要求比較大的sample size

定義required sample size(n)

在決定好 α, β (or power), effect size後，就可以根據統計模型計算出sample size了。

以常態分佈比較兩個群體平均值是否相同為例：

假設你知道兩個獨立群體的標準差(σ_1, σ_2)，平均數(μ_1, μ_2),且定好了 α, β ，那麼n所求就是：

$$n = \frac{(\sigma_1^2 + \sigma_2^2)(Z_{1-\alpha/2} + Z_{1-\beta})^2}{(\mu_1 - \mu_2)^2}$$

Remark:

- 1. 不同的假設檢定，會有不同的sample size計算方法, 例如two proportion z-test(比例檢定)。
- 2. 做power, effect size, type I error, required sample size的分析就是 Power analysis。
- 3. 可以使用G*Power軟體做power analysis。

參考資料：

- 1. [科學Online](#)
- 2. [iT邦幫忙](#)

In [1]:

```
# Example-two proportion
# Given alpha=0.05,power=0.8(beta=0.2),ratio=1,p1=0.15,p2=0.12
# Find the required number of sample size

from statsmodels.stats import api
from math import ceil
p1=0.15
p2=0.12
effect_size=api.proportion_effectsize(p1,p2)
required_n=api.NormalIndPower().solve_power(effect_size,alpha=0.05,power=0.8,ratio=1)
required_n=ceil(required_n)
print("Rquired number of sample size is:",required_n)
```

Rquired number of sample size is: 2031

In [2]:

```
# Example-t-test
# Given alpha=0.05,power=0.8(beta=0.2),ratio=1,effect_size=3
# Find the required number of sample size

from statsmodels.stats import power
effect_size=3 #d(meand)/S.D
required_n=power.tt_ind_solve_power(effect_size=effect_size,alpha=0.05,power=0.8,ratio=1)
required_n=ceil(required_n)
print("Rquired number of sample size is:",required_n)
```

Rquired number of sample size is: 4

In [4]:

```
# Example-find effect size
# Given alpha=0.05, power=0.8, nobs1=100
# Fin minimum effect_size
effect_size=power.tt_ind_solve_power(alpha=0.05,power=0.8,nobs1=100)
print(effect_size)
```

0.39813911739112756