数学Ｂ　統計的な推測「統計的な推測」

~高校数学のまとめ~

母集団分布

大きさ N の母集団において，

変量 x の異なるとりうる値を x_1, x_2, $\cdots$, x_r とし，

それぞれの個体の個数を f_1, f_2, $\cdots$, f_r とする．

ここで $f_1+f_2+\cdots+f_r=N$

この母集団から1個の個体を無作為抽出するとき

その個体における変量 x の値 X は偶然に決まるが

$X=x_k$ となる確率は $\dfrac{f_k}{N}$ $(k=1,\ 2,\ \cdots,\ r)$

つまり，次のような確率分布となる．

階級値	度数
x_1	f_1
x_2	f_2
$\vdots$	$\vdots$
x_r	f_r
計	N

X	x_1	x_2	$\cdots$	x_k	$\cdots$	x_r	計
P	$\dfrac{f_1}{N}$	$\dfrac{f_2}{N}$	$\cdots$	$\dfrac{f_k}{N}$	$\cdots$	$\dfrac{f_r}{N}$	1

この確率分布は，母集団の相対度数の分布と一致する．

母集団における変量 x の平均を m，標準偏差を σ

確率変数 X の期待値を $E(X)$，標準偏差を $\sigma(X)$ とすると

$$E(X)=m,\ \ \sigma(X)=\sigma$$

一般に 母集団における変量 x の分布を **母集団分布**（ぼしゅうだんぶんぷ） といい，

その平均値を **母平均**（ぼへいきん），分散を**母分散**（ぼぶんさん），標準偏差を **母標準偏差**（ぼひょうじゅんへんさ） という．

大きさ1の無作為標本における変量 x の値 X は母集団分布に従う．

確率変数 X の期待値，分散，標準偏差は

それぞれ母平均，母分散，母標準偏差に一致する．

㊟補 この確率分布は，母集団において調査の対象となっている変量を特徴づけるものである．

復元抽出・非復元抽出と確率変数

母集団の中から大きさ n の標本を無作為に抽出し，

その n 個の個体における変量の値を X_1, X_2, $\cdots$, X_n とする．

$\boxed{1}$ 復元抽出のとき

母集団から大きさ1の標本を無作為に抽出する試行を n 回繰り返す

反復試行であるから，X_1, X_2, $\cdots$, X_n は

それぞれが母集団分布に従う互いに独立な確率変数となる．

$\boxed{2}$ 非復元抽出のとき

標本は n 個の確率変数 X_1, X_2, $\cdots$, X_n であるが，

これらは互いに独立ではない．

しかし，母集団の大きさが標本の大きさ n に比べて十分大きいときは

母集団の構成にはほとんど変化がないから復元抽出との違いは小さくなる．

すなわち，非復元抽出で取り出した標本も近似的に復元抽出で取り出した

標本とみなすことができ，X_1, X_2, $\cdots$, X_n は

それぞれが母集団分布に従う互いに独立な確率変数となる．

補 $\boxed{2}$ 母集団の大きさが標本の大きさ n に比べて十分大きいときの目安は
母集団の大きさを N として　n は N の10分の1以下 $\left(n \leqq \dfrac{N}{10}\right)$

標本平均

母集団から無作為抽出する大きさ n の標本の変量を X_1, X_2, $\cdots$, X_n とする.

これらの平均 $\overline{X}=\dfrac{X_1+X_2+\cdots+X_n}{n}$ を 標本平均（ひょうほんへいきん）という.

標本平均 $\overline{X}$ は，抽出される標本によって変化する確率変数である.

例 あるメーカーで生産されたポテトチップスの袋を無作為に 10 袋抽出すると，
重さが単位を g として

101，99，101，101，102，103，98，105，102，98

であった.

これらの平均は

$$\frac{101+99+101+101+102+103+98+105+102+98}{10}=101$$

これが標本平均である.

このことをふまえ，このメーカーのポテトチップスの袋の容量はどのくらい？
と考えるのが「母平均の推定」である.

標本平均の期待値と標準偏差

母平均 m，母標準偏差 σ の母集団から大きさ n の無作為標本を復元抽出するとき，その標本平均を $\overline{X}$ とすると，次が成り立つ.

1 $\overline{X}$ の期待値は $E\left(\overline{X}\right)=m$

2 $\overline{X}$ の分散は $V\left(\overline{X}\right)=\dfrac{\sigma^2}{n}$

3 $\overline{X}$ の標準偏差は $\sigma\left(\overline{X}\right)=\dfrac{\sigma}{\sqrt{n}}$

考 母平均 m，母標準偏差 σ の母集団から大きさ n の無作為標本の変量を X_1，X_2，…，X_n とする.

1
$$\begin{aligned}E\left(\overline{X}\right)&=E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)\\&=\frac{1}{n}\{E(X_1)+E(X_2)+\cdots+E(X_n)\}\\&=\frac{1}{n}(\underbrace{m+m+\cdots+m}_{n\text{ 個}})\\&=\frac{1}{n}\cdot nm\\&=m\end{aligned}$$

2
$$\begin{aligned}V\left(\overline{X}\right)&=V\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)\\&=\left(\frac{1}{n}\right)^2\{V(X_1)+V(X_2)+\cdots+V(X_n)\}\\&=\frac{1}{n^2}(\underbrace{\sigma^2+\sigma^2+\cdots+\sigma^2}_{n\text{ 個}})\\&=\frac{1}{n^2}\cdot n\sigma^2\\&=\frac{\sigma^2}{n}\end{aligned}$$

3
$$\sigma\left(\overline{X}\right)=\sqrt{V\left(\overline{X}\right)}=\sqrt{\frac{\sigma^2}{n}}=\frac{\sigma}{\sqrt{n}}$$

補 非復元抽出の場合も，n に比べて母集団の大きさが十分大きいならば復元抽出と同様に扱える.

例 母平均20，母分散4の母集団から大きさ25の標本を復元抽出するときの標本平均 $\overline{X}$ について，

平均 $E(\overline{X})=20$

分散 $V(\overline{X})=\dfrac{4}{25}=0.16$

標準偏差 $\sigma(\overline{X})=\dfrac{2}{\sqrt{25}}=\dfrac{2}{5}=0.4$

大数の法則

母平均 m の母集団から大きさ n の無作為標本を抽出するとき

その標本平均 $\overline{X}$ は n を大きくすると母平均 m に近づく.

すなわち $\displaystyle\lim_{n\to\infty}\overline{X}=m$

㊯ 標本平均 $\overline{X}$ の分散を $V(\overline{X})$ として

$$\lim_{n\to\infty}V(\overline{X})=\lim_{n\to\infty}\frac{\sigma^2}{n}=0$$

すなわち，n を大きくすると $\overline{X}$ は母平均 m の近くに限りなく集中して分布する.

㊫ n を大きくすると $\overline{X}$ が m に近い値をとる確率が 1 に近づく.

㊬ 直観的にも n が大きければ大きいほど，標本平均は母平均（全体の平均）に落ち着く.

標本平均の分布の正規分布による近似

母平均 m，母分散 σ^2 の母集団から

無作為抽出された大きさ n の標本平均 $\overline{X}$ の分布は n が十分大きいとき

近似的に正規分布 $N\left(m,\ \dfrac{\sigma^2}{n}\right)$ に従う．

このことから

$$Z=\frac{X-m}{\dfrac{\sigma}{\sqrt{n}}}$$

とおくと 確率 Z は標準正規分布 $N(0,\ 1)$ に従う．

例 母平均 20，母分散 4 の母集団から大きさ 25 の標本を復元抽出するときの標本平均 $\overline{X}$ について，25 が十分大きいとして $N(20,\ 0.16)$ に従うとみなすことができる．

母平均に対する信頼度 95 %の信頼区間

母平均 m，母標準偏差 σ の母集団から

抽出された大きさ n の無作為標本の標本平均 $\overline{X}$ を考える．

n が大きいとき，母平均 m に対する信頼度 95 %の信頼区間は

$$\overline{X}-1.96\cdot\frac{\sigma}{\sqrt{n}} \leqq m \leqq \overline{X}+1.96\cdot\frac{\sigma}{\sqrt{n}}$$

すなわち

$$\left[\overline{X}-1.96\cdot\frac{\sigma}{\sqrt{n}}\ ,\ \overline{X}+1.96\cdot\frac{\sigma}{\sqrt{n}}\right]$$

㊟ 「母平均 m に対して信頼度 95 %の信頼区間を求める」ことを

「母平均 m を信頼度 95 %で推定する」ということがある．

㊫ n が大きいとき，$\overline{X}$ の確率分布は近似的に正規分布 $N\left(m,\ \dfrac{\sigma^2}{n}\right)$ に従う．

$Z=\dfrac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}}$ とおくと Z は標準正規分布 $N(0,\ 1)$ に従う．

正規分布表から　$P\left(|Z|\leqq 1.96\right)=2\times P(0\leqq Z\leqq 1.96)=2\times 0.4750=0.95$

すなわち，母平均 m に対する信頼度 95 %の信頼区間は　$|Z|\leqq 1.96$ として

$$\left|\frac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}}\right|\leqq 1.96 \iff \overline{X}-1.96\cdot\frac{\sigma}{\sqrt{n}}\leqq m\leqq \overline{X}+1.96\cdot\frac{\sigma}{\sqrt{n}}$$

㊬ $P\left(|Z|\leqq 2.58\right)=2\times P(0\leqq Z\leqq 2.58)=2\times 0.4951=0.9902\fallingdotseq 0.99$

これより，母平均 m に対する信頼度 99 %の信頼区間は

$$\overline{X}-2.58\cdot\frac{\sigma}{\sqrt{n}}\leqq m\leqq \overline{X}+2.58\cdot\frac{\sigma}{\sqrt{n}}$$

㊡ 標準偏差が 10 の母集団から抽出した大きさ 25 の標本の平均が 150 とする．

母平均 m に対する信頼度 95 %の信頼区間は

$$100-1.96\cdot\frac{10}{\sqrt{25}}\leqq m\leqq 100+1.96\cdot\frac{10}{\sqrt{25}}$$

よって　$96.08\leqq m\leqq 103.92$

母平均に対する信頼度 D %の信頼区間の幅

母平均 m に対する信頼度 D %の信頼区間を $A \leqq m \leqq B$ とするとき

この信頼区間の幅は　$B - A$

考 $\boxed{\text{母平均に対する信頼度 95 \%の信頼区間}}$の幅は

$$\overline{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} - \left(\overline{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 2 \cdot 1.96 \cdot \frac{\sigma}{\sqrt{n}} = 3.92 \cdot \frac{\sigma}{\sqrt{n}}$$

例 標準偏差が 10 の母集団から抽出した大きさ 25 の標本の平均が 150 とする．
母平均 m に対する信頼度 95 %の信頼区間の幅は

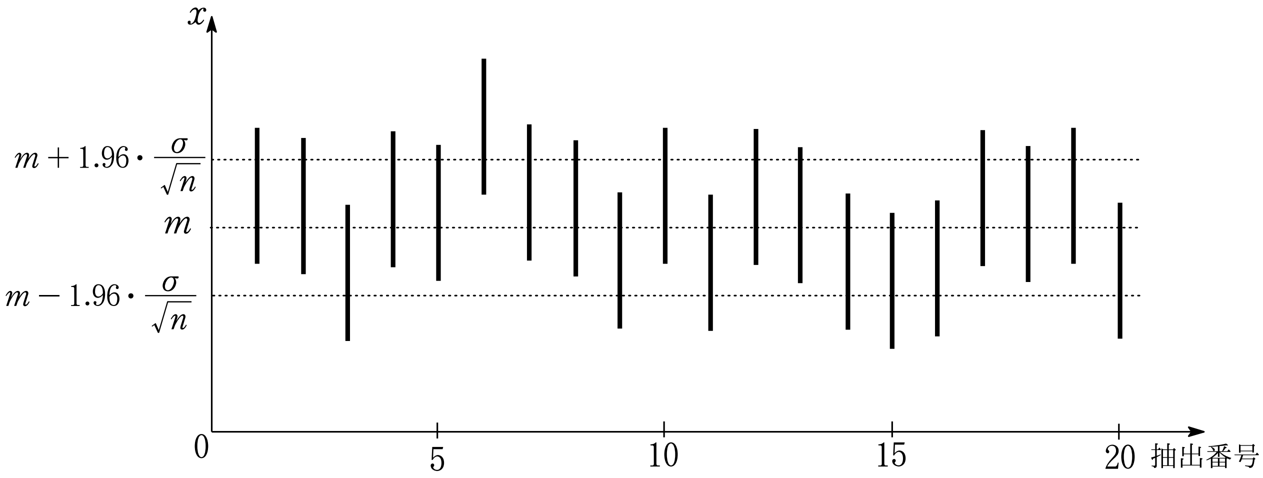
$$2 \cdot 1.96 \cdot \frac{10}{\sqrt{25}} = 3.92 \cdot 2 = 7.84$$

母平均に対する信頼度 95 ％の信頼区間の意味

母平均 m に対する信頼度 95 ％の信頼区間とは

無作為抽出を繰り返し，このような区間を，例えば 100 個作ると，

m を含む区間が 95 個位あることを意味する．



㊓ 上図では抽出番号 20 までの 20 個の区間を作り，抽出番号 6 の区間以外の 19 個の区間は m を含んでいる．

母比率と標本比率

母集団の中で，ある特性 A をもつ個体の割合を p とする．

この p を特性 A をもつ個体の母集団における **母比率**（ぼひりつ）という．

標本の中で特性 A をもつ個体の割合を **標本比率**（ひょうほんひりつ）という．

㊠ ある工場で，製品の中から無作為に100個選んで調べてたら，10個の不良品があった．
このとき，不良品の比率は $\dfrac{10}{100}=0.1$
これが標本比率である．
このことをふまえ，この工場での不良品の比率はどのくらい？
と考えるのが「母比率の推定」である．

標本比率の分布の正規分布による近似

特性 A をもつ個体の母比率が p である母集団から

大きさ n の標本を無作為抽出し

その標本に含まれる特性 A をもつ個体の個数を X とする.

このとき 確率変数 X は 二項分布 $B(n,\ p)$ に従い, n が大きいとき

近似的に正規分布 $N\big(np,\ np(1-p)\big)$ に従う.

また 標本比率を $\dfrac{X}{n}=R$ とすると

確率変数 R は正規分布 $N\left(p,\ \dfrac{p(1-p)}{n}\right)$ に従う.

このことから

$$Z=\frac{R-p}{\sqrt{\dfrac{p(1-p)}{n}}}$$

とおくと 確率 Z は標準正規分布 $N(0,\ 1)$ に従う.

考 確率変数 X は $B(n,\ p)$ に従うので $E(X)=np,\ V(X)=np(1-p)$

確率変数 R の平均は $E(R)=E\left(\dfrac{X}{n}\right)=\dfrac{1}{n}E(X)=\dfrac{1}{n}\cdot np=p$

確率変数 R の分散は $V(R)=V\left(\dfrac{X}{n}\right)=\left(\dfrac{1}{n}\right)^2V(X)=\dfrac{1}{n^2}\cdot np(1-p)=\dfrac{p(1-p)}{n}$

母比率に対する信頼度95％の信頼区間

母比率 p の母集団から

抽出された大きさ n の無作為標本の標本比率 R を考える．

n が大きいとき，母平均 p に対する信頼度95％の信頼区間は

$$R-1.96\sqrt{\frac{R(1-R)}{n}} \leqq p \leqq R+1.96\sqrt{\frac{R(1-R)}{n}}$$

すなわち

$$\left[R-1.96\sqrt{\frac{R(1-R)}{n}}\ ,\ R+1.96\sqrt{\frac{R(1-R)}{n}}\ \right]$$

注 「母比率 p に対して信頼度95％の信頼区間を求める」ことを
「母比率 p を信頼度95％で推定する」ということがある．

考 n が大きいとき，R の確率分布は近似的に正規分布 $N\left(p,\ \dfrac{p(1-p)}{n}\right)$ に従う．

$Z=\dfrac{R-p}{\sqrt{\dfrac{p(1-p)}{n}}}$ とおくと Z は標準正規分布 $N(0,\ 1)$ に従う．

正規分布表から　$P\big(|Z|\leqq 1.96\big)=2\times P(0\leqq Z\leqq 1.96)=2\times 0.4750=0.95$

すなわち，母比率 p に対する信頼度95％の信頼区間は　$|Z|\leqq 1.96$ として

$$\left|\frac{R-p}{\sqrt{\dfrac{p(1-p)}{n}}}\right|\leqq 1.96$$

$$\iff\ R-1.96\sqrt{\frac{p(1-p)}{n}}\leqq p\leqq R+1.96\sqrt{\frac{p(1-p)}{n}}$$

n が十分大きいときは R を p とみなしてよいので，根号内の p を R で置きかえて

$$R-1.96\sqrt{\frac{R(1-R)}{n}}\leqq p\leqq R+1.96\sqrt{\frac{R(1-R)}{n}}$$

補 $P\big(|Z|\leqq 2.58\big)=2\times P(0\leqq Z\leqq 2.58)=2\times 0.4951=0.9902\fallingdotseq 0.99$

これより，母平均 m に対する信頼度99％の信頼区間は

$$R-2.58\sqrt{\frac{R(1-R)}{n}}\leqq p\leqq R+2.58\sqrt{\frac{R(1-R)}{n}}$$

例 ある工場で，製品の中から無作為に100個選んで調べてたら，10個の不良品があった．

この工場で作られる製品の不良率（不良品の比率）p に対する信頼度95％の信頼区間は

$$0.1-1.96\sqrt{\frac{0.1\cdot 0.9}{100}}\leqq p\leqq 0.1+1.96\sqrt{\frac{0.1\cdot 0.9}{100}}$$

よって　$0.0412\leqq p\leqq 0.1588$

母比率に対する信頼度 D %の信頼区間の幅

母比率 p に対する信頼度 D %の信頼区間を $A \leqq p \leqq B$ とするとき

この信頼区間の幅は $B-A$

例 $\boxed{\text{母比率に対する信頼度 95 %の信頼区間}}$ の幅は

$$R+1.96\sqrt{\frac{R(1-R)}{n}}-\left(R+1.96\sqrt{\frac{R(1-R)}{n}}\right)=2\cdot1.96\sqrt{\frac{R(1-R)}{n}}$$

$$=3.92\sqrt{\frac{R(1-R)}{n}}$$

例 ある工場で，製品の中から無作為に100個選んで調べてたら，10個の不良品があった．

母平均 m に対する信頼度95 %の信頼区間の幅は

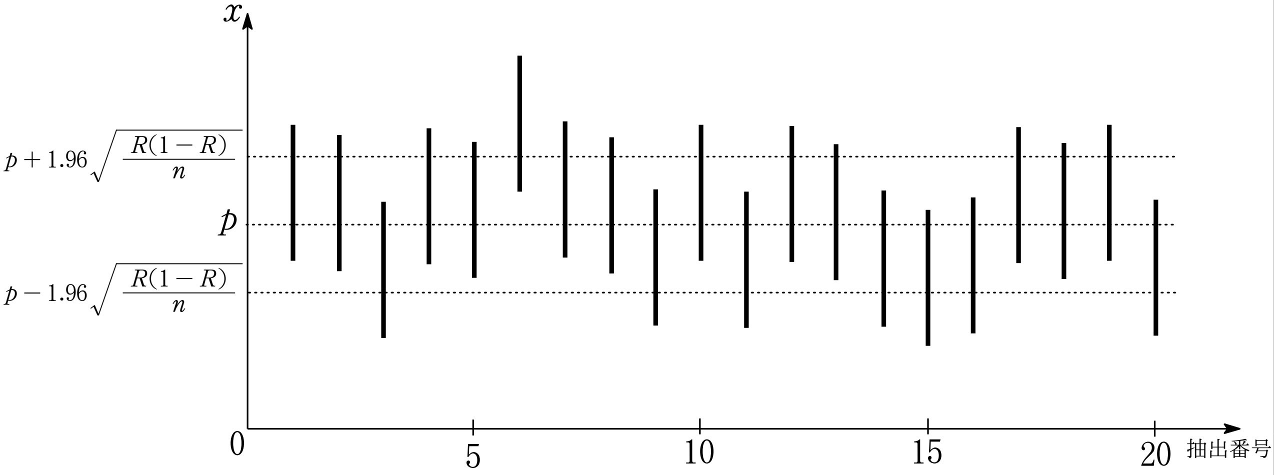
$$2\cdot1.96\sqrt{\frac{0.1\cdot0.9}{100}}=3.92\cdot0.03=0.1176$$

母比率に対する信頼度 95 %の信頼区間の意味

母比率 p に対する信頼度 95 %の信頼区間とは

無作為抽出を繰り返し，このような区間を，例えば 100 個作ると，

p を含む区間が 95 個位あることを意味する．



㊓ 上図では抽出番号 20 までの 20 個の区間を作り，抽出番号 6 の区間以外の 19 個の区間は p を含んでいる．

仮説検定

標本調査の結果をもとに，母集団に関する主張が正しいことを判断するのに，

仮説を棄却（ききゃく）することで正しいと判断する

または

仮説を棄却しないで正しいとは判断できない

ことを決める方法を **仮説検定**（かせつけんてい） という．

(補)「棄却する」とは「棄（す）て去（さ）る」こと

(話) 数学 Ｉ では仮説検定を実験から考えたが，数学Ｂでは正規分布を用いて考える．

仮説検定の考え方

標本調査の結果をもとに，母集団に関する主張 H_1 が正しいと判断するのに，

主張 H_1 に反する仮説の主張 H_0 を立てる.

ここで，H_0 を帰無（きむ）仮説，H_1 を対立（たいりつ）仮説という.

帰無仮説 H_0 のもとで，棄却すべき確率 p を求める.

このとき，起こりやすさの基準となる確率 α を定めておく.

この α を有意水準（ゆういすいじゅん）といい，0.05(5 %) とする場合が多い.

p と α の大小関係で，次の 2 つの場合になる.

$\boxed{1}$ $p < \alpha$ ならば

p が小さすぎるので，H_0 は棄却され，H_1 が正しいと判断できる.

$\boxed{2}$ $p > \alpha$ ならば

p が小さくなく，H_0 は棄却されず，H_1 が正しいと判断できない.

補 $\boxed{2}$ H_1 が正しくないと判断するわけではない.

補 おおざっぱな説明だが，対立仮説は「本当はこうではないのかという仮説」
帰無仮説は「無かったことにしたい仮説」

補 有意水準 α で仮説検定を行うことを「有意水準 α で検定する」ということがある.

棄却域

有意水準 α の仮説検定について，

仮説が棄却されるような確率変数の値が定まる．

この範囲を有意水準 α の**棄却域**（ききゃくいき）という．

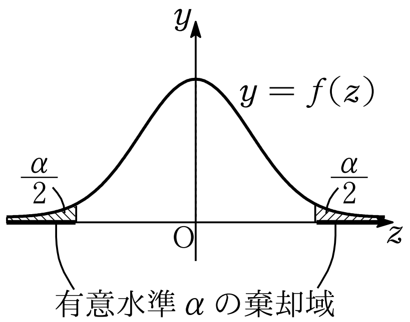
とくに

棄却域を両側にとっている検定を **両側検定** という．

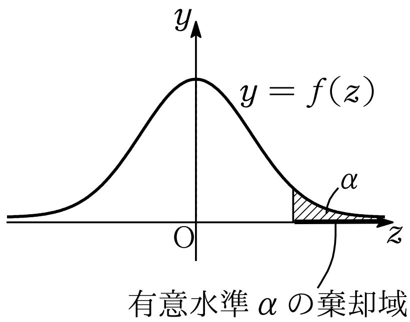
棄却域を片側にとっている検定を **片側検定** という．

正規分布を用いるときは，次のようになる．

〔両側検定〕



〔片側検定〕



㊥ 片側検定は上の図で右側のみであるが，左側の場合もある．

標本平均の検定（両側検定）

母平均 m_1，標準偏差 σ の正規分布に従うはずの母集団から大きさ n の標本を取り出したとき，その標本平均が a で，m_1 とは異なる値とする．

その結果から 母平均は m_1 ではないと判断できるかを仮説検定する．

そこで，母平均を m として

帰無仮説 H_0 ： $m = m_1$

を立てると

対立仮説 H_1 ： $m \neq m_1$

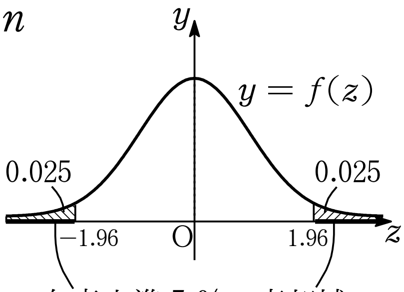
ここで，有意水準を 5 ％とし，n は十分大きいとする．

帰無仮説 H_0 が正しいとすると標本平均 $\overline{X}$ は，$N(m,\ \sigma^2)$ に従うので

$$Z = \frac{\overline{X} - m}{\dfrac{\sigma}{\sqrt{n}}}$$

として，Z の分布は $N(0,\ 1)$ に従う．

$\overline{X} = a$ としたときの Z の値を z_0 とすると $z_0 = \dfrac{a - m}{\dfrac{\sigma}{\sqrt{n}}}$



有意水準 5 ％の棄却域

$\boxed{1}$ 確率 $P\left(|Z| \geqq |z_0|\right) < 0.05$ ならば

H_0 は棄却され，H_1 が正しいと判断できる．

とくに $P\left(|Z| \geqq 1.96\right) = 0.05$ であるから

$|z_0| > 1.96$ であれば $P\left(|Z| \geqq |z_0|\right) < 0.05$

すなわち，有意水準 5 ％の棄却域は

$$|Z| > 1.96 \iff Z < -1.96 \text{ または } 1.96 < Z$$

$\boxed{2}$ 確率 $P\left(|Z| \geqq |z_0|\right) > 0.05$ ならば

H_0 は棄却されず，H_1 が正しいと判断できない．

例題　ある会社ではお菓子を1個平均200g，標準偏差4gの正規分布に従うように製造している．その中から16個を無作為抽出して調査すると，平均が198gであった．
この結果から「お菓子の1個平均は200gではない」と判断できるか．有意水準5％で仮説検定せよ．

解 製造されるお菓子の母平均を m とする．
このとき

帰無仮説 H_0：$m = 200$

対立仮説 H_1：$m \neq 200$

帰無仮説 H_0 が正しいとすると，標本平均 $\overline{X}$ は $N\left(200,\ \dfrac{4^2}{16}\right)$ に従う．

$$Z = \frac{\overline{X} - 200}{\frac{4}{\sqrt{16}}}$$

として，Z は $N(0,\ 1)$ に従う．

$\overline{X} = 198$ としたときの Z の値を z_0 とすると　$z_0 = \dfrac{198 - 200}{\frac{4}{\sqrt{16}}} = -2$

これは棄却域 $|Z| > 1.96$ に入っているから 確率 $P(|Z| \geqq 2) < 0.05$ であり，H_0 は棄却される．

よって，**お菓子の1個平均は200gではないと判断できる．**

標本平均の検定（片側検定）

母平均 m_1，標準偏差 σ の正規分布に従うはずの母集団から大きさ n の標本を取り出したとき，その標本平均が a で，m_1 より大きな値とする．

その結果から 母平均は m_1 より大きいと判断できるかを仮説検定する．

そこで，母平均を m として

帰無仮説 H_0 ： $m = m_1$

を立てると

対立仮説 H_1 ： $m > m_1$

ここで，有意水準を５％とし，n は十分大きいとする．

帰無仮説 H_0 が正しいとすると標本平均 $\overline{X}$ は，$N(m,\ \sigma^2)$ に従うので

$$Z = \frac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}}$$

として，Z の分布は $N(0,\ 1)$ に従う．

$\overline{X} = a$ としたときの Z の値を z_0 とすると　$z_0 = \dfrac{a - m}{\frac{\sigma}{\sqrt{n}}}$ (> 0)

1 確率 $P\left(Z \geqq z_0\right) < 0.05$ ならば

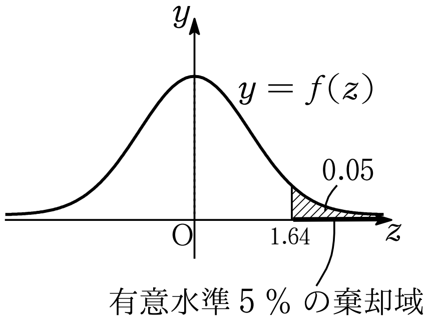
H_0 は棄却され，H_1 が正しいと判断できる．

とくに $P\left(Z \geqq 1.64\right) \fallingdotseq 0.05$ であるから

$z_0 > 1.64$ であれば $P\left(Z \geqq z_0\right) < 0.05$

すなわち，有意水準５％の両側検定での棄却域は

$Z > 1.64$



2 確率 $P\left(Z \geqq z_0\right) > 0.05$ ならば

H_0 は棄却されず，H_1 が正しいと判断できない．

補 $P(Z \geqq 1.64) = P(Z \geqq 0) - P(0 \leqq Z \leqq 1.64) = 0.5 - 0.4495 = 0.0505 \fallingdotseq 0.05$

補 帰無仮説「$m = m_1$」に対し対立仮説は「$m \neq m_1$」「$m > m_1$」「$m < m_1$」などがある．

母比率の検定（両側検定）

特性 A の母比率 p_1 の正規分布に従うはずの母集団から大きさ n の標本を取り出したとき，その比率が q で，p_1 とは異なる値とする．

その結果から 母比率は p_1 ではないと判断できるかを仮説検定する．

そこで，母比率を p として

帰無仮説 H_0 ： $p = p_1$

を立てると

対立仮説 H_1 ： $p \neq p_1$

ここで，有意水準を 5 ％とし，n は十分大きいとする．

帰無仮説 H_0 が正しいとすると

特性 A の個数 X は，$N\big(np,\ np(1-p)\big)$ に従うので

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

として，Z の分布は $N(0,\ 1)$ に従う．

$\dfrac{X}{n} = q$ としたときの Z の値を z_0 とすると

$$z_0 = \frac{nq - np}{\sqrt{np(1-p)}} = \frac{q-p}{\sqrt{\dfrac{p(1-p)}{n}}}$$

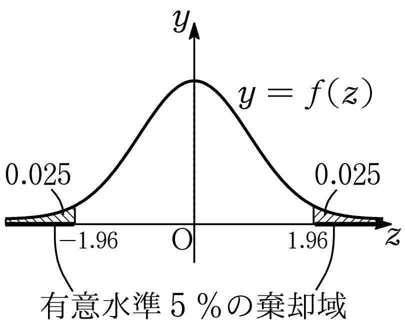
1 確率 $P\big(|Z| \geqq |z_0|\big) < 0.05$ ならば

H_0 は棄却され，H_1 が正しいと判断できる．

とくに $P\big(|Z| \geqq 1.96\big) = 0.05$ であるから

$|z_0| > 1.96$ であれば $P\big(|Z| \geqq |z_0|\big) < 0.05$

すなわち，有意水準 5 ％の棄却域は



$$|Z| > 1.96 \iff Z < -1.96 \text{ または } 1.96 < Z$$

2 確率 $P\big(|Z| \geqq |z_0|\big) > 0.05$ ならば

H_0 は棄却されず，H_1 が正しいと判断できない．

補 個数 X は二項分布 $B(n,\ p)$ に従うから標準正規分布 $N\big(np,\ np(1-p)\big)$ に従う．

標本平均と検定（片側検定）

特性 A の母比率 p_1 の正規分布に従うはずの母集団から大きさ n の標本を取り出したとき，その比率が q で，p_1 より大きい値とする．

その結果から 母比率は p_1 より大きいと判断できるかを仮説検定する．

そこで，母比率を p として

　　帰無仮説 H_0 ： $p = p_1$

を立てると

　　対立仮説 H_1 ： $p > p_1$

ここで，有意水準を 5 ％とし，n は十分大きいとする．

帰無仮説 H_0 が正しいとすると

特性 A の個数 X は，$N\big(np,\, np(1-p)\big)$ に従うので

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

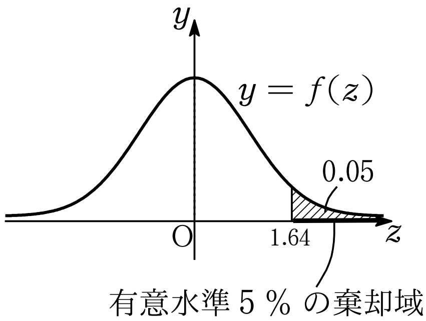
として，Z の分布は $N(0,\, 1)$ に従う．

$\dfrac{X}{n} = q$ としたときの Z の値を z_0 とすると

$$z_0 = \frac{nq - np}{\sqrt{np(1-p)}} = \frac{q - p}{\sqrt{\dfrac{p(1-p)}{n}}} \quad (> 0)$$

$\boxed{1}$ 確率 $P\big(Z \geqq z_0\big) < 0.05$ ならば

　　H_0 は棄却され，H_1 が正しいと判断できる．



とくに $P\big(Z \geqq 1.64\big) \fallingdotseq 0.05$ であるから

$z_0 > 1.64$ であれば $P\big(Z \geqq z_0\big) < 0.05$

すなわち，有意水準 5 ％の両側検定での棄却域は

　　$Z > 1.64$

$\boxed{2}$ 確率 $P\big(Z \geqq z_0\big) > 0.05$ ならば

　　H_0 は棄却されず，H_1 が正しいと判断できない．

例題　ある選挙で，候補者 A，B のいずれかに投票し，白票はないものとする．投票した人から 625 人を無作為抽出して出口調査を行ったところ，候補者 A に投票した人が 340 人であった．この調査の結果から，有権者の過半数が候補者 A を支持していると判断できるか．有意水準 5 %で仮説検定せよ．

㊫ 投票者全体における A の支持率を母比率 p とする．
このとき

帰無仮説 H_0 : $p=0.5$

対立仮説 H_1 : $p>0.5$

帰無仮説 H_0 が正しいとし，A に投票したと答えた人の人数を X とすると
X は $N\big(625p,\ 625p(1-p)\big)$ に従う．

$$Z=\frac{X-625p}{\sqrt{625p(1-p)}}=\frac{X-625\cdot0.5}{\sqrt{625\cdot0.5\cdot0.5}}=\frac{X-312.5}{25\cdot0.5}$$

として，Z は $N(0,1)$ に従う．
$X=340$ としたときの Z の値を z_0 とすると　$z_0=\dfrac{340-312.5}{25\cdot0.5}=2.2$
これは棄却域 $Z>1.64$ に入っているから 確率 $P(Z>2.2)<0.05$ であり，H_0 は棄却される．
よって，**有権者の過半数が候補者 A を支持していると判断できる．**

㊥ 標本比率は $\dfrac{X}{625}$