平均、分散（離散変数の場合）

平均値 $\bar{x}$

$x_k,\ k=1,2,3,...n$ のとき、x の平均値 $\bar{x}$ は、

$$\bar{x}=\frac{1}{n}\left(x_1+x_2+x_3+...+x_k+...+x_n\right)=\frac{1}{n}\sum_{k=1}^{n}x_k$$

分散 σ_x^2（ばらつきの度合いを 表す）

$$\sigma_x^2=\frac{1}{n}\sum_{k=1}^{n}\left(x_k-\bar{x}\right)^2$$

x と y の和、差の平均値はそ れぞれの

平均値の和、差、すな わち $\overline{x\pm y}=\bar{x}\pm\bar{y}$

$$\therefore \overline{x\pm y}=\frac{1}{n}\sum_{k=1}^{n}\left(x_k\pm y_k\right)=\frac{1}{n}\sum_{k=1}^{n}x_k\pm\frac{1}{n}\sum_{k=1}^{n}y_k=\bar{x}\pm\bar{y}$$

x と y が相互に独立なとき、 $x\pm y$ の分散はそれ

ぞれの分散の和 $\sigma_{x\pm y}^2=\sigma_x^2+\sigma_y^2$

変数の差に対しても分 散の和になることに注 意。

$$\therefore \sigma_{x\pm\mathrm{y}}^2=\frac{1}{n}\sum_{k=1}^{n}\left\{\left(x_k\pm y_k\right)-\left(\bar{x}\pm\bar{y}\right)\right\}^2$$

$$=\frac{1}{n}\sum_{k=1}^{n}\left\{\left(x_k-\bar{x}\right)\pm\left(y_k-\bar{y}\right)\right\}^2$$

$$=\frac{1}{n}\sum_{k=1}^{n}\left\{\left(x_k-\bar{x}\right)^2+\left(y_k-\bar{y}\right)^2\pm 2\left(x_k-\bar{x}\right)\left(y_k-\bar{y}\right)\right\}$$

$$=\frac{1}{n}\sum_{k=1}^{n}\left(x_k-\bar{x}\right)^2+\frac{1}{n}\sum_{k=1}^{n}\left(y_k-\bar{y}\right)^2$$

$$\pm\frac{2}{n}\sum_{k=1}^{n}\left(x_k-\bar{x}\right)\left(y_k-\bar{y}\right)$$

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

x、y が互いに独立なとき、 共分散 $\sigma_{xy}=0$

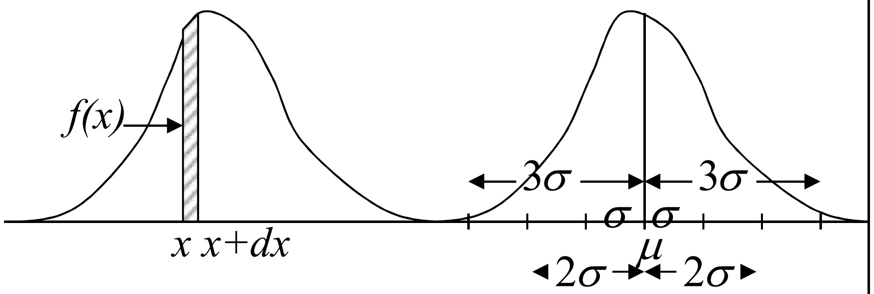
$$\therefore \sigma_{\mathrm{x\pm y}}^2=\sigma_x^2+\sigma_y^2$$

分散の平方根 $\sqrt{\sigma_Z^2}=\sigma_Z$ を標準偏差という。

$$\sigma_x=\sqrt{\sigma_x^2}=\sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(x_k-\bar{x}\right)^2}$$

$$\sigma_{x\pm y}=\sqrt{\sigma_x^2+\sigma_y^2}$$

平均、分散（連続変数の場合）



期待値 E

変量X がxと$x+dx$との間にある確率が$f(x)\,dx$ で表されるとき、$f(x)$ を確率密度関数といい

$$\mu = \int_{-\infty}^{\infty} x f(x) dx$$ を

$$= E(X)$$

と書き、Xの期待値あるいは平均値と呼ぶ。

分散σ^2 は、次式で定義される。

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

$$= E\{(X-\mu)^2\} (= E(X-\mu)^2 \text{とも書く})$$

同じ母集団から任意に二つのサンプル X_1, X_2 を取るとき、それぞれの平均値は $E(X_1) = E(X_2) = \mu$ であり、分散も $E(X_1-\mu)^2 = E(X_2-\mu)^2 = \sigma^2$ であり、両者は等しい。

いま､2つの互いに独立な母集団（μ_1, σ_1^2),（μ_2, σ_2^2）から１つずつのサンプルX_1, X_2 をとり,その和と差をつくり、その分布を調べる。

平均値 μ は、

$$\mu = E(X_1 \pm X_2) = E(X_1) \pm E(X_2) = \underline{\mu_1 \pm \mu_2}$$

分散σ^2 は、

$$\sigma^2 = E\{(X_1 \pm X_2) - (\mu_1 \pm \mu_2)\}^2$$

$$= E\{(X_1-\mu_1) \pm (X_2-\mu_2)\}^2$$

$$= E\{(X_1-\mu_1)^2 + (X_2-\mu_2)^2\} \pm 2E\{(X_1-\mu_1)(X_2-\mu_2)\}$$

2つの母集団は互いに独立であるから、

$E\{(X_1-\mu_1)(X_2-\mu_2)\} = 0$ であり、したがって

$$E\{(X_1 \pm X_2) - (\mu_1 \pm \mu_2)\}^2$$

$$= E\{(X_1-\mu_1)^2 + (X_2-\mu_2)^2\}$$

$$= E\{(X_1-\mu_1)^2\} + E\{(X_2-\mu_2)^2\}$$、すなわち、

$\underline{\sigma^2 = \sigma_1^2 + \sigma_2^2}$となる。

すなわち、和の分布も差の分布も平均値はそれぞれの平均値の和、差になるが、分散はいずれも、それぞれの分散の和になる。

標準偏差も、いずれも、$\underline{\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}}$ となる。

正規分布と平均、標準偏差、信頼水準

連続的な確率変数 x について、
関数 $\varphi(x)$ を次のように定義する 。
$\varphi(x)$ を正規確率密度関数と 言う。

$$\varphi(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$$ 、ここで、

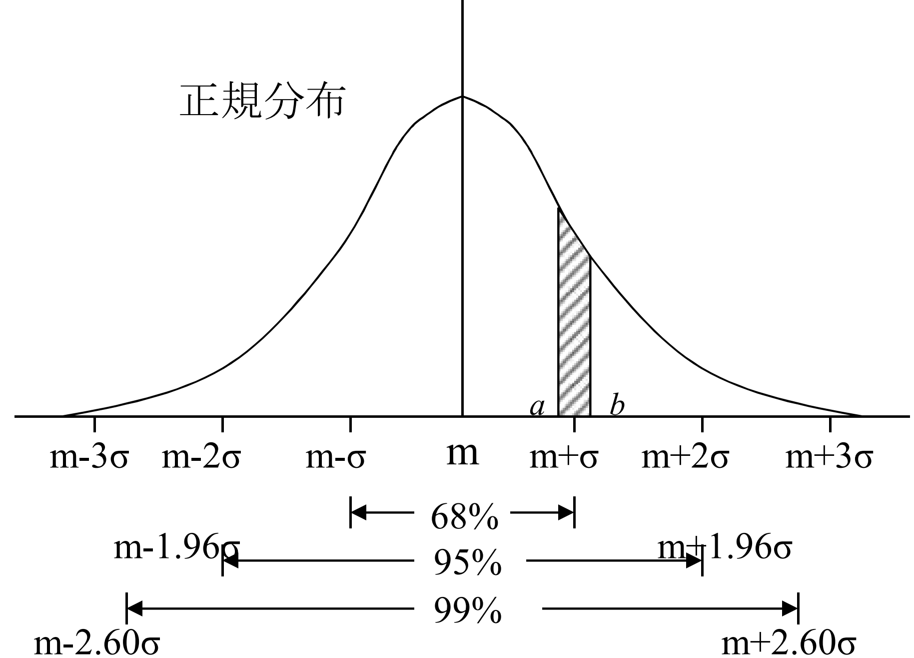
m はx の平均値、σ^2 は分散、
σ は標準偏差である。

なお、$\int_{-\infty}^{\infty}\varphi(x)\,dx=1$ である。

また、区間 $[a, b]$ の間の値をとる確率
P は、$P=\int_a^b \varphi(x)\,dx$ である(右図 a,b 間)。

多くの確率変数がこの 正規分布で表されることが知られている 。

なお、母集団が (平均m、分散 σ^2)の分布に従うとき、そこから n 個のサンプルを取ったときの、サンプ ルの平均値の分散は σ^2/n、標準偏差は $\sigma/\sqrt{n}$ となることが知られている。(p.7 参照)



正規分布では、
$[m-\sigma, m+\sigma]$ の確率が 68.3%、
$[m-2\sigma, m+2\sigma]$ の確率が に95.4%、
$[m-3\sigma, m+3\sigma]$ の確率が に99.7%、
平均値のまわりに 95%が入るのは $m\pm 1.96\sigma$
平均値のまわりに 99%が入るのは $m\pm 2.6\sigma$
であることが知られて いる。

あるサンプルが、この グラフの両側の端に近 い値になったとき、その サンプルがこの集団に 属しないとしたときの誤 差はその確率 p 程度（ 棄却確率という）であり 、「信頼水準 1-p でそのサンプルは母集団外と 判定する」という。

例題１（平成21年、基礎）

ある材料に生ずる応力 R がその材料の強度 S より小さくなるように設計する。ここで、$Z=R-S$ と定義する。いま、R が平均 μ_R、分散 σ_R^2 の正規分布に、応力 S が平均 μ_S 、分散 σ_S^2 の正規分布に従う互いに独立な確率変数とみなせるとき、Z の記述として正しいものを①～⑤の中から選べ。

①Z は平均 $(\mu_R-\mu_S)$、標準偏差 $\sqrt{\sigma_R^2-\sigma_S^2}$ の正規分布に従う。

②Z は平均 $(\mu_R-\mu_S)$、標準偏差 $\sqrt{\sigma_R^2+\sigma_S^2}$ の正規分布に従う。

③Z は平均 $(\mu_S-\mu_R)$、標準偏差 $(\sigma_R^2-\sigma_S^2)$ の正規分布に従う。

④Z は平均 $(\mu_R-\mu_S)$、標準偏差 $(\sigma_R^2+\sigma_S^2)$ の正規分布に従う。

①Z は平均 $(\mu_S-\mu_R)$、標準偏差 $\sqrt{\sigma_R^2-\sigma_S^2}$ の正規分布に従う。

確率変数の差の分散はそれぞれの分散の和になる。
分散の平方根が標準偏差（p.1～3参照）。
よって、正解は②

例題2（平成20年、基礎）

あるプラントにおいて、正常に稼働した場合の1日当りの生産量は平均値 $\mu = 1{,}200\ kg$、標準偏差 $\sigma = 48\,kg$ の正規分布で表される母集団であることが過去のデータからわかっている。生産量を $n = 36$ 日間計測したところ、1日当りの生産量の平均値は $\bar{x} = 1{,}176\ kg$ であった。信頼水準 95% で、このプラントが正常に稼働しているかどうかを統計学的に検定するための検討を行った。次の記述の中から、適切なものを選べ。なお、信頼水準 95% に対する標準正規分布における両側信頼限界は1.96 としてよい

①$\left|\dfrac{\bar{x}-\mu}{\sigma^2/n}\right| = \left|\dfrac{1176-1200}{48^2/36}\right| = 0.375 < 1.96$となるので、少なくとも95%の確率でプラントは正常に稼働している。

②$\left|\dfrac{\bar{x}-\mu}{\sigma}\right| = \left|\dfrac{1176-1200}{48}\right| = 0.5 < 1.96$となるので、少なくとも95%の確率でプラントは正常に稼働している。

③$\left|\dfrac{\bar{x}-\mu}{\sigma/n}\right| = \left|\dfrac{1176-1200}{48/36}\right| = 18 > 1.96$となるので、少なくとも95%の確率でプラントは正常に稼働していない。

④$\left|\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right| = \left|\dfrac{1176-1200}{48/\sqrt{36}}\right| = 3 > 1.96$となるので、少なくとも95%の確率でプラントは正常に稼働していない。

⑤正常に稼働しているかどうか統計学的に検定できない。

正解　標本（サンプル）平均の標準偏差は $\sigma/\sqrt{n}$ から④が正しい（$p.7 \sim 8$参照)。

例題3（平成17年、基礎）

凸部と凹部（たとえば 軸と穴）が組み合う場 合、組み合う部分の幅 寸法の分布が次式で表される正規分布で 近似できる。

凸部の幅（例えば 、軸の外径）：$N(\mu_1, \sigma_1^2)$

凹部の幅（例えば、軸 の内径）：$N(\mu_2, \sigma_2^2)$

ここで、μ_i は平均値、σ_i は標準偏差である。（$i = 1, 2$）

この２つの部品を組み 合わせた場合に生じる 隙間寸法の分布を表す ものを選べ。ただし隙間がある場合は隙間 寸法は正であるとする 。

①$\left(\mu, \sigma^2\right) = \left(\dfrac{\mu_1 + \mu_2}{2}, \dfrac{\sigma_1^2 + \sigma_2^2}{2}\right)$

②$\left(\mu, \sigma^2\right) = \left((\mu_1 + \mu_2), (\sigma_1^2 + \sigma_2^2)^{\frac{1}{2}}\right)$

③$\left(\mu, \sigma^2\right) = \left(\dfrac{\mu_1}{\mu_2}, \dfrac{\sigma_1^2}{\sigma_2^2}\right)$

④$\left(\mu, \sigma^2\right) = \left((\mu_2 - \mu_1), (\sigma_1^2 + \sigma_2^2)\right)$

⑤$\left(\mu, \sigma^2\right) = \left((\mu_1 - \mu_2), (\sigma_1^2 + \sigma_2^2)^{\frac{1}{2}}\right)$

正解　差の分布の平均 値は平均値の差、差の 分布の分散は分散の和 から④が正しい（$p.1 \sim 3$参照)。

標本平均の分布（平均、分散）

平均値が m、分散が σ^2 である母集団から n 個の独立な標本（サンプル）を取ってその平均値(標本平均)を作る。標本平均を(多数)作って、その分布を表す平均と分散を求めると、それぞれ m 、σ^2/n となることを説明する。確率変数を X 、期待値を E で表す。

1.平均値

$m=E(X)$ であり、下式のように、標本平均 m_S は m に等しいことがわかる。

標本平均 $$m_S=E\left\{\frac{1}{n}\left(\sum_{k=1}^{n}X_k\right)\right\}=\frac{1}{n}\sum_{k=1}^{n}E(X_k)=\frac{1}{n}\sum_{k=1}^{n}m=\frac{1}{n}nm=m$$

2.分散

母集団では $\sigma^2=E(X-m)^2$、標本平均の分散を σ_S^2 とすると、

$$\sigma_S^2=E\left\{\frac{1}{n}\left(\sum_{k=1}^{n}X_k\right)-m\right\}^2=E\left\{\frac{1}{n}\left(\sum_{k=1}^{n}X_k-nm\right)\right\}^2=E\left\{\frac{1}{n}\left(\sum_{k=1}^{n}(X_k-m)\right)\right\}^2$$

$$=E\left\{\frac{1}{n^2}\sum_{k=1}^{n}(X_k-m)^2\right\}$$

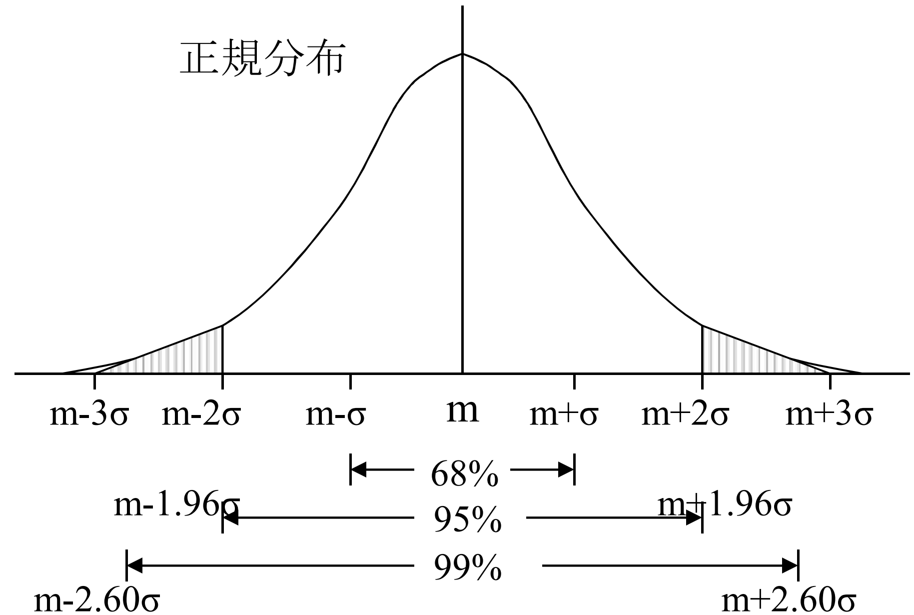
$\Big($ X_k、$X_{j\neq k}$は相互に独立なので $E(X_k-m)(X_j-m)=(E(X_k)-m)(E(X_j)-m)=0$ $\Big)$

$$=\frac{1}{n^2}\sum_{k=1}^{n}E(X_k-m)^2=\frac{1}{n^2}\times n\sigma^2=\frac{\sigma^2}{n}$$

これから、標本平均の標準偏差は $\sigma/\sqrt{n}$ となる。

すなわち平均値の標準偏差は母集団のそれの $1/\sqrt{n}$ と母集団より小さくなる。

統計的仮設とその検定



母集団は正規分布に従い、その平均値が m、
分散が σ^2 であるとする。

1 個のサンプル x_1 について、$T = \left|\dfrac{x_1 - m}{\sigma}\right|$ を計算

する。この値が k より大きければ、

$\left|\dfrac{x_1 - m}{\sigma}\right| > k$ すなわち、

$(x_1 - m) < -k\sigma \ or \ (x_1 - m) > +k\sigma$、すなわち、

x_1 は m から $k\sigma$ 以上離れた所にある。

仮設1.　x_1 はこの母集団に属するデータである。
仮設2.　x_1 はこの母集団に属しないデータである。

仮設 1 が正しい確率は、右図から $m \pm k\sigma$ の外側で、$k > 2$ なら 5% に満たない。
仮設 2 が正しい確率は、$m \pm k\sigma$ の内側で、$k > 2$ なら 95% 以上である。

仮設の検定とはこのようにして、あるデータがある母集団に属するかどうか、
通常は、属しないことを説明するために使われる。
k は、$k = 1.96$ で確率 5%(両側計)、$k = 2.60$ で確率1%(同)であるから、サンプルがこの範囲に
入ったら、この母集団には属しないものと判断し、それに伴う(誤りの) 確率を「棄却確率 5%
(信頼水準 95%)」あるいは「棄却確率 1%､(信頼水準 99%)」と呼んでいる。