誤差論

測定値の信頼性と平均法

誤差の定義

- 誤差(ε)とは測定値(x)から真値（μ、X）を引いた値とする。しかし、真値を求めることは可能でないから、それを予測する値が必要になる。これが最確値($\tilde{x}$)である。

(a)誤差　$\varepsilon = x - X$ …(1.1)

(b)残差　$v = x - \tilde{x}$ …(1.2)

分類

(a)原因による分類

1）器械的誤差：セオドライトのエンコーダ誤差

2)外部影響による誤差：温度変化による光波の遅延

3）個人的誤差：人的観測誤差（偏り）

(b)性質による分類

1）過失：大きな誤差、角度の読み間違い

2）系統誤差：システマチックな誤差、気差、球差

3）偶然誤差：観測値から過失、系統誤差を除いた後に依然として残る小さなばらつきの誤差

(c)数学上の定義による分類

1）真の誤差

$$\varepsilon = x - X \quad \ldots(1.1)$$

2）平均誤差

$$d = \frac{[|\varepsilon|]}{n} \approx \frac{[|v|]}{n} \quad \ldots(2.1)$$

ここで、n：測定回数

3）単位重さ当たりの標準偏差

$$\sigma = \lim_{n \to \infty} \sqrt{\frac{[\varepsilon\varepsilon]}{n}} \quad \ldots(2.2)$$

式(2.2)を予測する値は

$$\sigma = \sqrt{\frac{[vv]}{n-1}} \quad \ldots(2.3)$$

から計算される。

重み付き観測値の場合

$$\sigma = \sqrt{\frac{[pvv]}{n-1}} \quad \ldots(2.3')$$

4）**平均値の標準偏差（統計学では「標準誤差」という）**

$$\sigma_m = \sqrt{\frac{[vv]}{n(n-1)}} \quad \ldots(2.4)$$

重み付き観測値の場合

$$\sigma_m = \sqrt{\frac{[pvv]}{[p](n-1)}} \quad \ldots(2.4')$$

(問題①)標準偏差

次の式を作れ。

$$\sigma=\sqrt{\frac{[vv]}{n-1}}$$

ここで、[　]＝Σ、v：残差、n：測定値の数

（解説）

誤差$\varepsilon = x - X$…①

-)残差$v = x - \tilde{x}$…②

$$\varepsilon - v = \tilde{x} - X$$

または

$$\varepsilon = (\tilde{x} - X) + v \quad …③$$

これを平方すると

$$\varepsilon^2 = (\tilde{x} - X)^2 + 2v(\tilde{x} - X) + v^2 \quad \text{…④}$$

n個の観測を行えば、

$$\varepsilon_1^2 = (\tilde{x} - X)^2 + 2v_1(\tilde{x} - X) + v_1^2$$

$$\varepsilon_2^2 = (\tilde{x} - X)^2 + 2v_2(\tilde{x} - X) + v_2^2$$

….

$$+)\varepsilon_n^2 = (\tilde{x} - X)^2 + 2v_n(\tilde{x} - X) + v_n^2$$

$$[\varepsilon^2] = n(\tilde{x} - X)^2 + 2(\tilde{x} - X)[v] + [v^2] \quad \text{…⑤}$$

そこで、標準偏差の定義は$\sigma=\sqrt{[\varepsilon\varepsilon]/n}$,それぞれの観測値$x_1, x_2, \ldots, x_n$の標準偏差は$\sigma_1, \sigma_2, \ldots, \sigma_n$とすれば、平均値の標準偏差（標準誤差）$\sigma_m$は、平均値

$$\tilde{x}=\frac{[x]}{n}=\frac{x_1}{n}+\frac{x_2}{n}+\cdots+\frac{x_n}{n}$$

から

$$\sigma_m^2=\frac{\sigma_1^2}{n^2}+\frac{\sigma_2^2}{n^2}+\cdots+\frac{\sigma_n^2}{n^2}$$

通常$\sigma=\sigma_1=\sigma_2=\cdots=\sigma_n$で表されるので

$$\sigma_m^2=n\frac{\sigma^2}{n^2}=\frac{\sigma^2}{n}$$

$\therefore\ \sigma_m=\frac{\sigma}{\sqrt{n}}$ …⑥

また誤差あるいは残差の総和はゼロであるので、

$[v]=v_1+v_2+\cdots+v_n=0$ …⑦

そこで、式⑤で⑦を用い、nで割れば

$$[\varepsilon^2]=n(\tilde{x}-X)^2+2(\tilde{x}-X)+[v^2]$$

$\frac{[\varepsilon^2]}{n}=(\tilde{x}-X)^2+\frac{[v^2]}{n}=\sigma^2$…⑧

この式で$(\tilde{x}-X)^2$は平均値($\tilde{x}$)の分散なので

$(\tilde{x}-X)^2=\frac{\sigma^2}{n}$ …⑨

これを⑧に適用して$\frac{[\varepsilon^2]}{\mathrm{n}}=\frac{\sigma^2}{n}+\frac{[\mathrm{v}^2]}{\mathrm{n}}=\sigma^2$

$$\frac{\sigma^2}{\mathrm{n}}+\frac{[\mathrm{v}^2]}{\mathrm{n}}=\frac{n\sigma^2}{n}$$

又は

$$\frac{[\mathrm{v}^2]}{\mathrm{n}}=\frac{n\sigma^2}{n}-\frac{\sigma^2}{\mathrm{n}}=\left(\frac{\mathrm{n}-1}{\mathrm{n}}\right)\sigma^2$$

$$\sigma^2=\left(\frac{\mathrm{n}}{\mathrm{n}-1}\right)\frac{[\mathrm{v}^2]}{\mathrm{n}}$$

$$\therefore\ \sigma^2=\frac{[\mathrm{v}^2]}{\mathrm{n}-1}$$

$\sigma=\sqrt{\frac{[\mathrm{vv}]}{\mathrm{n}-1}}$…⑩

偶然誤差の法則

偶然誤差とは、測定値から過失、定誤差、系統誤差を取り除いた後、依然として残る小さなバラツキ誤差である。

そこで、偶然誤差の法則は、次のとおりとする。

(1)正の誤差と負の誤差は等偶然に起こり、同じ大きさの観測値の度数は等しい。

(2)大数の観測において、その誤差は限界をもち、観測値は±ℓ間に存在する。

(3)測定値から誤差を取り除いた値が最確値である。

•

図1.1のような左右対称な曲線をガウスの正規分布曲線(Gaussian Normal Distribution Curve)という。また、誤差密度曲線とも呼ばれる。その曲線式は

$$y = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\varepsilon}{\sigma}\right)^2} \quad \ldots(3.1)$$

$h = {}^{1}\!/_{\sqrt{2}\sigma}$とおけば

$$y = \frac{h}{\sqrt{\pi}} e^{-h^2\varepsilon^2} \ldots(3.2)$$

と表すことができる。

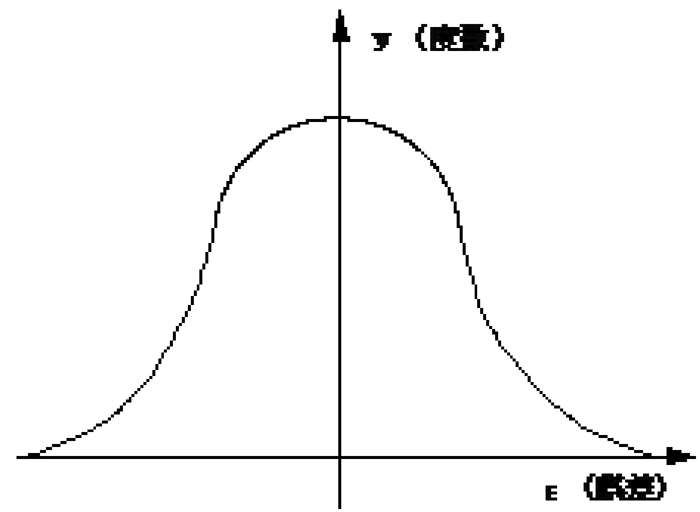
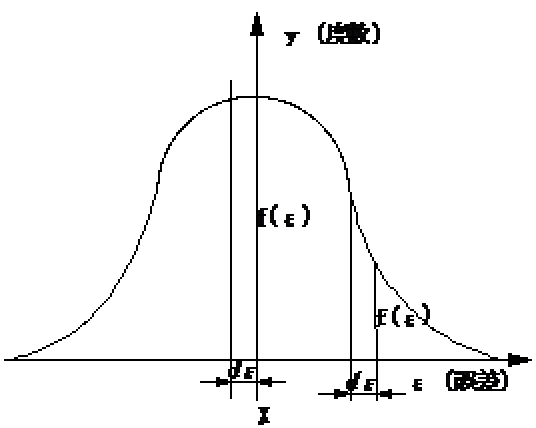


図1.1

正規分布曲線の誘導

（問題）

$y = \frac{h}{\sqrt{\pi}} e^{-h^2 \varepsilon^2}$ の式を誘導しよう。



真の誤差をε、それが起こる確率をf(ε)とすると

$y = f(\varepsilon)$ …①

と書ける。n個の測定値に関して、それぞれの誤差の起こる確率は,

$f(\varepsilon_1), f(\varepsilon_2), \dots, f(\varepsilon_n)$…②

また誤差はそれぞれ

$\varepsilon_1 = x_1 - X$

$\varepsilon_2 = x_2 - X$

……

$\varepsilon_n = x_n - X$…③

式②から誤差$\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$がそれぞれ同時に起こる確率Pは

$$P = f(\varepsilon_1) f(\varepsilon_2) \dots f(\varepsilon_n) \dots ④$$

自然対数をとると

$$\ln P = \ln f(\varepsilon_1) + \ln f(\varepsilon_2) + \cdots + \ln f(\varepsilon_n) \dots ⑤$$

真値Xの最確値は確率Pを最大にするから

$$\frac{d(\ln P)}{dX} = \frac{d(\ln f(\varepsilon_1))}{d\varepsilon_1} \cdot \frac{d\varepsilon_1}{dX} + \frac{d(\ln f(\varepsilon_2))}{d\varepsilon_2} \cdot \frac{d\varepsilon_2}{dX} + \cdots + \frac{d(\ln f(\varepsilon_n))}{d\varepsilon_n} \cdot \frac{d\varepsilon_n}{dX} = 0 \dots ⑥$$

$$\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_n = [\varepsilon] = 0 \dots ⑦$$

誤差εをXで微分すると

$$\frac{d\varepsilon_1}{dX} = \frac{d\varepsilon_2}{dX} = \cdots = \frac{d\varepsilon_n}{dX} = -1 \dots ⑧$$

式⑥に式⑧を代入すると

$$\frac{d(\ln P)}{dX} = \frac{d(\ln f(\varepsilon_1))}{d\varepsilon_1} + \frac{d(\ln f(\varepsilon_2))}{d\varepsilon_2} + \cdots + \frac{d(\ln f(\varepsilon_n))}{d\varepsilon_n} = 0 \dots ⑨$$

式⑨を次のように変形する。

$$\varepsilon_1 \cdot \frac{d(\ln f(\varepsilon_1))}{\varepsilon_1 \cdot d\varepsilon_1} + \varepsilon_2 \cdot \frac{d(\ln f(\varepsilon_2))}{\varepsilon_2 \cdot d\varepsilon_2} + \cdots + \varepsilon_n \cdot \frac{d(\ln f(\varepsilon_n))}{\varepsilon_n \cdot d\varepsilon_n} = 0 \dots ⑩$$

式⑦と式⑩は同時に満足すべきなので、

$$\frac{d(\ln f(\varepsilon_1))}{\varepsilon_1 \cdot d\varepsilon_1} = \frac{d(\ln f(\varepsilon_2))}{\varepsilon_2 \cdot d\varepsilon_2} = \cdots = \frac{d(\ln f(\varepsilon_n))}{\varepsilon_n \cdot d\varepsilon_n} = \text{一定} \ldots ⑪$$

$$\frac{d(\ln f(\varepsilon))}{\varepsilon \cdot d\varepsilon} = k \ldots ⑫$$

$$\frac{d(\ln f(\varepsilon))}{d\varepsilon} = \varepsilon k \ldots ⑫'$$

とおき、式⑫'を積分すると

$$\int \frac{d(\ln f(\varepsilon))}{d\varepsilon} \cdot d\varepsilon = \int k\varepsilon \cdot d\varepsilon$$

または

$$\ln f(\varepsilon) = \frac{k}{2}\varepsilon^2 + c = Ce$$

又は

$$f(\varepsilon) = e^{\frac{k}{2}\cdot\varepsilon^2 + c} = Ce^{\frac{k}{2}\varepsilon^2} \ldots ⑬$$

また、誤差εが増加すれば、その確率f(ε)は小さくなるので、

$$\frac{k}{2} = -h^2$$

とおけば、

$$f(\varepsilon) = Ce^{-h^2\varepsilon^2} \ldots ⑭$$

誤差曲線の確率面積は

$\int_{-\infty}^{+\infty} f(\varepsilon)\cdot d\varepsilon = 1$…⑮

であり、$f(-\varepsilon) = f(\varepsilon)$であることから、

$\int_{-\infty}^{+\infty} f(\varepsilon)\cdot d\varepsilon = \int_{-\infty}^{+\infty} C\cdot e^{-h^2\varepsilon^2} d\varepsilon = 2C\int_{0}^{+\infty} e^{-h^2\varepsilon^2}\, d\varepsilon = 1$

…⑯

または

$$\frac{1}{2C} = \int_{0}^{+\infty} e^{-h^2\varepsilon^2}\, d\varepsilon = \int_{0}^{\infty} e^{-t^2}\, \frac{dt}{t}$$

ここで、$h\varepsilon = t, hd\varepsilon = dt$とおいた。

$\frac{h}{2C} = \int_{0}^{+\infty} e^{-h^2\varepsilon^2}\, d\varepsilon = \int_{0}^{\infty} e^{-t^2} dt$…⑰

そこでこの定積分を求めるには、

$$\int_0^{\infty} e^{-t^2} dt = \int_0^{\infty} e^{-v^2} dv = A$$

ここで、v=tu、あるいはdv=t.duなので

$$A^2 = \int_0^{\infty} \int_0^{\infty} e^{-(t^2+v^2)}\, dv \cdot dt$$

$$= \int_0^{\infty} du \int_0^{\infty} e^{-t^2(1+u^2)} t \cdot dt$$

そこで

$$\int_0^{\infty} e^{-t^2(1+u^2)} t \cdot dt = \left[\frac{e^{-t^2(1+u^2)}}{2(1+u^2)}\right]_0^{\infty} = \frac{1}{2(1+u^2)}$$

なので、

$$A^2 = \frac{1}{2}\int_0^{\infty} \frac{du}{1+u^2} = \frac{1}{2}(\tan^{-1}\infty - \tan^{-1}0) = \frac{\pi}{4}$$

そして

$$\int_0^{\infty} e^{-t^2} dt = \int_0^{\infty} e^{-v^2} dv = \frac{\sqrt{\pi}}{2} \cdots ⑱$$

を得る。式⑰と式⑱は等しいので、

$$\frac{h}{2C} = \frac{\sqrt{\pi}}{2}$$

又は

$$C = \frac{h}{\sqrt{\pi}} \quad \text{....⑲}$$

これを式⑭に代入すると

$$f(\varepsilon) = \frac{h}{\sqrt{\pi}} e^{-h^2\varepsilon^2} \quad \text{...⑳}$$

を得る。

標準偏差

誤差曲線式は$y = \frac{h}{\sqrt{\pi}} e^{-h^2 \varepsilon^2}$であり、これを$y = \frac{h}{\sqrt{\pi}} e^{-h^2 x^2}$とおき、xに関して微分すると

$\frac{dy}{dx} = -2h^2 xy$…(1.4.1)

を得る。すなわち、その曲線の接線を示しているから、

x=0のとき $\frac{dy}{dx} = 0$

y=0のとき $\frac{dy}{dx} = 0$

x=0のとき、その接線はx軸（すなわち誤差軸）に平行であり、y=0のときその接線はx軸に平行になることを表している。

そこで、$\frac{d^2y}{dx^2}$を求めると

$\frac{d^2y}{dx^2} = 2h^2 y(2h^2 x^2 - 1)$…(1.4.2)

すなわち$\frac{d^2y}{dx^2}=0$はその曲線の変曲点を表すので、y=0、又は$2h^2x^2-1=0$から、

$$x=\pm\frac{1}{h\sqrt{2}}....(1.4.3)$$

また$h=\frac{1}{\sqrt{2}\sigma}$なので、

$$x=\pm\frac{1}{\frac{1}{\sqrt{2}\sigma}\cdot\sqrt{2}}=\pm\sigma$$

...(1.4.4)

つまり、変曲点の位置は標準偏差の値に等しい位置になる。

標準偏差は誤差の二乗の総和の平均であることから、平均二乗誤差(mean square error)と呼ばれ、その標準偏差の二乗は分散(variance)である。

ヒストグラム

同じ対象で一定の分類できる事象の起こる度数と事象との関係はヒストグラムで表すのが最も有力な方法である。

表2-1は一つのクラスの個人個人の体重測定を行った結果、最低の体重53kg、最高の体重85kgであった。そこで体重の級の幅(class interval)を決定しなければならないから

測定範囲＝86-53＝33kg

表2-1

53	87	85	60	78	70	68
55	75	83	57	63	74	79
58	81	76	61	65	64	80
68	70	62	73	65	68	65
62	73	68	81	72	71	64
65	67	67	69	71		

（級の数）

①10～25

②測定単位より大きくする

級の数10：33/10＝3.3

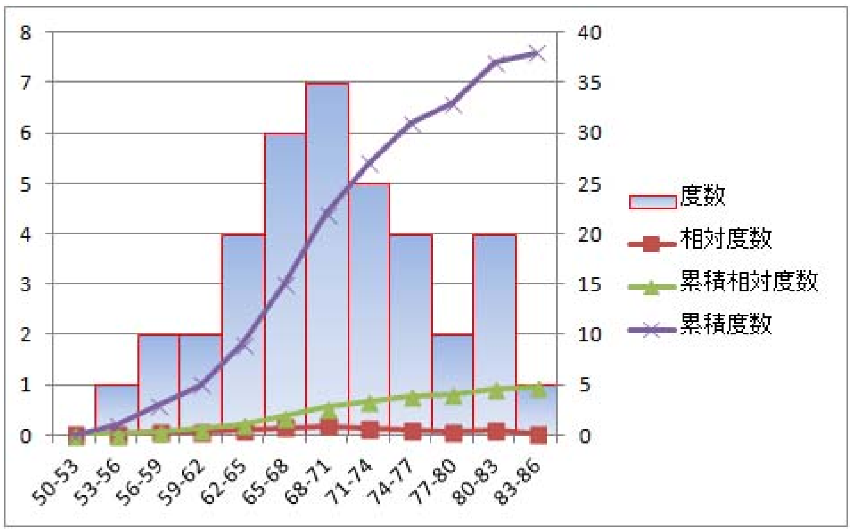
級の数25：33/25＝1.3

∴級の幅＝3.3kg

階級名

52-55,55-58、…、85-88

階級値	階級名	度数	累積度数	相対度数	累積相対度数
51.5	50-53	0	0	0	0
54.5	53-56	1	1	0.5	0.5
57.5	56-59	2	3	1	1.5
60.5	59-62	2	5	1	2.5
63.5	62-65	4	9	2	4.5
66.5	65-68	6	15	3	7.5
69.5	68-71	7	22	3.5	11
72.5	71-74	5	27	2.5	13.5
75.5	74-77	4	31	2	15.5
78.5	77-80	2	33	1	16.5
81.5	80-83	4	37	2	18.5
84.5	83-86	1	38	0.5	19
87.5	86-89	2	40		
	合計	40			



（問）表2.1の体重のヒストグラムを用いて、2000人の学生の70kg以下の数を求めよ。

22/40×2000＝1100人

パラメータ

(1)位置の測定

- 算術平均

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \frac{[x]}{n}$$

- 重量平均

$$\bar{x} = \frac{w_1x_1 + w_2x_2 + \cdots + w_nx_n}{w_1 + w_2 + \cdots + w_n} = \frac{[wx]}{[w]}$$

(2)散布測定

ある観測値の値域はその集合の最大値と最小値の差で表される。

散布測定としては、分散、又はその平方根の標準偏差が用いられる。

$$\sigma^2 = \frac{[(x - \bar{x})^2]}{n - r}$$

ここで

σ^2 : 標準偏差(平均二乗誤差)の平方値

x : 観測値

$\bar{x}$: 平均値

n : 測定回数

r : 自由度（未知数の数）

平均値・メジアン・モード

算術平均値

一連の数$x_1,x_2,\ldots,x_n$の平均値（算術平均値）

$$\bar{x} = \frac{x_1+x_2+\cdots+x_n}{n} = \frac{\sum x}{n} \quad (1)$$

(例)8,3,5,12,10の算術平均

$$\bar{x} = \frac{8+3+5+12+10}{5}$$

$$= \frac{38}{5} = 7.6$$

重量平均

数$x_1,x_2,\ldots,xn$はそれぞれ$f_1,f_2,\ldots,f_n$回起こるならば、平均値は

$$\bar{x} = \frac{x_1 f_1+x_2 f_2+\cdots+x_n f_n}{f_1+f_2+\cdots+f_n} = \frac{\sum xf}{\sum f} \quad (2)$$

で計算される。

(例)データ5,8,6,2は、それぞれ3,2,4,1回起こる。

重量平均を求めよ。

$$\bar{x} = \frac{5 \times 3 + 8 \times 2 + 6 \times 1}{3 + 2 + 4 + 1} = \frac{57}{10} = 5.7$$

メジアン（中央値）

- 小さな数から大きな数へと順番に並べた一連の数のメジアン（中央値）は、奇数の場合その中央の値、偶数の場合二つの数の平均である。

（例1）3,4,4,5,6,8,8,8,10：

メジアン＝6

（例2）

5,5,7,9,11,12,15,18：

メジアン＝(9+11)/2＝10

（例3）次の40名の身長の級間（級幅）の級数と度数についてメジアンを求める。

身長cm	度数f	累計
118-126	3	3
127-135	5	8
136-144	9	17
145-153	12	29
154-162	5	
163-171	4	
172-180	2	
合計	40	

方法①　補間法

1）N/2＝40/2＝20：メディアンはf＝20にある。

2）最初の3つの級数の合計3+5+9＝17であり、

20-17＝3足りない。

3）これは4番目の級数の12の度数の内で3足りないので、

メジアン＝144.5+3/12(153.5-144.5)＝146.8

方法②　公式による

N/2=40/2=20にメジアンがある。

最初の3つの合計17

最初の4つの合計29 ⇨ ここにメジアンがある。

メジアンの最小境界値L_1＝144.5

全度数N=40

$(\Sigma f)_1$＝メディアンを含まない下位の級数の全度数＝3+5+9＝17

f_m＝メジアンの度数＝12

c＝メジアンの級幅＝153.5-144.5＝9

メジアン＝L_1+[(N/2-$(\Sigma f)_1$)/f_m]c…(4)

＝144.5+[(40/2-17)/12]×9＝144.5+2.25＝146.8

モード（最頻値）

（例a）
平均、メジアン、モードの計算

2,2,3,5,5,5,6,6,8,9

平均＝
$\frac{1}{10}\begin{pmatrix}2+2+3+5+5+5\\+6+6+8+9\end{pmatrix}=5.1$

メジアン＝2つの中央値の平均＝(5+5)/2＝5

モード＝最も多く現れる値＝5

(例b)平均、メジアン、モードの計算

51.6,48.7,50.3,49.5,48.9

平均＝
$\frac{1}{5}\begin{pmatrix}48.7+48.9+49.5\\+50.3+51.6\end{pmatrix}=49.8$

メジアン＝中央値＝49.5

モード＝存在しない

幾何平均

$\mathrm{G} = \sqrt[n]{x_1 x_2 \ldots x_n}$ ….(5)

（例）2,3,6の幾何平均

$\mathrm{G} = \sqrt[3]{2 \times 3 \times 6} = 3.3$

調和平均

$$\mathrm{H} = \frac{1}{\frac{1}{n}\sum \frac{1}{x}} = \frac{n}{\sum \frac{1}{x}}$$

$\frac{1}{\mathrm{H}} = \frac{\sum \frac{1}{x}}{n} = \frac{1}{n}\sum \frac{1}{x}$…(6)

(例)2,3,6の調和平均

$$\frac{1}{\mathrm{H}} = \frac{1}{3}\sum\left(\frac{1}{2} + \frac{1}{3} + \frac{1}{6}\right) = \frac{1}{3}$$

H=3

算術平均(平均)、幾何平均と調和平均の関係

$$\mathrm{H} \leq \mathrm{G} \leq \bar{\mathrm{x}}$$

(問題①)次の数値はデジタル図化機で測定したX座標値の集合を示す。平均値と標準偏差を求めよ。

観測値	140. 33	140.34	140.35	140.36	140.37	140.38	140.40 (m)
度数	1	0	0	5	9	9	2

Δx	w	Δx・w	v	vv	wvv
0.33	1	0.33	-0.04231	0.00179	0.00179
0.34	0	0	-0.03231	0.001044	0
0.35	0	0	-0.02231	0.000498	0
0.36	5	1.8	-0.01231	0.000151	0.000757
0.37	9	3.33	-0.00231	5.33E-06	4.79E-05
0.38	9	3.42	0.007692	5.92E-05	0.000533
0.4	2	0.8	0.027692	0.000767	0.001534
合計	26	9.68	-0.07615	0.004314	0.004662

重量平均

$$\bar{x} = 140\mathrm{mm} + \frac{[wx]}{[w]} = 140 + \frac{9.68}{26}$$
$$= 140..37\mathrm{mm}$$

分散

$$\sigma^2 = \frac{[wvv]}{n-1} = \frac{0.00466}{7-1}$$
$$= 0.00078\mathrm{mm}^2$$

標準偏差

$$\sigma = 2.8\mathrm{mm}$$

理論的確率分布

二項分布

任意の実験(trial)において、ある事象が起こる確率p、起こらない確率q=1-pならば、その事象は確実にn回の実験においてX回起こる。

- $\mathrm{p(X)} = \mathrm{n}\mathrm{C_x}\mathrm{p^x}\mathrm{q^{n-x}} = \frac{\mathrm{n!}}{\mathrm{X!(n-X)!}}\mathrm{p^x}\mathrm{q^{n-x}}$

ここで、

$\mathrm{X} = 1,2,\ldots,\mathrm{n}$

$\mathrm{n!} = \mathrm{n(n-1)(n-2)}\ldots 1$

$0! = 1$

- 上の式は17世紀後半にJames Bernoulliによって発見されたベルヌーイ分布である。

(問題②)

- 銅貨の6回の投において、正確に2個の表の出る確率を求めよ。

- $p = q = \frac{1}{2}$
- n=6, X=2なので
- $6C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{6-2} = \frac{6!}{2!4!}\left(\frac{1}{2}\right)^6 = \frac{15}{64}$

- また、ベルヌーイは起こる確率p、そうでない確率をqとするとき、n回の試みならば
- $(p+q)^n = p^n + np^{n-1}q + \frac{n(n-1)}{2!}p^{n-2}q^2 + \cdots + q^n$

それぞれ(p,q)の係数はPascal’s triangle(パスカルの三角数)で表される。

n	$(\mathrm{p}+\mathrm{q})^{\mathrm{n}}$の係数
1	1 1
2	1 2 1
3	1 3 3 1
4	1 4 6 4 1
5	1 5 10 10 5 1
6	1 6 15 20 15 6 1
7	1 7 21 35 35 21 7 1
8	1 8 28 56 70 56 28 8 1

…

(問題③)4人の子供をもつ家庭について

(a)少なくとも1人が男子、(b)少なくとも1人が男子で1人が女子である確率を求めよ。ただし、男子、女子の生まれる確率は1/2とする。

$$(p+q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$$

$$= \frac{1}{16} + \frac{1}{4} + \frac{3}{8} + \frac{1}{4} + \frac{1}{16}$$

- $\frac{1}{16}$:男４人、$\frac{1}{4}$:男３人女１人、$\frac{3}{8}$: 男２人女２人、$\frac{1}{4}$:男１人女３人、$\frac{1}{16}$:男０人女４人

(a)少なくとも1人が男子

$$\frac{1}{16}+\frac{1}{4}+\frac{3}{8}+\frac{1}{4}=\frac{1+4+6+4}{16}$$

$$=\frac{15}{16}$$

(b)少なくとも1人が男子で1人が女子

$$\frac{1}{4}+\frac{3}{8}+\frac{1}{4}=\frac{7}{8}$$

(問題④)4人の家族である2000世帯の家庭について

(a)少なくとも1人が男子、(b)2人が男子

(c)2人が女子、(d)女子なしは、それぞれ何世帯であるか推定せよ。

$$(\mathrm{p}+\mathrm{q})^4$$
$$=\frac{1}{16}+\frac{1}{4}+\frac{3}{8}+\frac{1}{4}+\frac{1}{16}$$

(a)少なくとも1人が男子

$2000\left(\frac{1}{16}+\frac{1}{4}+\frac{3}{8}+\frac{1}{4}\right)=$

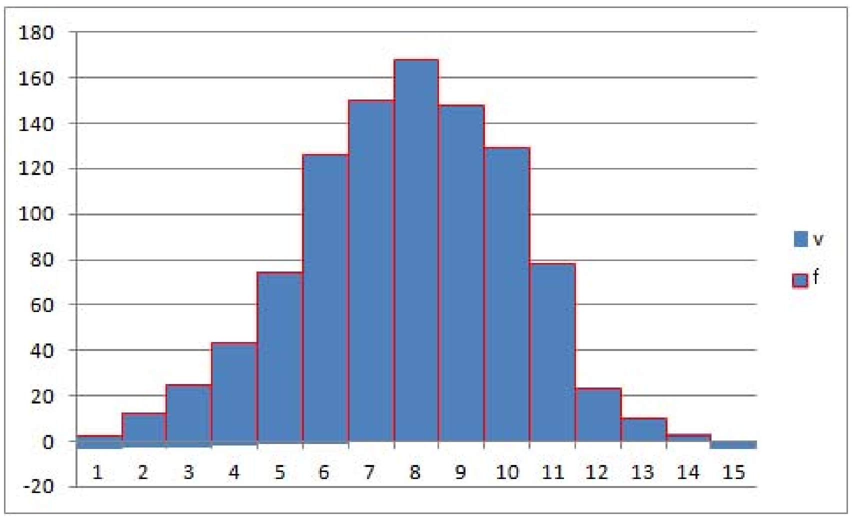
$2000\cdot\frac{15}{16}=1875$ 世帯

(b)$2000\cdot\frac{3}{8}=750$世帯

(c) $2000\cdot\frac{3}{8}=750$世帯

(d) $2000\cdot\frac{1}{16}=125$世帯

正規分布

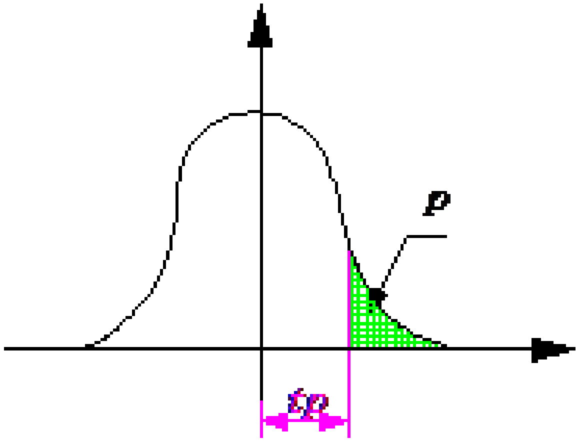


- この曲線をスムージングすると、正規分布曲線になる。その平均値から変曲点までの長さが標準偏差σに一致する。
- また曲線の式は、
- $y = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2}$

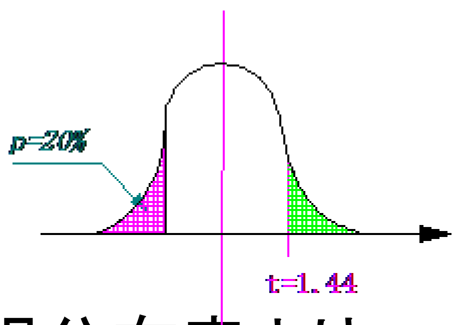
そこで、$\bar{x} = 0$すなわち誤差に関する正規分布曲線における全面積を1(100%)とし、σ=1と考えるならば、これは標準正規分布(standard normal distribution)であるから、

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$$

ここで、$t = \frac{x-\bar{x}}{\sigma}$



(問題①)t＝1.44より上方の確率及びp=20%の境界値を求めよ。



正規分布表より

t＝1.44⇒p＝0.0749

p＝7.49%

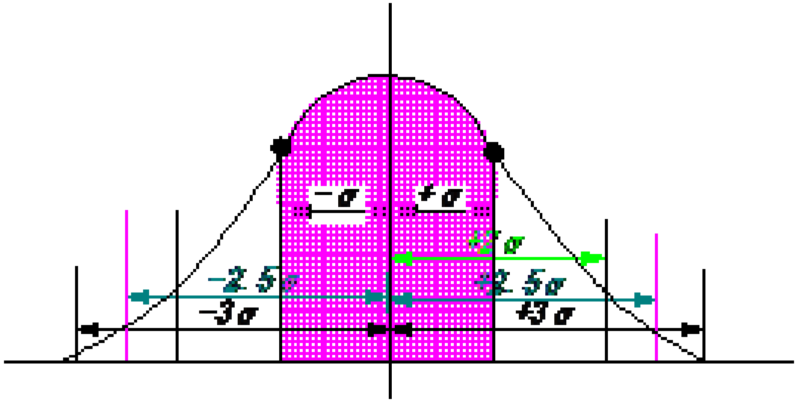
p1＝20.05⇒t1＝0.84

p2＝19.77⇒t2＝0.85

Δp=0.28、Δt=0.01

- t＝0.84-0.01/0.28×0.05
  ＝0.838

- **(問題②)**t=1、2、2.5、3のそれぞれの確率面積を求めよ。
- t=1→p=15.87%　1σ=100%-2×15.87%＝68.26%
- t=2→p=2.28%　　2σ＝100%-2×2.28%＝95.44%
- t=2.5→p=0.62%　2.5σ＝100%-2×0.62%＝98.76%
- t=3→p=0.13%　3σ＝100%-2×0.13%＝99.74%

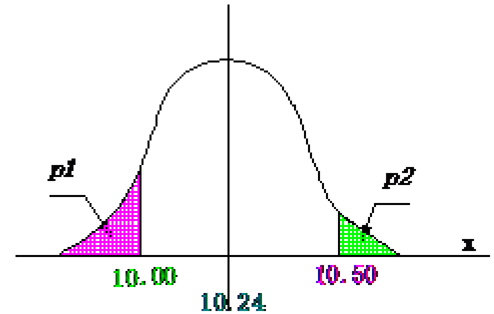


(問題③)正規分布に従う
200個のx視差観測を行ったところ、標準偏差σ=0.12mm、平均値$\bar{x} = 10.24\mathrm{mm}$である。

(1)x=10.00mmと10.50mm間の測定値の数

(2)10.30mmを超える観測値の数

(3)観測値の90%以下の下方限界値を推定せよ。



(1)$t = \frac{x - \bar{x}}{\sigma}$

$t_1 = \frac{10.00 - 10.24}{0.12} = -2.0$

→p=2.28%

$t_2 = \frac{10.50 - 10.24}{0.12} = 2.17$

→p=0.15%

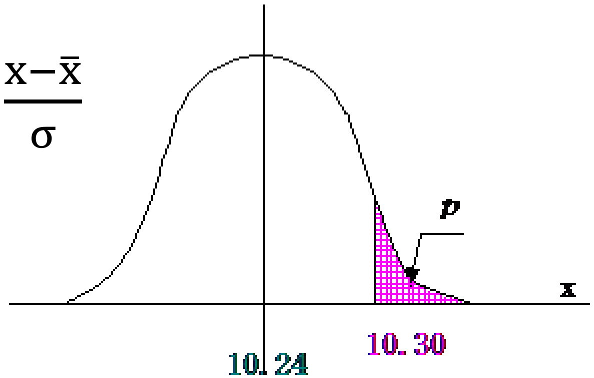
したがって、その境界内の面積は

100%-(2.28+0.15)%＝96.22%

であり、その範囲の観測値の数は

∴0.9622×200＝192個

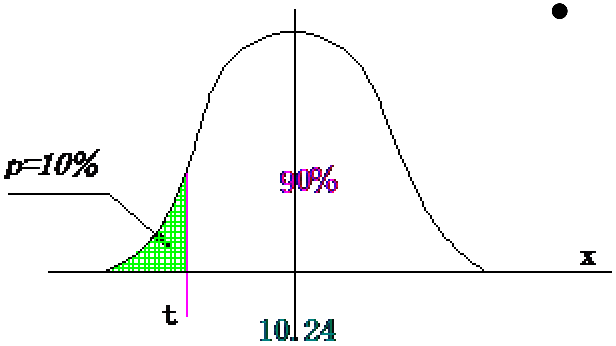
(２) $t=\frac{x-\bar{x}}{\sigma}$



$t=\frac{10.30-10.24}{0.12}=0.50$ →p＝30.85%

したがって、観測値の数は、

0.3085×200＝62個



(3) 観測値の90%以下の下方限界値を推定せよ。

p=-10%→t=-1.28

t	p
1.28	10.03%
1.29	9.85%
差0.01	0.18

t=1.28+0.01×0.03/0.18＝1.28

$$t=\frac{x-\bar{x}}{\sigma}$$

- $x=\bar{x}-t\sigma=10.24-1.28\cdot 0.12=10.09\mathrm{mm}$

(問題④)街路に1000個の防犯灯をつける。そのLEDの平均寿命は4万時間であり、その標準偏差は8000時間とする。

(a)LED灯を取り付けた後1万時間で何個が切れるか？

(b)LED灯の10%が切れるのは何時間後か？

(a)

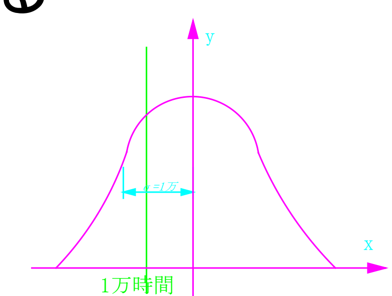
$$t = \frac{x - \bar{x}}{\sigma}$$

$$= \frac{10000 - 40000}{10000}$$

$$= -3.0$$

→p=0.13%

∴10000×0.0013＝13個



(b) 水銀灯の10%が切れるのは何時間後か？

p=-0.1000→t＝-1.28

1003→1.28

$$t = \frac{x - \bar{x}}{\sigma}$$

$x = \bar{x} + t\sigma = 4万 - 1.28 \times 1万 = 2.7万$時間後

(問題⑤)作業規程の準則より検査高の99%が主曲線間隔の1/4以内にその誤差が超えないように規定されているとき、等高線間隔1mの地形図を作成するには、最低いくらの精度にすればいいか？

$$t = \frac{x - \bar{x}}{\sigma}$$

$$x - \bar{x} = t\sigma = \Delta$$

$$\Delta = \frac{1}{4} \times 1\mathrm{m} = 0.25\mathrm{m}$$

標準正規分布表より

p=0.5%→t=3.32

$$\therefore \sigma = \frac{\Delta}{t} = \frac{0.25\mathrm{m}}{3.32}$$

$$= 0.075\mathrm{m}$$

(問題⑥)

5mmの精度の光波測距儀で、

(a)2mm、(b)3mm、(c)5mm

の精度を得るには、最低何回測定すればいいか。

(a)σ =5mm,$\sigma_{\mathrm{m}} = 2\mathrm{mm}$より

$$\sigma_{\mathrm{m}} = \frac{\sigma}{\sqrt{\mathrm{n}}}$$

$$\mathrm{n} = \frac{\sigma^2}{\sigma_{\mathrm{m}}^2} = \frac{25}{4} = 7(回)$$

(b) $\mathrm{n} = \frac{\sigma^2}{\sigma_{\mathrm{m}}^2} = \frac{25}{9} = 3(回)$

(c) $\mathrm{n} = \frac{\sigma^2}{\sigma_{\mathrm{m}}^2} = \frac{25}{25} = 1(回)$

(問題⑦)

図化機の実験で次のようなx座標の観測値を75個得た。(a)最小値と最大値のt値を求めよ。

(b)3σ=99.74%のt値を許容値とすれば、どの値を棄却するか？

座標(μm)	75	83	87	88	89	90	91	92	93	94	95	96	98	99
度数	1	2	2	4	6	9	10	12	9	7	5	5	2	1

番号	x	f(度数)	xf	v	vv	fvv
1	75	1	75	-16.6	275.56	275.56
2	83	2	166	-8.6	73.96	147.92
3	87	2	174	-4.6	21.16	42.32
4	88	4	352	-3.6	12.96	51.84
5	89	6	534	-2.6	6.76	40.56
6	90	9	810	-1.6	2.56	23.04
7	91	10	910	-0.6	0.36	3.6
8	92	12	1104	0.4	0.16	1.92
9	93	9	837	1.4	1.96	17.64
10	94	7	658	2.4	5.76	40.32
11	95	5	475	3.4	11.56	57.8
12	96	5	480	4.4	19.36	96.8
13	98	2	196	6.4	40.96	81.92
14	99	1	99	7.4	54.76	54.76
合計		75	6870	-12.4	527.84	936

$$\bar{x} = \frac{[fx]}{[f]} = \frac{6870}{75} = 91.6$$

$$\sigma^2 = \frac{[fvv]}{n-1} = \frac{936}{14-1}$$

$$= 72$$

$$\sigma = 8.5$$

(a)$t_1 = \frac{75-91.6}{8.5} = -1.95$

$$t_2 = \frac{99-91.6}{8.5} = 0.87$$

(b)有意水準$\bar{x} \pm 3\sigma$

$$\bar{x} \pm 3\sigma = 91.6 \pm 3 \times 8.5$$

$$= 66.1 - 117.1$$

∴棄却するデータはない。

多次元観測

- 未知数（真値）の個数が1個ならば一次元観測、2個ならば二次元観測である。したがって、二次元以上の未知数をもつ観測を多次元観測と言う。
- また、多次元観測においても、一次元観測と同じ標準偏差をもつが、「相関」という現象が起こってくる。つまり、異次元観測値間で相関するパラメータが存在することになる。

相関

変数(x、y)間の相関は確率変数間の相互の関係を描くものであって、それは共分散(co-variance)によって表される。

共分散

$$\sigma_{xy} = \frac{1}{n-r}[(x-\bar{x})(y-\bar{y})]$$

分散

$$\sigma x^2 = \frac{1}{n-r}[(x-\bar{x})^2]$$

$$\sigma y^2 = \frac{1}{n-r}[(y-\bar{y})^2]$$

相関係数

２つのデータ（変数）が、かなりの程度の**規則性**をもって、**同時に変化していく性質**を相関という。

相関を示す度合は、次の相関係数rで表される。

$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$$

または

$$r_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma x^2 \sigma y^2}}$$

$$= \frac{[(x-\bar{x})(y-\bar{y})]}{\sqrt{[(x-\bar{x})^2][(y-\bar{y})^2]}}$$

$$= \frac{[v_x v_y]}{\sqrt{[v_x^2][v_y^2]}}$$

r_{xy}は-1～+1の間に存在する。

F.N.Davidによる相関係数

観測数	3	4	5	6	7	8	10	12	15	20	25
r	0.99	0.95	0.88	0.82	0.75	0.7	0.63	0.58	0.52	0.45	0.39
観測数	50	100	200	400	∞						
r	0.28	0.2	0.14	0.1	0.01						

母集団の相関係数ρ=0で5%有意水準において、観測数nと最小標本相関係数を与えている。

(問題①)

- ある100回測定したXYZ座標の集合から抽出した標本を次表に示す。それぞれの相関係数を求め、5%有意水準において観測値が信頼域にあるかどうか判定せよ。

番号	座標		
	X	Y	Z
10	21	39	73
20	24	46	78
30	20	38	74
40	21	43	76
50	18	37	71
60	21	44	80
70	22	41	77
80	18	39	73
90	23	41	75
100	19	40	76

（単位μm）

番号	座標			vx	vy	vz	vx vy	vy vz	vz vx	vx vx	vy vy	vz vz
	X	Y	Z									
10	21	39	73	0.3	−1.8	−2.3	−0.54	4.14	−0.7	0.09	3.24	5.29
20	24	46	78	3.3	5.2	2.7	17.16	14	8.91	10.89	27	7.29
30	20	38	74	−0.7	−2.8	−1.3	1.96	3.64	0.91	0.49	7.84	1.69
40	21	43	76	0.3	2.2	0.7	0.66	1.54	0.21	0.09	4.84	0.49
50	18	37	71	−2.7	−3.8	−4.3	10.26	16.3	11.6	7.29	14.4	18.5
60	21	44	80	0.3	3.2	4.7	0.96	15	1.41	0.09	10.2	22.1
70	22	41	77	1.3	0.2	1.7	0.26	0.34	2.21	1.69	0.04	2.89
80	18	39	73	−2.7	−1.8	−2.3	4.86	4.14	6.21	7.29	3.24	5.29
90	23	41	75	2.3	0.2	−0.3	0.46	−0.1	−0.7	5.29	0.04	0.09
100	19	40	76	−1.7	−0.8	0.7	1.36	−0.6	−1.2	2.89	0.64	0.49
合計	207	408	753	0	0	0	37.4	58.6	28.9	36.1	71.6	64.1

$$\bar{x} = \frac{207}{10} = 20.7\mu m$$

$$\bar{y} = \frac{408}{10} = 40.8\mu m$$

$$\bar{z} = \frac{753}{10} = 75.3\mu m$$

$$r_{xy} = \frac{[v_x v_y]}{[v_x^2][v_y^2]} = \frac{37.4}{36.1 \times 71.6} = 0.74 > 0.63$$

$$r_{yz} = \frac{[v_y v_z]}{[v_y^2][v_z^2]} = \frac{58.4}{71.6 \times 64.1} = 0.86 > 0.63$$

$$r_{zx} = \frac{[v_z v_x]}{[v_z^2][v_x^2]} = \frac{28.9}{64.1 \times 36.1} = 0.60 < 0.63$$

- したがって、X,Y,Zにおける5%有意水準においてZ,X間の相関係数のみが有意でないことを示しているが、XYやYZ間の関係が一概に悪いわけではない。上記の例はステレオ図化機での観測結果であり、測標（メスマーク）を測点に合わせる際、無作為性を保つことが難しいことを示す例である。

観測値の表現

- 観測値はそのまま利用することはなく、何らかの方法で最確値を求め、これを利用するか、又は何らかの条件が適用できる場合、それを用いて調整して、最終的な最確値とする。
- また、測定した値や最確値は必ずその信頼性を示す誤差(標準偏差)や精度を添付して表現することが必要である。

(1)相加平均

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \frac{[x]}{n}$$

(2)測定値の標準偏差

$$\sigma = \sqrt{\frac{[vv]}{n-1}}$$

(3)平均値の標準偏差(標準誤差)

$$\sigma_m = \sqrt{\frac{[vv]}{n(n-1)}} = \frac{\sigma}{\sqrt{n}}$$

(4)母平均

$$\mu \approx \mu_m = \frac{1}{n}(\tilde{x}_1 + \tilde{x}_2 + \cdots + \tilde{x}_n) = \frac{[\tilde{x}]}{n}$$

(5)標準偏差の標準偏差

$$\sigma_s = \frac{\sigma}{\sqrt{2n}}$$

(問題①)

- 異なる3台のTS(トータルステーション）で距離をそれぞれ50個ずつ観測し、次表の結果を得た。観測値の99.74%信頼水準における平均値の精度と標準偏差の標準偏差を求めよ。

TS	観測数	平均値	標準偏差
A	50	38ｍｍ	7ｍｍ
B	50	44	10
C	50	46	12

p=99.74%→(100%-99.74%)/2＝0.13→標準正規分布表よりt=3
平均値の信頼限界$\tilde{x} \pm t\sigma = \tilde{x} \pm 3\sigma$で表され、標準偏差の信頼限界も$\sigma \pm 3\sigma_s$で表される。
TS-A

$$\sigma_m = \sigma/\sqrt{n} = 7/\sqrt{50} = 0.99mm$$

$$\sigma_s = \sigma/\sqrt{2n} = 7/\sqrt{100} = 0.7\mathrm{mm}$$

$$\tilde{x} \pm 3\sigma = 38 \pm 2.97 = 35.03, 40.97\mathrm{mm}$$

$$\sigma \pm 3\sigma_s = 7 \pm 2.1 = 4.9, 9.1\mathrm{mm}$$

TS-B

$$\sigma_m = \sigma/_{\sqrt{n}} = 10/_{\sqrt{50}} = 1.41\text{mm}$$

$$\sigma_s = \sigma/_{\sqrt{2n}} = 10/_{\sqrt{100}} = 1.0\text{mm}$$

$$\tilde{x} \pm 3\sigma = 44 \pm 4.23 = 39.77,48.23\text{mm}$$

$$\sigma \pm 3\sigma_s = 10 \pm 3.0 = 7.0,13.0\text{mm}$$

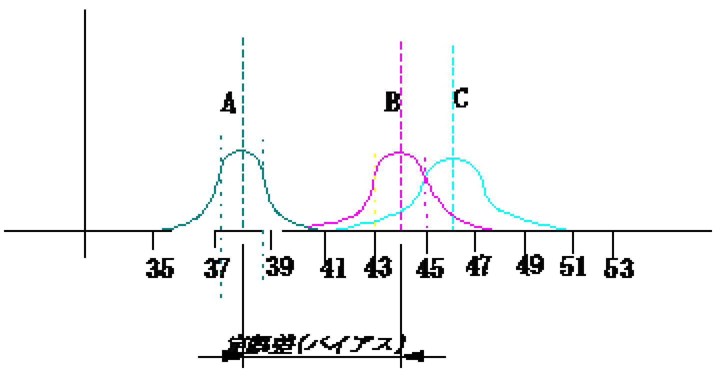
TS-合C

$$\sigma_m = \sigma/_{\sqrt{n}} = 12/_{\sqrt{50}} = 1.70\text{mm}$$

$$\sigma_s = \sigma/_{\sqrt{2n}} = 12/_{\sqrt{100}} = 1.2\text{mm}$$

$$\tilde{x} \pm 3\sigma = 46 \pm 5.1 = 40.9,51.1\text{mm}$$

$$\sigma \pm 3\sigma_s = 12 \pm 3.6 = 8.4,15.6\text{mm}$$



図から、TS-BとCはほぼ同族であるが、TS-Aには系統誤差（バイアス）が含まれることが推測できる。

(註)有意水準(Level of significant)と信頼水準(confidence level)

10%(90%)→たぶん有効

5%(95%)→有効

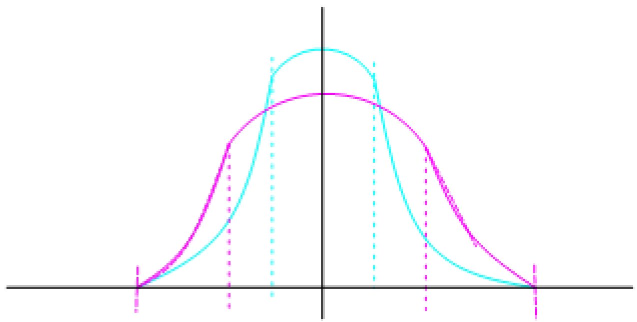
1%(99%)～0.1%(99.9%)→高い有効性

(　)内は信頼水準

検定

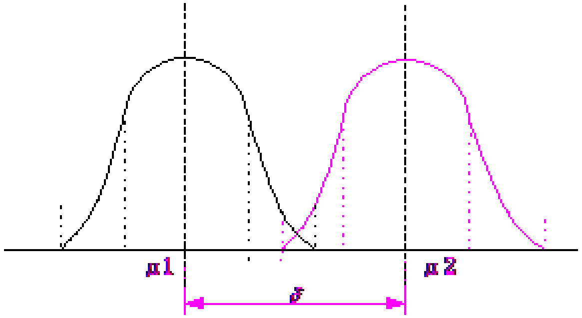
- 実際の問題において、母集団観測値の分析を、標本を基にして行わなければならない場合がある。
- 観測値の抽出標本は、確率変動をもつか、系統変動をもつかのいずれかであり、その2つについて母集団を予測する。

(1)確率変動



- それぞれの標本集合の平均値は同じであるが、その度数に変化がある。つまり、標準偏差が異なる集合であるとき、確率変動があるという。

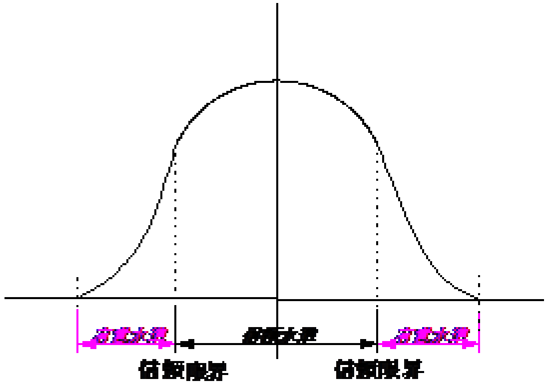
(2)系統変動(systematic variation)



- 標本集合の標準偏差は同じであるが、それぞれの平均値が異なった観測値の分布状態に場合、これを系統変動があるという。すなわち、平均値の差は$\delta = \mu_2 - \mu_1$である。

- そこで、検定により、δが有意かどうかを判定するのであるが、そこには確率変動と系統変動が存在する。そこで仮説を立てる必要性がある。
- 一つの仮説は**帰無仮説**(null hypothesis)であり、これを棄却にする基準は有意水準(α = 0.1%,1%,5%)で、この仮説は初期仮説である。
- もう一つの仮説は**対立仮説**(alternative hypothesis)である。
- 信頼水準(1-α)は帰無仮説が真実であるとき許される。(すなわち、(1-α)＝99.9%、99%、95%)

- 相対的に系統誤差をもたない確率の最大確率をβ(力関数)とし、(1-β)が存在するとき、系統的確率誤差がその集合に存在する。



U-検定

- U-検定は平均値の差が有意かどうかを検定するもので、母平均μと標本標準偏差σ、平均値x̃とを与えると、

$$U = \frac{\tilde{x} - \mu}{\sigma / \sqrt{n}} = \frac{\tilde{x} - \mu}{\sigma_m}$$

で表される。

- UとU$_\alpha$(標準正規分布曲線の面積表のt値)を比較して合格･不合格判定する。

- **(問題①)**母平均μ=45m、標本標準偏差σ=5m、標本平均x̃=49mが10回観測における一つの集合であるか、U-検定において5%有意水準で判定せよ。

$$\mathrm{U} = \frac{\tilde{\mathrm{x}} - \mu}{\sigma/\sqrt{\mathrm{n}}} = \frac{49 - 45}{5/\sqrt{10}}$$
$$= 2.5 > 1.64 (= \mathrm{U}_{\alpha} = t)$$

∴U>U_{α}なので帰無仮説を棄却して、対立仮説を採用し、有意差があると結論付ける。

スチューデントtテスト

母分散が既知の場合

母平均μ、母分散σ、標本平均$\bar{x}$とすると

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

これは平均0、分散1の標準正規分布に従うことを利用して検定する。帰無仮説はH0：$\bar{x} = \mu$である。
n：標本数

母分散が未知の場合、

$$t = \frac{(\tilde{x} - \mu)\sqrt{n-1}}{s}$$

また、スチューデントtの分布は対称であり、測定回数n≧30のとき有利である。

(問題③)
54人の男性の身長データの平均が175.8148cm，標準偏差は8.014であった.
身長の平均は170cmであると判断してもよいか．ただし，有意水準は5%とする.

$$t = \frac{175.8148 - 170}{8.014/\sqrt{54-1}} = 5.28 < 2.007$$ (t表より)

2.5%より

n	t
40	2.021
60	2.000

t=2.021-13×0.021/20＝2.007

- このとき，自由度53，有意水準5%のtの境界値は表より2.007なので，
「検定統計量(の絶対値)＞境界値」より，
帰無仮説は5%の有意水準で棄却される．これにより，対立仮説が採択され，このデータにおける身長の平均は170cmとは異なる，となります.

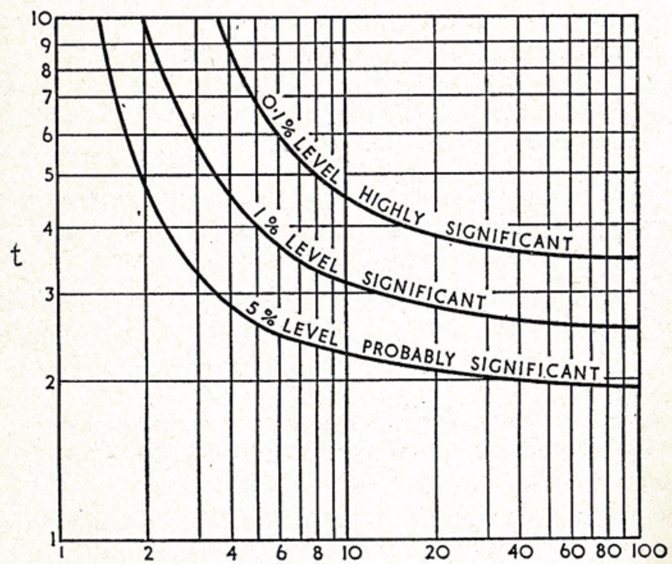


Fig. 81. Graphs of Student's t for 50%, 1%, and 0·1% significance level. If the calculated value for t is greater than the value shown above for the appropriate number of degrees of freedom, the indicated level of significance is reached*

カイ二乗検定(χ^2-テストChi-Square)

- χ^2-テストは非対称分布であり、30個以上の自由度の正規分布で、任意分布の期待度数が用いられる。

$$\chi^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \cdots + \frac{(o_n - e_n)^2}{e_n} = \sum_{j=1}^{n} \frac{(o_j - e_j)^2}{e_j}$$

ここで、

$o_1, o_2, \ldots, o_n$:測定度数

$e_1, e_2, \ldots, e_n$:事象$E_1, E_2, \ldots, E_n$の期待度数

χ^2-テストは経験的分布に一致する正規分布や二項分布のような理論的分布をテストするのに非常に有利である。

(問題④)

コインの40回の投において、表の度数を次のように観測した。ただし、3枚の銅貨を同時に投げたものとする。

- 有意水準1%、5%において合否判定を行え。ただし、χ^2分布を用いるものとする。

表の数	3	2	1	0
度数	9	11	19	1

二項分布で推測される表の数は5,15,15,5であるから

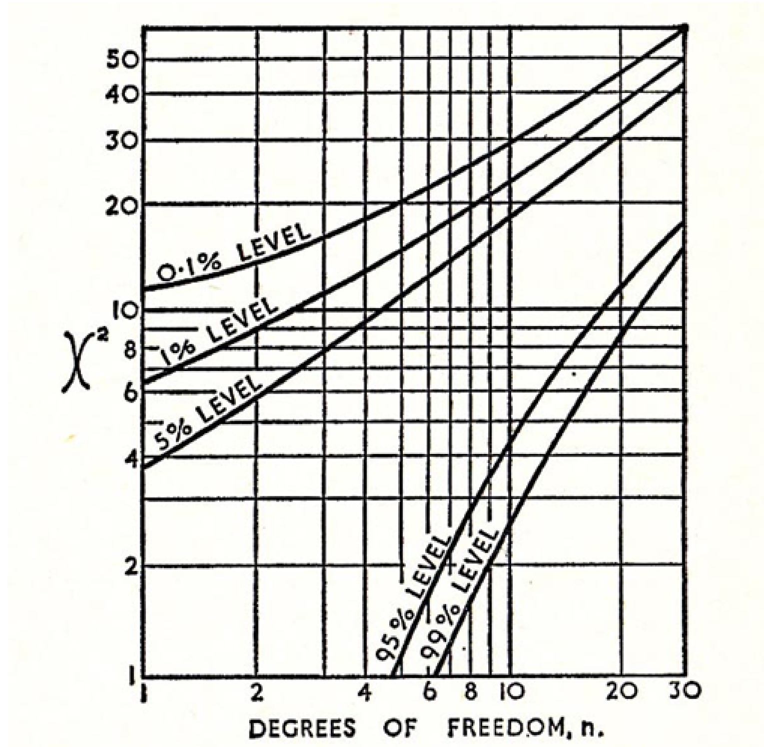
$$\chi^2 = \frac{(9-5)^2}{5} + \frac{(11-15)^2}{15} + \frac{(19-15)^2}{15} + \frac{(1-5)^2}{5} = 8.5$$

そこで、級は4個であり、
自由度は4-1＝3であるから、

$\chi^2|p = 5\%, n = 3 \rightarrow 7.8 < 8.5$

$\chi^2|p = 1\%, n = 3 \rightarrow 11 > 8.5$

$\therefore \chi^2|p = 5\%, n = 3 \rightarrow 7.8 < 8.5$ を不合格とする。



F-テスト(F-test)

- U-検定やスチューデントt分布は平均値に関して、同じ母集団に属しているかの検定であった。
- **F-検定(テスト)は標本分散間の差の有意性を決定するものである。**
- $\mathrm{F} = \frac{\text{比較的大の標準偏差}^2}{\text{比較的小の標準偏差}^2} = \frac{\sigma_{\mathrm{max}}^2}{\sigma_{\mathrm{min}}^2}$
- 通常、F値は1より大きい数になることは明らかである。

- **F検定(等分散の検定)**
  まず「等分散」とは、等しく分散しているということであり、それぞれの群の分布の形が似ているということである。
- 独立2群の差の検定の場合、2つの標本のt検定には「正規分布」「等分散」の二つの条件が必要である。そのため、たとえ正規分布していても等分散でなければ2つの標本t検定を使ってはいけない。

- この等分散かどうかを調べるためにF検定がある。２つの標本においてt検定をする前にF検定をして等分散であることを確認する必要がある。
- （**仮説**）
  帰無仮説(H_0)：「2群間の分散に差がない(等分散である)」と仮定する。
- 対立仮説(H_1)：「2群間の分散に差がある(等分散でない)」と仮定する。
- 1≦F（計算値）≦F_αのとき、P＞0.05となる→帰無仮説を棄却できない→等分散である。
  F(計算値)＞F_αのとき、P＜0.05となる→帰無仮説を棄却する→不等分散である。

(問題⑥)

- 2つの標本があり、1つは11個の測定値、もう一つは6個の測定値である。その標準偏差はそれぞれ5.0mmと8.0mmであった。それぞれの分散の差を5%有意水準でF-検定せよ。

- $n_1 = 6, \sigma_1 = 8.0\mathrm{mm},$ $\sigma_1^2 = \left(\frac{n_1}{n_1-1}\right) s_1^2 = \left(\frac{6}{6-1}\right) \times 64 = 76.8$
- $n_2 = 11, \sigma_2 = 5.0\mathrm{mm},$ $\sigma_2^2 = \left(\frac{n_2}{n_2-1}\right) s_2^2 = \left(\frac{11}{11-1}\right) \times 25 = 27.5$
- $F = \frac{76.8}{27.5} = 2.8$

- σ_2 ＝ 5.0mmと σ_1 ＝ 8.0mmの自由度は10と5なので、
- 5%有意水準の表より
- F_α＝3.33>F=2.8、帰無仮説を棄却できないので、

∴有意でない(等分散である)。

1% LEVEL OF VARIANCE RATIO*

Number of degrees of freedom in lesser variance estimate	Number of degrees of freedom in the greater variance estimate: 1	2	3	4	5	10	20	∞
1	4,100	5,000	5,400	5,600	5,800	6,000	6,200	6,400
2	98	99	99	99	99	99	99	99
3	34	31	29	29	28	27	27	26
4	21	18	17	16	16	15	14	13
5	16	13	12	11	11	10	9·6	9·0
10	10	7·6	6·6	6·0	5·6	4·8	4·4	3·9
20	8·1	5·8	4·9	4·4	4·1	3·4	2·9	2·4
∞	6·6	4·6	3·8	3·3	3·0	2·3	1·9	1·0

N.B. The symbol ∞ denotes ‘infinitely great’, i.e., in practice, ‘very large’.

5% LEVEL OF VARIANCE RATIO

Number of degrees of freedom in lesser variance estimate	Number of degrees of freedom in the greater variance estimate: 1	2	3	4	5	10	20	∞
1	161	200	216	225	230	242	248	254
2	18·5	19	19·2	19·2	19·3	19·4	19·4	19·5
3	10·1	9·6	9·3	9·1	9·0	8·8	8·7	8·5
4	7·7	6·9	6·6	6·4	6·3	6·0	5·8	5·6
5	6·6	5·8	5·4	5·2	5·0	4·7	4·6	4·4
10	5·0	4·1	3·7	3·5	3·3	3·0	2·8	2·5
20	4·3	3·5	3·1	2·9	2·7	2·3	2·1	1·8
∞	3·8	3·0	2·6	2·4	2·2	1·8	1·6	1·0

*See acknowledgements, p. viii.

（**問題⑦**）Aクラスには7人、Bクラスには9人の生徒がいる。この生徒たちに物理のテストを実施した。AクラスとBクラスの物理のテストの点は5%有意水準で等分散かどうかを検定せよ。

番号	Bクラス	Aクラス
1	60	49
2	52	40
3	68	52
4	55	37
5	65	55
6	47	38
7	45	45
8	62	
9	53	

帰無仮説(H_0)：2群間の分散に差がない(等分散である)
対立仮説(H_1)：2群間の分散に差がある(等分散でない)

F-検定: 2 標本を使った分散の検定		
	Bクラス	Aクラス
平均	56.33	45.14
分散	63	50.48
観測数	9	7
自由度	8	6
観測された分散比	1.248	
P(F<=f) 片側	0.405	
F 境界値 片側	4.147	

- 分散63（B）と分散47.0（A）の自由度は8と6なので
- 5%有意水準の表より
- F_α＝3.5>F=1.25、P=0.4＞0.05であり、帰無仮説を棄却できないので、

∴有意でない(等分散である)。

観測論

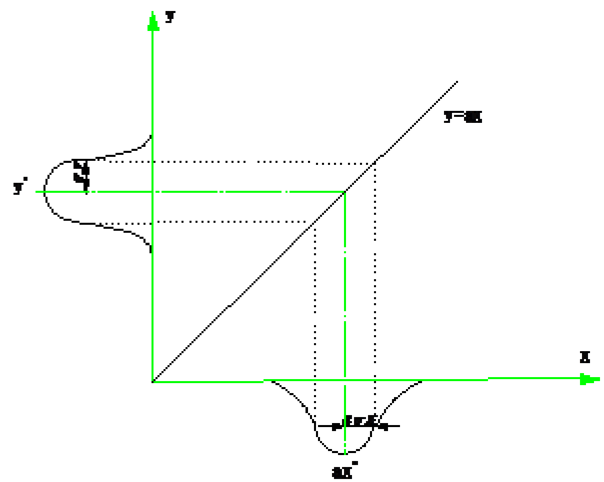
観測と測定は同意義である。これには記録が含まれる。

観測法を決め、機械器具の装置を用いて測定し、記録する。

直接観測値から必然的変動を分析、評価する。これが観測論である。

- 観測値の確率分布

（1）単独観測



観測値間に

$y = ax$…（3.1）

の関係があり、その平均値が

$\bar{y} = a\bar{x}$ …(3.2)

であり、標準偏差は

$\sigma_y = a\sigma_x$ …(3.3)

であらわされる。

これは次のように証明できる。

xの確率密度は

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma_x}\right)^2} \text{…(3.4)}$$

と書ける。

そこで、

$$y = ax$$

または

$x = \frac{y}{a}$…(3.5)

あるいは

$$f(y) = f(x)\frac{dx}{dy} \dots (3.6)$$

とおける。そこで、yの密度は

$$f(y) = \frac{1}{\sqrt{2\pi}a\sigma_x} e^{-\frac{1}{2}\left(\frac{\frac{y}{a}-\bar{x}}{\sigma_x}\right)^2}$$

あるいは

$$f(y) = \frac{1}{\sqrt{2\pi}a\sigma_x} e^{-\frac{1}{2}\left(\frac{y-a\bar{x}}{a\sigma_x}\right)^2} \dots (3.7)$$

を得る。

平均値の関係：$\bar{y} = a\bar{x}$

標準偏差の関係：$\sigma_y = a\sigma_x$

また、観測値の関係が
$y = ax + b$という関係で、
(a, b)が定数ならば、上と同様
にして

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma_x}\right)^2} \dots (3.4)$$

$y = ax + b$から$x = \frac{y}{a} - \frac{b}{a}$

$$f(y) = \frac{1}{\sqrt{2\pi}a\sigma_x} e^{-\frac{1}{2}\left(\frac{\frac{y}{a}-b/a\bar{x}}{\sigma_x}\right)^2}$$

または

$f(y) =$

$$\frac{1}{\sqrt{2\pi}a\sigma_x} e^{-\frac{1}{2}\left(\frac{y-b-a\bar{x}}{a\sigma_x}\right)^2} \quad \ldots(3.8)$$

すなわち$y = ax + b$関係において

平均値:$\bar{y} = a\bar{x} + b$

標準偏差:$\sigma_y = a\sigma_x$

を得る。

(問題①)次の関係のyの分散を求めよ。

y=3z+4、$\sigma_z = 3\mathrm{mm}$

(解答)

dy=3dzより、$\sigma_y = 3\sigma_z$

$$\sigma_y^2 = 9\sigma_z^2 = 9 \times 9$$
$$= 81\mathrm{mm}^2$$

を得る。

多次元観測

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

- Y=AXで表せる。

変数Xは確率変数をもてば、それは次に示す分散共分散行列で表せる。

$$\mathrm{V_{XX}} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \sigma_{n3} & \cdots & \sigma_{nn} \end{bmatrix}$$

Yの分散共分散行列は分散伝播規則から

$$\mathrm{V_{YY}} = \mathrm{AV_{XX}A^T} \quad \text{…(3.9)}$$

(問題②)座標変換式において、任意の座標系(x,y)からもう一つの座標系(N,E)になると

$$\left.\begin{array}{r} N = ax + by + c \\ E = -bx + ay + d \end{array}\right\}$$

…(3.10)

が与えられれば、N,Eの分散はいくらになるか。

コファクタg^{ii}を用いると

- $$\left.\begin{array}{r} g^{N} = ag^{x} + bg^{y} \\ g^{E} = -bg^{x} + ag^{y} \end{array}\right\}$$

両式を平方及び掛け合わせると

$$g^{NN} = a^2 g^{xx} + 2ab g^{xy} + b^2 g^{yy}$$

$$g^{EE} = b^2 g^{xx} - 2ab g^{xy} + a^2 g^{yy}$$

$$g^{NE} = -ab g^{xx} + (a^2 - b^2) g^{xy} + ab g^{yy}$$

同様に分散は次のように書ける。

$$\sigma_N^2 = a^2 \sigma_x^2 + 2ab\sigma_{xy} + b^2 \sigma_y^2$$

$$\sigma_E^2 = b^2 \sigma_x^2 - 2ab\sigma_{xy} + a^2 \sigma_y^2$$

$$\sigma_{NE} = -ab(\sigma_x^2 - \sigma_y^2) + (a^2 - b^2)\sigma_{xy}$$

$$g^{xx} = \frac{\sigma_x^2}{\sigma_0^2} = \frac{1}{w^{xx}}$$

$$g^{xy} = \frac{\sigma_{xy}}{\sigma_0^2} = \frac{1}{w^{xy}}$$

$$g^{yy} = \frac{\sigma_y^2}{\sigma_0^2} = \frac{1}{w^{yy}}$$

ただし、g：コファクタ、w：重量、σo^2:不偏分散、σx^2:分散、σ_{xy}:共分散

座標変換式を行列で表すと

$$\begin{bmatrix} N \\ E \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c \\ d \end{bmatrix}$$

…(3.11)

又は

Y=AX+C…（3.12）

で表されるので、分散共分散伝播式（3.9）

$V_{YY} = AV_{XX}A^T$から

$$V_{YY} = \begin{bmatrix} a & a \\ -b & a \end{bmatrix} \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

n個の関数Y=AXと置ける場合、Xの確率密度は

$$f(X) = \frac{1}{(2\pi)^{n/2}|V^{1/2}|} e^{-\frac{1}{2}(X-\bar{X})^T V^{-1} (X-\bar{X})} \dots (3.12)$$

ここで、VはXの分散共分散行列である。また、Y=AXの行列関数から、Aが正則で|A|≠0ならば、$\mathrm{X} = \mathrm{A}^{-1}\mathrm{Y}$が成立する。そこで、式（3.12）に代入すると

$$\mathrm{f(Y)} = \frac{|\mathrm{A}^{-1}|}{(2\pi)^{n/2}|\mathrm{V}^{1/2}|} \mathrm{e}^{-\frac{1}{2}\left(A^{-1}\mathrm{Y}-\bar{\mathrm{X}}\right)^{\mathrm{T}}\mathrm{V}^{-1}\left(A^{-1}\mathrm{Y}-\bar{\mathrm{X}}\right)}$$

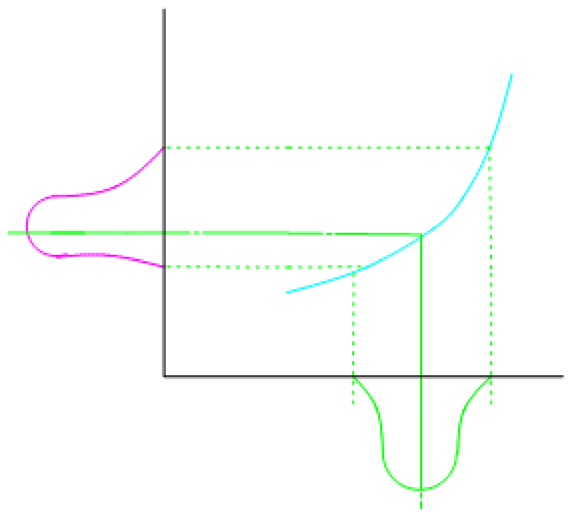
これを整理すると

$$\mathrm{f(Y)} = \frac{|\mathrm{A}^{-1}|}{(2\pi)^{n/2}|\mathrm{V}^{1/2}|} \mathrm{e}^{-\frac{1}{2}(\mathrm{Y}-\mathrm{A}\bar{\mathrm{X}})^{\mathrm{T}}(\mathrm{AVA}^{\mathrm{T}})^{-1}(\mathrm{Y}-\mathrm{A}\bar{\mathrm{X}})} \quad \text{…(3.13)}$$

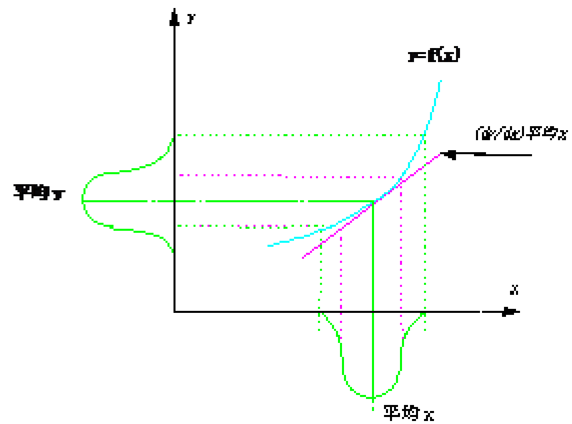
- 関数Y=AXは平均値$\overline{\mathrm{Y}} = \mathrm{A}\overline{\mathrm{X}}$と分散共分散行列$\mathrm{V_{YY}} = \mathrm{AV_{XX}A^T}$をもつことが証明できた。

非線形関数（non-linear function）

- 関数y=f(x)が非線形であり、xが正規分布であっても、yは正規分布しない場合がこの関数である。



- そこでその関係をそのまま解くことは、その理論上法則に従わないから、近似法を導入することが適当であると仮定する。それは線形化(linearization)と呼ばれるものである。この手法によると、導かれた量も正規分布するようになる。
- すなわち、その初期の関数において$x = \bar{x}$の接線（導関数）とみなすもので、yの平均値$\bar{y}$はこの近似解によって乱されない。



(問題③)次の関数の線形化を行え。

(1) $y = ax^2$

(2) $y = \sin x$

(3) $y = ax^3 + bx^2 + cx + d$

(4) $y = \ln x$

(5) $z = x^2 + y^2$

(6) $z = \sin x \cos y$

•

(解答)

(1)この関数の導関数は$\frac{dy}{dx} = 2ax_0$、またはdy $= 2ax_0 dx$を得る。あるいは、y=axをテーラー展開すると、

$$y = (ax_0^2) + \frac{2a_0x_0}{1!} \cdot \Delta x + \cdots$$

であり、また$y = y_0 + \Delta y$とおけるので

$$y_0 + \Delta y = ax_0^2 + 2a_0x_0\Delta x + \cdots$$

また、$y_0 = ax_0^2$なので

$$\Delta y = 2a_0x_0\Delta x$$

を得る。これが線形化である。

•

(2) $\mathrm{dy} = \cos x_0 \cdot \mathrm{dx}$

(3)

$\mathrm{dy} =$
$(3ax_0^2 + 2bx_0 + c)\mathrm{dx}$

(4) $\mathrm{dy} = \frac{1}{x_0}\mathrm{dx}$

(5) $\mathrm{dz} = 2x_0\mathrm{dx} + 2y_0\mathrm{dy}$

(6) $\mathrm{dz} = \cos x_0 \cos y_0 \mathrm{dx} - \sin x_0 \cos y_0 \mathrm{dy}$

•

(d) 非線形関数における分散伝播法則

上記の問題において、それぞれの分散をユニークな方法で以下のように解く。

(1)

$$\mathrm{dy} = 2a_0x_0\mathrm{dx}$$

一般に分散は$\overline{\mathrm{dy},\mathrm{dy}^{\mathrm{T}}}$と書き

$$\overline{\mathrm{dy},\mathrm{dy}^{\mathrm{T}}}$$
$$=\overline{2\mathrm{a}_0\mathrm{x}_0\mathrm{dx},(2\mathrm{a}_0\mathrm{x}_0\mathrm{dx})^{\mathrm{T}}}$$
$$=2\mathrm{a}_0\mathrm{x}_0\overline{\mathrm{dx},\mathrm{dx}^{\mathrm{T}}}2\mathrm{a}_0\mathrm{x}_0$$
$$=4\mathrm{a}_0^2\mathrm{x}_0^2\overline{\mathrm{dx},\mathrm{dx}^{\mathrm{T}}}$$

となるから、

$$\sigma_{\mathrm{y}}^2=4\mathrm{a}_0^2\mathrm{x}_0^2\sigma_{\mathrm{x}}^2$$

(問題④)$\mathbf{x}=\frac{\mathrm{e}\sin\varphi}{S}$におけるSの分散を求めよ。

ただし、$\sigma_{\mathrm{e}}=0.01\mathrm{m},\sigma_S=0.02\mathrm{m},\sigma_\varphi=20'',e=5\mathrm{m},\mathrm{S}=2000\mathrm{m},\varphi=90^\circ$とし、共分散は無視する。

(解答)

$$\Delta \mathrm{x} = \frac{\sin\varphi}{S}\Delta \mathrm{e} + \frac{e\ \cos\varphi}{S}\Delta\varphi - \frac{e\ \sin\varphi}{S^2}\Delta S$$

$$\sigma x^2 = \frac{sin^2\varphi}{S^2}\sigma e^2 + \frac{e^2cos^2\varphi}{S^2}\sigma\varphi^2 + \frac{e^2sin^2\varphi}{S^4}\sigma S^2$$

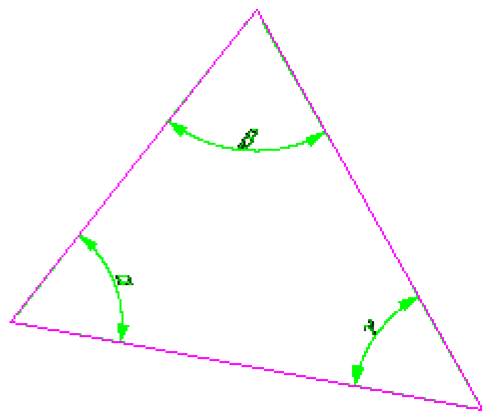
$$\sigma x^2 = \frac{1^2}{2000^2}0.01^2 + \frac{5^2\cdot 0}{2000^2}(20/206265)^2 + \frac{5^2\cdot 1^2}{2000^4}0.02^2$$

$$\sigma_x = 0.00223rad$$

$$= 461'' = 7.7'$$

観測論における標準偏差の取り扱い

- **応用問題(1)**　平面三角形において内角を測定した。それぞれの標準偏差が10”ならば、内角の和の精度はいくらになるか。



T=α+β+γ＝180°

$$\sigma T^2 = \sigma\alpha^2 + \sigma\beta^2 + \sigma\gamma^2$$
$$= 3 \times 10''^2 = 300$$

$\sigma_T = 10\sqrt{3}$=18”

行列によると

$$V=\begin{bmatrix}100 & 0 & 0\\ 0 & 100 & 0\\ 0 & 0 & 100\end{bmatrix}''^2$$

式(3.2.1)の期待値は

$$E(T)=E(\alpha)+E(\beta)+E(\gamma)$$

$$=\bar{\alpha}+\bar{\beta}+\bar{\gamma}=\bar{T}=180^{\circ}\ldots(3.2.2)$$

であるから、

$$\bar{T}=UX=\begin{bmatrix}1 & 1 & 1\end{bmatrix}\begin{bmatrix}\bar{\alpha}\\ \bar{\beta}\\ \bar{\gamma}\end{bmatrix}\ldots(3.2.3)$$

とも書ける。したがって、Tに関する分散はUVU^Tから、

UVU^T

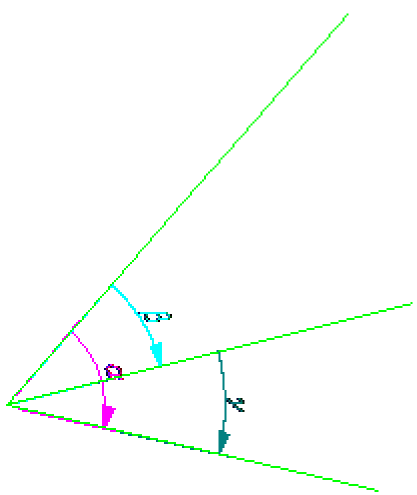
$$=\begin{bmatrix}1 & 1 & 1\end{bmatrix}\begin{bmatrix}100 & 0 & 0\\ 0 & 100 & 0\\ 0 & 0 & 100\end{bmatrix}\begin{bmatrix}1\\ 1\\ 1\end{bmatrix}$$

$$=300''^2$$

すなわち、

$$\sigma_T=10\sqrt{3}=18''$$

問題(2) 水平角α、βを観測した。γの精度を求めよ。ただし、$\sigma_\alpha = \sigma_\beta = 10''$, $\sigma_{\alpha\beta} = 0$とする。



(解答)

$$\alpha - \beta = \gamma$$

において分散伝播法則を適用すると、

$$\sigma_\gamma^2 = \sigma_\alpha^2 + \sigma_\beta^2 = 200''$$

$$\sigma_\gamma = 10\sqrt{2} = 15''$$

(別解)

分散行列$V = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}''^2$

$$\gamma = \alpha - \beta$$

から

$$\gamma = [1 \quad -1]\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = UX$$

したがって、

$UVU^T = [1 \quad -1]\begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

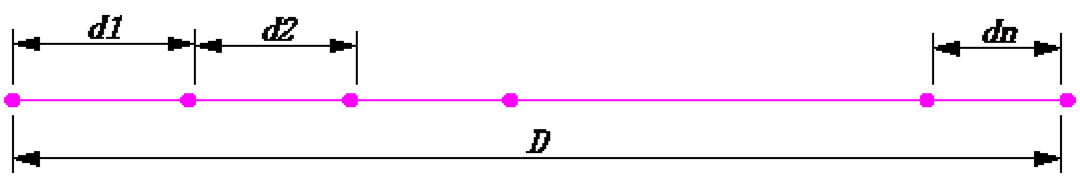
$= 200''^2$

すなわち、

$$\sigma_\gamma = 10\sqrt{2} = 15''$$

問題(3)　長さd(m)のテープで距離測量を行ったところD(m)を得た。Dの精度を求めよ。ただし、各観測は独立等精度である。

(解答)



$$D = d_1 + d_2 + \cdots + d_n$$

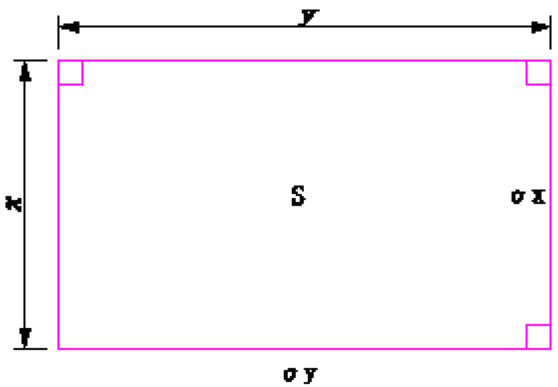
$$\sigma_D^2 = \sigma_{d1}^2 + \sigma_{d2}^2 + \cdots + \sigma_{dn}^2 \approx n\sigma_d^2$$

そして

$$n = \frac{\sigma_D^2}{\sigma_d^2} =\approx \frac{D}{d}$$

$$\therefore \sigma_D = \sqrt{\frac{D}{d}}\sigma_d$$

問題(4) 長方形の土地の縦横x、yを測定した。その精度をそれぞれ$\sigma_x, \sigma_y, \sigma_{xy} = 0$とするとき、面積Sの精度を求めよ。



(解答)

$$S = xy$$

これをテーラー展開すると

$$S = S_0 + y_0(x - x_0) + x_0(y - y_0)$$

又は

$$S = x_0 y_0 + x y_0 - x_0 y_0 + x_0 y - x_0 y_0$$

又は

$$S = xy_0 + x_0 y - x_0 y_0$$

これを行列にすると

$$S = \begin{bmatrix} y_0 & x_0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x_0 y_0 \end{bmatrix}$$

$$= UX - S_0$$

したがって、分散行列は

UVU^Tから

$$\sigma_S^2$$

$$= \begin{bmatrix} y_0 & x_0 \end{bmatrix} \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} \begin{bmatrix} y_0 \\ x_0 \end{bmatrix}$$

$$= y_0^2 \sigma_x^2 + x_0^2 \sigma_y^2 + 2 x_0 y_0 \sigma_{xy}$$

そこで、$x_0 = 20\mathrm{m}, y_0 = 10\mathrm{m}, \sigma_x = 5\mathrm{cm}, \sigma_y = 3\mathrm{cm}, \sigma_{xy} = \sigma_{yx} = 0$とすれば、

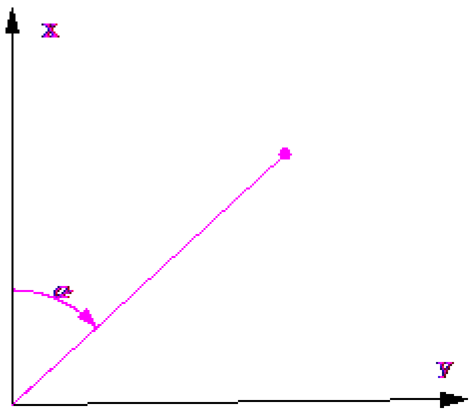
$$\sigma_S^2 = y_0^2\sigma_x^2 + x_0^2\sigma_y^2$$

$$= (10\mathrm{m})^2(5\mathrm{cm})^2+(20\mathrm{m})^2(3\mathrm{cm})^2$$

$$= 6100$$

$$\sigma_S = 10\sqrt{61}\mathrm{m}\cdot\mathrm{cm} = 0.7\mathrm{m}^2$$

問題(5) トラバース測量において、距離dと角αを測定した。座標の精度を求めよ。ただし、$d = 10.02\mathrm{m}, \alpha = 45^o00'20", \sigma_d = 2\mathrm{cm}, \sigma_\alpha = 10", \sigma_{\alpha d} = 0$とする。



（解答）

$$x = d\cos\alpha, y = d\sin\alpha$$

すなわち、

$$\Delta x = \cos\alpha_0 \Delta d + d_0(-\sin\alpha_0)\Delta\alpha$$

$$\sigma_x^2 = \cos^2\alpha_0\sigma_d^2 + d_0^2(\sin^2\alpha_0)^2\sigma_\alpha^2 - 2d_0\sin\alpha_0\cos\alpha_0\sigma_{\alpha d}$$

また、

$$\cos\alpha_0 \approx \sin\alpha_0 \approx \sqrt{2}/2$$

$$\sigma_x^2 = \frac{1}{2} \times (4\mathrm{cm})^2 + (10.02\mathrm{m})^2\frac{1}{2} \times \left(\frac{10''}{206265''}\right)^2$$

$$= 8\mathrm{cm}^2$$

$$\sigma_x = 2.9\mathrm{cm}$$

$$\Delta y = \sin\alpha_0 \Delta d + d_0(\cos\alpha_0)\Delta\alpha$$

$$\sigma_y^2 = \sin^2\alpha_0 \sigma_d^2 + d_0^2(\cos^2\alpha_0)^2\sigma_\alpha^2 + 2d_0\sin\alpha_0\cos\alpha_0\sigma_{\alpha d}$$

$$\sigma_y^2 = \frac{1}{2} \times (4\text{cm})^2 + (10.02\text{m})^2\frac{1}{2} \times (\frac{10''}{206265})^2 = 8\text{cm}^2$$

$$\sigma_y = 2.9\text{cm}$$

$$\Delta x = \cos\alpha_0 \Delta d + d_0(-\sin\alpha_0)\Delta\alpha$$

$$\Delta y = \sin\alpha_0 \Delta d + d_0(\cos\alpha_0)\Delta\alpha$$

$$\sigma_{xy} = \sin\alpha_0\cos\alpha_0\left(\sigma_d^2 - d_0^2\sigma_\alpha^2\right) + (\cos^2\alpha_0 - \sin^2\alpha_0)\sigma_{d\alpha}$$

$$= \frac{1}{2}\left[(2\text{cm})^2 - (1002\text{cm})^2 \times \left(\frac{10''}{206265''}\right)^2\right] = 2\text{cm}^2$$