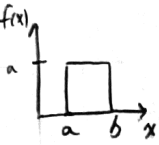
# PD連續分佈

① 均勻分佈 (Uniform Dis)

x在區間[a,b]內發生機率皆相同

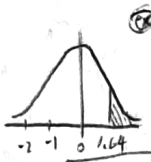



ex 公車 20min 一班，等5到10分上公車？

$P(5 \le Y \le 10) = \frac{10}{20} - \frac{5}{20} = \frac{1}{4}$

② 常態分佈 (Normal Dis) (Z-table)

將X轉化為Z再查表，即可得出所求的機率(面積)

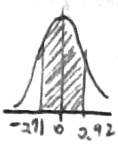



ex $P(Z \ge 1.64)$

$= 1 - P(1.64)$

$= 0.05$

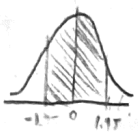
$$Z = \frac{x-\mu}{\sigma}$$




ex $P(-2.91 \le Z \le 2.92)$

$= pnorm(2.92) - pnorm(-2.91)$

$= 0.58$




ex 大學附考 $\mu = 500$, $\sigma = 100$, 分於325-675間

$P(325 < x < 675)$

$= P\left(\frac{325-500}{100} < Z < \frac{675-500}{100}\right)$

$= P(-1.75 < Z < 1.75)$

③ 抽樣分佈 (Sampling Dis)

[a] 樣本平均數 (Sampling mean)

ex 抽75個高中生，$\mu = 100$, $\sigma = 15$ 求大於165，

$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

[b] 樣本母體數 (Sampling Proportion)

ex 抽100人，若女性比例為0.6，則小於0.56機率？

$$Z = \frac{\bar{p}-p}{\sqrt{p(1-p)/n}}$$

0.56 0.6 0.6 ∞

## Z-test (大樣本 $n \le 30$)

母體 $\sigma$ 已知

## T-test (小樣本 $n > 30$)

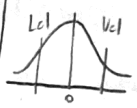
母體 $\sigma$ 未知、$\mu$ 未知

以樣本 $s$ 代替 $\sigma$，自由度 $n-1$

## HTE 區間估計

$Z(\frac{0.05}{2}) = qnorm(\frac{0.05}{2})$

① Z 信賴區間 (Confidence Interval)




ex 抽900個水果，$\mu = 60$, $\sigma = 6$ 則95%信賴區間？ (0.05)

$$CI = \bar{x} \pm Z\left(\frac{\alpha}{2}\right) \times \frac{\sigma}{\sqrt{n}}$$ abs

60 60 900

$Z(\frac{0.01}{2}) = qnorm(\frac{0.01}{2})$

② 樣本估計 (Sample Size)

ex 抽取n個樣本，討論欲信賴99%誤差(within)為2以內，$\sigma = 9$

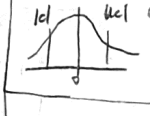
$$n = \left(\frac{Z(\frac{\alpha}{2}) \times \sigma}{B}\right)^2$$

9 2

③ T區間估計 (T-test CI)

$t(\frac{0.05}{2}) = qt(\frac{0.05}{2}, df = n-1)$

224

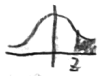



ex 抽225人，$\mu = 150$, $\sigma = 30$ 則95%信賴區間 (0.05)

$$CI = \bar{x} \pm t\left(\frac{\alpha}{2}\right) \times \frac{s}{\sqrt{n}}$$ abs

150

## HS 假設檢定



| | 左尾 (left-tail) | 右尾 Right-tail | 雙尾 Two-tail |
|---|---|---|---|
| ① 標準設定 | $H_0: \mu \ge 3.5$<br>$H_1: \mu < 3.5$<br>(5×sig, n=25, σ=0.6, 0.05α, $\bar{x}=2$, $\mu=3.5$) | $H_0: \mu \le 50$<br>$H_1: \mu > 50$<br>(5×sig, n=36, σ=10, 0.05α, $\bar{x}=60$, $\mu=50$) | $H_0: \mu = 250$<br>$H_1: \mu \ne 250$<br>(5×sig, n=16, σ=5, 0.05α, $\bar{x}=246$, $\mu=250$) |
| $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ | $Z = -12.5$ | $Z = 6$ | $Z = -3.2$ |
| ② Z檢定<br>雙尾除2 | 若 $Z < -Z_{0.05}$<br>$= Z < qnorm(\alpha)$<br>Reject $H_0$ | 若 $Z > Z_{0.05}$<br>$= Z > qnorm(\alpha)$<br>Reject $H_0$ | 若 $Z \begin{cases} < -Z(\frac{0.05}{2}) \\ > Z(\frac{0.05}{2}) \end{cases}$<br>$= Z \begin{cases} < qnorm(\frac{\alpha}{2}) \\ > qnorm(\frac{\alpha}{2}) \end{cases}$<br>Reject $H_0$ |
| ③ P值法<br>雙尾×2 | 若 $P < \alpha$<br>= pnorm(z, lower.tail = TRUE)<br>Reject $H_0$ | 若 $P < \alpha$<br>= pnorm(z, lower.tail = FALSE)<br>Reject $H_0$ | 若 $2P < \alpha$<br>$= 2 \times pnorm(z) < \alpha$<br>Reject $H_0$ |
| $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ | $t = -12.5$ | $t = 6$ | $t = -3.2$ |
| ④ P值法 | 若 $P < \alpha$<br>= pt(t, df = n-1)<br>Reject $H_0$ | 若 $P < \alpha$<br>= pt(t, df = n-1, lower.tail = FALSE)<br>Reject $H_0$ | 若 $2P < \alpha$<br>= 2pt(t, df = n-1)<br>Reject $H_0$ |
| ⑤ Import<br>mydata = data.frame(Xrls) | t.test(mydata$Time, alternative = "less", mu = 3.5) (mean)<br>claim at least 75% soti<br>survey 200, 60% soti<br>$H_0: p \ge 0.75$<br>$H_1: p < 0.75$<br>$Z = -4.89$ | "greater"<br>exceeds 超過<br>claim at most 7% soti<br>surveyed 150, 95% soti<br>$H_0: p \le 0.1$<br>$H_1: p > 0.1$<br>$Z = 1.33$ | "two.sided"<br>claim 8% satisfied<br>surveyed 100, 93% satis<br>$H_0: p = 0.8$<br>$H_1: p \ne 0.8$<br>$Z = -1.75$ |
| $Z = \frac{\bar{p}-p}{\sqrt{p(1-p)/n}}$ | prop.test(95, 200, p=0.9, alternative = "less", conf.level = 0.95, correct = FALSE) | 若 $P < \alpha$<br>= pnorm(z, lower.tail = TRUE)<br>Reject $H_0$<br>"greater" | 若 $2P < \alpha$<br>= 2pnorm(z, lower.tail = TRUE)<br>Reject $H_0$<br>"two.sided" |
| Import<br>str(mydata) 找統數<br>table(mydata) 找數量 | (越低越?) | | |

| | 左尾 (left-tail) | 右尾 (Right-tail) | 雙尾 (Two-tail) |
|---|---|---|---|
| | Variance decreased 250<br>Select 10, 近=14.4 | Variance lager 0.05<br>Select 25, Sd=0.06<br>(Sample Variance) | Sample 20, Sd=12.5<br>Is diff from 100? |
| 卡方<br>$chi = \frac{(n-1)^2 svar}{pvar}$ | $H_0: \sigma^2 \ge 250$<br>$H_1: \sigma^2 < 250$<br>(Svar=(14.4)^2, n=10, Pvar=250, α=0.9)<br>Chi = 7.5<br>若 $P < \alpha$<br>= pchisq(chi, df=n-1)<br>Reject $H_0$ | $H_0: \sigma^2 \le 0.05$<br>$H_1: \sigma^2 > 0.05$<br>(Svar=0.06, n=25, Pvar=0.05, α=0.05)<br>Chi = 28.8<br>若 $P < \alpha$<br>= pchisq(chi, df=n-1, lower.tail = FALSE)<br>Reject $H_0$ | $H_0: \sigma^2 = 100$<br>$H_1: \sigma^2 \ne 100$<br>(Svar=(12.5)^2, n=20, Pvar=100, α=0.05)<br>chi = 29.68<br>若 $2P < \alpha$<br>= 2 × pchisq(chi, df=n-1, lower.tail = FALSE)<br>Reject $H_0$ |
| Import | VarTest(mydata$x, alternative = "less", conf.level = 0.9, sigma.squared = 250, data.name = NULL) | "greater" | VarTest(mydata$x, alternative = "two.sided", conf.level = 0.95, sigma.squared = 100, data.name = NULL) |

Poisson – dpois(x, lambda, log = FALSE) avg of 3

Binomial – dbinom(x, size, prob, log = FAL) 95% die, 6 random 4

Hypergeo – dhyper(x, m, n, k, log = FALSE) select 5 without replace

Uniform – punif(q, min=0, max=1, lower.tail=T) find 25 to 30 min land 30.