

## 7章 区間推定

通常の初心者向けの教科書では検定よりもむしろこの章で学ぶ区間推定に重点を置く場合があります。それは初心者にとって直感的に理解しやすいということですが、理論に正規性を仮定しているため、どんな場合にでも使えるというものではありません。このテキストでは実用性を重視したため正規分布でなくても使える検定を中心に話を進めてきました。しかし区間推定は基本的な方法なので、最後にどうしても取り上げておかなければなりません。

区間推定は標本から推測される母比率や母平均などがどの位の値の範囲に入るかを区間で表す方法です。推定した区間に母比率や母平均などが入る確率を信頼係数と言います。通常は%で表されることが多く、95%か99%がよく使われます。1-信頼係数の値は検定での有意水準に相当します。

### 7.1 母比率の区間推定

最初は母比率の区間推定についてです。以下の例を見て下さい。

#### 例

ある制度についてのアンケート調査をランダムに抽出された100人に対して行ったところ、賛成65人、反対35人であった。母集団の賛成の比率を、信頼係数95%（有意水準5%に相当）で推定せよ。また、調査数1000人で同じ比率ではどうか。

メニュー「分析-基本統計-区間統計-比率の推定」で分析メニューを表示し、この場合、「集計から」ラジオボタンを選択して、データ数と標本比率を書き込み、「母比率の推定」ボタンをクリックすると図7.1.1のようになります。



図 7.1.1 母比率の区間推定

これより母比率の 95% 信頼区間は以下となります。

$$0.55652 \leq p \leq 0.74348$$

また調査数 1000 人では、以下のように精度が上がります。

$$0.62044 \leq p \leq 0.67956$$

信頼係数の値は、「変数選択」ボタンで変更できます。元データから直接推定を行う場合は、「データから」ラジオボタンを選択し、指定カテゴリ名テキストボックスにどの分類について推定を行うかを記入し、「母比率の推定」ボタンをクリックします。

この分析で利用された理論は、検定に利用されるものと同じで、以下の通りです。

### 理論

データ数  $n$ 、標本比率  $\hat{p}$  の標本から、母比率  $p$  を信頼係数  $(1 - \alpha) \times 100\%$  で推定する。

$z_0 = \text{normsinv}(1 - \alpha / 2)$  として、信頼区間は以下で与えられる。

$$\hat{p} - \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} z_0 \leq p \leq \hat{p} + \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} z_0$$

この理論から、信頼係数を大きく取ると信頼区間は拡がり、調査数が多くなれば信頼区間は狭まることがあります。

### 7.2 正規母集団の母平均と母分散の区間推定

ここでは量的データについて標本データからの母平均と母分散の推定方法を学びます。但し、このデータについて正規性があるものと仮定します。正規性がない場合は使えません。以下の例を見て下さい。

## 例

ある標本データから所得について集計したところ以下の結果を得た。母集団は正規分布するとして母平均と母分散を信頼係数 95% で推定せよ。

データ数 30, 平均 620, 標準偏差 90

また、データ数を 100 にすると結果はどう変わるか？

メニュー [分析ー基本統計ー区間統計ー平均と分散の推定] で分析メニューを表示し、この場合、「集計から」ラジオボタンを選択して、データ数、平均、不偏分散または標準偏差を書き込み、「母平均の推定」ボタンをクリックすると図 7.2.1 のようになります。

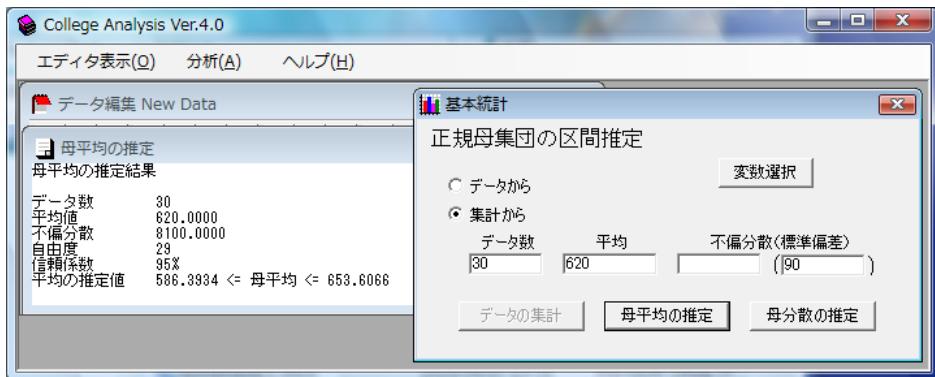


図 7.2.1 母平均の区間推定

これより母平均の 95% 信頼区間は以下となります。

$$586.3934 \leq \text{母平均} \leq 653.6066$$

同様にして、「母分散の推定」ボタンをクリックすると図 7.2.2 のようになります。

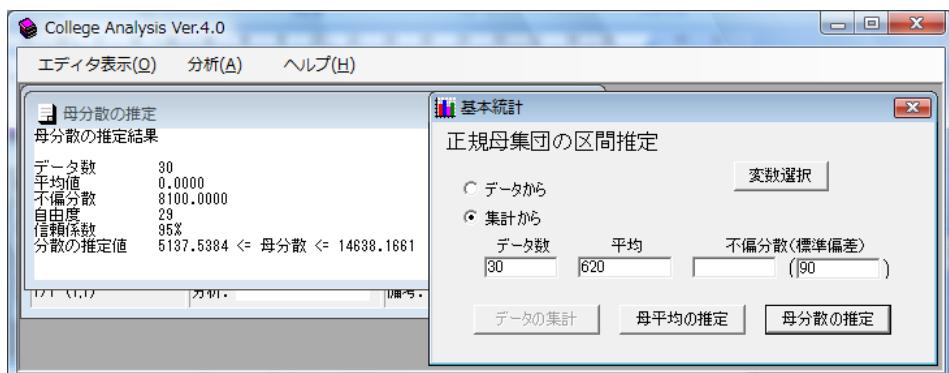


図 7.2.2 母分散の区間推定

これより母分散の 95% 信頼区間は以下となります。

$$5138 \leq \text{母分散} \leq 14638$$

データ数を 100 にすると、以下のように精度が向上することが分ります。

$$602.142 \leq \text{母平均} \leq 637.858$$

$$6244 \leq \text{母分散} \leq 10932$$

ここで用いた理論は t 検定や F 検定と同じですが、書き方が異なります。

### 理論

正規分布する母集団から得られた標本より、母平均  $\mu$  と母分散  $\sigma^2$  を信頼係数  $(1-\alpha) \times 100\%$  で推定する。データ数を  $n$ ，標本平均を  $\bar{x}$ ，不偏分散を  $u^2$ ，  
 $t_0 = tinv(\alpha, n-1)$ ， $x_1 = chiinv(1-\alpha/2, n-1)$ ， $x_2 = chiinv(\alpha/2, n-1)$  として、各信頼区間は以下で与えられる。

$$\text{母平均} : \bar{x} - \frac{u}{\sqrt{n}} t_0 \leq \mu \leq \bar{x} + \frac{u}{\sqrt{n}} t_0$$

$$\text{母分散} : \frac{(n-1)u^2}{x_2} \leq \sigma^2 \leq \frac{(n-1)u^2}{x_1}$$

### 問題 1

ある 500 人に対する調査で支持 205 人、不支持 295 人という結果を得た。母集団における支持の比率を信頼係数 95% で推定せよ。

$$\text{信頼区間は } [\quad] \leq \text{母比率} \leq [\quad]$$

### 問題 2

正規分布を仮定して、以下の身長データ(cm)から母平均と母分散を信頼係数 95% で推定せよ。

184, 170, 164, 176, 177, 170, 171, 159, 174, 170,

165, 170, 171, 183, 175, 169, 181, 172, 171, 164

$$[\quad] \leq \text{母平均} \leq [\quad]$$

$$[\quad] \leq \text{母分散} \leq [\quad]$$

### 問題 3

Samples¥テキスト 9.txt のデータを用いて以下の問いに答えよ。

1) ここで用いた区間推定の手法は、支出に利用できるか。[可能・不可能]

可能な場合は支出の母平均と母分散を信頼係数 95% で推定せよ。

$$[\quad] \leqq \text{母平均} \leqq [\quad]$$

$$[\quad] \leqq \text{母分散} \leqq [\quad]$$

2) 上の結果を用いて、支出の平均は 50 (万円) と差があるかどうか有意水準 5% で判定したい。信頼区間 [内・外] なので、差があると [いえる・いえない]。

## 8章 アンケート調査

ここではアンケート調査票の作り方について少しだけ、まとめておきます。詳しくは多くの本が出ていますので、そちらを参考にして下さい。注意事項とそれに基づくごく簡単なアンケート票の例を示します。

### アンケート注意事項

1) アンケートは次の順序で作る。

    タイトル, あいさつ文, 調査団体または代表者名,  
    アンケート本文, 謝辞

2) 何を知りたいか十分検討し、アンケート対象者や項目を選ぶ。

    アンケートの対象は、全数調査か、調べたい対象の中から無作為に抽出した標本とする。但し、年齢構成などで層別に抽出する場合もある。

    質問に漏れがないか十分注意する。

    例えば意見の男女差を知りたければ、当然性別を聞いておく必要がある。  
    最初に区分けのための質問、続いて具体的な意見などを聞く方が答え易い。  
    集計のことを頭に置いて質問項目を考える。

    不必要なことはできるだけ聞かずに、アンケートをコンパクトにまとめる。

3) 質問は答え易い形で書く。

    数字を書かせる場合と自由記述を除いては、番号を選ぶのが無難。

    例 あなたの性別は        1) 男        2) 女

    集計と統計処理の簡単化のため、番号選択は1つか、いくつでもかが無難。

    例 あなたの最も大切にしていることはなんですか。以下から1つだけ  
    選んで下さい。

    あなたの大切にしているものはなんですか。以下の該当するものす  
    べてを選んで下さい。

    明らかな場合を除いて、選択肢の中には「その他」の項目を設け、具体的な  
    内容を書く欄を添える。

    例 1) 製造業    2) 流通業    3) サービス業  
        4) その他 [ ]

    具体的な数字を書かせる場合は、単位を明確に。(千円はやめておくべき)

    例 あなたの年収は \_\_\_\_\_ 万円

    質問項目の右側に回答欄を設けると集計に便利であるが、利用しない人もいるので注意する。

    回答者を絞って答えてもらう場合は、分かり易さを心掛ける。

    例 前問で「1) はい」と答えた人のみ回答して下さい。その他の人には

設問5へ進んで下さい。

4) その他

予め集計用のフォームを考えておく。(大規模でなければExcelは有力)  
あらかじめ少數の人で試し、集計までをシミュレーションしておく。  
回収後、回答用紙には必ず整理番号を振っておく。

## 学生生活アンケート調査

この度情報処理論Ⅱの授業において、アンケートの作成法とその集計方法を学ぶために仮想的なアンケート調査を実施することになりました。個人のプライバシー等につきましては十分な注意を払うことはもちろんですが、このアンケートをその他の目的に使用することはありません。どうかご協力をお願い致します。

福山平成大学 福井正康

質問1 あなたの性別は？

- 1) 男性      2) 女性

質問2 あなたは自宅通学ですか？

- 1) 自宅通学      2) 自宅通学でない

質問3 あなたの自由に使えるお金(生活費を除く)は1ヶ月におよそいくらですか？

[\_\_\_\_\_円]

質問4 あなたはアルバイトをしていますか？

- 1) している      2) していない

前問で1) していると答えた人だけ回答して下さい。その他の人は質問7へ進んで下さい。

質問5 どれ位の頻度でアルバイトをしていますか？1つ選んで下さい。

- 1) 週5日以上      2) 週3, 4日      3) 週1, 2日  
4) 長期休業時のみ      5) その他 [ ]

質問6 あなたのアルバイトの収入は1ヶ月におよそいくらですか？不定期にやっている人は、1ヶ月にならしてお答え下さい。

[\_\_\_\_\_円]

**質問7** あなたの現在の悩みに当てはまるものがあればいくつでも選択して下さい。

- 1) 特にない
- 2) 勉学上の問題
- 3) 金銭問題
- 4) 異性問題
- 5) 健康上の問題
- 6) 就職・進路の問題
- 7) その他 [\_\_\_\_\_]

ご協力有難うございました。

最後にアンケート報告書を書く際の簡単な注意事項を書いておきます。

### アンケート報告書注意事項

- 1) タイトル、調査団体名または代表者名及び住所等（ここまで表紙にしてもよい）を最初に示す。
- 2) アンケートの実施時期と実施方法、対象数と回収数・回収率を明記する。
- 3) アンケート集計結果は以下の点に注意する。

単純集計から始めて、次にクロス集計をする。

図表には番号とタイトルを付け（通し番号または章ごと）、文中で指定して説明を加える。例 図1に設問3のヒストグラムを示す。

図表番号とタイトルを付ける位置として、表は上側、図は下側が多い。  
必要があれば、その他を選んだ場合の内容を紹介してもよい。

質問用紙を最後に掲載するのもよい。

#### 4) 集計・検定結果の表示

集計値の桁数は、平均・標準偏差等でデータ桁数より1桁か2桁程度多く表示する。

例：171, 173, 174, … → 平均 172.7

検定の際、t検定とか Wilcoxon の順位和検定とか、手法の名前は明らかにした方がよいが、t統計量の値や自由度などは書かない。

有意水準・検定確率値・判定については必要に応じて流れの中で記述する。  
検定確率値については、小数点以下3桁か4桁で表示する。

### 問題

上のアンケートの結果である Samples¥テキスト 10.txt のデータを用いて、以下の空欄を埋めて、報告書を完成させよ。

## 学生生活アンケート調査報告書

福山平成大学 福井正康

福山平成大学では 20XX 年 11 月 28 日に、本学情報処理論 II の授業で受講生 53 名を対象に「学生生活アンケート調査」を対面して記述させる方式で実施した。調査結果の回収数は □ で回収率は □ % であった。この報告書で行なった検定については有意水準を 5% としている。

男女別にみると男 □ 名、女 □ 名であり、自宅通学かどうかをみると自宅通学 □ 名、自宅通学以外は □ 名であった。アルバイトをしている学生は □ 名、していない学生は □ 名で、アルバイトをしている割合は、□ % であった。アルバイトをしているかどうか通学区分別に見ると、表 1 のようになった。

表 1 通学区分によるアルバイト状況

	している	していない
自宅		
自宅外		

これから通学区分によるアルバイト状況の有意差は見られなかった。また、アルバイトの頻度は、週 5 回以上 □ 名、3~4 回 □ 名 1~2 回 □ 名であった。

自由に使える 1 ヶ月の金額は、平均 □ 万円、標準偏差 □ 万円であり、そのヒストグラムを描くと、図 1 のようになった。

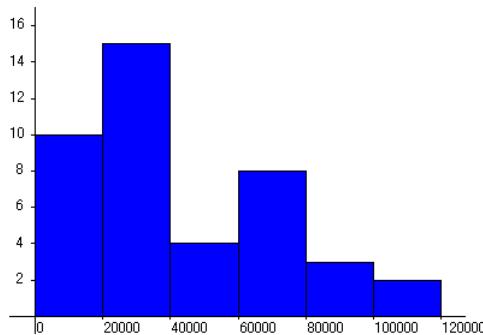


図 1 自由に使える金額

性別、通学別、アルバイト状況別の自由に使える金額の平均は表 2 のようになった。

表 2 各分類別平均 (万円)

性別		通学		アルバイト	
男	女	自宅	自宅外	している	していない
□	□	□	□	□	□

図 1 のヒストグラムの形から、データが正規分布していると考えにくいので、これら

の差を Wilcoxon の順位和検定で調べたところ、アルバイトをしているかどうかで有意な差が見られたが ( $p=\square$ )、その他については有意な差は見られなかった。もう少しデータ数を増やして、男女間の差について検討するのも興味深い。

アルバイト収入の平均は  $\square$  万円、標準偏差は  $\square$  万円であった。また、自由に使える金額とアルバイト収入の関係は、図 2 で与えられ、アルバイト収入がないものを除いた相関係数は  $\square$  であった。このことからアルバイト収入と自由に使える金額には相関関係があると思われる。

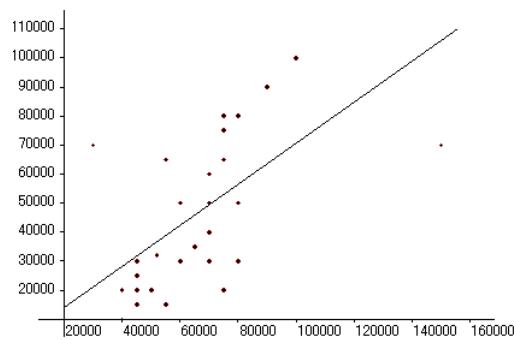


図 2 アルバイト収入（横軸）と使える金額（縦軸）の相関

自由に使える金額を目的変数、アルバイト収入を説明変数として回帰分析を行なったところ、寄与率  $\square$  で、 $y=\square x+\square$  という結果が得られた。回帰直線は図 2 に記入している。

悩みについては「なし」が  $\square$  名、項目のどれかにチェックをした学生は  $\square$  名であった。全体の中で悩みの種類毎の比率は、図 3 のようになる。不況を反映してであろうか、金銭と就職の問題の比率が高いように思われる。

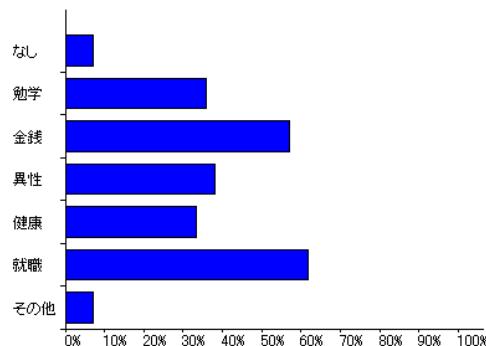


図 3 憂みの種類の割合