確率・統計

＜個数の処理＞

基本的には樹形図で数え上げげする

その他の方法として、

順列：${}_nP_r$

組合せ：${}_nC_r$

重複順列：n^r

同じものを含む順列：$\dfrac{n!}{p!q!r!\cdots}$

重複組合せ：${}_nH_r = {}_{n+r-1}C_r$

円順列：$(n-1)!$

じゅず順列：$\dfrac{(n-1)!}{2}$

組分けについて…同じ数ずつに分けるとき、A，B，Cのような区別があるかないか注意

＜二項定理＞

$$(a+b)^n = {}_nC_0a^n + {}_nC_1a^{n-1}b + \cdots + {}_nC_ra^{n-r}b^r + \cdots + {}_nC_nb^n$$

パスカルの三角形を利用できること

多項定理：$(a+b+c+\cdots)^n$ の展開式で、$a^pb^qc^r\cdots$ の係数は、$\dfrac{n!}{p!q!r!\cdots}$ である。

＜確率＞

$P(A) = \dfrac{n(A)}{n(U)}$　　$0 \le \boldsymbol{P(A)} \le 1$

余事象の確率

$\boldsymbol{P(\overline{A}) = 1 - P(A)}$

加法定理（場合分け）

$\boldsymbol{P(A \cup B) = P(A) + P(B)}$

注意：背反ではない場合は

$\boldsymbol{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$

事象 $\boldsymbol{A}$ が起こるが、事象 $\boldsymbol{B}$ が起こらない確率

$\boldsymbol{P(A \cap \overline{B}) = P(A) - P(A \cap B)}$

独立試行の確率（引き続き起こる）

$P(C) = P(A)P(B)$

反復試行の確率

${}_nC_rp^rq^{n-r}$　ここで、$\boldsymbol{q = 1 - p}$

＜条件付き確率＞

事象 $\boldsymbol{A}$ が起こったときの、事象 $\boldsymbol{B}$ が起こる確率

$$\boldsymbol{P_A(B) = \frac{P(A \cap B)}{P(A)}}$$

したがって、乗法定理

$\boldsymbol{P(A \cap B) = P(A)P_A(B)}$ を得る

＜確率分布と統計的な推測＞

基本は確率変数を読み取り確率分布表

X	$\boldsymbol{x_1\ x_2\ \cdots\ x_{n-1}\ x_n}$
P	$\boldsymbol{p_1\ p_2\ \cdots\ p_{n-1}\ p_n}$

（$\boldsymbol{p_1 + p_2 + \cdots + p_{n-1} + p_n = 1}$）

から、

期待値（平均）：

$$\boldsymbol{E(X) = m = x_1p_1 + x_2p_2 + \cdots + x_{n-1}p_{n-1} + x_np_n}$$

$$\boldsymbol{= \sum_{k=1}^{n} x_kp_k}$$

分散：

$$\boldsymbol{V(X) = E((X-m)^2) = (x_1-m)^2p_1 + \cdots + (x_n-m)^2p_n}$$

$$\boldsymbol{= \sum_{k=1}^{n}(x_k-m)^2p_k = E(X^2) - \{E(X)^2\}}$$

（分散は、二乗の平均から平均の二乗を引く）

標準偏差：$\boldsymbol{\sigma(X) = \sqrt{V(X)}}$

＜確率変数の変換＞（旧数C→数B）

$\boldsymbol{a}$,$\boldsymbol{b}$ は定数として

$\boldsymbol{Y = aX + b}$ のとき期待値（平均）、分散、標準偏差は

$\boldsymbol{E(Y) = aE(X) + b}$

$\boldsymbol{V(Y) = a^2V(X)}$

$\boldsymbol{\sigma(Y) = |a|\sigma(X)}$

＜確率変数の和と積＞（旧数C→数B）

$\boldsymbol{E(X+Y) = E(X) + E(Y)}$

$\boldsymbol{E(aX+bY) = aE(X) + bE(Y)}$

ここでさらに、$\boldsymbol{X}$,$\boldsymbol{Y}$ が互いに独立ならば

$\boldsymbol{E(XY) = E(X)E(Y)}$

$\boldsymbol{V(X+Y) = V(X) + V(Y)}$

$\boldsymbol{V(aX+bY) = a^2V(X) + b^2V(Y)}$

＜二項分布＞

二項分布：試行を $\boldsymbol{n}$ 回繰り返し、同じ確率 $\boldsymbol{p}$ で $\boldsymbol{r}$ 回起こるとき $\boldsymbol{(r = 0,1,2,\cdots,n)}$

二項分布 $\boldsymbol{B(n,p)}$ と書く、$\boldsymbol{P(X=r) = {}_nC_rp^r(1-p)^{n-r}}$

期待値（平均）$\boldsymbol{E(X) = np}$、分散 $\boldsymbol{V(X) = np(1-p)}$

＜確率変数が連続のとき＞（旧数C→数B）

連続型確率変数 $\boldsymbol{X}$ の確率密度関数を $\boldsymbol{f(x)\ (\alpha \le x \le \beta)}$ とは

①　$\boldsymbol{f(x)} \ge 0$（常に正）

②　$\displaystyle\int_\alpha^\beta \boldsymbol{f(x)dx} = 1$　（全面積は１）

③　$\displaystyle\boldsymbol{P(a \le X \le b) = \int_a^b f(x)dx}$

（$\boldsymbol{a \le X \le b}$ となる確率は面積を計算して求める）

期待値（平均） $E(X) = m = \int_\alpha^\beta xf(x)dx$、

分散 $V(X) = \int_\alpha^\beta (x-m)^2 f(x)dx$

＜正規分布＞（旧数C→数B）

最も有名な確率分布で、正規分布 $N(m, \sigma^2)$ と書く

ここで、$m = E(X)$ 平均、$\sigma^2 = V(X)$ 分散である。

通常は標準正規分布 $N(0,1)$ に変換して扱われる。

標準化：$Z = \dfrac{X-m}{\sigma}$ とおく

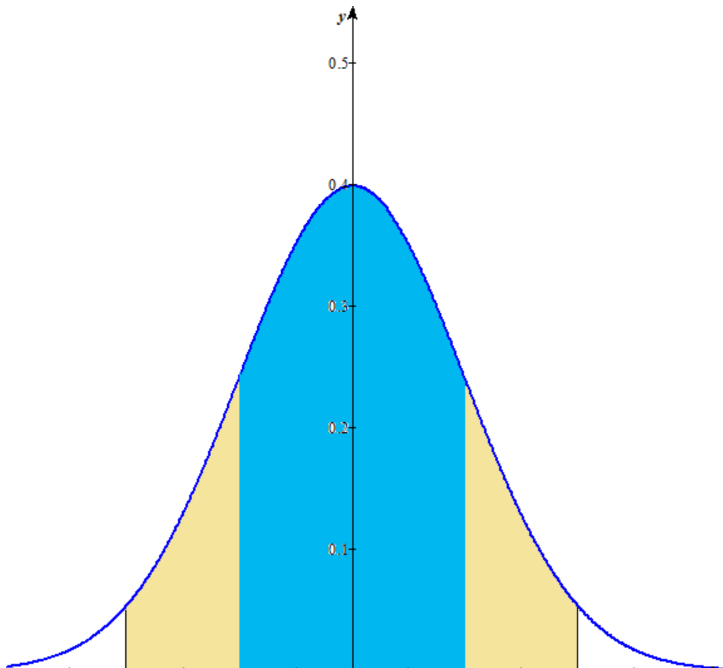
入学試験の偏差値（合格可能性の判定）等々、統計（仮説検定・区間推定）の分野でお馴染みである。

参考１：$N(0,1)$ は、$f(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$N(m,\sigma^2)$ は、$f(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$

（ここで e はネピア数と呼ばれ自然対数の底として有名な無理数である。$e = 2.718\cdots$ である。）

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$



$$P(-1 \le X \le 1) = \int_{-1}^{1} f(x)dx \cong 0.682$$

$$P(-2 \le X \le 2) = \int_{-2}^{2} f(x)dx \cong 0.955$$

$$P(-3 \le X \le 3) = \int_{-3}^{3} f(x)dx \cong 0.997$$

標準偏差の ± 1 の間に入る確率は６８．２％

標準偏差の ± 2 の間に入る確率は９５．５％

標準偏差の ± 3 の間に入る確率は９９．７％である。

また、

標準偏差の ± 1.96 の間に入る確率は９５％

標準偏差の ± 2.58 の間に入る確率は９９％

$X < -1.64$ または $1.64 < X$ に入る確率は５％

を用いて推定や検定という手法が使われる。

参考２：区間推定と仮説検定

かつて、高校でも指導していた内容をまとめる

＜区間推定＞（旧数C→数B）

（Ⅰ）母平均の推定（信頼区間を求める）

ア）信頼度95%のとき

区間は $\left(\bar{x} - 1.96 \cdot \dfrac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \cdot \dfrac{\sigma}{\sqrt{n}}\right)$

区間の幅は $2 \times 1.96 \cdot \dfrac{\sigma}{\sqrt{n}}$

イ）信頼度99%のとき

区間は $\left(\bar{x} - 2.58 \cdot \dfrac{\sigma}{\sqrt{n}}, \bar{x} + 2.58 \cdot \dfrac{\sigma}{\sqrt{n}}\right)$

区間の幅は $2 \times 2.58 \cdot \dfrac{\sigma}{\sqrt{n}}$

ここで、$\bar{x}$：標本平均の値、n：標本の大きさ（個数）

σ：母標準偏差の値（未知なら標本標準偏差も使用可）

を代入して使用する。

（Ⅱ）母比率の推定

ア）信頼度95%のとき

区間は $\left(\bar{p} - 1.96 \cdot \sqrt{\dfrac{\bar{p}(1-\bar{p})}{n}}, \bar{p} + 1.96 \cdot \sqrt{\dfrac{\bar{p}(1-\bar{p})}{n}}\right)$

区間の幅は $2 \times 1.96 \cdot \sqrt{\dfrac{\bar{p}(1-\bar{p})}{n}}$

イ）信頼度99%のとき

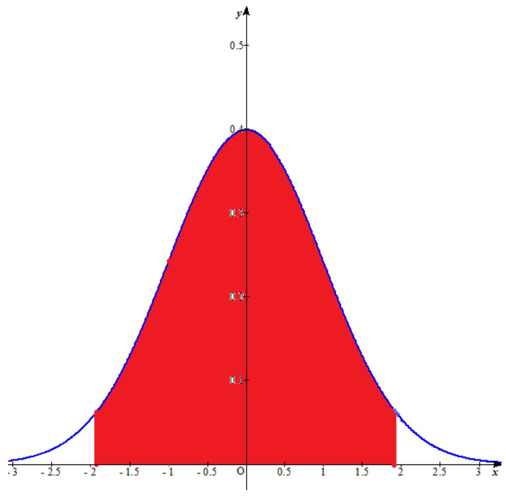
区間は $\left(\bar{p} - 2.58 \cdot \sqrt{\dfrac{\bar{p}(1-\bar{p})}{n}}, \bar{p} + 2.58 \cdot \sqrt{\dfrac{\bar{p}(1-\bar{p})}{n}}\right)$

区間の幅は $2 \times 2.58 \cdot \sqrt{\dfrac{\bar{p}(1-\bar{p})}{n}}$

ここで、$\bar{p}$：標本比率の値、n：標本の大きさ（個数）

を代入して使用する。

※推定は信頼区間の幅を変えることにより、%を変えることが可能



上図の赤い部分に確率変数が入るのが通常であり、
その意味で信頼区間と呼ばれている

＜仮説検定＞（旧数C→数B）

（Ⅰ）母平均の検定

① 仮説を立てる（母平均$m = m'$とする）

② 有意水準（危険率）を５％または１％とする

③ 棄却域：$|X| > 1.96 \Rightarrow 5\%$または$|X| > 2.58 \Rightarrow 1\%$

④ 正規分布$N(0,1)$に従うとして$u = \dfrac{\bar{x} - m'}{\dfrac{\sigma}{\sqrt{n}}}$を計算する

ここで、m'が仮説平均　$\bar{x}$：標本平均の値、n：標本の大きさ（個数）

⑤ 定めた棄却域に入るか調べる

・入っていれば仮説を棄却する
（$m = a$とは言えないと結論付ける）

・入っていなければ仮説は棄却できない
（$m = a$でないとは言い切れない）

（Ⅱ）母比率の検定

① 仮説を立てる（母比率$p = p'$とする）

② 有意水準（危険率）を５％または１％とする

③ 棄却域：$|X| > 1.96 \Rightarrow 5\%$または$|X| > 2.58 \Rightarrow 1\%$

④ 正規分布$N(0,1)$に従うとして$u = \dfrac{x - np'}{\sqrt{np'(1-p')}}$を計算する

ここで、xは標本の値、p'が仮説比率　n：標本の大きさ（個数）

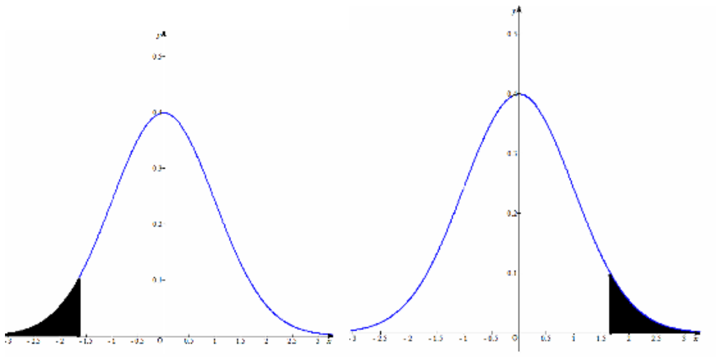
⑤ 定めた棄却域に入るか調べる

・入っていれば仮説を棄却する
（$p = p'$とは言えないと結論付ける）

・入っていなければ仮説は棄却できない
（$p = p'$でないとは言い切れない）

※初めから大きい値に外れるか小さい値に外れるかが明らかな場合は、片側検定を行うこともある

例えば棄却域を$X < -1.64$または$1.64 < X$として、ここに入る確率は５％を用いる



※また、正規分布でなくｔ分布やχ^2分布（ｶｲｼﾞｼﾞｮｳﾌﾞﾝﾌﾟ）を用いるばあいもある。