何人に調査すればいいの？～区間推定について～

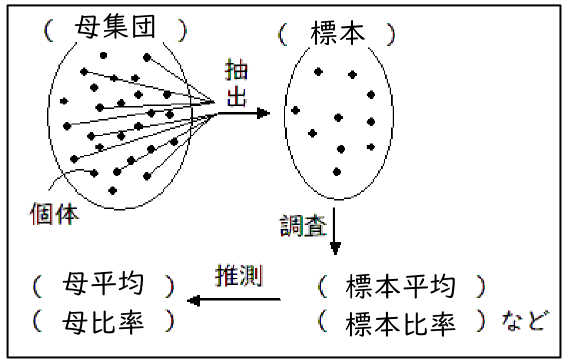
１．前提のお話について

(１)　標本調査

　→ 対象全体からその一部を抜き出して調査すること。

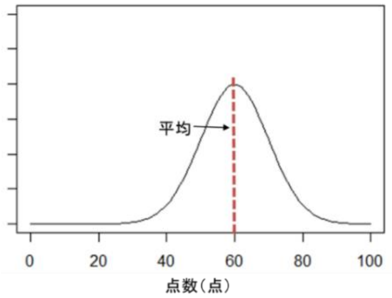
　取り出した標本の個体の個数を標本の大きさと言う。

　★母集団の性質を正しく推測するために，標本の選び方に“イカサマ（作為）”があってはいけません！



(2)　正規分布 → 左右対称で平均値が山の頂点となるような分布のこと。
分布が右の図のような形になる。

★ある程度大量に調査をしてその分布を図にしたとき，山が１つであれば正規分布とみなしても大丈夫。山が２つになるとダメ。



(3)　今回のお話の２つの前提

①標本が無作為抽出（ランダムサンプリング）されていること。

②標本の分布の山が１つとなっている（正規分布とみなせる）こと。

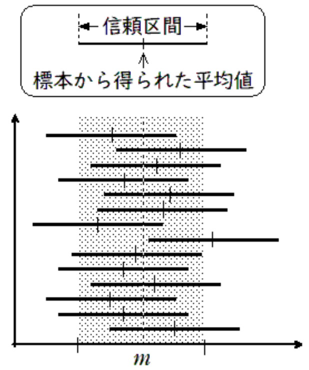
　特に，母平均を調べたいときはよく確認すること。

※これ以降出てくる公式の成り立ちは省略します。知りたい場合は数学Ｂの教科書第４章を参照。
数学を極めたい人以外公式は覚える必要はありません。課題研究で必要なときに，このプリントを引っ張り出してください。

２．区間推定とは…？

→標本から母集団の傾向を「幅」を持たせて推定すること。

例えば，標本調査から母平均を正確に求めることはできないが，「**母平均を含む確率が 95%である区間**」であれば求めることはできる。この区間を**信頼度 95%の信頼区間**という。



(１)　母平均 m の区間推定

公式　標本平均 $\bar{X}$，母標準偏差 σ，標本の大きさが n のとき，母平均 m に対する信頼度 95%の信頼区間は，

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

※標本の大きさ n が大きいとき（概ね $n \geqq 30$ なら），σ は標本の標準偏差を用いてもよい。

（例）M 県の 12 歳男子を 100 名無作為抽出して身長を測定したところ，平均 152.0cm，標準偏差 8.1cm であった。M 県の 12 歳男子全体の平均身長（母平均）m はどのくらいと推定できるか？

→ 上の公式で，$\sigma =$ (　　　)，$n =$ (　　　)なので，

$$1.96 \cdot \frac{\sigma}{\sqrt{n}} =$$

また，$\bar{X} =$(　　　　)であるから，平均身長 m に対する信頼度 95%の信頼区間は，

(　　　　　　　　　) $\leqq m \leqq$ (　　　　　　　　　)

したがって，(　　　　) cm 以上（　　　　）cm 以下である。

⇒この区間に M 県の平均身長（母平均）が含まれる確率が 95%と考えてよい！

Ｑ１：大量生産されたある工業製品から標本 400 個を無作為抽出して重さを測ったら，平均の重さは 98.8g，標準偏差は 2.0g であった。この工業製品の平均重量 m に対して，信頼度 95%の信頼区間を求めてみよう。

２年　ＳＳ探究Ⅱ　統計学②

(2)　母比率 p の区間推定

公式 標本比率を R とし，標本の大きさ n が十分大きいとき，母比率 p に対する信頼度 95%の信頼区間は，

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

（例）ある地域で無作為に抽出された 600 台のテレビを調査すると，ある番組「P」が 99 台で視聴されていた。この地域における番組「P」の視聴率（母比率）p に対する 95%信頼区間はどうなるか。

→上の公式で，$n=$（　　　），$R=$（　　　　　　　　）なので，

$$1.96\sqrt{\frac{R(1-R)}{n}}=$$

よって，番組「P」の視聴率 p に対する 95%信頼区間は，

（　　　　　　　　）$\leqq p \leqq$（　　　　　　　　）

したがって，（　　　）以上（　　　）以下である。

⇒番組「P」視聴率が（　　　）%から（　　　）%の間にある確率が 95%だと考えてよい！

Ｑ２：ある街で下水道の臭いについて住民の意識調査を行ったところ，回答のあった 400 人のうち 144 人が臭いが気になると回答した。この街の住民全体で臭いが気になる人の割合 p に対して，信頼度 95%の信頼区間を求めてみよう。

３．標本はどれだけ調べるべき？

区間推定では，**標本の大きさが大きいほどその幅は小さくなる（つまり，信憑性が高くなる）。**

公式 ①母標準偏差 σ で，母平均 m に対する 95%信頼区間の幅を h 以内にしたいとき，標本の大きさ n は，

$$n \geqq \left(\frac{2\times1.96\times\sigma}{h}\right)^2$$ であればよい。

②母比率 p に対する 95%信頼区間の幅を h 以内にしたいとき，標本の大きさ n は，$n \geqq \left(\frac{1.96}{h}\right)^2$ …Ⓐ

であればよい。ただし，標本比率が R と予想できるときは，$n \geqq \left(\frac{2\times1.96}{h}\right)^2 R(1-R)$ …Ⓑ　でよい。

※②は，ⒶよりⒷの方が小さいので，標本比率が予測できるならⒷを使うと標本が小さくて済む。

（例）畑で育てた大量の玉ねぎがある。この中から無作為抽出するとき，

(1)　玉ねぎの重さの母平均が知りたい。母標準偏差が 30(g)であるとき，母平均の 95%信頼区間の幅を 3g 以下にしたい。このとき必要な玉ねぎの個数 n は，$\sigma=$（　　　），$h=$（　　）なので，

$n \geqq$　　　　　　　　　　　　よって，（　　　）個以上である。

(2)　不良品の玉ねぎがどのくらいの割合であるか知りたい。母比率の 95%信頼区間の幅を 0.07（7%）以下にしたい。このとき必要な玉ねぎの個数 n は，$h=$（　　　）なので，

$n \geqq$　　　　　　　　　　　　よって，（　　　）個以上である。

Ｑ３：ある県で実施された試験の平均点を推定したい。信頼度 95%で誤差が 2 点以内になるようにするには，何人以上の受験者の得点を無作為抽出して調べればよいか。ただし，従来の傾向から，標準偏差は 20 点としてよい。　（注）誤差が 2 点以内ということは，信頼区間の幅は 4 点ということ。

Ｑ４：ある県の 20 歳女性の運転免許保有率を推定したい。信頼度 95%の信頼区間の幅が 2%（0.02）以下になるように推定するには，何人以上の標本を無作為抽出すればよいか。また，この保有率がほぼ 30%であると予想できるなら，何人以上の標本を無作為抽出すればよいか。

【おまけ】信頼度を 95%以外で考えたいとき

★区間推定でよく使われる信頼度は，90%・95%・99%の 3 つです。

※90%なら信頼度は小さいが区間の幅は小さくなり，99%なら信頼度は高い分区間の幅も大きくなる。

→信頼度を変えたいときは，それぞれの公式の「1.96」の部分を，

信頼度 90%で考えたいなら「1.64」に，信頼度 99%で考えたいなら「2.58」に置き換えるだけで OK！

何人に調査すればいいの？～区間推定について～

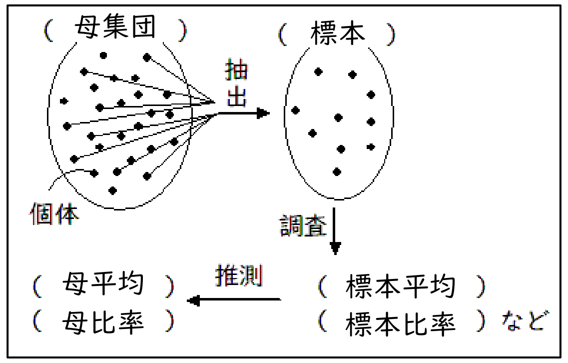
Ⅰ．前提のお話について

(1)　標本調査

→ 対象全体からその一部を抜き出して調査すること。

取り出した標本の個体の個数を標本の大きさと言う。

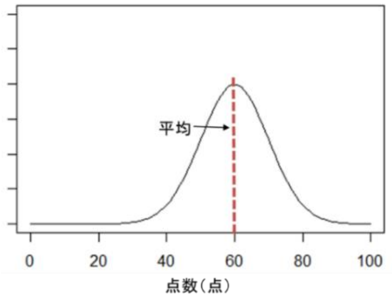
★母集団の性質を正しく推測するために，標本の選び方に“イカサマ（作為）”があってはいけません！



(2)　正規分布 → 左右対称で平均値が山の頂点となるような分布のこと。

分布が右の図のような形になる。

★ある程度大量に調査をしてその分布を図にしたとき，山が１つであれば正規分布とみなしても大丈夫。山が２つになるとダメ。



(3)　今回のお話の２つの前提

①標本が無作為抽出（ランダムサンプリング）されていること。

②標本の分布の山が１つとなっている（正規分布とみなせる）こと。

特に，母平均を調べたいときはよく確認すること。

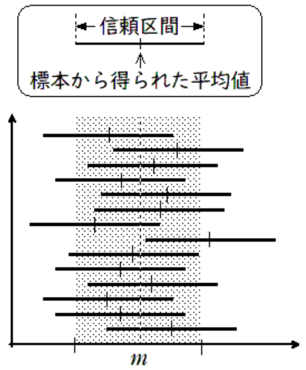
※これ以降出てくる公式の成り立ちは省略します。知りたい場合は数学Ｂの教科書第４章を参照。

数学を極めたい人以外公式は覚える必要はありません。課題研究で必要なときに，このプリントを引っ張り出してください。

２．区間推定とは…？

→標本から母集団の傾向を「幅」を持たせて推定すること。

例えば，標本調査から母平均を正確に求めることはできないが，「**母平均を含む確率が 95%である区間**」であれば求めることはできる。この区間を**信頼度 95%の信頼区間**という。



(1)　母平均 m の区間推定

公式　標本平均 $\bar{X}$，母標準偏差 σ，標本の大きさが n のとき，母平均 m に対する信頼度 95%の信頼区間は，

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

※標本の大きさ n が大きいとき（概ね $n \geqq 30$ なら），σ は標本の標準偏差を用いてもよい。

(例) M 県の 12 歳男子を 100 名無作為抽出して身長を測定したところ，平均 152.0cm，標準偏差 8.1cm であった。M 県の 12 歳男子全体の平均身長（母平均） m はどのくらいと推定できるか？

→ 上の公式で，$\sigma =$ (　8.1　)，$n =$ (　100　)なので，

$$1.96 \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{8.1}{\sqrt{100}} = 1.96 \cdot 0.81 \fallingdotseq 1.59$$

また，$\bar{X} =$ (　152.0　)であるから，平均身長 m に対する信頼度 95%の信頼区間は，

(　152.0 − 1.59　) $\leqq m \leqq$ (　152.0 + 1.59　)

したがって，(　150.41　) cm 以上 (　153.59　) cm 以下である。

⇒この区間に M 県の平均身長（母平均）が含まれる確率が 95%と考えてよい！

Ｑ１：大量生産されたある工業製品から標本 400 個を無作為抽出して重さを測ったら，平均の重さは 98.8g，標準偏差は 2.0g であった。この工業製品の平均重量 m に対して，信頼度 95%の信頼区間を求めてみよう。

上の公式で，$\sigma = 2.0$，$n = 400$ なので，

$$1.96 \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{2.0}{\sqrt{400}} = 1.96 \cdot 0.1 \fallingdotseq 0.2$$

また，$\bar{X} = 98.8$ であるから，平均重量 m に対する信頼度 95%の信頼区間は，

$$98.8 - 0.2 \leqq m \leqq 98.8 + 0.2$$

したがって，98.6 g 以上 99.0 g 以下である。

(2)　母比率 p の区間推定

公式　標本比率を R とし，標本の大きさ n が十分大きいとき，母比率 p に対する信頼度 95%の信頼区間は，

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

（例）ある地域で無作為に抽出された 600 台のテレビを調査すると，ある番組「P」が 99 台で視聴されていた。この地域における番組「P」の視聴率（母比率）p に対する 95%信頼区間はどうなるか。

→上の公式で，$n=$（　600　），$R=$（　$\frac{99}{600}=0.165$　）なので，

$$1.96\sqrt{\frac{R(1-R)}{n}}=1.96\cdot\sqrt{\frac{0.165(1-0.165)}{600}}≒1.96\cdot 0.0152≒0.0298$$

よって，番組「P」の視聴率 p に対する 95%信頼区間は，

（　$0.165-0.0298$　）$\leqq p \leqq$（　$0.165+0.0298$　）

したがって，（　0.135　）以上（　0.195　）以下である。

⇒番組「P」視聴率が（　13.5　）%から（　19.5　）%の間にある確率が 95%だと考えてよい！

Ｑ２：ある街で下水道の臭いについて住民の意識調査を行ったところ，回答のあった 400 人のうち 144 人が臭いが気になると回答した。この街の住民全体で臭いが気になる人の割合 p に対して，信頼度 95%の信頼区間を求めてみよう。

上の公式で，$n=400$，$R=\frac{144}{400}=0.36$ なので，

$$1.96\sqrt{\frac{R(1-R)}{n}}=1.96\cdot\sqrt{\frac{0.36(1-0.36)}{400}}=1.96\cdot\frac{0.6\times0.8}{20}=1.96\times0.024≒0.047$$

よって，臭いが気になる人の割合 p に対する 95%信頼区間は，

$$0.36-0.047\leqq p\leqq 0.36+0.047$$

したがって，0.313 以上 0.407 以下である。（31.3%以上 40.7%以下）

３．標本はどれだけ調べるべき？

区間推定では，**標本の大きさが大きいほどその幅は小さくなる（つまり，信憑性が高くなる）。**

公式　①母標準偏差 σ で，母平均 m に対する 95%信頼区間の幅を h 以内にしたいとき，標本の大きさ n は，

$$n\geqq\left(\frac{2\times1.96\times\sigma}{h}\right)^2$$ であればよい。

②母比率 p に対する 95%信頼区間の幅を h 以内にしたいとき，標本の大きさ n は，$n\geqq\left(\frac{1.96}{h}\right)^2$ …Ⓐ

であればよい。ただし，標本比率が R と予想できるときは，$n\geqq\left(\frac{2\times1.96}{h}\right)^2 R(1-R)$ …Ⓑ　でよい。

※②は，ⒶよりⒷの方が小さいので，標本比率が予測できるならⒷを使うと標本が小さくて済む。

（例）畑で育てた大量の玉ねぎがある。この中から無作為抽出するとき，

(1)　玉ねぎの重さの母平均が知りたい。母標準偏差が 30(g)であるとき，母平均の 95%信頼区間の幅を 3g 以下にしたい。このとき必要な玉ねぎの個数 n は，$\sigma=$（　30　），$h=$（　3　）なので，

$$n\geqq\left(\frac{2\times1.96\times30}{3}\right)^2=39.2^2=1536.64$$

よって，（　1537　）個以上である。

(2)　不良品の玉ねぎがどのくらいの割合であるか知りたい。母比率の 95%信頼区間の幅を 0.07（7%）以下にしたい。このとき必要な玉ねぎの個数 n は，$h=$（　0.07　）なので，

$$n\geqq\left(\frac{1.96}{0.07}\right)^2=28^2=784$$

よって，（　784　）個以上である。

Ｑ３：ある県で実施された試験の平均点を推定したい。信頼度 95%で誤差が 2 点以内になるようにするには，何人以上の受験者の得点を無作為抽出して調べればよいか。ただし，従来の傾向から，標準偏差は 20 点としてよい。　（注）誤差が 2 点以内ということは，信頼区間の幅は 4 点ということ。

①の公式で，$\sigma=20$，$h=4$ なので，$n\geqq\left(\frac{2\times1.96\times20}{4}\right)^2=19.6^2=384.16$

よって，385 人以上である。

Ｑ４：ある県の 20 歳女性の運転免許保有率を推定したい。信頼度 95%の信頼区間の幅が 2%（0.02）以下になるように推定するには，何人以上の標本を無作為抽出すればよいか。また，この保有率がほぼ 30%であると予想できるなら，何人以上の標本を無作為抽出すればよいか。

②のＡの公式で，$h=0.02$ なので，　$n\geqq\left(\frac{1.96}{0.02}\right)^2=98^2=9604$

よって，9604 人以上である。

また，保有率がほぼ 30%と予想できるとき，②のＢの公式で $R=0.3$ なので，

$$n\geqq\left(\frac{2\times1.96}{0.02}\right)^2\times0.3\times(1-0.3)=196^2\times0.21=8067.36$$

よって，8068 人以上である。

【おまけ】信頼度を 95%以外で考えたいとき

★区間推定でよく使われる信頼度は，90%・95%・99%の 3 つです。

※90%なら信頼度は小さいが区間の幅は小さくなり，99%なら信頼度は高い分区間の幅も大きくなる。

→信頼度を変えたいときは，それぞれの公式の「1.96」の部分を，

信頼度 90%で考えたいなら「1.64」に，信頼度 99%で考えたいなら「2.58」に置き換えるだけで OK！