統計的な推測(5)

ここでは,

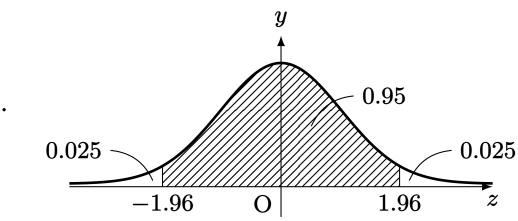
- 正規分布を用いた母平均，母比率の推定
- 仮説検定
- 正規分布を用いた母平均，母比率の仮説検定

について理解していきましょう。

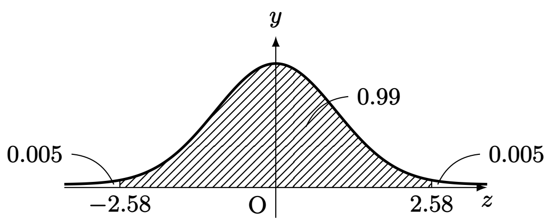
Z が $N(0,\ 1)$ に従うとき，正規分布の表から，次のことがいえます。

見慣れておきましょう。

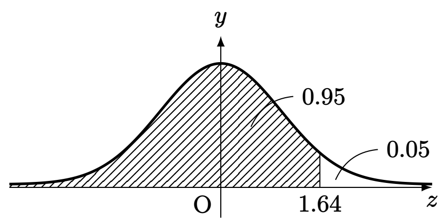
- $P(-1.96 \leqq Z \leqq 1.96) = 0.95$ ....



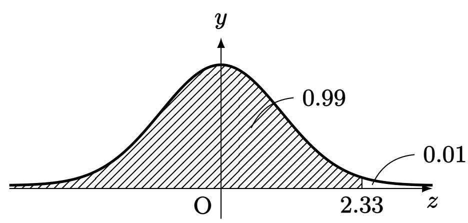
- $P(-2.58 \leqq Z \leqq 2.58) = 0.99$ ....



- $P(Z \leqq 1.64) = 0.95$ ....



- $P(Z \leqq 2.33) = 0.99$ ....



基本１　仮説検定

(1)　仮説検定

　母集団分布に関する仮定を仮説といい，標本から得られた結果によって，この仮説が正しいか正しくないかを判断する方法を仮説検定という。また，仮説が正しくないと判断することを，仮説を棄却するという。

　また，仮説検定において，正しいかどうか判断したい主張に反する仮定として立てた仮説を帰無仮説といい，もとの主張を対立仮説という。

(2)　有意水準と棄却域

　仮説検定では，どの程度小さい確率の事象が起こると仮説を棄却するか，という基準をあらかじめ定めておく。この基準となる確率 α を有意水準または危険率という。

　有意水準 α に対して，立てた仮説のもとでは実現しにくい確率変数の値の範囲を棄却域という。

(3)　両側検定と片側検定

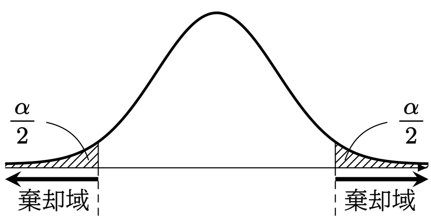
　棄却域を分布の両側に設定して行う仮説検定を両側検定といい，

　棄却域を分布の片側に設定して行う仮説検定を片側検定という。

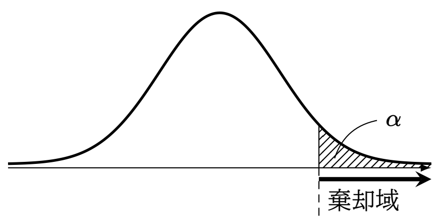
　対立仮説として「帰無仮説で仮定した値ではない」を立てる場合には両側検定を，

　対立仮説として「帰無仮説で仮定した値より大きい (または，小さい)」を立てる場合には片側検定を利用する。

(両側検定の棄却域)



(片側検定の棄却域)



(4)　仮説検定の手順

(i)　事象が起こった状況や原因を推測し，仮説を立てる。

(ii)　有意水準 α を定め，仮説に基づいて確率分布を特定し，棄却域を求める。

(iii)　標本から得られた確率変数の値が棄却域に入れば仮説を棄却し，棄却域に入らなければ仮説を棄却しない。

基本 2　正規分布を用いた母平均，母比率の検定

(1)　母平均の検定

母標準偏差が σ である母集団から，大きさ n の無作為標本を抽出する。

その標本の平均値を $\overline{X}$，標本標準偏差を S とする。

帰無仮説として「母平均が m である」を立てる。

n が大きいとき，$\overline{X}$ の棄却域は，

- 有意水準 5 %，両側検定の場合は，$\overline{X} \leqq m - 1.96 \cdot \dfrac{\sigma}{\sqrt{n}}$，$m + 1.96 \cdot \dfrac{\sigma}{\sqrt{n}} \leqq \overline{X}$
- 有意水準 5 %，片側検定の場合は，$\overline{X} \leqq m - 1.64 \cdot \dfrac{\sigma}{\sqrt{n}}$，$m + 1.64 \cdot \dfrac{\sigma}{\sqrt{n}} \leqq \overline{X}$

など。母標準偏差が不明なときは，σ の代わりに標本標準偏差 S を用いて検定する。

証　母標準偏差 σ をもつ母集団から，大きさ n の無作為標本を抽出する。

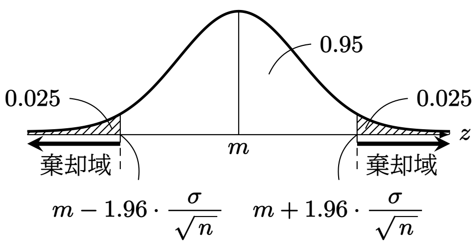
「母平均が m である」という仮説の下，

→　標本平均 $\overline{X}$ は，n が大きいとき近似的に $N\left(m, \dfrac{\sigma^2}{n}\right)$ に従う。

→　$Z = \dfrac{\overline{X} - m}{\dfrac{\sigma}{\sqrt{n}}}$ は，n が大きいとき近似的に $N(0, 1)$ に従う。

→　正規分布の表から，$P(Z \leqq -1.96,\ 1.96 \leqq Z) = 0.05$ がいえるので，

$$P\left(\overline{X} \leqq m - 1.96 \cdot \frac{\sigma}{\sqrt{n}},\ m + 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leqq \overline{X}\right) = 0.05$$



これより，有意水準 5 %の両側検定の場合，$\overline{X}$ の棄却域は，

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

実際には，$\dfrac{\overline{X} - m}{\dfrac{\sigma}{\sqrt{n}}}$ の値を 1.96 や 1.64 と比べると計算が易しい。

(2)　母比率の検定

ある母集団から，大きさ n の無作為標本を抽出する。

その標本の標本比率を R とする。

帰無仮説として「母比率が p である」を立てると，n が大きいとき，R の棄却域は，

- 有意水準 5 %，両側検定の場合，$R \leqq p-1.96\sqrt{\dfrac{p(1-p)}{n}}$，$p+1.96\sqrt{\dfrac{p(1-p)}{n}} \leqq R$
- 有意水準 5 %，片側検定の場合，$R \leqq p-1.64\sqrt{\dfrac{p(1-p)}{n}}$，$p+1.64\sqrt{\dfrac{p(1-p)}{n}} \leqq R$

など。

証　母集団から，大きさ n の無作為標本を抽出する。

「母比率が p である」という仮説の下，特性 A をもつとき 1 を，もたないとき 0 の値をとる確率変数を X とすると，

- 標本比率 R は X の標本平均
- X の母比率は p，母標準偏差は $\sqrt{p(1-p)}$

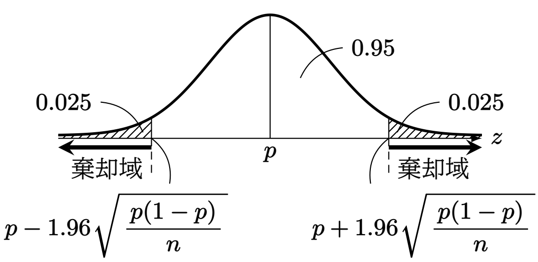
がいえるのでした。

よって，(1) の「母平均の検定」において，

$$\overline{X} \to R \qquad m \to p \qquad \sigma \to \sqrt{p(1-p)}$$

とすると，n が大きいとき，

$$P\left(R \leqq p-1.96\sqrt{\frac{p(1-p)}{n}},\ p+1.96\sqrt{\frac{p(1-p)}{n}} \leqq R\right) = 0.05$$



これより，有意水準 5 %の両側検定の場合，R の棄却域は，

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

実際には，$\dfrac{R-p}{\sqrt{\dfrac{p(1-p)}{n}}}$ の値を 1.96 や 1.64 と比べると計算が易しい。

基本 3　正規分布を用いた母平均，母比率の推定

(1)　母平均の推定

母平均 m，母標準偏差 σ である母集団から，大きさ n の無作為標本を抽出する。

その標本の平均値を $\overline{X}$，標本標準偏差を S とする。

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，$\left[\overline{X}-1.96\cdot\dfrac{\sigma}{\sqrt{n}},\ \overline{X}+1.96\cdot\dfrac{\sigma}{\sqrt{n}}\right]$
- 信頼度 99 %の信頼区間は，$\left[\overline{X}-2.58\cdot\dfrac{\sigma}{\sqrt{n}},\ \overline{X}+2.58\cdot\dfrac{\sigma}{\sqrt{n}}\right]$

母標準偏差 σ が不明なときは，σ の代わりに標本標準偏差 S を用いて計算する。

証

母平均 m，母標準偏差 σ をもつ母集団から，大きさ n の無作為標本を抽出する。

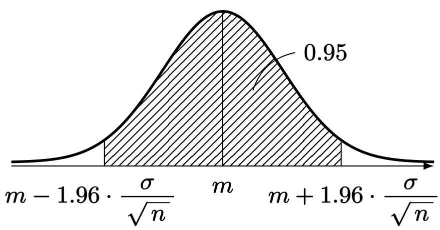
→　標本平均 $\overline{X}$ は，n が大きいとき，近似的に正規分布 $N\left(m,\ \dfrac{\sigma^2}{n}\right)$ に従う。

→　$Z=\dfrac{\overline{X}-m}{\dfrac{\sigma}{\sqrt{n}}}$ は，n が大きいとき，近似的に標準正規分布 $N(0,\ 1)$ に従う。

→　正規分布表から，$P(-1.96\leqq Z\leqq 1.96)=0.95$ (95 %) がいえるので，

$$P\left(-1.96\cdot\frac{\sigma}{\sqrt{n}}\leqq\overline{X}-m\leqq 1.96\cdot\frac{\sigma}{\sqrt{n}}\right)=0.95$$

$$P\left(m-1.96\cdot\frac{\sigma}{\sqrt{n}}\leqq\overline{X}\leqq m+1.96\cdot\frac{\sigma}{\sqrt{n}}\right)=0.95 \qquad \cdots\cdots\cdots\cdots \quad ①$$



①より，

$$P\left(\overline{X}-1.96\cdot\frac{\sigma}{\sqrt{n}}\leqq m\leqq\overline{X}+1.96\cdot\frac{\sigma}{\sqrt{n}}\right)=0.95$$

であるから，区間 $\overline{X}-1.96\cdot\dfrac{\sigma}{\sqrt{n}}\leqq x\leqq\overline{X}+1.96\cdot\dfrac{\sigma}{\sqrt{n}}$ が母平均 m の値を含むことが約 95 %の確からしさで期待されることがわかる。

(2) 母比率の推定

母比率が p である母集団から，大きさが n の標本を無作為抽出する。

その標本の標本比率を R とする。

n が大きいとき，母比率 p に対する

- 信頼度 95 %の信頼区間は，$\left[R-1.96\cdot\sqrt{\dfrac{R(1-R)}{n}},\ R+1.96\cdot\sqrt{\dfrac{R(1-R)}{n}}\right]$
- 信頼度 99 %の信頼区間は，$\left[R-2.58\cdot\sqrt{\dfrac{R(1-R)}{n}},\ R+2.58\cdot\sqrt{\dfrac{R(1-R)}{n}}\right]$

証

基本 86・3 で履修した内容から，

母比率が p である母集団において，特性 A をもつとき 1 を，もたないとき 0 の値をとる確率変数を X とすると，

- 標本比率 R は X の標本平均
- X の母標準偏差は $\sqrt{p(1-p)}$

がいえるのでした。

よって，(1) の「母平均の推定」において，

$$\overline{X}\to R \qquad \sigma\to\sqrt{p(1-p)}$$

とすると，n が大きいとき，母比率 p に対する

- 信頼度 95 %の信頼区間は，$\left[R-1.96\cdot\sqrt{\dfrac{p(1-p)}{n}},\ R+1.96\cdot\sqrt{\dfrac{p(1-p)}{n}}\right]$
- 信頼度 99 %の信頼区間は，$\left[R-2.58\cdot\sqrt{\dfrac{p(1-p)}{n}},\ R+2.58\cdot\sqrt{\dfrac{p(1-p)}{n}}\right]$

が得られる。

n が大きいとき，大数の法則から，R は p に近いので，p を R に置き換える。

1・A

新型テレビを視聴して画質がよくなったと回答した人が 10 人中 7 人であった。

このデータから，新型テレビの画質がよくなったと判断してよいか。有意水準を 0.05 として答えよ。

ただし，10 枚のコインを投げて表が出た枚数を記録することを 100 回行ったところ，次の表になったとし，この結果を用いよ。

表の枚数	0	1	2	3	4	5	6	7	8	9	10
回数	0	2	7	11	16	20	19	15	7	2	1

2・A

内容量が 255 g と表示されている商品がある。この商品から 64 個を抽出して内容量を調べたところ，平均値が 252 g，標準偏差が 9.6 g であった。このとき，この商品 1 個あたりの内容量は表示通りでないと判断してよいか。有意水準 5 %で検定せよ。

3・A

ある県の高校 1 年生の男子 1600 人を無作為に抽出して身長を調べたところ，平均身長が 164 cm，標準偏差が 6 cm であった。

この県の高校 1 年生男子の平均身長 m を信頼度 95 %で推定せよ。

4・B

ある 1 個のさいころを 720 回投げたところ，1 の目が 95 回出た。このさいころは，1 の目が出る確率が $\frac{1}{6}$ ではないとい判断してよいか。有意水準 5 %で検定せよ。

5・B

ある種子の発芽率は，従来 80 %であったが，発芽しやすいように品種改良した。品種改良し種子から無作為に 400 個を抽出して種をまいたところ 334 個が発芽した。品種改良によって発芽率が上がったと判断してよいか。

(1) 有意水準 5 %で検定せよ。

(2) 有意水準 1 %で検定せよ。

6・B

(1) ある工場の製品 400 個について検査したところ，不良品が 8 個あった。全製品における不良率を，信頼度 95 %で推定せよ。

(2) ある工場で生産される製品の不良率は 5 %前後であると考えられている。この製品の不良率を 99 %で推定するとき，信頼区間の幅を 0.02 以下にするためには，何個以上抽出して調べればよいか。

7・C

頂点 A，B，C，D の正四面体がある。Q は，ある頂点から他の頂点に移動し，どの観測時刻においてもこれらの頂点のいずれかに存在している。Q が頂点 D に存在する確率が $\dfrac{1}{4}$ であるという仮説を立てた。この仮説の正しさを調べるために，Q の位置を異なる時刻で 10 回観測したところ，そのうち 6 回頂点 D に存在した。この仮説の正しさを有意水準 5 %で検定せよ。

8・C

ある種類のねずみは，生まれてから 3 か月後の体重が平均 65 g，標準偏差 4.8 g の正規分布に従うという。いまこの種類のねずみ 10 匹を特別な飼料で飼育し，3 か月後の体重を測定したところ，次の結果を得た。

67，71，63，74，68，61，64，80，71，73

この飼料はねずみの体重に異常な変化を与えたと考えられるか。有意水準 5 %で検定せよ。

9・C

発芽して一定期間後の，ある花の苗の高さの分布は，母平均 m cm，母標準偏差 1.5 cm の正規分布であるとする。大きさ n の標本を無作為抽出して，信頼度 95 %の m に対する信頼区間を求めたところ，$[9.81,\ 10.79]$ であった。標本平均 $\overline{x}$ と n の値を求めよ。