Lecture 8 (II): Hypothesis Testing 2

Hypothesis Testing
假設檢定

Example:

丟一個銅板, X (隨机变数) $\begin{cases} X=1, & \text{if 正面} \\ X=0, & \text{if 反面} \end{cases}$

$\Rightarrow X \sim \text{Bernoulli}(p)$ $\begin{cases} Pr(X=1)=p \\ Pr(X=0)=1-p \end{cases}$

$EX = p$, $Var\, X = p(1-p)$

- 实验 100 次

結果: $\underbrace{X_1, X_2, \cdots, X_{100}}$

55 個正面 i.e. 55 個 $X_i = 1$
45 個反面 45 個 $X_i = 0$

$\Rightarrow \hat{p} = \frac{55}{100} = 0.55$ (point estimate of p)

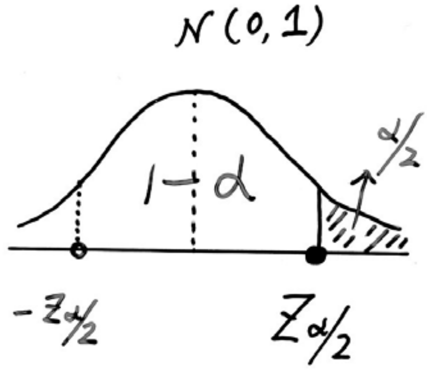
p 之 95% confidence interval (C.I.) =

2

$$\hat{p} \pm \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \cdot Z_{\alpha/2} \quad (\alpha=0.05)$$

$$= 0.55 \pm \sqrt{\frac{0.55 \times 0.45}{100}} \cdot 1.96$$

$$= (0.452,\ 0.648)$$



• 95% C.I. 之解釋：

剛剛的那個实验得到的 $\hat{p}$，即 $\bar{X}_{①}$，$\left(\bar{X} \pm Z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right)_{①}$

在同樣狀況下，实验無窮多次，得 $\bar{X}_{②}$，$(\quad)_{②}$

(frequentist)

$\vdots$

$\bar{X}_{\infty}$，$(\quad)_{\infty}$

$\alpha=0.05$

平均每 100 个 $(\quad)_{i}$，

会有 95 个 cover 住真正的 p

前面简略複習 estimation $<$ point estimation / interval "

以下谈 testing .

Null hypothesis $\mathcal{H}_0: \ p = \frac{1}{2}$

Alternative hypothesis $\mathcal{H}_a: \ p \neq \frac{1}{2}$

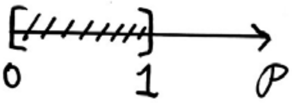
- 先简化问题，以丢銅板為例：

假設只丢一次，並且 p 只有两种可能

$\mathbb{P} = \{ 0.1 , 0.9 \}$ parameter space 参数空间

結果 $X=1$? $\to$ $H_0: \ p = 0.9$ / $H_a: \ p \neq 0.9 \ (p = 0.1)$

- 实际情況裡，$\mathbb{P} = [0,1]$



丢 100 次，得 $X_1, X_2, \cdots X_{100}$ （$n=100$，算不算 $\uparrow\infty$?）

由 CLT，$\bar{X} \sim N\left(p, \frac{p(1-p)}{n}\right)$ (if $n \uparrow \infty$)

或 $\dfrac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$ (if $n \uparrow \infty$)

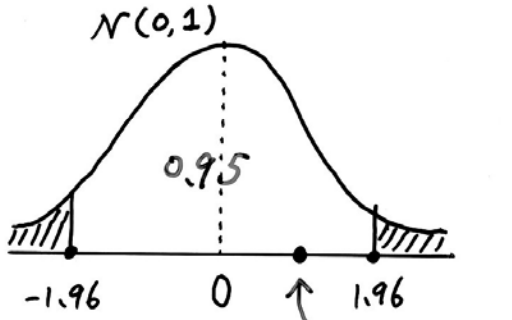
Under H_0 (i.e. if $p = 0.5$ is true)

$$\Rightarrow \quad \frac{\bar{X} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{100}}} \sim N(0,1)$$

如果 $X_1, \cdots, X_{100}$ 是一組 random sample

(大假設)

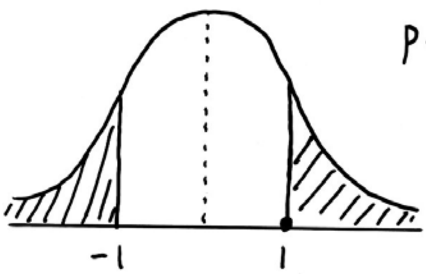
$$\bar{X} \text{ 应接近 } p \begin{cases} = 0.5 \\ \neq 0.5 \end{cases}$$



two-sided test

if $\bar{X} = 0.55$ (承前例) $\Rightarrow$ under H_0

$$\frac{\bar{X} - 0.5}{\sqrt{\frac{0.5 \cdot 0.5}{100}}} = \frac{0.05}{0.05} = 1$$



p-value

> 0.05

Example： 設 CMU 男生的平均身高 $\sim N(\mu, \sigma^2)$

随机抽取 10 人，得如下之"樣本"：

170	184	179	157	175
164	162	170	180	182

Q1: $\mu = ?$

$\sigma = ?$

Q2: $H_0: \mu \leq 165$

$H_a: \mu > 165$

Q1:

$\Rightarrow \hat{\mu} = 172.3 \quad \bar{X}$

$\hat{\sigma} = 9.20 \quad \Sigma(X_i - \bar{X})^2 / n-1 = \hat{\sigma}^2 \equiv S^2$

又，$\because X_1, \cdots X_n \sim N(\mu, \sigma^2)$

我们有 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$, for any n

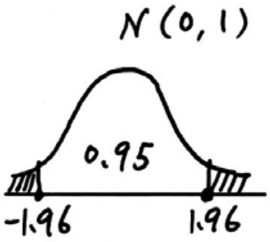
$\Rightarrow \mu$ 之 95% C.I. =

(i) $\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$, if σ is known

(ii) $\bar{X} \pm t_{\alpha/2, n-1} \frac{\hat{\sigma}}{\sqrt{n}}$, 〃 〃 unknown

!

(i) $\dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$



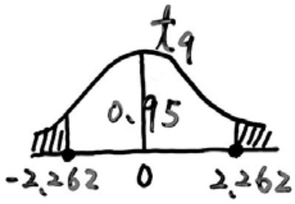
$\Rightarrow \Pr\left\{-1.96 < \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} < +1.96\right\} = 0.95$

or $\Pr\left\{-\bar{x}-1.96\frac{\sigma}{\sqrt{n}} < -\mu < -\bar{x}+1.96\frac{\sigma}{\sqrt{n}}\right\} = 0.95$

or $\Pr\left\{\bar{x}-1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x}+1.96\frac{\sigma}{\sqrt{n}}\right\} = 0.95$

Gosset ; Pearson

(ii) $\dfrac{\bar{x}-\mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \sim t_{n-1}$ (Student's t- distribution with degree of freedom $n-1$)



$\Rightarrow \Pr\left\{-2.262 < \dfrac{\bar{x}-\mu}{\frac{\hat{\sigma}}{\sqrt{n}}} < 2.262\right\} = 0.95$

$\Rightarrow \cdots\cdots \Rightarrow \Pr\left\{\bar{X}-2.262\frac{\hat{\sigma}}{\sqrt{n}} < \mu < \bar{X}+2.262\frac{\hat{\sigma}}{\sqrt{n}}\right\} = 0.95$

此题，μ 之 95% C.I. 為：

(165.72 , 178.88)

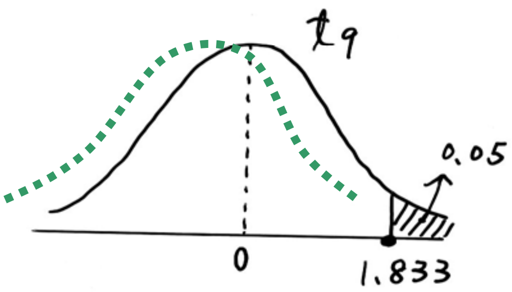
Q2 :

$\because \quad \dfrac{\bar{x}-\mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \sim t_9$

one-sided test !

if $\mu = 165 \Rightarrow$

$\dfrac{\bar{x}-165}{\frac{\hat{\sigma}}{\sqrt{n}}} \sim t_9$



In this case, the decision you have made will result in a smaller type I error than α (=0.05, say).

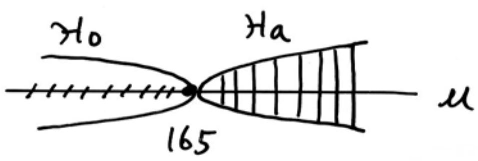
$\dfrac{172.3-165}{\frac{9.20}{\sqrt{10}}} = 2.51 > 1.833$

$\therefore$ reject H_0

$0.01 < \text{p-value} < 0.025$

$\boxed{\text{if } \mu < 165}$, the p-value gets smaller, and the result (difference between $\bar{x}$ & μ) gets more significant !

- In terms of μ,



d.f.	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$
1	3.078	6.3138	12.706	31.821
2	1.886	2.9200	4.3027	6.965
3	1.638	2.3534	3.1825	4.541
4	1.533	2.1318	2.7764	3.747
5	1.476	2.0150	2.5706	3.365
6	1.440	1.9432	2.4469	3.143
7	1.415	1.8946	2.3646	2.998
8	1.397	1.8595	2.3060	2.896
9	1.383	1.8331	2.2622	2.821
10	1.372	1.8125	2.2281	2.764
11	1.363	1.7959	2.2010	2.718
12	1.356	1.7823	2.1788	2.681
13	1.350	1.7709	2.1604	2.650
14	1.345	1.7613	2.1448	2.624

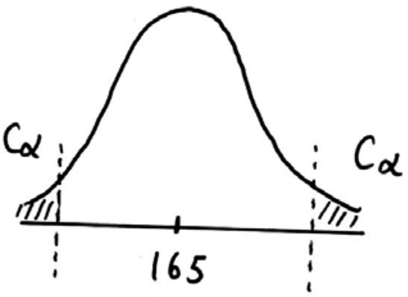
- Type I error and Type II error

decision	Reject H_0	Do not Reject H_0
H_0 IS true	Type I Error (**α**)	
H_0 is NOT true		Type II Error (**β**)

(i) $X_1, \cdots X_n \sim N(\mu, \sigma^2)$

$H_0: \mu = 165$

$H_a: \mu \neq 165$



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

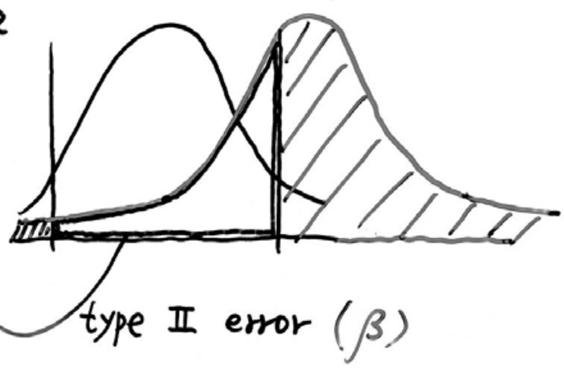
or $\dfrac{\bar{X}-\mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \sim t_{n-1}$, $\bar{X} \sim \mu + \dfrac{\hat{\sigma}}{\sqrt{n}} t_{n-1}$

C_α : critical region (拒斥域)

type I error $= \alpha$

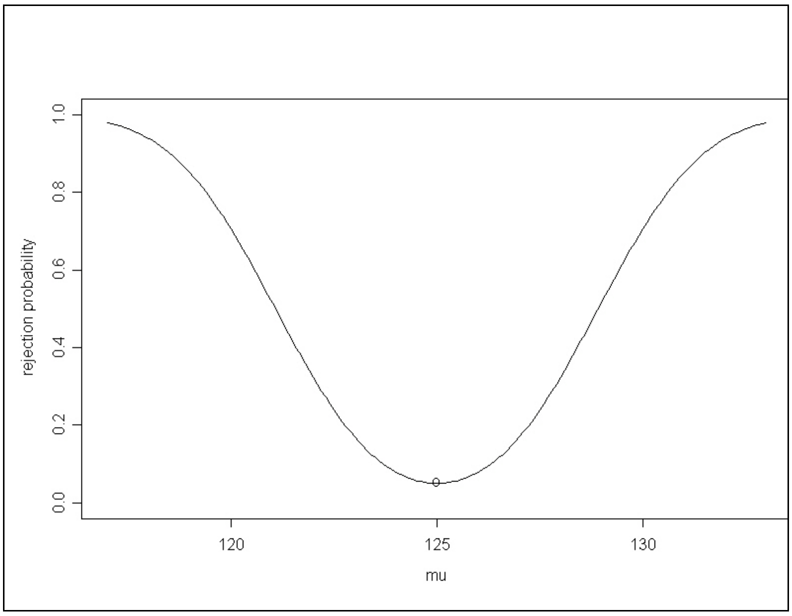
Now, if $\mu = 170$, for example.

the probability that (under $\mu = 170$) $\bar{X}$ will fall into ⊛ and you do not reject H_0. (power $= 1-\beta$)

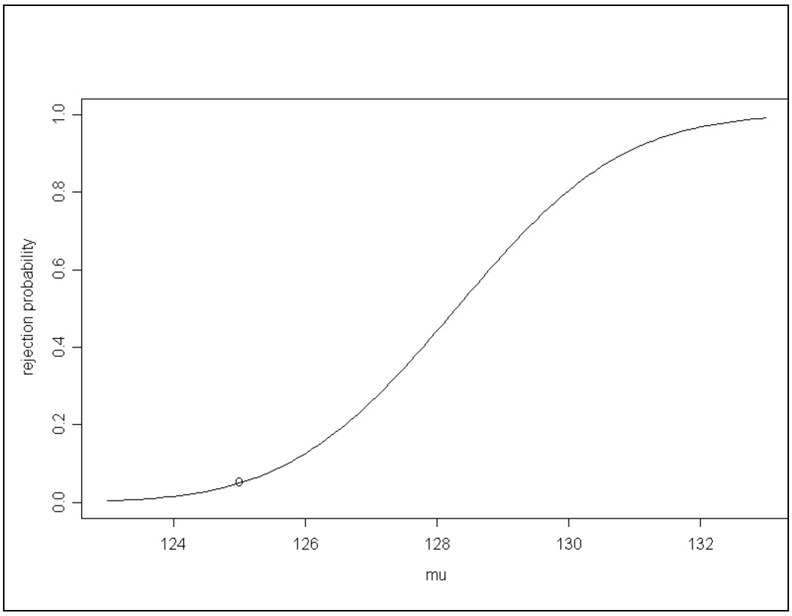


Power curve

Two-sided



One-sided



關於 Power (of test) 的一些說明
～以糖尿病的例子來說明～

在糖尿病的例子裡，我们假設了 $X_1, X_2, \cdots, X_{100} \sim N(\mu, \sigma^2)$ [$X_1, \cdots, X_{100}$ 表 100 個糖尿病人的血糖值]，並設 $\sigma^2 = 400$ ($\sigma = 20$) 為已知。故有：

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \text{，或即}$$

$$\bar{X} \sim N(\mu, 2^2) \text{，} \because n = 100 \text{，} \sigma^2 = 400$$

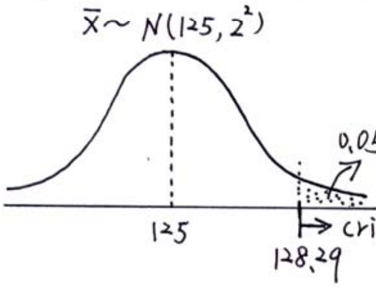
A. One-Sided Test

對於以下的 hypotheses：

$$H_0: \mu \le 125 \quad \text{v.s.} \quad H_a: \mu > 125 \text{，}$$

如何定下你的 test 呢？

(A.1) 先考慮 H_0 所对应的参数空间（指「参數的可能值所構成的集合」）与 H_a 所对应的参数空间"交会"的那一点：$\mu = 125$ ($\mu \in H_0$)



若定 $\alpha = 0.05$，可以計算出當 $\bar{X} \ge 128.29$ 时，則 reject H_0 (why?)

所以你定的 test，$\tau(X_1, X_2, \cdots, X_n)$，如下：

$$\begin{cases} \tau = 1 \text{, if } \bar{X} \ge 128.29 \\ \tau = 0 \text{, otherwise} \end{cases} \Leftrightarrow \begin{cases} \tau = 1 \Leftrightarrow \text{reject } H_0 \\ \tau = 0 \Leftrightarrow \text{do not reject } H_0 \end{cases}$$

$\tau(x_1,\cdots,x_n)$ 是一个随机变数 (random variable)，这樣子定是為了表達的方便！据此，則：

$$\alpha = Pr\{\tau=1 \mid H_0\} = E\{\tau \mid H_0\}$$

$$\begin{aligned} power &= 1-\beta \\ &= Pr\{\tau=1 \mid H_a\} = E\{\tau \mid H_a\} \end{aligned}$$

$$\beta = 1 - power = Pr\{\tau=0 \mid H_a\}$$

(A.2) 考慮 $\mu \in H_0$ 中之所有点（在这裡 H_0 表 $(0, 125]$）。

我们要計算 $Pr\{\tau=1 \mid H_0\}$，$\forall\, \mu \in H_0$.

比如：

$\mu = 125$

$\mu = 124.9$

$\vdots$

$\mu = 123$

$\vdots$

时，我们要計算：

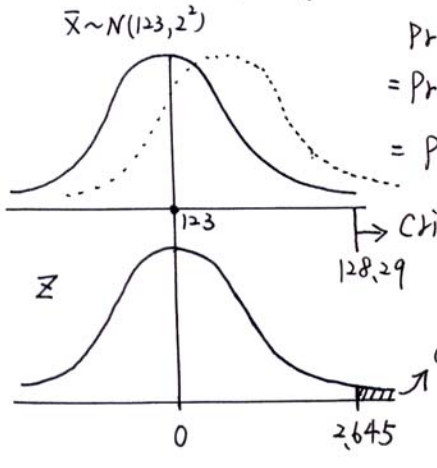
$$Pr\{\tau=1 \mid \mu\} = ?$$

（即所有可能的 μ，$\mu \le 125$）

以 $\mu = 123$ 為例，則 $\bar{x}$ 会落在 $C_\alpha \equiv [128.29, \infty)$ 间之机率為：

$$\begin{aligned} & Pr\{\bar{x} \ge 128.29 \mid \mu=123\} \\ &= Pr\left\{\frac{\bar{x}-123}{2} \ge \frac{128.29-123}{2}\right\} \\ &= Pr\{z \ge 2.645\} \approx 0.0041 \end{aligned}$$

（参見表一之"H_0"）



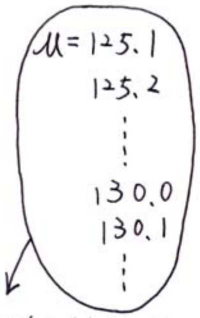
critical region（前面定出之 critical region，C_α）

[其餘的情況，亦請参見表一"H_0".]

(A.3) 考慮 $\mu \in \mathcal{H}_a$ 中之所有点（此时，$\mathcal{H}_a$ 表 $(125, \infty)$），

我们要計算 $Pr\{\tau = 1 \mid \mathcal{H}_a\}$，$\forall\, \mu \in \mathcal{H}_a$.

比如：

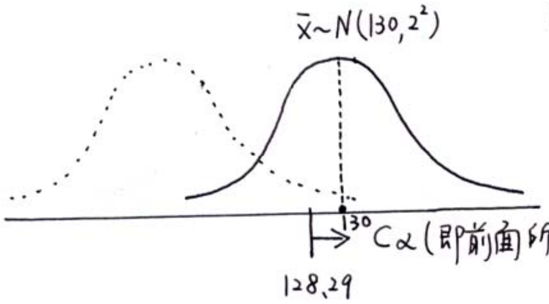


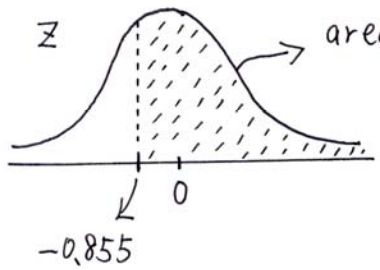
(即所有可能的 μ，而 $\mu > 125$)

時，我们要計算：

$$Pr\{\tau = 1 \mid \mu\} = ?$$

以 $\mu = 130$ 為例，則 $\bar{x}$ 会落在 C_α，$(128.29, \infty)$，间之机率為何？（見下面(*)）





(*)：
$$\begin{aligned} Pr\{\tau = 1 \mid \mu = 130\} &= Pr\{\bar{x} > 128.29 \mid \mu = 130\} \\ &= Pr\left\{\frac{\bar{x} - 130}{2} > \frac{128.29 - 130}{2}\right\} \\ &= Pr\{Z > -0.855\} \approx 0.8037 \end{aligned}$$

［其餘的情况，亦請参見"表一：$\mathcal{H}_a$"的結果.］

(A.4) power curve：图一！ (power function)

B. Two-Sided Test

對於如右之 Hypotheses： $\begin{cases} H_0: \mu=125 & \text{v.s.} \\ H_a: \mu\neq 125 & , \end{cases}$

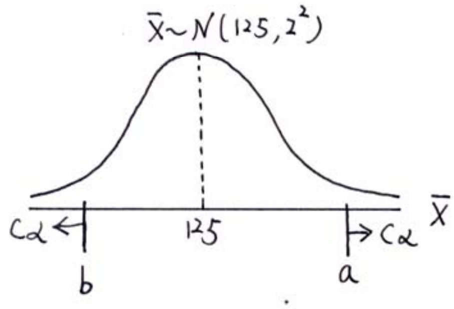
這时，对应於"H_0"的参数空间只有一个点，即 $\mu=125$，
而，对应於"H_a" " " " " 為 $\underbrace{(0,125)\cup(125,\infty)}$

亦即"$\mu=125$"以外之所有 positive value.

在这个例子下，如何定下你的 test 呢？

(B.1) 在 $H_0: \mu=125$ 之下，$\bar{X}\sim N(125, 2^2)$，

若定 $\alpha=0.05$，則 C_α (critical region) 以下图表示：



只要 $\bar{X}$ 落在图中之 C_α 区域，則推论上，即不利於"$\mu=125$"，而有利於"$\mu\neq 125$"！

亦即 $\begin{cases} \tau=1, & \text{if } \bar{X}>a \text{ or } \bar{X}<b \\ \tau=0, & \text{otherwise.} \end{cases}$

a, b 兩点之計算：

$$125 \pm 1.96\frac{\sigma}{\sqrt{n}}$$

$$= 125 \pm 1.96\times 2 = 125 \pm 3.92$$

$$\Rightarrow \quad a = 128.92, \quad b = 121.08$$

以下 (B.2) & (B.3) 考慮 $H_a: \mu\neq 125$ 之情形；(B.2) 考慮 $\mu<125$，而 (B.3) 考慮 $\mu>125$.

(B.2) $\mu < 125$ 时，$Pr\{\tau = 1 \mid \mathcal{H}_a\} = ?$

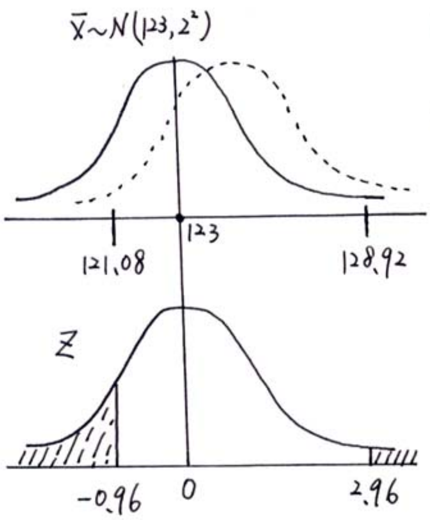
满足 "$\mu \in \mathcal{H}_a$"

當 $\mu = 124.9$, 124.8, $\vdots$, 123.0, $\vdots$ 时，我们要根据 (B.1) 中所定下之 Test ($\tau(x_1, \cdots, x_n)$) 来计算

$$Pr\{\tau = 1 \mid \mu\} = ?$$

(即所有可能的 μ，而 $\mu < 125$)

以 $\mu = 123$ 为例，此时 $\bar{X} \sim N(123, 2^2)$ $\Rightarrow$



$$Pr\{\tau = 1 \mid \mu = 123\}$$

$$= Pr\{\bar{X} < 121.08 \text{ or } \bar{X} > 128.92 \mid \mu = 123\}$$

$$= Pr\{\bar{X} < 121.08 \mid \mu = 123\} + Pr\{\bar{X} > 128.92 \mid \mu = 123\}$$

$$= Pr\left\{\frac{\bar{X} - 123}{2} < \frac{121.08 - 123}{2}\right\} + Pr\left\{\frac{\bar{X} - 123}{2} > \frac{128.92 - 123}{2}\right\}$$

$$= Pr\{Z < -0.96\} + Pr\{Z > 2.96\}$$

$$= 0.1685 + 0.0015 = 0.1700$$

(对照"表二"！)

(B.3) $\mu > 125$时，$Pr\{\tau=1 \mid \mathcal{H}_a\} = ?$

满足 $\mu \in$ "$\mathcal{H}_a$"

3大应当

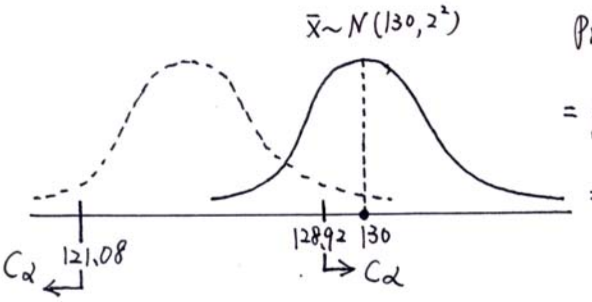
$\mu = 125.1$
125.2
⋮
130.0
130.1
⋮

时，亦根据(B.1)中所定之 Test (τ)来計算

$$Pr\{\tau=1 \mid \mu\} = ?$$

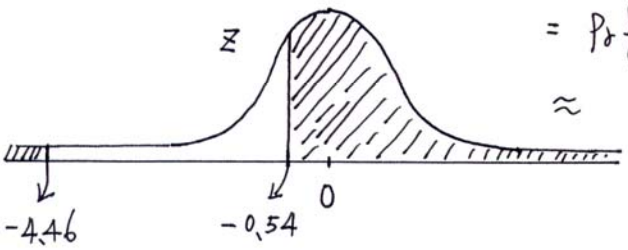
(即所有可能之μ，而$\mu > 125$)

以 $\mu = 130$ 为例，此时 $\bar{X} \sim N(130, 2^2) \Rightarrow$



$$Pr\{\tau=1 \mid \mu=130\}$$
$$= Pr\{\bar{X} < 121.08 \text{ or } \bar{X} > 128.92 \mid \mu=130\}$$
$$= Pr\{\bar{X} < 121.08 \mid \mu=130\} + Pr\{\bar{X} > 128.92 \mid \mu=130\}$$
$$= Pr\{Z < -4.46\} + Pr\{Z > -0.54\}$$
$$\approx 0 + 0.7054 = 0.7054$$

(对照表二)



(B.4) Power curve (or power function)：参见图二！

Homework 8.1

Tossing a coin with distribution: X=1 (if head, probability=p) and X=0 (if tail, probability=1-p). Let $X_1 \ldots X_{100}$ be 100 realizations. Please (1) give the point estimate; (2) derive the formula of 95% confidence interval of p and justify your derivation. (3) Explain: What is the distribution of $\sum_{i=1}^{100} X_i$?

Homework 8.2

Explain the meaning of 95% (or 100(1-α)%) confidence interval.

Homework 8.3

Let $X_1 \ldots X_{10}$ and $Y_1 \ldots Y_5$ be distributed as $N(\mu_1, 4)$ and $N(\mu_2, 4)$ respectively. If you use the following Z-statistic to test the hypotheses, H_0: $\mu_1<\mu_2$ versus H_a: $\mu_1>\mu_2$,

$$Z=\frac{\bar{X}-\bar{Y}}{\sigma\sqrt{1/5+1/10}} \quad \text{with} \quad \sigma=2.$$

Please give a plot (better by a computer package like Excel or others) of the 'power curve' over a reasonable range of μ_1-μ_2.