

人生好難

Normal distribution (Gaussian distribution)

常態分布 (高斯分布)

- 將一連續變項之觀察值發生機率以圖呈現其分布情形
- 中線會是平均值
- 範圍：負無窮大 ~ 無窮大

References

homepage.ntu.edu.tw/~clhsieh/biostatistic/4/4-1.htm

P-value

p值其實就是在分佈常態分佈下 $\geq t$ 值的機率密度值(probability density value , $P(x \geq t)$)

T-value

越接近 0 代表差異越沒有差異，相反的，越遠離差異就越大

$$t = \frac{(X_1 - X_2)}{\sqrt{\frac{(S_1)^2}{n_1} + \frac{(S_2)^2}{n_2}}}$$

(平均數1 - 平均數2) / 兩組資料的標準差

How to calculate P-value

從 T-value 積分到無窮大之面積，就為 P-value

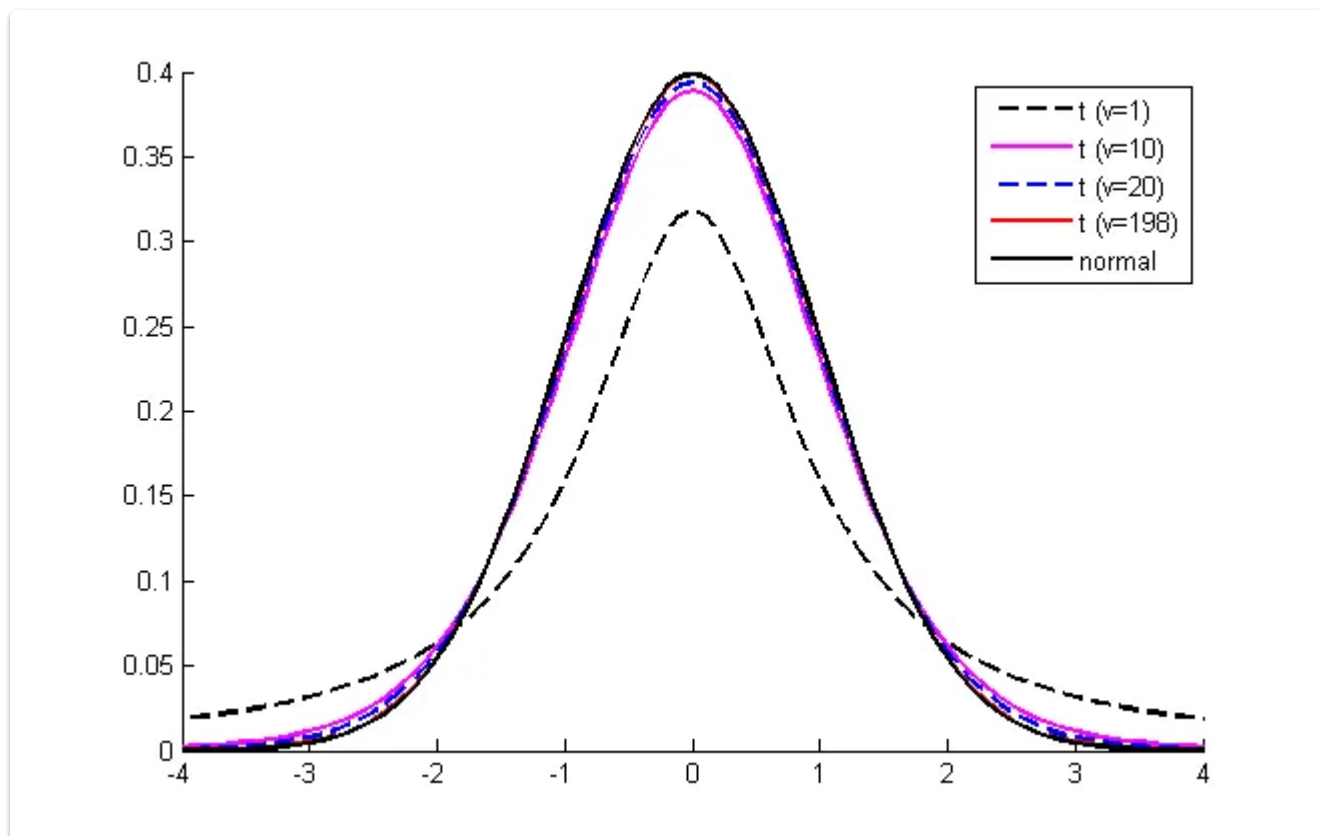
有兩邊無窮大，所以正常來說兩邊會加起來，也可以面積 * 2

T分布 (Student's t-distribution)

抽樣下的常態分佈

自由度越高，會越接近該樣本的常態分佈

$v \rightarrow$ 自由度 ($N1 + N2 - 2$)



Null hypothesis

- 又稱**虛無假說**，通常被標寫為 H_0
- 假設兩筆資料集相似、差不多

P-value < 0.05

意義就在於要當下 T-value 到無窮大的面積大於 0.05 的自由度
而就是說我們**接受虛無假說**，又可以說兩筆資料是有差異的

↔

意義就在於要當下 T-value 到無窮大的面積小於 0.05 的自由度
而就是說我們**拒絕虛無假說**，有可以說兩筆資料差異不大

References

[統計學:大家都喜歡問的系列-p值是什麼. 常常看到有人說p-value... | by Tommy Huang | Medium](#)

你真的懂甲基化嗎？

甲基化

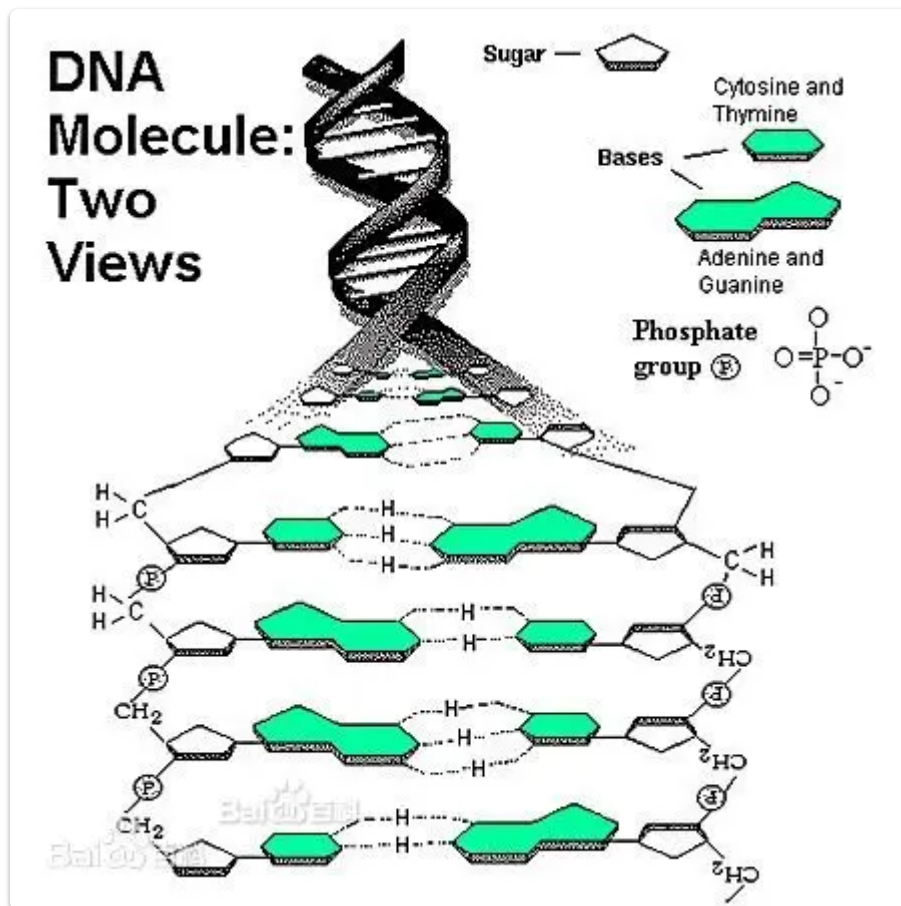
簡單說：DNA 分子上多了一個甲基

而這個甲基只會結合在一個特定位置 → 胞嘧啶 (Cytosine)

CpG

胞嘧啶 (C · Cytosine)、磷酸 (p · phosphoric acid)、鳥嘌呤 (G · Guanine)

其實就是配對的核苷酸



CpG位點 (CpG site)：就是每一個這樣的核苷酸

CpG島 (CpG island)：包含有很多個CpG位點的DNA片段。

CpG

人類 CpG 特徵：

- 有 25000 個 CpG 島
- 50% 在啟動子區域，25% 在基因內，剩下....我也不知道
- 人體內約 60~70% 基因的啟動子中都具有 CpG 島

啟動子

啟動子未必在起始密碼子的上游

所以有些甲基化位點會在基因內部

甲基化如何影響基因表達

1. 基因啟動子區域存在 CpG 島
2. CpG 島上的 CpG 位點通常會有甲基
→
3. CpG 島被甲基化時會抑制基因的表達
4. CpG 島上有很多 CpG 位點，CpG 島上的多少 CpG 位點被甲基化，才算這個 CpG 島被甲基化

轉錄相當於表達？！