บทที่ 2
การแจกแจงของตัวอย่าง และการประมาณค่า
(Sampling Distribution and Estimation)

Probability Density Function (p.d.f.) คือ รูปการแจกแจงของความน่าจะเป็นของตัวแปร**เชิงสุ่มแบบต่อเนื่อง ซึ่งมีคุณสมบัติดังนี้**

1. **พื้นที่ภายใต้ density curve = 1**
2. **$Pr[a \leq X \leq b]$ = พื้นที่ภายใต้ curve ซึ่งอยู่ระหว่าง a และ b**
3. $f(x) \geq 0$

สำหรับตัวแปรเชิงสุ่มแบบต่อเนื่อง $Pr[X = x] = 0$ **ค่าความน่าจะเป็นจะมีค่าเมื่อกล่าวถึงความน่าจะเป็นที่ตัวแปรเชิงสุ่ม X จะตกอยู่ในช่วงใดช่วงหนึ่งเท่านั้น และ** $Pr[a \leq X \leq b] = Pr[a < X \leq b] = Pr[a \leq X < b] = Pr[a < X < b]$

2.1 การแจกแจงแบบปกติ (Normal distribution)

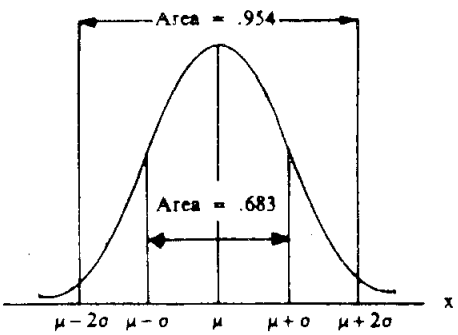
การแจกแจงแบบปกติมีรูปการแจกแจงเป็นรูประฆัง ถ้า X มีการแจกแจงแบบปกติ**ที่มี mean = μ และ variance = σ^2 เราเขียนแทนด้วย**

$X \sim N(\mu, \sigma^2)$ **และ**

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

$E(X) = \mu$

$Var(X) = \sigma^2$



Normal distribution

จากรูป $\Pr[\mu-\sigma < X < \mu+\sigma] = .683$

$\Pr[\mu-2\sigma < X < \mu+2\sigma] = .954$

นอกจากนั้น $\Pr[\mu-3\sigma < X < \mu+3\sigma] = .997$

ถ้า $X \sim N(\mu, \sigma^2)$ แล้ว $Z = \dfrac{X-\mu}{\sigma} \sim N(0, 1)$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$

1) $\Pr[X < b] = \Pr\left[\dfrac{X-\mu}{\sigma} < \dfrac{b-\mu}{\sigma}\right] = \Pr\left[Z < \dfrac{b-\mu}{\sigma}\right]$

2) $\Pr[a < X < b] = \Pr\left[\dfrac{a-\mu}{\sigma} < \dfrac{X-\mu}{\sigma} < \dfrac{b-\mu}{\sigma}\right]$

$= \Pr\left[\dfrac{a-\mu}{\sigma} < Z < \dfrac{b-\mu}{\sigma}\right]$

โดยที่ Pr สำหรับ Z เราอ่านได้จากตาราง Standard Normal (Z - table) นั่นคือ ตาราง A3

ถ้า $X \sim N(\mu, \sigma^2)$ แล้ว $Y = a + bX \sim N(a + b\mu, b^2\sigma^2)$

$[E(Y) = a + b\,E(X) = a + b\mu$ และ $Var(Y) = b^2 Var(X) = b^2\sigma^2]$

ผลบวกของตัวแปรเชิงสุ่มแบบปกติ (Normal random variable) 2 ตัว ที่เป็นอิสระต่อกัน (independent) จะมีการแจกแจงเป็นแบบปกติด้วย **นั่นคือ ถ้า $X_1 \sim N(\mu_1, \sigma_1^2)$ และ $X_2 \sim N(\mu_2, \sigma_2^2)$ และ X_1 และ X_2 เป็นอิสระต่อกัน**

$(X_1 + X_2) \sim N(\mu, \sigma^2)$ **โดยที่** $\mu = \mu_1 + \mu_2$,

$\sigma^2 = \sigma_1^2 + \sigma_2^2$

และ $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$

2.2 คำจำกัดความ (Definitions)

ตัวอย่างสุ่มขนาด n (a random sample of size n) จากประชากรหนึ่งซึ่งมีการแจกแจงเป็น f(x) ประกอบด้วยตัวแปรเชิงสุ่ม n ตัว ที่เป็นอิสระต่อกัน และแต่ละตัวมีการแจกแจงเป็น f(x)

ตัวสถิติ (Statistic) เป็นฟังก์ชันของค่าสังเกตจากตัวอย่าง ตัวสถิติเองก็เป็นตัวแปรเชิงสุ่ม การแจกแจงของความน่าจะเป็นของมันเราเรียกว่า Sampling distribution ของตัวสถิติ

Mean และ Standard diviation ของ $\bar{X}$: การแจกแจงของ $\bar{X}$ ซึ่งขึ้นอยู่กับตัวอย่างสุ่ม **ขนาด n มี mean $E(\bar{X}) = \mu$ (population mean), $Var(\bar{X}) = \frac{\sigma^2}{n}$ (population variance/n) และ s.d.$(\bar{X}) = \frac{\sigma}{\sqrt{n}}$ (population s.d.$/\sqrt{n}$)** (ดูหมายเหตุ)

<u>หมายเหตุ</u> ตัวอย่างสุ่มขนาด n จาก $N(\mu, \sigma^2)$ นั่นคือ $X_i \sim N(\mu, \sigma^2)$, $i = 1, \ldots, n$

$$\because \bar{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

$$E(\bar{X}) = \frac{1}{n} E(X_1 + \ldots + X_n) = \frac{1}{n}(n\mu) = \mu$$

$$Var(\bar{X}) = \frac{1}{n^2} Var(X_1 + \ldots + X_n) = \frac{1}{n^2}\left|Var(X_1) + \ldots + Var(X_n)\right| = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

$\bar{X}$ จาก Normal population จะมีการแจกแจงเป็น Normal **ด้วย ($\bar{X}$ จากตัวอย่างสุ่มขนาด n จาก $N(\mu, \sigma^2)$) นั่นคือ ถ้า $X \sim N(\mu, \sigma^2)$ แล้ว $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$**

2.3 ทฤษฎีขีดจำกัดส่วนกลาง [Central Limit Theorem (CLT)]

ในการสุ่มตัวอย่างจากประชากรใด ๆ ที่มี population mean $= \mu$ และ population variance $= \sigma^2$ การแจกแจงของ $\bar{x}$ เมื่อ n ขนาดใหญ่เราประมาณได้ด้วยการแจกแจงแบบปกติ (Normal distribution) **ที่มี mean $= \mu$ และ variance $= \frac{\sigma^2}{n}$ หรืออาจกล่าวได้ว่า เราประมาณ**

การแจกแจงของ $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ **ว่าเป็น N(0, 1)**

2.4 การตรวจสอบข้อสมมุติเกี่ยวกับการแจกแจงแบบปกติ (Checking the assumption of a normal population)

เราสามารถทำการตรวจสอบได้หลายวิธีการเช่น

2.4.1 การทดสอบ Goodness of fit โดยใช้ Chi-square test (ดูหัวข้อ 3.6)

2.4.2 ตรวจค่า proportion ของข้อมูลที่อยู่ในช่วงห่างรอบ ๆ mean

สำหรับ Normal distribution ที่มี mean μ และ variance σ^2 เราทราบว่า ด้วยความน่าจะเป็น .683, .954 และ .997 ตัวแปรเชิงสุ่มจะตกอยู่ในช่วง (interval) $(\mu - \sigma, \mu + \sigma)$, $(\mu - 2\sigma, \mu + 2\sigma)$ และ $(\mu - 3\sigma, \mu + 3\sigma)$ ตามลำดับ นั่นคือประมาณความน่าจะเป็นนอกช่วงเหล่านี้เป็น $\frac{1}{3}$, $\frac{1}{20}$ และ $\frac{1}{300}$ ตามลำดับ โดยทั่ว ๆไปเรามักไม่ทราบค่า μ และ σ^2 แต่ถ้า n ค่อนข้างใหญ่ เราคาดว่าค่าของ $\bar{X}$ จะใกล้ค่าของ μ และ s.d. s จะใกล้ σ ดังนั้นจากตัวอย่างเราคำนวณ $\bar{x}$ และ s แล้วนับจำนวนค่าสังเกตที่อยู่นอกช่วง $(\bar{x} - s, \bar{x} + s)$, $(\bar{x} - 2s, \bar{x} + 2s)$ และ $(\bar{x} - 3s, \bar{x} + 3s)$ แล้วหารจำนวนที่นับได้ด้วยขนาดของตัวอย่าง (= n) เราจะได้ relative frequency 3 ค่า $(\hat{p})$ เอาค่าทั้ง 3 เปรียบเทียบกับ $\frac{1}{3}$, $\frac{1}{20}$ และ $\frac{1}{300}$ (เรียก expected frequency : p) แล้วถ้า $|\hat{p} - p| / \sqrt{\frac{p(1-p)}{n}} > 3$ เราจะสรุปว่า การแจกแจงไม่เป็น Normal distribution (lack of normality)

2.4.3 ใช้ Normal Probability Paper

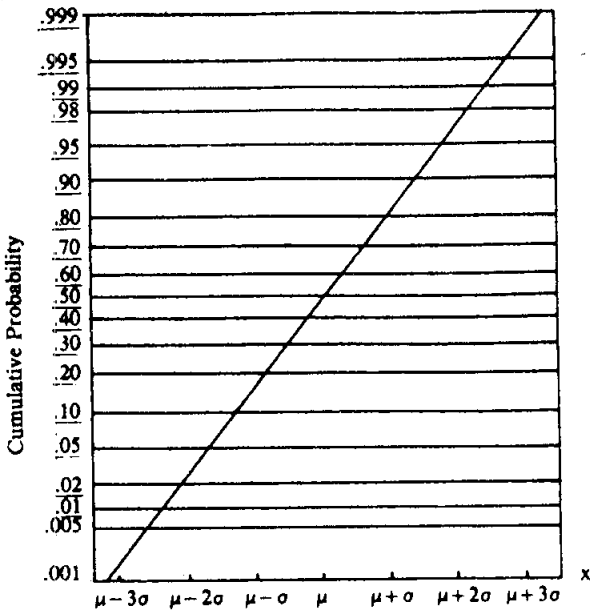
วิธีทำ

1. เรียงลำดับ n observations จากน้อยไปหามาก (i = ลำดับที่ = 1, ..., n)

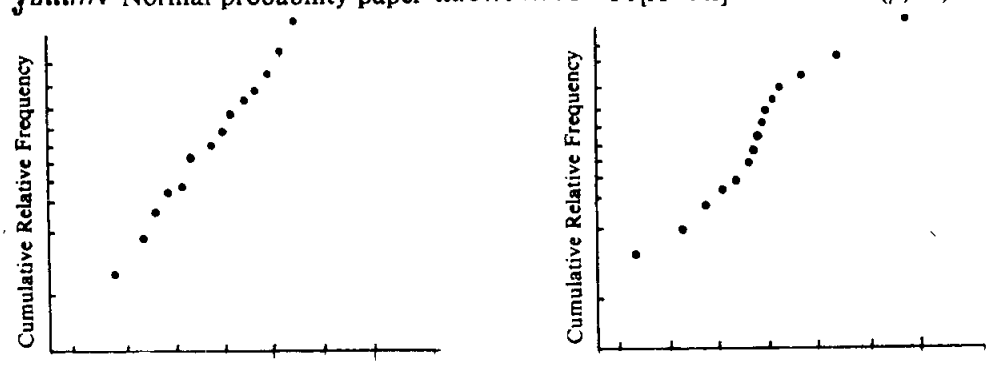
2. เขียนแกนนอนให้สามารถ plot ค่าของทุก observation ได้

3. plot (modified cumulative relative frequency $(i - \frac{1}{2})/n$ บนแกนตั้ง, ค่าของค่าสังเกตที่ i)

4. สำรวจว่าจุด plot ห่างไปจากเส้นตรงหรือไม่ ถ้าจุดห่างจากเส้นตรงอย่างมีระบบ เราจะสรุปว่าการแจกแจงไม่เป็น Normal distribution



รูปแสดง Normal probability paper *และกราฟของ* $Pr[X \leq x]$ เมื่อ $X \sim N(\mu, \sigma^2)$



ก) เป็น normal ได้ ข. ไม่เป็น normal

รูปแสดงการ plot *ลงบน* Normal probability paper

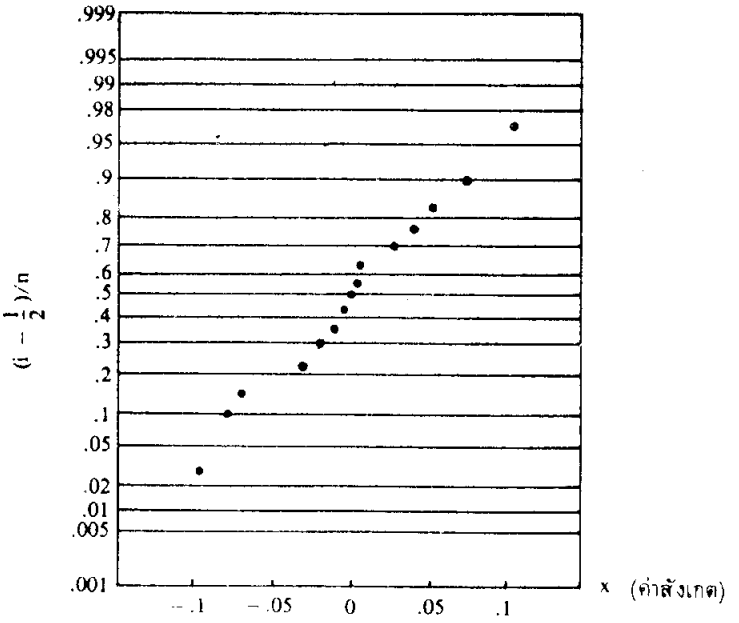
ตัวอย่างที่ 2.1 จากข้อมูลชุดหนึ่งซึ่งประกอบด้วย 15 ค่าสังเกต (observations)

.075	−.077	.001
−.028	.050	−.022
.005	.003	−.009
−.095	−.007	.101
.042	.030	−.072

จงตรวจสอบว่าข้อมูลชุดนี้มีการแจกแจงเป็น Normal distribution หรือไม่ โดยให้ใช้วิธีการที่อธิบายในหัวข้อ 2.4.3 นั่นคือใช้ Normal probability paper

n = 15

ค่าสังเกต : x	i	$(i - \frac{1}{2})/n$
−.095	1	.033
−.077	2	.1
−.072	3	.17
−.028	4	.23
−.022	5	.3
−.009	6	.37
−.007	7	.45
.001	8	.5
.003	9	.57
.005	10	.63
.030	11	.7
.042	12	.77
.050	13	.83
.075	14	.9
.101	15	.97



จากรูป การแจกแจงของข้อมูลชุดนี้พอบอกได้ว่าเป็น Normal distribution

2.5 การประมาณค่า (Estimation)

θ: พารามิเตอร์ (Parameter) คือ ค่าคงที่ที่แสดงคุณลักษณะของประชากร ซึ่งเรามักไม่ทราบค่าที่แท้จริงของมัน เราจึงสุ่มตัวอย่าง แล้วใช้ข้อมูลจากตัวอย่างเพื่อสรุปถึงลักษณะของพารามิเตอร์ที่เราสนใจ

2.5.1 การประมาณค่าพารามิเตอร์แบบจุด (Point estimation) คือ การใช้ค่า 1 ค่าจากตัวอย่างเป็นค่าประมาณของพารามิเตอร์

Estimator : $\hat{\Theta}$ เป็นฟังก์ชันของค่าสังเกตในตัวอย่าง เราใช้มันเป็นตัวประมาณค่าพารามิเตอร์

Estimate : $\hat{\theta}$ เป็นค่าค่าหนึ่งของ $\hat{\Theta}$ ซึ่งคำนวณได้จากตัวอย่างหนึ่ง

เรากล่าวว่า Estimator $\hat{\Theta}$ เป็น Unbiased estimator ของพารามิเตอร์ θ ถ้า $E(\hat{\Theta}) = \theta$

ในบรรดา Unbiased estimator ของพารามิเตอร์ θ จะเรียกตัวใดตัวหนึ่งว่าเป็น Minimum variance unbiased estimator (MVUE) ของพารามิเตอร์ θ ถ้า estimator ตัวนั้นมี variance น้อยที่สุด

เราเรียก Standard deviation ของ estimator $\hat{\Theta}$ ว่า Standard error ของ $\hat{\Theta}$ หรือเขียน S.E. ($\hat{\Theta}$)

ตัวอย่างที่ 2.2 Point estimation ของ Population mean

Parameter : Population mean μ

Data : $X_1, \ldots, X_n$ (ตัวอย่างสุ่มขนาด n)

Estimator : Sample mean $\bar{X}$

S.E.($\bar{X}$) = $\sigma/\sqrt{n}$, estimate S.E.($\bar{X}$) = $\frac{s}{\sqrt{n}}$ โดยที่ s = sample standard deviation

สำหรับ n โต ๆ ด้วยความเชื่อมั่นประมาณ 95.4% อาจกล่าวได้ว่า $\bar{X}$ จะต่างจาก μ ได้มากที่สุด $\pm 2s/\sqrt{n}$ (นั่นคือค่าประมาณของ 95.4% error bound)

$$\Pr[-2s/\sqrt{n} < \mu - \bar{X} < 2s/\sqrt{n}] = .954 \Rightarrow \Pr[\bar{X} - 2s/\sqrt{n} < \mu < \bar{X} + 2s/\sqrt{n}] = .954$$

ตัวอย่างที่ 2.3 Point estimation ของ Binomial parameter

Parameter : Population proportion p [Pr (s)]

Data : X = จำนวนของสิ่งที่มีลักษณะที่สนใจจากของทั้งหมด n สิ่ง

Estimator : $\hat{P} = \frac{X}{n}$ และ $\hat{Q} = 1 - \hat{P}$

S.E. ($\hat{P}$) $= \sqrt{\frac{pq}{n}}$, estimate S.E. ($\hat{P}$) $= \sqrt{\frac{\hat{p}\hat{q}}{n}}$

สำหรับ n โต ๆ ด้วยความเชื่อมั่นประมาณ 95.4% อาจกล่าวได้ว่า $\hat{P}$ ห่างจาก p ได้มากที่สุด $\pm 2\sqrt{\frac{\hat{p}\hat{q}}{n}}$ (ค่าประมาณของ 95.4% error bound)

$$\Pr[-2\sqrt{\hat{p}\hat{q}/n} < p - \hat{p} < 2\sqrt{\hat{p}\hat{q}/n}] = .954 \Rightarrow \Pr[\hat{P} - 2\sqrt{\hat{p}\hat{q}/n} < p < \hat{P} + 2\sqrt{\hat{p}\hat{q}/n}] = .954$$

2.5.2 การประมาณค่าเป็นช่วง (Interval estimation or Estimation by Confidence Interval)

ถ้า $(1 - \alpha)$ เป็นค่าของความน่าจะเป็นที่มีค่าสูง และ L และ U เป็นฟังก์ชันของ $X_1, \ldots, X_n$ ซึ่งทำให้ $\Pr[L < \theta < U] = 1 - \alpha$ แล้ว เราจะเรียก interval (L, U) ว่า $100(1 - \alpha)\%$ ช่วงของความเชื่อมั่น (Confidence interval) ของพารามิเตอร์ θ และเรียก $(1 - \alpha)$ ว่าเป็นระดับของความเชื่อมั่นของ interval นี้

$\Pr[L < \theta < U] = 1 - \alpha$ หมายความว่า ความน่าจะเป็นที่ random interval (L, U) จะรวมค่าของพารามิเตอร์ θ อยู่ด้วย เป็น $(1 - \alpha)$

ตัวอย่างที่ 2.4 ช่วงของความเชื่อมั่นของ μ จากตัวอย่างขนาดใหญ่ (C.I. ของ μ)

จาก CLT (เราไม่ต้องมีข้อสมมุติเกี่ยวกับรูปการแจกแจงของประชากร นอกจากค่าของ $\sigma^2 \neq \infty$ (มีค่า finite)) เราจะได้ว่า

> เมื่อ n มีขนาดใหญ่ และเราไม่ทราบค่า Population variance σ^2
> $100(1-\alpha)\%$ C.I. ของ μ คือ $(\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}})$ โดยที่
> s คือ sample s.d. และค่า $z_{\alpha/2}$ อ่านได้จากตาราง A3

ตัวอย่างที่ 2.5

> ช่วงของความเชื่อมั่นของ p เมื่อตัวอย่างมีขนาดใหญ่
> $100(1-\alpha)\%$ C.I. ของ p คือ $(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$
> $\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}})$

2.6 การแจกแจงแบบ t(Student's t-distribution)

ถ้า $X_1, \ldots, X_n$ เป็นตัวอย่างสุ่มขนาด n (n มีขนาดเล็ก ซึ่งในทางปฏิบัติ $n < 30$ เราถือว่า n มีขนาดเล็ก) จาก Normal population $N(\mu, \sigma^2)$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{และ} \quad S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}} \text{ แล้ว}$$

ตัวแปรเชิงสุ่ม $T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ จะมีการแจกแจงเป็น Student's t distribution ที่มี degrees of freedom (d.f.) = $n - 1$

ถ้า d.f. ของ t-Distribution $\to \infty$, t-distribution จะเข้าใกล้ N(0, 1) distribution

หมายเหตุ ถ้า n มีขนาดเล็ก และประชากรไม่เป็น normal population การแจกแจงของ T ไม่ใช่ t distribution

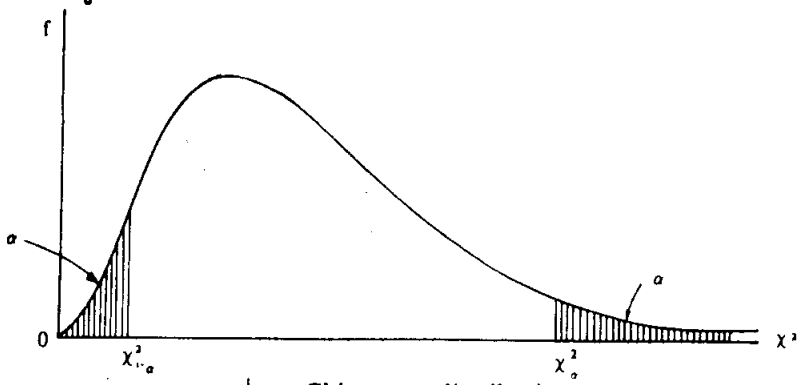
ตัวอย่างที่ 2.6

> ช่วงของความเชื่อมั่นของ μ จากตัวอย่างขนาดเล็กและไม่ทราบค่า σ^2
> $100(1-\alpha)\%$ C.I. ของ μ คือ $(\bar{x} - t_{n-1,\alpha/2}\, s/\sqrt{n}, \bar{x} + t_{n-1,\alpha/2}\, s/\sqrt{n})$
> โดยที่ $t_{n-1,\alpha/2}$ อ่านได้จากตาราง A4

2.7 การแจกแจงแบบ Chi-square (Chi-square distribution หรือ χ^2 distribution)

ถ้า $X_1, \ldots, X_n$ เป็นตัวอย่างสุ่มขนาด n จาก Normal population $N(\mu, \sigma^2)$ แล้วตัวแปรเชิงสุ่ม $X^2 = \dfrac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2} = \dfrac{(n-1)S^2}{\sigma^2}$ จะมีการแจกแจงเป็น Chi-square distribution ที่มี d.f. = n − 1

รูป Probability density curve ของ Chi-square ไม่เป็นรูปสมมาตร (symmetric curve) curve ทั้งหมดอยู่ใน guadrant ที่ 1 ซึ่งมีหางขวายาว รูปร่างของ curve ขึ้นอยู่กับค่าของ d.f. รูปข้างล่างนี้แสดงรูปทั่ว ๆ ไปของ Chi-square curve



รูปของ Chi-square distribution

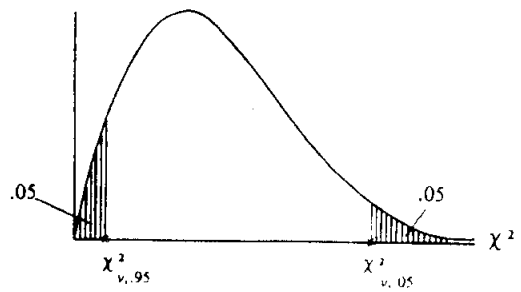
2.7.1 การใช้ตาราง Chi-square (ตาราง A5)

ตาราง A5 ให้ค่า upper α points ของ χ^2 distribution สำหรับค่าต่าง ๆ ของ α และ d.f.

χ^2_{α} หมายความถึงค่า χ^2 จากตารางที่มีพื้นที่ทางขวาของมัน $= \alpha$

$\chi^2_{1-\alpha}$ หมายความถึงค่า χ^2 จากตารางที่มีพื้นที่ทางขวาของมัน $= 1 - \alpha$

ถ้า $\alpha = .05$ และ d.f. $= \nu$

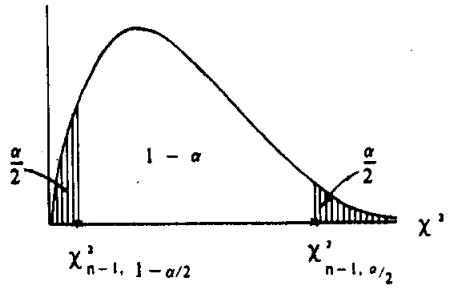


ค่าของ $\chi^2_{\nu,.95}$ และ $\chi^2_{\nu,.05}$ อ่านได้โดยตรงจากตาราง A5.

ตัวอย่างที่ 2.7 d.f. = 15, $\chi^2_{15,.95} = 7.261$

$\chi^2_{15,.05} = 24.996$

2.7.2 การหาช่วงของความเชื่อมั่นของ σ^2



จากรูป $\Pr[\chi^2_{n-1,\,1-\frac{\alpha}{2}} < X^2 < \chi^2_{n-1,\,\alpha/2}] = 1 - \alpha$

แทนค่า $X^2 = \dfrac{(n-1)S^2}{\sigma^2}$

ดังนั้น $\Pr\left[\chi^2_{n-1,1-\frac{\alpha}{2}} < \dfrac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1,\frac{\alpha}{2}}\right] = 1 - \alpha$

$$\Rightarrow \quad \Pr\left[\frac{1}{\chi^2_{n-1,\frac{\alpha}{2}}} < \frac{\sigma^2}{(n-1)S^2} < \frac{1}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right] = 1 - \alpha$$

$$\Rightarrow \quad \Pr\left[\frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right] = 1 - \alpha$$

ถ้า $X_1, \ldots, X_n$ เป็นตัวอย่างสุ่มจาก $N(\mu, \sigma^2)$ แล้ว $100(1-\alpha)\%$ C.I. ของ σ^2 คือ $\left(\dfrac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}}, \dfrac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right)$ โดยที่ $\chi^2_{n-1,\frac{\alpha}{2}}$ และ $\chi^2_{n-1,1-\frac{\alpha}{2}}$ เป็นค่าของ χ^2 ที่มี d.f. = n − 1 และมีพื้นที่ $\frac{\alpha}{2}$ และ $1 - \frac{\alpha}{2}$ อยู่ทางขวาของมันตามลำดับ

ดังนั้น $100(1-\alpha)\%$ C.I. ของ σ คือ $\left(\sqrt{\dfrac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}}}, \sqrt{\dfrac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}}\right)$

ตัวอย่างที่ 2.8 ผู้ผลิตนาฬิกาต้องการจะตรวจสอบความเที่ยงตรงของนาฬิกาที่เขาผลิตขึ้น เขาตกลงใจจะหา Confidence interval ของ σ (s.d. ของเวลาที่ต่างไปจากเวลามาตรฐาน) โดยสุ่มนาฬิกามา 10 เรือนจากนาฬิกาปริมาณมากที่ผ่านการตรวจสอบขั้นสุดท้ายแล้ว หลังจาก 1 เดือน เขาได้จดบันทึกเวลาของนาฬิกาแต่ละเรือนที่ต่างไปจากนาฬิกาเรือนมาตรฐาน จากข้อมูล 10 ค่า คำนวณได้ $\bar{x} = 0.7$ วินาที และ $s = 0.4$ วินาที สมมุติว่าข้อมูลที่จะวัดได้จากนาฬิกาทุกเรือนที่ผลิตได้มีการแจกแจงเป็น Normal distribution เราจะหา 90% C.I. ของ σ

$n = 10 \therefore$ d.f. $= n - 1 = 9$

$\alpha = .10$ จากตาราง A5 $\chi^2_{9,.95} = 3.325$ และ $\chi^2_{9,.05} = 16.919$

ดังนั้น 90% C.I. ของ σ^2 คือ $\left(\frac{9(.4)^2}{16.919}, \frac{9(.4)^2}{3.325} \right) = (.085, .433)$

$\therefore$ 90% C.I. ของ σ คือ $(\sqrt{.085}, \sqrt{.433}) = (.29, .66)$

หมายความว่าผู้ผลิตนาฬิกาสามารถเชื่อได้ 90% ว่าค่าของ s.d. ของเวลาของนาฬิกาที่เขาผลิตที่ต่างไปจากเวลามาตรฐาน จะอยู่ระหว่าง .29 และ .66 วินาที

2.8 การเปรียบเทียบ 2 วิธีการ (Comparing two treatments) เมื่อตัวอย่างมีขนาดใหญ่

ตัวอย่างสุ่ม 2 ตัวอย่าง จาก 2 ประชากร ตัวอย่างทั้ง 2 เป็นอิสระต่อกัน

2 ตัวอย่างที่เป็นอิสระกัน	สรุปค่า
$X_{11}, X_{12}, \ldots, X_{1n_1}$ จากประชากรที่ 1 ที่มี mean μ_1 และ variance σ_1^2	$\bar{X}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{1i}$ $S_1^2 = \frac{1}{(n_1 - 1)} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2$
$X_{21}, X_{22}, \ldots, X_{2n_2}$ จากประชากรที่ 2 ที่มี mean μ_2 และ variance σ_2^2	$\bar{X}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} X_{2i}$ $S_2^2 = \frac{1}{(n_2 - 1)} \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)^2$

ถ้าตัวอย่างขนาดใหญ่ (ทั้ง n_1 และ n_2 มากกว่าหรือเท่ากับ 30) เราจะใช้ CLT (หัวข้อ 2.3) ช่วยในการหา Confidence interval ของ $\mu_1 - \mu_2$

$E(\bar{X}_1) = \mu_1$ และ $Var(\bar{X}_1) = \sigma_1^2/n_1$

$E(\bar{X}_2) = \mu_2$ และ $Var(\bar{X}_2) = \sigma_2^2/n_2$

$E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = \mu_1 - \mu_2$ และ

$$Var(\bar{X}_1 - \bar{X}_2) = Var(\bar{X}_1) + Var(\bar{X}_2) \quad (\because \bar{X}_1 \text{ และ } \bar{X}_2 \text{ เป็นอิสระกัน})$$

$$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \approx \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}$$

ดังนั้น $$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0, 1)$$

ประมาณ $100(1 - \alpha)\%$ C.I. ของ $\mu_1 - \mu_2$ (n_1 และ n_2 มีขนาดใหญ่) คือ

$$\left((\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},\ (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)$$

2.9 การเปรียบเทียบ 2 วิธีการเมื่อตัวอย่างมีขนาดเล็ก

ข้อสมมุติเมื่อตัวอย่างมีขนาดเล็ก

ก) Population ทั้ง 2 เป็น Normal

ข) Population variances σ_1^2 และ σ_2^2 มีค่าเท่ากัน ($\sigma_1^2 = \sigma_2^2$)

หมายเหตุ เราอาจทำการทดสอบก่อนว่า $\sigma_1^2 = \sigma_2^2$ หรือไม่ (ดูวิธีการในหัวข้อ 3.5) ถ้า σ_1^2 เท่ากับ σ_2^2 จริง เราจึงใช้วิธีการในหัวข้อนี้ ในการสร้างตัวสถิติในการทดสอบสมมติฐาน $H_0 : \mu_1 = \mu_2$ (หัวข้อ 3.3)

ค) ตัวอย่างทั้ง 2 เป็นอิสระต่อกัน

$$\left.\begin{array}{l} X_{11}, \ldots, X_{1n_1} \text{ เป็นตัวอย่างสุ่มจาก } N(\mu_1, \sigma^2) \\ X_{21}, \ldots, X_{2n_2} \text{ เป็นตัวอย่างสุ่มจาก } N(\mu_2, \sigma^2) \end{array}\right\} \quad \because (\sigma_1^2 = \sigma_2^2 = \sigma^2)$$

$X_{11}, \ldots, X_{1n_1}$ และ $X_{21}, \ldots, X_{2n_2}$ เป็นอิสระต่อกัน

$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$

$$Var(\bar{X}_1 - \bar{X}_2) = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

Pooled estimator ของ Common σ^2 คือ

$$S^2_{pooled} = \frac{\sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)^2}{(n_1 - 1) + (n_2 - 1)}$$

$$= \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

ดังนั้น $\text{est.Var}(\bar{X}_1 - \bar{X}_2) = S^2_{pooled}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$

ตัวแปรเชิงสุ่ม $T = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{pooled}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ มีการแจกแจงเป็น t-distribution ที่มี

d.f. $= n_1 + n_2 - 2$

$100(1 - \alpha)\%$ C.I. ของ $\mu_1 - \mu_2$ (n_1 และ n_2 มีขนาดเล็ก) คือ

$$(\bar{x}_1 - \bar{x}_2) \pm t_{n_1+n_2-2,\frac{\alpha}{2}}\, s_{pooled}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

โดยที่ $s^2_{pooled} = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

2.10 Paired Comparison (ตัวอย่าง 2 ตัวอย่างไม่เป็นอิสระต่อกัน)

โครงสร้างของข้อมูลสำหรับ Paired Comparison

X_{ij} = ค่าสังเกตจากคู่ที่ i, วิธีการ (Treatment) ที่ j

$i = 1, \ldots, n$; $j = 1, 2$

Pair = คู่ที่	Treatment ที่ 1	Treatment ที่ 2	ผลต่าง
1	X_{11}	X_{12}	$D_1 = X_{11} - X_{12}$
2	X_{21}	X_{22}	$D_2 = X_{21} - X_{22}$
⋮	⋮	⋮	⋮
n	X_{n1}	X_{n2}	$D_n = X_{n1} - X_{n2}$

เพราะว่าระหว่างคู่ คือ $(X_{11}, X_{12}), \ldots, (X_{n1}, X_{n2})$ เป็นอิสระต่อกัน ดังนั้น $D_1, \ldots, D_n$ จะเป็นตัวอย่างสุ่มจากประชากรที่มี mean μ_D และ variance σ_D^2

$$D_i = X_{i1} - X_{i2}$$

$$\left.\begin{aligned} E(D_i) &= E(X_{i1} - X_{i2}) = \mu_D \\ Var(D_i) &= Var(X_{i1} - X_{i2}) = \sigma_D^2 \end{aligned}\right\} \forall i = 1, \ldots, n$$

$D_1, \ldots, D_n$ เป็นตัวอย่างสุ่มขนาดเล็กจาก $N(\mu_D, \sigma_D^2)$ ถ้า $\bar{D} = \frac{1}{n} \sum_{i=1}^{n} D_i$

และ $S_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - \bar{D})^2$ แล้ว

$100(1-\alpha)\%$ C.I. ของ μ_D คือ $\bar{d} \pm t_{n-1, \frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}$ โดยที่ $\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i$

และ $s_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})^2$

2.11 การเปรียบเทียบ 2 Binomial Proportions

โครงสร้าง

Parameter : $p_1 - p_2$ = (Proportion ของ Population 1) – (Proportion ของ Population 2)

ข้อมูล : ตัวอย่าง 2 ตัวอย่าง ซึ่งเป็นอิสระต่อกันขนาด n_1 และ n_2 ตามลำดับ

X_1 = จำนวน Success ในการกระทำ n_1 ครั้ง

X_2 = จำนวน Success ในการกระทำ n_2 ครั้ง

Proportion จากตัวอย่าง :

$\hat{P}_1 = \frac{X_1}{n_1}$	$\hat{P}_2 = \frac{X_2}{n_2}$
$E(\hat{P}_1) = p_1$	$E(\hat{P}_2) = p_2$
$Var(\hat{P}_1) = \frac{p_1(1-p_1)}{n_1}$	$Var(\hat{P}_2) = \frac{p_2(1-p_2)}{n_2}$

เพราะว่า $\hat{P}_1$ และ $\hat{P}_2$ เป็นอิสระต่อกัน

$$Var(\hat{P}_1 - \hat{P}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

$$s.d.(\hat{P}_1 - \hat{P}_2) = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$est.\ s.d.(\hat{P}_1 - \hat{P}_2) = \sqrt{\frac{\hat{P}_1(1-(\hat{P}_1)}{n_1} + \frac{(\hat{P}_2(1-\hat{P}_2)}{n_2}}$$

ถ้า n_1 และ n_2 มีขนาดใหญ่ เราสามารถประมาณรูปการแจกแจงของ $\hat{P}_1 - \hat{P}_2$ **เป็น** Normal distribution

> **ประมาณ** $100(1-\alpha)\%$ C.I. **ของ** $p_1 - p_2$ **(เมื่อ** n_1 **และ** n_2 **มีขนาดใหญ่) คือ**
>
> $$(\hat{P}_1 - \hat{P}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}$$

2.12 การเปรียบเทียบ Variances ของ 2 Normal populations และการแจกแจงแบบ F (Comparing the variances of two Normal populations and F-distribution)

ข้อสมมุติ (Assumptions)

1) $X_{11}, \ldots, X_{1n_1}$ เป็นตัวอย่างสุ่มจาก $N(\mu_1, \sigma_1^2)$ $\therefore X_{(1)}^2 = \frac{(n_1-1)S_1^2}{\sigma_1^2}$ เป็น Chi-square random variable มี d.f. $= n_1 - 1$

2) $X_{21}, \ldots, X_{2n_2}$ เป็นตัวอย่างสุ่มจาก $N(\mu_2, \sigma_2^2)$ $\therefore X_{(2)}^2 = \frac{(n_2-1)S_2^2}{\sigma_2^2}$ เป็น Chi-square random variable มี d.f. $= n_2 - 1$

3) ตัวอย่างสุ่มทั้งสองเป็นอิสระต่อกัน $\therefore X_{(1)}^2$ และ $X_{(2)}^2$ เป็นอิสระต่อกัน

ถ้า $X_{11}, \ldots, X_{1n_1}$, และ $X_{21}, \ldots, X_{2n_2}$ เป็นตัวอย่างสุ่ม 2 ตัวอย่างที่เป็นอิสระต่อกัน จาก 2 Normal populations : $N(\mu_1, \sigma_1^2)$ และ $N(\mu_1, \sigma_2^2)$ ตามลำดับแล้ว distribution ของ

$$F = \frac{X_{(1)}^2/(n_1-1)}{X_{(2)}^2/(n_2-1)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}$$

มีการแจกแจงเป็น F - distribution ที่มี d.f. $= (n_1 - 1, n_2 - 1)$

ถ้า F มีการแจกแจงเป็น F - distribution ที่มี d.f. $= (\nu_1, \nu_2)$ แล้ว $F^* = \frac{1}{F}$ จะมีการแจกแจงเป็น F-distribution ที่มี d.f. $= (\nu_2, \nu_1)$

หมายเหตุ subscripts (1) และ (2) ที่ X^2 หมายถึง X^2 ตัวที่ 1 และตัวที่ 2 ตามลำดับ

2.12.1 การใช้ตาราง F (ตาราง A6)

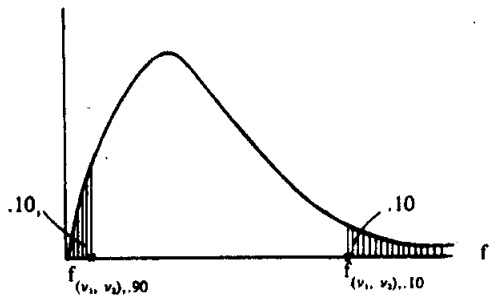
ตาราง A6 ให้ค่า upper α points ของ F - distribution สำหรับค่า $\alpha = .01, .05$ และ $.10$ สำหรับ d.f. $= (\nu_1, \nu_2)$

$f_{(\nu_1, \nu_2), \alpha}$ หมายความถึงค่า f จากตารางที่มีพื้นที่ทางขวาของจุด = α ค่านี้เราสามารถอ่านได้จากตารางทันที ถ้า α = .01, .05 หรือ .10

$f_{(\nu_1, \nu_2), 1-\alpha}$ หมายความถึงค่า f ที่มีพื้นที่ทางขวาของจุด = $1 - \alpha$ เราไม่สามารถอ่านค่าได้จากตารางโดยตรง แต่สามารถหาค่าได้จากสูตร

$$f_{(\nu_1, \nu_2), 1-\alpha} = \frac{1}{f_{(\nu_2, \nu_1), \alpha}}$$

ถ้า α = .10 และ d.f. = (ν_1, ν_2)

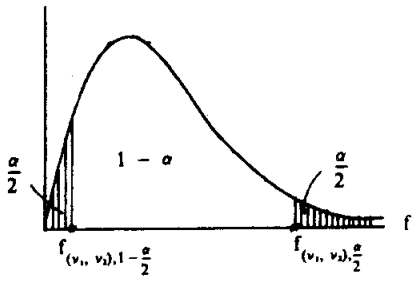


ตัวอย่างที่ 2.9 d.f. = (6, 12), $f_{(6,12),.10} = 2.3310$

$$f_{(6,12),.90} = \frac{1}{f_{(12,6),.10}}$$

$$= \frac{1}{2.9047} = .3443$$

2.12.2 การหาช่วงของความเชื่อมั่นของ $\frac{\sigma_1^2}{\sigma_2^2}$



จากรูป $\Pr\left[f_{(\nu_1,\nu_2),1-\frac{\alpha}{2}} < F < f_{(\nu_1,\nu_2),\frac{\alpha}{2}}\right] = 1 - \alpha$

แทนค่า $F = \dfrac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}$

ดังนั้น $\Pr\left[f_{(\nu_1,\nu_2),1-\frac{\alpha}{2}} < \dfrac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2} < f_{(\nu_1,\nu_2),\frac{\alpha}{2}}\right] = 1 - \alpha$

$$\Rightarrow \quad \Pr\left[\frac{1}{f_{(\nu_1,\nu_2),\frac{\alpha}{2}}} < \frac{S_2^2\sigma_1^2}{S_1^2\sigma_2^2} < \frac{1}{f_{(\nu_1,\nu_2),1-\frac{\alpha}{2}}}\right] = 1 - \alpha$$

$$\Rightarrow \quad \Pr\left[\frac{S_1^2}{S_2^2}\,\frac{1}{f_{(\nu_1,\nu_2),\frac{\alpha}{2}}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\,f_{(\nu_2,\nu_1),\frac{\alpha}{2}}\right] = 1 - \alpha$$

$100(1-\alpha)\%$ C.I. ของ $\dfrac{\sigma_1^2}{\sigma_2^2}$ คือ $\left(\dfrac{s_1^2}{s_2^2}\,\dfrac{1}{f_{(\nu_1,\nu_2),\frac{\alpha}{2}}},\ \dfrac{s_1^2}{s_2^2}\,f_{(\nu_2,\nu_1),\frac{\alpha}{2}}\right)$

ดังนั้น $100(1-\alpha)\%$ C.I. ของ $\dfrac{\sigma_1}{\sigma_2}$ คือ $\left(\sqrt{\dfrac{s_1^2}{s_2^2}\,\dfrac{1}{f_{(\nu_1,\nu_2),\frac{\alpha}{2}}}},\ \sqrt{\dfrac{s_1^2}{s_2^2}\,f_{(\nu_2,\nu_1),\frac{\alpha}{2}}}\right)$

ตัวอย่างที่ 2.10 โรงงานทอผ้าควบคุมปริมาณ fiber จากผ้าที่ทอจากเครื่องทอผ้า 2 เครื่อง แต่ละ observation คือ น้ำหนักของผ้ายาว 100 หลา สุ่มผ้ามา 12 ชิ้น จากเครื่อง A และ 10 ชิ้น จากเครื่อง B คำนวณค่า s.d. ของน้ำหนักของผ้าจากเครื่อง A ได้ 2.3 และ 1.5 จากเครื่อง B จงหา 90% C.I. ของ ratio ของ population standard deviations

เครื่อง A (เครื่องที่ 1) $n_1 = 12,\ \nu_1 = n_1 - 1 = 11,\ s_1^2 = (2.3)^2 = 5.29$
เครื่อง B (เครื่องที่ 2) $n_2 = 10,\ \nu_2 = n_2 - 1 = 9,\ s_2^2 = (1.5)^2 = 2.25$

90% C.I. ของ $\dfrac{\sigma_1^2}{\sigma_2^2}$ คือ $\left(\dfrac{s_1^2}{s_2^2}\,\dfrac{1}{f_{(11,9),.05}},\ \dfrac{s_1^2}{s_2^2}\,f_{(9,11),.05}\right)$

$$= \left(\frac{5.29}{2.25}\,\frac{1}{3.11},\ \frac{5.29}{2.25}\,2.9\right) = (.76,\ 6.82)$$

$\therefore$ 90% C.I. ของ $\frac{\sigma_1}{\sigma_2}$ คือ $(\sqrt{.76}, \sqrt{6.82}) = (.87, 2.61)$ นั่นคือ สรุปได้ด้วยความเชื่อมั่น 90% ว่า σ_1 อาจเล็กที่สุดถึง $.87\sigma_2$ หรืออาจใหญ่ที่สุดถึง $2.61\sigma_2$

2.13 ความสัมพันธ์ของ t, χ^2 และ F กับ Normal distribution

2.13.1 Chi-square distribution

2.13.1.1 ถ้า $Z_1, \ldots, Z_k$ เป็นตัวแปรเชิงสุ่มซึ่งเป็นอิสระต่อกัน แต่ละตัวมีการแจกแจงเป็น Standard Normal distribution นั่นคือ $Z_i \sim N(0, 1), \forall i = 1, \ldots, k$ แล้วการแจกแจงของตัวแปรเชิงสุ่ม $X^2 = Z_1^2 + Z_2^2 + \ldots + Z_k^2$ จะเป็น χ^2 distribution ที่มี d.f. = k

2.13.1.2 ถ้า $X_{k_1}^2$ และ $X_{k_2}^2$ เป็น Chi-square variable 2 ตัวที่เป็นอิสระต่อกัน มี d.f. k_1 และ k_2 ตามลำดับ แล้ว $X_{k_1}^2 + X_{k_2}^2$ จะมีการแจกแจงเป็น Chi-square distribution ที่มี d.f. $= k_1 + k_2$ (Additivity property of χ^2)

2.13.1.3 ใช้ Additivity property ของ χ^2 เพื่อแสดงว่า $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

พิจารณาตัวอย่างสุ่มขนาด n : $X_1, \ldots, X_n$ จาก $N(\mu, \sigma^2)$

$$Z_i = \frac{X_i - \mu}{\sigma}, \quad i = 1, \ldots, n$$

$Z_i \sim N(0, 1)$ และเป็นอิสระต่อกัน

จาก 2.13.1.1 : $\sum_{i=1}^{n} Z_i^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2 \sim \chi^2_n$

(1); $X_i - \mu = (X_i - \bar{X}) + (\bar{X} - \mu)$

$(1)^2$; $(X_i - \mu)^2 = (X_i - \bar{X})^2 + (\bar{X} - \mu)^2 + 2(\bar{X} - \mu)(X_i - \bar{X})$

$$\sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{n} (\bar{X} - \mu)^2 + 2(\bar{X} - \mu) \sum_{i=1}^{n} (X_i - \bar{X})$$

$$\because \sum_{i=1}^{n} (X_i - \bar{X}) = 0$$

ดังนั้น
$$\sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$$
$$= (n-1)S^2 + n(\bar{X} - \mu)^2$$

เอา σ^2 หารตลอด;

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 = \frac{(n-1)S^2}{\sigma^2} + \frac{n(\bar{X}-\mu)^2}{\sigma^2}$$

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

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

$$\therefore\ \frac{n(\bar{X}-\mu)^2}{\sigma^2} \sim \chi^2_1$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2_n$$ (ได้แสดงแล้วข้างต้น)

ดังนั้น $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

2.13.2 t - distribution

2.13.2.1 $Z \sim N(0, 1)$

$X^2_k \sim \chi^2$ distribution d.f. = k

ถ้าตัวแปรเชิงสุ่ม Z และ X^2_k เป็นอิสระต่อกัน แล้วการแจกแจงของ

$T = \frac{Z}{\sqrt{X^2_k/k}}$ หรือ $\frac{N(0, 1)}{\sqrt{\text{Chi-square/d.f.}}}$ คือ t - distribution ที่มี d.f. = k

2.13.2.2 ตัวอย่างสุ่มขนาด n : $X_1, \ldots, X_n$ จาก $N(\mu, \sigma^2)$

เราทราบว่า $Z = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0, 1)$ และเป็นอิสระกับ

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

จาก 2.13.2.1 : $$\frac{\sqrt{n}\,(\bar{X}-\mu)/\sigma}{\sqrt{\frac{(n-1)S^2}{\sigma^2} \Big/ (n-1)}} = \frac{(\bar{X}-\mu)}{S/\sqrt{n}} \sim t_{n-1}$$

2.13.2.3 ถ้ามีตัวอย่างสุ่ม 2 ตัวอย่างที่เป็นอิสระต่อกัน

$X_{11}, \ldots, X_{1n_1}$ จาก $N(\mu_1, \sigma^2)$

$X_{21}, \ldots, X_{2n_2}$ จาก $N(\mu_2, \sigma^2)$

เราทราบว่า $$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$$

และ $$X^2_{(1)} = \frac{(n_1-1)S_1^2}{\sigma^2} \sim \chi^2_{n_1-1}$$ ซึ่งเป็นอิสระกับ

$$X^2_{(2)} = \frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2_{n_2-1}$$

จาก 2.13.1.2 : $$X^2 = X^2_{(1)} + X^2_{(2)} = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{\sigma^2}$$

$$X^2 = \frac{(n_1+n_2-2)S^2_{pooled}}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$$

X^2 และ Z เป็นอิสระต่อกัน เพราะว่า $\bar{X}_1, \bar{X}_2, S_1^2$ และ S_2^2 ต่างก็เป็นอิสระต่อกัน

จาก 2.13.2.1

$$T = \frac{\dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}{\sqrt{\dfrac{(n_1+n_2-2)S^2_{pooled}}{\sigma^2(n_1+n_2-2)}}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{pooled}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

2.13.3 F - distribution

2.13.3.1 ถ้า $X^2_{k_1}$ และ $X^2_{k_2}$ เป็น Chi-square variable 2 ตัวที่เป็นอิสระต่อกัน มี d.f. = k_1 และ k_2 ตามลำดับ แล้วการแจกแจงของ

$$F = F_{(k_1,k_2)} = \frac{X^2_{k_1}/k_1}{X^2_{k_2}/k_2} \quad \text{หรือ} \quad \frac{\text{Chi-square/d.f.}}{\text{Chi-square/d.f.}}$$ คือ F - distribution ที่มี d.f. = (k_1, k_2)

2.13.3.2 ตัวอย่างสุ่ม 2 ตัวอย่างที่เป็นอิสระต่อกัน

$X_{11}, \ldots, X_{1n_1}$ จาก $N(\mu_1, \sigma_1^2)$

$X_{21}, \ldots, X_{2n_2}$ จาก $N(\mu_2, \sigma_2^2)$

เราทราบว่า $X^2_{(1)} = \dfrac{(n_1 - 1)S_1^2}{\sigma_1^2} \sim X^2_{n_1-1}$ เป็นอิสระกับ

$$X^2_{(2)} = \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \sim X^2_{n_2-1}$$

ดังนั้น

$$F = \frac{X^2_{(1)}/(n_1 - 1)}{X^2_{(2)}/(n_2 - 1)} = \frac{\dfrac{(n_1 - 1)S_1^2}{\sigma_1^2(n_1 - 1)}}{\dfrac{(n_2 - 1)S_2^2}{\sigma_2^2(n_2 - 1)}} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2\,\sigma_2^2}{S_2^2\,\sigma_1^2} \sim F_{(n_1-1,\,n_2-1)}$$

2.13.4 ความสัมพันธ์ระหว่าง t และ F

ถ้า $Z \sim N(0, 1)$

$Z^2 = X_1^2 \sim \chi_1^2$

จาก 2.13.2.1 $T = \dfrac{Z}{\sqrt{X_k^2/k}} \sim t_k$

$$T^2 = \frac{Z^2}{X_k^2/k} = \frac{X_1^2/1}{X_k^2/k}$$

จาก 2.13.3.1 แสดงว่า $T^2 \sim F_{(1,k)}$

ตัวอย่าง $\nu = 15$, $t_{15,.025} = 2.131$, $(2.131)^2 = 4.541161$

$f_{(1,15),.05} = 4.54 = t^2_{15,.025}$

แบบฝึกหัด

2.1 ใช้ตาราง Chi-square (ตาราง A5) เพื่อหา

ก) $\chi^2_{5,.05}$, $\chi^2_{5,.95}$ ข) $\chi^2_{9,.01}$, $\chi^2_{9,.99}$

ค) $\chi^2_{16,.975}$, $\chi^2_{16,.025}$ ง) $\chi^2_{10,.99}$, $\chi^2_{10,.01}$

2.2 ได้ขุดหินตัวอย่างจากแหล่งน้ำมันแห่งหนึ่ง และนำมาวิเคราะห์ทางเคมีเพื่อดูเปอร์เซนต์ของ cadmium ที่มีอยู่ หลังจากวิเคราะห์หินตัวอย่าง 25 ก้อน ได้ค่าเฉลี่ย 10.2 และ s.d. 3.1 มาตรการอย่างหนึ่งที่ใช้ตัดสินคุณภาพของน้ำมันก็คือ ความไม่แตกต่างกันของส่วนประกอบภายในสารน้ำมัน ถ้าคุณภาพของน้ำมันขึ้นอยู่กับว่า s.d. ของเปอร์เซนต์ของ cadmium ไม่เกิน 4 ในบทที่ 3 เราจะศึกษาวิธีการเพื่อทดสอบที่ $\alpha = .01$ ว่ามีเหตุผลพอที่จะสนับสนุนว่า σ น้อยกว่า 4 หรือไม่ ในที่นี้จงหา 98% C.I. ของ σ^2 และ 98% C.I. ของ σ

2.3 โรงงานผลิตแบตเตอรี่รถยนต์อ้างว่าอายุการใช้งานของแบตเตอรี่ที่ผลิตจากโรงงานของเขามี s.d. เป็น 0.9 ปี ถ้าเราสุ่มแบตเตอรี่มา 10 ตัวใช้จนหมดอายุ และคำนวณหา s.d. ได้เป็น 1.2 ปี จงหา 95% C.I. ของ σ^2 และ 95% C.I. ของ σ

2.4 ใช้ตาราง F (ตาราง A6) เพื่อหา

ก) $f_{(7,9),.05}$, $f_{(7,9),.95}$ ข) $f_{(3,8),.10}$, $f_{(3,8),.90}$

ค) $f_{(12,7),.95}$, $f_{(12,7),.05}$ ง) $f_{(4,8),.90}$, $f_{(4,8),.10}$

2.5 ได้มีการสำรวจเพื่อดูว่าการให้ฮอร์โมนแก่หนูที่ตั้งครรภ์จะเพิ่มน้ำหนักของมันหรือไม่ ได้สุ่มหนูที่ตั้งครรภ์ 6 ตัวมาให้ฮอร์โมน และอีก 6 ตัวไม่ได้ให้ฮอร์โมน ต่อไปนี้เป็นข้อมูลที่สรุปได้

	หนูที่ได้ฮอร์โมน	หนูที่ไม่ได้ฮอร์โมน กลุ่มควบคุม (control group)
Mean	$\bar{x}_1 = 41.8$	$\bar{x}_2 = 60.8$
s.d.	$s_1 = 7.6$	$s_2 = 16.4$

จงหา 90% C.I. ของ ratio ของ population standard deviation ง 2 (นั่นคือ ของ σ_1 σ_2)

2.6 การทดลองทำขึ้นเพื่อเปรียบเทียบความแม่นยำของเครื่องมือตรวจสอบปริมาณสารปรอทในอากาศ 2 ชนิด คือชนิด A และ B ได้ใช้เครื่องมือชนิด A วัดความเข้มข้นของสารปรอท 7 ครั้ง และใช้เครื่องมือชนิด B วัด 6 ครั้ง ระหว่างเวลาเที่ยงในใจกลางเมืองใหญ่แห่งหนึ่ง ผลจากการวัด (หน่วยเป็นไมโครกรัม/ลูกบาศก์เมตรของอากาศ) ปรากฏดังนี้

เครื่อง A	.95	.82	.78	.96	.71	.86	.99
เครื่อง B	.89	.91	.94	.91	.90	.89	

จงหา 90% C.I. ของ σ_1/σ_2 (ratio ของ s.d. ของค่าที่วัดได้จากเครื่อง A และเครื่อง B)

2.7 ในการศึกษาถึงพฤติกรรมเกี่ยวกับเพศของลิงกับเวลาในการเล่นของลิงอายุ 1 ปี ได้นำเอาลิงเพศชายและหญิงอายุ 1 ปี อย่างละ 6 ตัว มาทดลองโดยให้อยู่รวมกลุ่มกับครอบครัวลิง 4 ครอบครัว ในช่วงเวลาช่วงละ 10 นาที ระยะเวลาเฉลี่ยที่ลิงแต่ละตัวเล่นกับลิงตัวอื่นได้ถูกบันทึกไว้ดังนี้

ตัวผู้	3.64	3.11	3.80	3.58	4.55	3.92
ตัวเมีย	1.91	2.06	1.78	2.00	1.30	2.32

จงหา 98% C.I. ของ σ_1^2/σ_2^2 (ratio ของ variance ของเวลาในการเล่นของลิงตัวผู้และของลิงตัวเมีย) และ 98% C.I. ของ σ_1/σ_2

2.8 เป็นที่คาดหมายว่าวิธีการสอนแบบใหม่จะมีประสิทธิภาพในการช่วยเพิ่มความสามารถในการอ่านของนักเรียนชั้นประถมศึกษามากกว่าวิธีเก่าที่กำลังใช้อยู่ ในการทดสอบคำกล่าวนี้ ได้แบ่งนักเรียน 16 คน ออกเป็น 2 กลุ่ม กลุ่มละ 8 คน โดยวิธีสุ่ม กลุ่มแรกสอนโดยวิธีเดิม อีกกลุ่มหนึ่งสอนโดยวิธีใหม่ คะแนนทดสอบเกี่ยวกับการอ่านของเด็กเป็นดังนี้

คะแนนการอ่าน

(1) วิธีเก่า	65	70	76	63	72	71	68	68
(2) วิธีใหม่	75	80	72	77	69	81	71	78

จงหา 98% C.I. ของ σ_1^2/σ_2^2 และ σ_1/σ_2

2.9 นักศึกษาหญิง 16 คน ชาย 25 คน ทำข้อสอบมาตรฐานวิชาคณิตศาสตร์เบื้องต้น ผลปรากฏว่านักศึกษาหญิงทำคะแนนเฉลี่ยได้ 78 คะแนน และ s.d. = 8 คะแนน ส่วนนักศึกษาชายทำคะแนนเฉลี่ยได้ 82 คะแนน และ s.d. = 9 คะแนน

ก) จงหา 95% C.I. ของ σ_1^2 (ชาย) และของ σ_2^2 (หญิง)

ข) จงหา 98% C.I. ของ σ_1^2/σ_2^2 และ σ_1/σ_2