

10 標本分散

10.1 標本分散と不偏分散

サンプルの平均値を実現値とする確率変数である標本平均を考えた様に、サンプルの分散を実現値とする確率変数を考えます：

定義 10.1.1 母集団 X からとった大きさ n の標本 $X_1, \dots, X_n$ と標本平均 $\bar{X}$ に対して次で定まる確率変数 $\bar{V}$ を、大きさ n の標本分散と言います：

$$\bar{V} = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2.$$

右辺の自乗を展開してから加えれば

$$\bar{V} = \frac{1}{n} \sum_{j=1}^n X_j^2 - \frac{2}{n} \sum_{j=1}^n X_j \bar{X} + \bar{X}^2 = \frac{1}{n} \sum_{j=1}^n X_j^2 - \bar{X}^2 \quad \dots (*)$$

とも変形される事に注意しておきます。また少し変形してから展開すれば

$$\begin{aligned} E[\bar{V}] &= E\left[\frac{1}{n} \sum_{j=1}^n (X_j - m - \bar{X} + m)^2\right] \\ &= \frac{1}{n} \sum_{j=1}^n \{E[(X_j - m)^2] - 2E[(X_j - m)(\bar{X} - m)] + E[(\bar{X} - m)^2]\} \\ &= \frac{1}{n} \sum_{j=1}^n v - 2E\left[\frac{1}{n} \sum_{j=1}^n (X_j - m)(\bar{X} - m)\right] + Var[\bar{X}] \end{aligned}$$

となりますから

$$E[\bar{V}] = v - 2Var[\bar{X}] + Var[\bar{X}] = v - \frac{v}{n} = \frac{n-1}{n}v$$

であり、標本分散の平均値は母集団の分散より若干小さくなります。

従ってサンプルを採取してその分散を計算しても、サンプルの取り方によるばらつき以前の話として、そもそも母分散から『ずれて分布している』わけですから注意が必要になります（標本のサイズが大きくなるほどそのずれは緩和されますが）。

しかしここで両辺に $\frac{n}{n-1}$ を掛ければ $E\left[\frac{n}{n-1}\bar{V}\right] = v$ ですから、

$$\tilde{V} = \frac{n}{n-1}\bar{V} = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2$$

と定義すれば $\tilde{V}$ の平均は母分散に一致し、『ずれ』は解消します。どういうわけか n で割るのではなく $n-1$ で割った方が上手くいくと云う事です。 $\tilde{V}$ は不偏標本分散（あるいは単に不偏分散）と呼ばれ、その実現値は不偏分散と呼ばれます。

10.2 母分散が不明な場合の母平均の区間推定

事実 10.2.1 母集団の分散を v としたとき、大きさ n の標本分散 $\bar{V}$ と不偏標本分散 $\tilde{V}$ の平均値はそれぞれ以下の通りです：

$$E[\bar{V}] = \frac{n-1}{n}v, \quad E[\tilde{V}] = v.$$

従ってサンプル数がある程度大きければ（諸説ありますが、この講義では 30 以上とします）、サンプルの分散あるいはサンプルの不偏分散はいずれも母集団の分散の代用品として使う事が出来ます。当然ですが不偏分散を使った方がベターです。

問題 10.2.2 母分散は不明であるとします。この中から大きさ 50 のサンプルを取って調べたところ平均値が 169.6、分散が 4.07^2 でした。

この時に母平均 m の信頼度 95 パーセントの信頼区間を求めて下さい。

母分布：	unknown
母平均：	m : unknown
母分散：	unknown
サンプルサイズ：	50 (large)
サンプル平均：	169.6
サンプル分散：	4.07^2

大きさ 50 の大きなサンプルを取っているので母分散はサンプルの分散で代用出来ます。更に中心極限定理によれば標本平均 $\bar{X}$ は正規分布 $N\left(m, \frac{4.07^2}{50}\right)$ で近似されます。そこで

$$P[|\bar{X} - m| \leq d] = 0.95$$

となる様な $d > 0$ を求めると（計算略） $d \approx 1.13$ が分かります。

従って今回のサンプル平均 169.6 に関して信頼度 95 パーセントで $|169.6 - m| \leq 1.13$ が成り立ちますので、これを m に関する条件に読み替えれば

$$169.6 - 1.13 \leq m \leq 169.6 + 1.13$$

であり、求める信頼区間は $[168.47, 170.73]$ になります。

【不偏分散を使った場合の計算】母分散をサンプル分散ではなくサンプルの不偏分散で代用するとどうなるのでしょうか。

大きさ 50 のサンプル分散が 4.07^2 であるとき、サンプル不偏分散は $\frac{50}{49}4.07^2$ ですから、母分散をこれで代用すると標本平均 $\bar{X}$ は正規分布 $N\left(m, \frac{4.07^2}{49}\right)$ で近似されます。他は同様にして $P[|\bar{X} - m| \leq d] = 0.95$ となる様な $d > 0$ を求めます。変形すれば

$$0.95 = P\left[\left|\frac{\bar{X} - m}{\sqrt{\frac{4.07^2}{49}}}\right| \leq \frac{d}{\sqrt{\frac{4.07^2}{49}}}\right] \approx P\left[|N(0, 1)| \leq \frac{d}{\sqrt{\frac{4.07^2}{49}}}\right]$$

であり、標準正規分布表に依れば、 $P[|N(0, 1)| \leq 1.96] \approx 0.95$ ですから、

$$\frac{d}{\sqrt{\frac{4.07^2}{49}}} = 1.96, \quad d = \frac{1.96 \times 4.07}{7} \approx 1.1396 \approx 1.14$$

が分かります。従って $P[|\bar{X} - m| \leq 1.14] \approx 0.95$ が分かりました。

以上から、同様に 95 パーセントの信頼度で $|169.6 - m| \leq 1.14$ 、すなわち

$$169.6 - 1.14 \leq m \leq 169.6 + 1.14$$

であり、信頼区間は $[168.46, 170.74]$ になります。

□

10.3 なぜ $n - 1$ で割ると自然な結果となるのか

母集団 X からとった大きさ 3 の標本 X_1, X_2, X_3 に対して、標本分散 $\bar{V}$ は

$$\bar{V} = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + (X_3 - \bar{X})^2}{3}$$

ですが、『なぜ 3 で割るのか』と云うと、『3 つのものを足したから』です。平均からのズレの平均なんだからそりゃそうです。標本分散の各具体値は、具体的サンプルの分散値でした。

これを 3 で割る代わりに 2 で割ったものが標本不偏分散 $\tilde{V}$ です：

$$\tilde{V} = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + (X_3 - \bar{X})^2}{2}$$

当然ですが標本不偏分散の各具体値は具体的サンプルの分散値ではなく、具体的サンプルの不偏分散値です。

これは『3 つのもの』を足して（なぜか）2 で割っています。その 3 つを具体的に書くと

$$\begin{aligned}(X_1 - \bar{X})^2 &= X_1 - \frac{X_1 + X_2 + X_3}{3} = \frac{2}{3}X_1 - \frac{1}{3}X_2 - \frac{1}{3}X_3 \\(X_2 - \bar{X})^2 &= X_2 - \frac{X_1 + X_2 + X_3}{3} = -\frac{1}{3}X_1 + \frac{2}{3}X_2 - \frac{1}{3}X_3 \\(X_3 - \bar{X})^2 &= X_3 - \frac{X_1 + X_2 + X_3}{3} = -\frac{1}{3}X_1 - \frac{1}{3}X_2 + \frac{2}{3}X_3\end{aligned}$$

ですが、これは本当に『3 つのもの』なのでしょう？

『かさいはどうやら完全に毫碌して、数も数えられなくなったようだ』あるいは、『老眼が進んだのか』『確かに髪の毛の方は相当進んで・・・』いやいやいや、そう云う話ではありません。

もともと 3 つのものがありました。 X_1, X_2, X_3 です。これらはそれぞれが独立に採取された標本ですから、いわば『3 次元の自由度』をもっていました。3 つそれぞれが勝手にいろいろな値を取ると云うことです。

で、こいつらを『加工して』、あるいは、『変換して』

$$Y_1 = \frac{2}{3}X_1 - \frac{1}{3}X_2 - \frac{1}{3}X_3, \quad Y_2 = -\frac{1}{3}X_1 + \frac{2}{3}X_2 - \frac{1}{3}X_3, \quad Y_3 = -\frac{1}{3}X_1 - \frac{1}{3}X_2 + \frac{2}{3}X_3$$

を作ったわけですが、この子達も『3 次元の自由度をもつ』のでしょうか？ 3 つの値はそれぞれ完全に自由な値を取るのでしょうか？ 実はこれがそうではないのです。

本来は確率論の文脈で『確率論的な独立性』の話をしなければならないのですが、いろいろ難しいので、話を単純化して一次変換の話で置き換えて見てみましょう。

3 次元ユークリッド空間から 3 次元ユークリッド空間への点を点に移す写像として一次変換 M ：

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

を考えます（この係数行列を M とします。対称行列ですね）。

計算によれば固有値は $0, 1$ であり、それぞれに対応する固有ベクトルは

固有値	固有ベクトル
0	$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$
1	$\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$

となっています。

この 3 本のベクトルは一次独立ですから、3 次元空間内の任意の点はこれらの線型結合で表すことができます：

$$p \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + q \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + r \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

3 次元ユークリッド空間全体を、2 本のベクトル $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ の作る（原点を通る）平面 H と、それに直交する原点を通り方向ベクトルが $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ である直線 L に分

解して考えてみましょう。

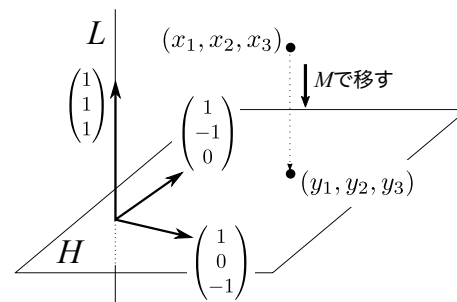
平面 H 上の点はこの一次変換によって動きません。固有値が 1 だったからです。また、直線 L 上の点は全て原点に移ります（固有値が 0 なので、全部原点に行ってしまう）。

合わせて考えれば、3 次元空間内の点は全て平面 H 内の点に移ることになります。

$$M \underbrace{\left\{ p \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + q \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + r \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}}_{\text{3 次元空間全体}} = pM \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + qM \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + rM \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

$$= p \cdot 0 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + q \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + r \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

$$= \underbrace{q \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + r \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}}_{\text{平面 } H \text{ 全体}}$$



実はこの一次変換 M は平面 H への正射影変換なのです。計算してみると分かる通り、 $M^2 = M$ です：

$$M^2 = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} = M$$

一般に、 $M^2 = M$ を満たす行列は射影行列と呼ばれ、射影変換（正射影、つまり直交射影とは限りません）の表現行列です。射影行列であって、かつ対称行列であるものが正射影に対応します（固有ベクトルが直交することを思い出して下さい）。

つまり、入力は 3 次元でしたが、出力は（1 次元分潰れて）2 次元になってしまっているのです。たとえ X_1, X_2, X_3 がそれぞれ自由な値を取ったとしても、 Y_1, Y_2, Y_3 の 3 つがそれぞれ自由な値を取ることはできないのです。気づけば簡単なことですが、 $Y_1 + Y_2 + Y_3$ は必ず 0 になってしまいますよね？ ということは、 Y_1, Y_2 が決まれば自動的に Y_3 も決まってしまうのです。

だから一見 3 つあるように見える Y_1, Y_2, Y_3 も、自由度の観点から見れば 2 つであると考えられるわけです。『だから』3 で割るのではなく、2 で割った方が『自然』なのです。

正確な確率論的観点から言うと、 X_1, X_2, X_3 は独立ですが、 Y_1, Y_2, Y_3 は独立ではありません。『ランダム』と云う言葉を聞くと、つい『独立』であるような印象をもってしましますが、別の概念ですから注意する必要があります。

10.4 問題演習

事実 10.4.1 [標本平均] 平均 m ・分散 v の母集団からとった大きさ n の標本平均 $\bar{X}$:

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}, \quad E[\bar{X}] = m, \quad Var[\bar{X}] = \frac{v}{n}.$$

事実 10.4.2 [標本分散] 分散 v の母集団からの大きさ n の標本分散 $\bar{V}$ と標本不偏分散 $\tilde{V}$:

$$\bar{V} = \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n}, \quad \tilde{V} = \frac{n}{n-1} \bar{V},$$

$$E[\bar{V}] = \frac{n-1}{n} v, \quad E[\tilde{V}] = v$$

サンプル数が 30 以上であればサンプルの分散あるいは不偏分散をもって母分散の代用とする事が出来ます。

事実 10.4.3 [正規母集団からの標本平均] 正規分布 $N(m, v)$ に従う母集団からとった大きさ n の標本平均は正規分布 $N(m, \frac{v}{n})$ に従います。

事実 10.4.4 [the central limit theorem] 平均値 m 、分散 $v \neq 0$ の母集団からとった十分大きなサイズ n ($n \geq 50$) の標本平均は正規分布 $N(m, \frac{v}{n})$ で近似されます。

基本演習 10.1 [教科書 問題 16.8 改題] 過去の記録から、ある高専 3 年男子学生の 100 メートル走の記録の度数分布は正規分布に従うことがわかっています。

学生の中から 40 人を無作為に選び、100 メートル走の記録をとったところ平均値が 15.3 秒、分散が $(1.88 \text{ 秒})^2$ でした。全学生の平均値を信頼度 95 % で推定して下さい。ただし母分散はサンプルの分散で代用して下さい。また、サンプルの不偏分散で代用した場合にどうなるかも計算してみてください。

基本演習 10.2 [問題集 5.16] ある動物用の新しい飼料を試作し、任意抽出された 100 匹にこの飼料を毎日与えて 1 週間後に体重の変化を調べました。増加量の平均は 2.57kg、標準偏差は 0.35kg でした。この増加量について以下の問いに答えて下さい。

(1) 母平均を信頼度 95 % で区間推定して下さい。

(2) 標本平均と母平均の差を 95 % の確率で 0.05kg 以下にするには標本数をいくらにすれば良いでしょうか。

基本演習 10.3 全国一斉にある教科のテストが行われました。受験生から 100 名を抽出し、その得点の平均と標準偏差を求めたところそれぞれ 58.3、12.4 でした。全受験生の得点の平均の 95 % 信頼区間を求めて下さい。

基本演習 10.4 ある学校の生徒 50 人を無作為に選び、1 週間に数学を何時間勉強するか聞いたところ、50 人の平均は 18.2 時間、不偏分散は 30.25 時間でした。この学校の生徒の 1 週間あたりの平均数学学習時間の 95 % 信頼区間を求めて下さい。

基本演習 10.5 ある年度の学校保健統計調査によると、全国 2 万人の 17 歳女子の身長は 157.9cm、不偏分散は 5.35^2 でした。17 歳女子の身長の平均の 95 % 信頼区間を求めて下さい。

基本演習 10.6 自動車衝突事故の対物保険についてのある研究によると、ある特殊の破損を受けた 120 台の車体を無作為に選んだところ、それらの修理費の平均は 14.4 万円で、標準偏差は 1.7 万円でした。この種の修理の平均費用を信頼度 99 パーセントで区間推定してください。ただし $\sqrt{120} \approx 10.95$ としてください。

基本演習 10.7 ある母集団から大きさ 50 の無作為サンプルを抽出して、そのサンプルの平均値が 54.3、分散が 24.5 でした。母平均の信頼度 99 % の信頼区間を求めて下さい。

問題演習 解答例

基本演習 10.1 [教科書 問題 16.8 改題] 過去の記録から、ある高専 3 年男子学生の 100 メートル走の記録の度数分布は正規分布に従うことがわかっています。

学生の中から 40 人を無作為に選び、100 メートル走の記録をとったところ平均値が 15.3 秒、分散が $(1.88 \text{ 秒})^2$ でした。全学生の平均値を信頼度 95 % で推定して下さい。ただし母分散はサンプルの分散で代用して下さい。また、サンプルの不偏分散で代用した場合にどうなるかも計算してみてください。

【解答例】母平均を m とし、母分散はサンプルの分散で代用する事にします。するとこの母集団からとった大きさ 40 の標本平均 A は正規分布 $N\left(m, \frac{1.88^2}{40}\right)$ で近似されます。

まず

$$P[|A - m| \leq d] = 0.95$$

となるような d を求めますが、

$$0.95 \approx P\left[\left|N\left(m, \frac{1.88^2}{40}\right) - m\right| \leq d\right] = P\left[|N(0, 1)| \leq \frac{d}{\sqrt{\frac{1.88^2}{40}}}\right]$$

$$0.475 = P\left[0 \leq N(0, 1) \leq \frac{d}{\sqrt{\frac{1.88^2}{40}}}\right]$$

ですから標準正規分布表によれば

$$\frac{d}{\sqrt{\frac{1.88^2}{40}}} \approx 1.96, \quad \text{すなわち} \quad d \approx 0.5826$$

が得られます。従って今回のサンプル平均 15.3 について、信頼度 95 % で

$$|15.3 - m| \leq 0.5826, \quad \text{従って} \quad 14.7174 \leq m \leq 15.8826$$

が成立すると言えますから、求める信頼区間は $[14.72, 15.88]$ です。

また、母分散をサンプル不偏分散 $\frac{40}{39}1.88^2$ で代用した場合は、この母集団からとった大きさ 40 の標本平均 B は正規分布 $N\left(m, \frac{1.88^2 \cdot 40}{39 \cdot 40}\right)$ で近似されます。まず

$$P[|B - m| \leq d] = 0.95$$

となるような d を求めますが、

$$0.95 \approx P\left[\left|N\left(m, \frac{1.88^2}{39}\right) - m\right| \leq d\right] = P\left[|N(0, 1)| \leq \frac{d}{\sqrt{\frac{1.88^2}{39}}}\right]$$

$$0.475 = P\left[0 \leq N(0, 1) \leq \frac{d}{\sqrt{\frac{1.88^2}{39}}}\right]$$

ですから標準正規分布表によれば

$$\frac{d}{\sqrt{\frac{1.88^2}{39}}} \approx 1.96, \quad \text{すなわち} \quad d \approx 0.59$$

が得られます。従って今回のサンプル平均 15.3 について、信頼度 95 % で

$$|15.3 - m| \leq 0.59, \quad \text{従って} \quad 14.71 \leq m \leq 15.89$$

が成立すると言えますから、求める信頼区間は $[14.71, 15.89]$ です。

基本演習 10.2 [問題集 5.16] ある動物用の新しい飼料を試作し、任意抽出された 100 匹にこの飼料を毎日与えて 1 週間後に体重の変化を調べました。増加量の平均は 2.57kg、標準偏差は 0.35kg でした。この増加量について以下の問いに答えて下さい。

(1) 母平均を信頼度 95 % で区間推定して下さい。

(2) 標本平均と母平均の差を 95 % の確率で 0.05kg 以下にするには標本数をいくらにすれば良いでしょうか。

【解答例】(1) 母平均を m とします。大きさ 100 の大きなサンプルを取っているので母分散は標本分散で代用出来、従って中心極限定理によれば標本平均 $\bar{X}$ は正規分布 $N\left(m, \frac{0.35^2}{100}\right)$ で近似されます。

そこでまず

$$P[|\bar{X} - m| \leq d] = 0.95$$

となる様な $d > 0$ を求めます。少し変形すれば

$$0.95 = P\left[\left|\frac{\bar{X} - m}{\sqrt{\frac{0.35^2}{100}}}\right| \leq \frac{d}{\sqrt{\frac{0.35^2}{100}}}\right] \approx P\left[|N(0, 1)| \leq \frac{d}{\sqrt{\frac{0.35^2}{100}}}\right]$$

となりますが、標準正規分布表に依れば、

$$P[|N(0, 1)| \leq 1.96] \approx 0.95$$

ですから、

$$\frac{d}{\sqrt{\frac{0.35^2}{100}}} = 1.96, \quad d = \frac{1.96 \times 0.35}{10} = 0.0686$$

であれば良い事が分かります。従って

$$P[|\bar{X} - m| \leq 0.0686] \approx 0.95$$

が分かりました。

標本平均の具体値が分かっていますのでこれを m に関する条件に読み替えれば

$$|2.57 - m| \leq 0.0686 \quad \text{すなわち} \quad 2.57 - 0.0686 \leq m \leq 2.57 + 0.0686$$

であることが 98 パーセントの信頼度で正しいわけです。

従って元データの平均値 m の、信頼度 98 % の信頼区間は $[2.501, 2.639]$ になります。

(2) 母平均を m とします。(1) の結果から見て、サンプル数 100 のときで標本平均と母平均の差は 95 % の確率で 0.0686 以下ですから、題意を満たすためにはもっと大きなサンプルをとる必要があります。仮にサンプル数を n とした場合、 $n > 100$ ですから母分散は標本分散で代用出来、また中心極限定理によれば標本平均 $\bar{X}$ は正規分布 $N\left(m, \frac{0.35^2}{n}\right)$ で近似されます。

まず

$$P[|\bar{X} - m| \leq 0.05] = 0.95$$

となる様な n を求めます。少し変形すれば

$$0.95 = P\left[\left|\frac{\bar{X} - m}{\sqrt{\frac{0.35^2}{n}}}\right| \leq \frac{0.05}{\sqrt{\frac{0.35^2}{n}}}\right] \approx P\left[|N(0, 1)| \leq \frac{0.05}{\sqrt{\frac{0.35^2}{n}}}\right]$$

となりますが、標準正規分布表に依れば、

$$P[|N(0, 1)| \leq 1.96] \approx 0.95$$

ですから、

$$\frac{0.05}{\sqrt{\frac{0.35^2}{n}}} = 1.96, \quad n \approx 188.24$$

であれば良い事が分かります。 n が大きくなれば信頼区間の幅は小さくなりますから、結局題意を満たすためには標本数は 189 以上にすれば良い事が分かります。□

基本演習 10.3 全国一斉にある教科のテストが行われました。受験生から 100 名を抽出し、その得点の平均と標準偏差を求めたところそれぞれ 58.3、12.4 でした。全受験生の得点の平均の 95 % 信頼区間を求めて下さい。

【解答例】母平均を m とします。サンプルサイズが大きいのので母分散はサンプル分散で代用します。この母集団からとった大きさ 100 の標本平均 A は、中心極限定理によれば正規分布 $N\left(m, \frac{12.4^2}{100}\right)$ で近似されます。

そこでまず

$$P[|A - m| \leq d] = 0.95$$

となるような $d > 0$ を求めますが、

$$\begin{aligned} 0.95 &\approx P\left[\left|N\left(m, \frac{12.4^2}{100}\right) - m\right| \leq d\right] \\ &= P\left[|N(0, 1)| \leq \frac{d}{1.24}\right] \end{aligned}$$

$$0.475 = P\left[0 \leq N(0, 1) \leq \frac{d}{1.24}\right]$$

ですから、標準正規分布表から

$$\frac{d}{1.24} \approx 1.96, \quad \text{従って} \quad d \approx 2.43$$

が得られます。

従って今回のサンプル平均 58.3 について、信頼度 95 % で

$$|58.3 - m| \leq 2.43, \quad \text{すなわち} \quad 55.87 \leq m \leq 60.73$$

が成り立っていると言えますから、求める信頼区間は $[55.87, 60.73]$ です。□

基本演習 10.4 ある学校の生徒 50 人を無作為に選び、1 週間に数学を何時間勉強するか聞いたところ、50 人の平均は 18.2 時間、不偏分散は 30.25 時間でした。この学校の生徒の 1 週間あたりの平均数学学習時間の 95 % 信頼区間を求めて下さい。

【解答例】母平均を m とします。サンプルサイズが大きいのので母分散はサンプル不偏分散 30.25 で代用します。この母集団からとった大きさ 50 の標本平均 A は、中心極限定理によれば正規分布 $N\left(m, \frac{30.25}{50}\right)$ で近似されます。
そこでまず

$$P[|A - m| \leq d] = 0.95$$

となるような $d > 0$ を求めますが、

$$\begin{aligned} 0.95 &\approx P\left[\left|N\left(m, \frac{30.25}{50}\right) - m\right| \leq d\right] \\ &= P\left[|N(0, 1)| \leq \frac{d}{0.778}\right] \\ 0.475 &= P\left[0 \leq N(0, 1) \leq \frac{d}{0.778}\right] \end{aligned}$$

ですから、標準正規分布表から

$$\frac{d}{0.778} \approx 1.96, \quad \text{従って} \quad d \approx 1.525$$

が得られます。

従って今回のサンプル平均 18.2 について、信頼度 95 % で

$$|18.2 - m| \leq 1.525, \quad \text{すなわち} \quad 16.675 \leq m \leq 19.725$$

が成り立っていると言えますから、求める信頼区間は $[16.675, 19.725]$ です。□

基本演習 10.5 ある年度の学校保健統計調査によると、全国 2 万人の 17 歳女子の身長平均は 157.9cm、不偏分散は 5.35^2 でした。17 歳女子の身長平均の 95 %信頼区間を求めて下さい。

【解答例】母平均を m とします。サンプルサイズが大きいのので母分散はサンプル不偏分散 5.35^2 で代用します。この母集団からとった大きさ 20000 の標本平均 A は、中心極限定理によれば正規分布 $N\left(m, \frac{5.35^2}{20000}\right)$ で近似されます。

そこでまず

$$P[|A - m| \leq d] = 0.95$$

となるような $d > 0$ を求めますが、

$$0.95 \approx P\left[\left|N\left(m, \frac{5.35^2}{20000}\right) - m\right| \leq d\right] = P\left[|N(0, 1)| \leq \frac{d}{0.0378}\right]$$

$$0.475 = P\left[0 \leq N(0, 1) \leq \frac{d}{0.0378}\right]$$

ですから、標準正規分布表から

$$\frac{d}{0.0378} \approx 1.96, \quad \text{従って} \quad d \approx 0.074$$

が得られます。

従って今回のサンプル平均 157.9 について、信頼度 95 %で

$$|157.9 - m| \leq 0.074, \quad \text{すなわち} \quad 157.826 \leq m \leq 157.974$$

が成り立っていると言えますから、求める信頼区間は $[157.83, 157.97]$ です。□

基本演習 10.6 自動車衝突事故の対物保険についてのある研究によると、ある特殊の破損を受けた 120 台の車体を無作為に選んだところ、それらの修理費の平均は 14.4 万円で、標準偏差は 1.7 万円でした。この種の修理の平均費用を信頼度 99 パーセントで区間推定してください。ただし $\sqrt{120} \approx 10.95$ としてください。

【解答例】この種の修理の費用全体を母集団 X とし、そこからとった大きさ 120 の標本平均を $\bar{X}$ とします。

大きなサイズの標本をとっているのので、 X の標準偏差はサンプルの標準偏差 1.7 で代用することが出来、また中心極限定理によれば、母集団の平均を m として $\bar{X}$ は $N\left(m, \frac{1.7^2}{120}\right)$ で近似されます。

まず

$$0.99 = P[|\bar{X} - m| \leq d]$$

となるような $d > 0$ を求めます。

$$0.99 = P[|\bar{X} - m| \leq d] \approx P\left[\left|N\left(m, \frac{1.7^2}{120}\right) - m\right| \leq d\right] = P\left[|N(0, 1)| \leq \frac{d}{\frac{1.7}{\sqrt{120}}}\right]$$

$$0.495 = P\left[0 \leq N(0, 1) \leq \frac{d}{\frac{1.7}{\sqrt{120}}}\right]$$

標準正規分布表によれば

$$\frac{d}{\frac{1.7}{\sqrt{120}}} \approx 2.575$$

$$d \approx \frac{2.575 \times 1.7}{10.95} = \frac{4.3775}{10.95} \approx 0.3998$$

ですから、

$$0.99 = P[|\bar{X} - m| \leq 0.3998]$$

が成り立っています。

従って今回のサンプル値 14.4 について、信頼度 99 パーセントで不等式：

$$|14.4 - m| \leq 0.3998$$

つまり、

$$14.0002 \leq m \leq 14.7998$$

が成り立ちますから、これが求める推定区間です：

$$[14.0, 14.8]$$

□

基本演習 10.7 ある母集団から大きさ 50 の無作為サンプルを抽出して、そのサンプルの平均値が 54.3、分散が 24.5 でした。母平均の信頼度 99 %の信頼区間を求めて下さい。

【解答例】サンプルサイズが大きいのので、サンプルの分散で母分散の代用とします。また、中心極限定理によれば、母平均を m とするとこの母集団からとった大きさ 50 の標本平均 $\bar{X}$ は正規分布 $N\left(m, \frac{24.5}{50}\right)$ で近似されます。

まず

$$0.99 = P[|\bar{X} - m| \leq d]$$

となるような $d > 0$ を求めます。

$$0.99 = P[|\bar{X} - m| \leq d]$$

$$\approx P\left[\left|N\left(m, \frac{24.5}{50}\right) - m\right| \leq d\right]$$

$$\begin{aligned} &= P[|N(0, 0.49)| \leq d] \\ &= P\left[|N(0, 1)| \leq \frac{d}{0.7}\right] \\ 0.495 &= P\left[0 \leq N(0, 1) \leq \frac{d}{0.7}\right] \end{aligned}$$

すると正規分布表より

$$2.575 \approx \frac{d}{0.7} \quad \text{従って} \quad d \approx 1.8025$$

が得られます。

従って

$$0.99 \approx P\left[|\bar{X} - m| \leq 1.80\right]$$

が成り立ちますので、今回のサンプル値 54.3 について、信頼度 99 %で

$$\begin{aligned} |54.3 - m| &\leq 1.80 \\ 54.3 - 1.80 &\leq m \leq 54.3 + 1.80 \\ 52.50 &\leq m \leq 56.10 \end{aligned}$$

が成り立ちますから、求める信頼区間は [52.5, 56.1] です。

□