第 7 回：二標本検定

北村　友宏

2020 年 6 月 19 日

本日の内容

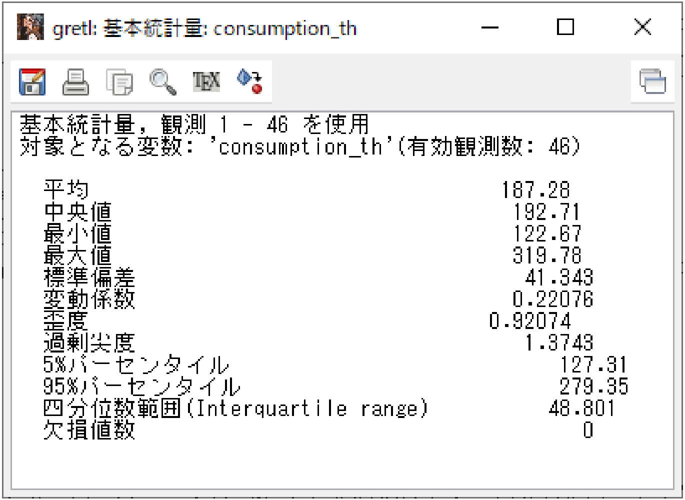
1. 異なるグループの間での平均の差
2. 統計的推測
3. 二標本検定（母分散が等しいと仮定する場合）
4. 二標本検定（母分散が異なると仮定する場合）

実習 1

消費支出（千円単位）の記述統計を，男女別に出力してみる．まずは男性の可処分所得（千円単位）の記述統計を出力する．そのために，一時的にデータセットの観測値の範囲を男性のみに限定する．

1. gretl を起動．
2. 「ファイル」→「データを開く」→「ユーザー・ファイル」と操作．
3. 消費 2009.gdt を選択し，「開く」をクリック．
4. gretl のメニューバーから「標本」→「基準に基づいて制限する」と操作．
5. 「ダミー変数を使用する:」をクリックし，その右のプルダウンメニューから male を選択し，OK をクリック．

6. 「46 個の観測を落としました」というメッセージが表示されるので，「閉じる」をクリック．この段階で，絶対に上書き保存しない！
   - データセットの観測値の範囲が，male というダミー変数の値が 1 になっているグループ（男性）のみに限定された．
7. gretl のウィンドウで，consumption_th を左クリックして選択してから右クリック→「基本統計量」と操作すると，男性のみに限定された，consumption_th の記述統計が出力される．



このような画面が表示されれば成功．まだウィンドウを閉じない！

男性の可処分所得（千円単位）の記述統計から分かること

- 平均値：187.28
  - 千円単位なので，18.728 万円（約 18 万 7280 円）
- 標準偏差：41.343
- 観測値数：46
  - 「有効観測数」に表示されている．

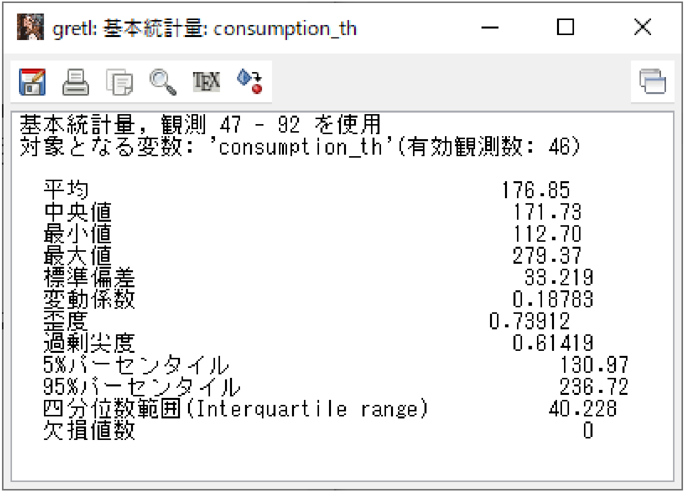
これらの数字を，後で使用する．

実習 2

続いて，女性の可処分所得（千円単位）の記述統計を出力する．そのために，データセットの観測値の範囲を一旦元に戻した後，一時的にデータセットの観測値の範囲を女性のみに限定する．

1. gretl のメニューバーから「標本」→「全範囲に戻す」と操作すると，先ほど男性のみに限定した観測値の範囲が元に戻る．
2. gretl のメニューバーから「標本」→「基準に基づいて制限する」と操作．
3. 「ダミー変数を使用する:」をクリックし，その右のプルダウンメニューから female を選択し，OK をクリック．

4. 「46 個の観測を落としました」というメッセージが表示されるので，「閉じる」をクリック．この段階で，絶対に上書き保存しない！
    - データセットの観測値の範囲が，female というダミー変数の値が 1 になっているグループ（女性）のみに限定された．
5. gretl のウィンドウで，consumption_th を左クリックして選択してから右クリック→「基本統計量」と操作すると，女性のみに限定された，consumption_th の記述統計が出力される．



このような画面が表示されれば成功．まだウィンドウを閉じない！

女性の可処分所得（千円単位）の記述統計から分かること

- 平均値：176.85
  - 千円単位なので，17.685 万円（約 17 万 6850 円）
- 標準偏差：33.219
- 観測値数：46
  - 「有効観測数」に表示されている．

これらの数字を，後で使用する．

平均の差

現在分析中のデータによれば，

- 男性の消費支出額の平均は，約 18 万 7280 円
- 女性の消費支出額の平均は，約 17 万 6850 円

⇒ 約 1 万 430 円の差がある．

⇒ しかし，このことから直ちに，「男性は女性よりも消費支出額が高い」と言うことはできない．

- 理由

  『全国消費実態調査』は，日本の世帯の一部を抽出して調査．

  ➥ どの世帯が抽出されるかによって，男女別平均とその差は異なる．

  ➥ 約 1 万 430 円の差は，たまたま出たもの．

  ➥ 本当は（日本人全体では）男女間では差がないかもしれない．

- 「約 1 万 430 円の差」は，統計的に意味のある差なのか？

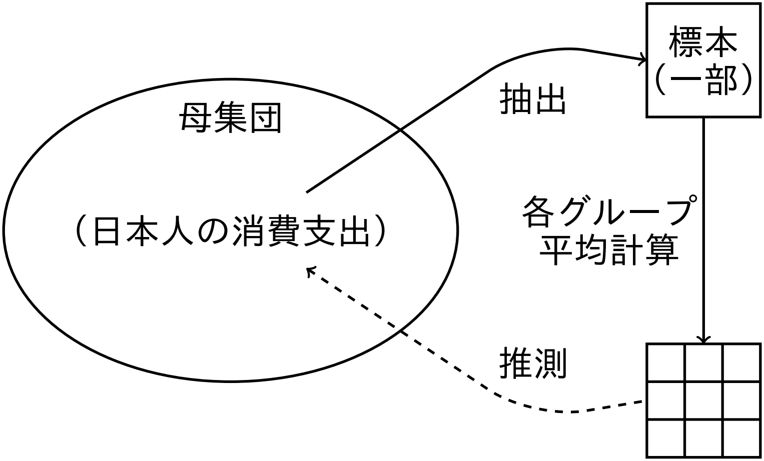
$$\Downarrow$$

二標本検定を実行し，ルールに基づいて検証.

統計的推測

- 調査対象の集団全体を母集団（population）という．
  - e.g., 日本全国の単身勤労世帯の消費支出額（全世帯を調べるのは大変）
- 母集団の中から個体を抽出して集めたものを標本（sample）という．
  - e.g., 日本全国の単身勤労世帯のうち，一部の世帯の１ヶ月間の消費支出額
- 標本から母集団の特徴を推測することを統計的推測（statistical inference）という．
  - e.g., 日本の単身勤労世帯の一部の１ヶ月間の消費支出の都道府県別・男女別平均データを用い，「日本全国の単身勤労世帯の１ヶ月間の消費支出額は，男女間で差があるのか」を推測する．

統計的推測のイメージ



仮説検定

- 母集団分布の平均を母平均（population mean）という．
- 母集団分布の分散を母分散（population variance）という．

⇒ 母集団全体を観測できない場合，どちらも現実には未知．

- 母集団の特性に関する仮説を，標本を用いて検証することを仮説検定（hypothesis testing）という．
  - e.g.,「男性の消費支出額の母平均 μ_M は女性の消費支出額の母平均 μ_F と変わらない」という仮説を，1ヶ月間の消費支出額の都道府県別・男女別平均の標本を用いて検証する．

帰無仮説

- とりあえず「真」であると想定する仮説を帰無仮説（null hypothesis）という.
  - H_0 と書くことが多い.
  - e.g., $H_0 : \mu_M = \mu_F$.
  - H_0 は必ず「=」または「$\leq$ や $\geq$」を使った式.
    「$\mu_M > \mu_F$」を H_0 とする検定は不可能.
- まずは H_0 が「真」であると仮定し，それを「偽」とするための証拠を探す.
  - 刑事裁判における推定無罪の原則と同様（被告人は，判決が出るまでは「無罪」として扱われ，「有罪」とするための証拠が探され裁判で提示される）.

$\Longrightarrow$ 具体的には，検定統計値を計算する.

- 標本の関数を統計量（statistic）という．
  - e.g., 標本平均，標本分散など
- 検定に用いる統計量を検定統計量（test statistic）といい，その実現値を検定統計値という．

- 仮に H_0 が真であれば，計算した検定統計値が5%や 1%のわずかな確率でしか生じえない値になっている

$$\Downarrow$$

それを証拠として H_0 を偽と判断し，H_0 を棄却する（reject）．

- 仮に H_0 が真であれば，計算した検定統計値が小さすぎない確率で生じうる値になっている

$$\Downarrow$$

H_0 を偽とする証拠が不十分であり，偽とはいえないと判断し，H_0 を採択する（accept）．
  - 15%や 20%は「小さすぎない」.
  - 「H_0 は真」という判断ではない.

対立仮説

- H_0 が偽のときに代わりに採択する仮説を対立仮説（alternative hypothesis）という．
  - H_1 と書くことが多い．
  - e.g., $H_1 : \mu_M > \mu_F$.
  - H_1 は「$\neq, <, >$」を使った式で設定できる．
- 両側検定（two-sided test）問題の定式化：

$$H_0 : \mu_M = \mu_F \quad \text{vs} \quad H_1 : \mu_M \neq \mu_F$$

➡ H_1 の意味は，「母平均は男女で異なる」．

- 片側検定（one-sided test）問題の定式化：

$$H_0 : \mu_M = \mu_F \quad \text{vs} \quad H_1 : \mu_M > (<) \mu_F$$

➡ H_1 の意味は，「母平均は男性（女性）のほうが大きい」．

有意水準

- H_0 が真なのに棄却することを第 1 種の誤り（type I error）という．
- H_0 を真としたときに，検定統計値が「わずかな確率でしか生じえない値」かの判断の基準となる確率，また，許容する第 1 種の誤りの確率を有意水準（significance level）という．
  - 通常は 10%，5%，1%に設定．
  - e.g., $\boxed{\text{「有意水準 5\%で } H_0 \text{ が棄却された」}}$
    ⇒ 仮に H_0 が真であれば，そんな検定統計値が出てくる確率は 5%以下に過ぎない（H_0 を偽とする証拠）ので H_0 を棄却．
    ⇒ 言い換えると，H_0 が真のとき，「そんな検定統計値」は 5%以下の確率で出現しうる．
    ⇒ H_0 を棄却する第 1 種の誤りを犯すことが，多くとも 5%の確率でありうる．

p 値

- 検定統計量（の絶対値）が実現値（検定統計値）を超える（以上になる）確率を p 値という.
  - p 値が 0.1 以下（未満）: 有意水準 10%で H_0 を棄却.
  - p 値が 0.05 以下（未満）: 有意水準 5%で H_0 を棄却.
  - p 値が 0.01 以下（未満）: 有意水準 1%で H_0 を棄却.

$\Rightarrow$ p 値を見て，帰無仮説の採択・棄却を判断できる.

※検定統計量が連続型の確率分布（正規分布，t 分布，カイ二乗分布，F 分布など）に従う場合，「以上」と「超える」,「以下」と「未満」は区別しなくて良い.

二標本検定（母分散が等しいと仮定する場合）

- 異なる 2 つの母集団から無作為抽出した標本を用いた，母平均の差の仮説検定を二標本検定（two-sample test）という．
  - ここでは，男性と女性が異なる 2 つの母集団に属しているとする．
- 男性の各都道府県の平均消費支出額の無作為標本を $(X_{M1}, X_{M2}, \cdots, X_{Mn_M})$，女性の各都道府県の平均消費支出額の無作為標本を $(X_{F1}, X_{F2}, \cdots, X_{Fn_F})$ とする．
  - 平成 21 年版の単身勤労世帯のデータでは，具体的には $n_M = n_F = 46$.

- 仮定
  - $\left(X_{M1}, X_{M2}, \cdots, X_{Mn_M}\right)$ と $\left(X_{F1}, X_{F2}, \cdots, X_{Fn_F}\right)$ は互いに独立.
  - 平均消費支出額は正規分布に従っており，母分散 σ^2 は男女間で共通.

$$X_{Mi} \sim N\left(\mu_M, \sigma^2\right) \quad \text{for } i = 1, 2, \cdots, n_M,$$
$$X_{Fi} \sim N\left(\mu_F, \sigma^2\right) \quad \text{for } i = 1, 2, \cdots, n_F.$$

  - μ_M : 男性の消費支出額の母平均
  - μ_F : 女性の消費支出額の母平均

- 「$\mu_M = \mu_F$（男性と女性の消費支出額の母平均が同じ）」という帰無仮説を検定する検定統計量は，

$$t = \frac{\overline{X_M} - \overline{X_F}}{S\sqrt{\frac{1}{n_M} + \frac{1}{n_F}}} \sim t(n_M + n_F - 2).$$

  - $\overline{X_M} = \frac{1}{n_M}\sum_{i=1}^{n_M} X_{Mi}, \overline{X_F} = \frac{1}{n_F}\sum_{i=1}^{n_F} X_{Fi},$

$$S = \sqrt{\frac{\sum_{i=1}^{n_M}\left(X_{Mi} - \overline{X_M}\right)^2 + \sum_{i=1}^{n_F}\left(X_{Fi} - \overline{X_F}\right)^2}{n_M + n_F - 2}}.$$

  - H_0 が真という仮定の下で，自由度 $n_M + n_F - 2$ の t 分布に従う．

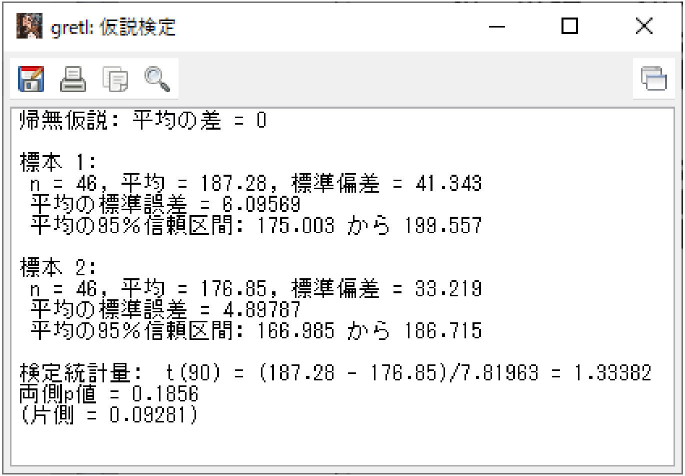
実習 3

「男性と女性の消費支出額は同じ」という帰無仮説を，母分散が男女間で等しいという仮定の下で，検定してみる．

1. gretl のメニューバーから「標本」→「全範囲に戻す」と操作すると，先ほど女性のみに限定した観測値の範囲が元に戻る．
2. gretl のメニューバーから「ツール」→「検定統計量計算機」と操作．

3. 「2 標本の平均」タブの,「標本 1 の平均」の入力ボックスに 187.28 と入力.
    - 男性の可処分所得（千円単位）の平均値を入力.
4. 「標本 1 の標準偏差」の入力ボックスに 41.343 と入力.
    - 男性の可処分所得（千円単位）の標準偏差を入力.
5. 「標本 1 のサイズ」の入力ボックスに 46 と入力.
    - 男性の可処分所得（千円単位）の観測値数を入力.
6. 「2 標本の平均」タブの,「標本 2 の平均」の入力ボックスに 176.85 と入力.
    - 女性の可処分所得（千円単位）の平均値を入力.

7. 「標本 2 の標準偏差」の入力ボックスに 33.219 と入力.
    - 女性の可処分所得（千円単位）の標準偏差を入力.
8. 「標本 2 のサイズ」の入力ボックスに 46 と入力.
    - 女性の可処分所得（千円単位）の観測値数を入力.
9. 「共通の母標準偏差を仮定する」にチェックが入っていることを確認. 入っていなければクリックしてチェックを入れる.
    - 母分散が男女間で等しいという仮定の下での検定となる.
10. 「帰無仮説 (H0): 平均の差=」の入力ボックスに 0 と入力されていることを確認. 入力されていなければ 0 と入力.
    - 「標本 1 の母平均から標本 2 の母平均を引いたものが 0」つまり「標本 1 の母平均と標本 2 の母平均が等しい」という帰無仮説の検定となる.
11. OK をクリック.



このような画面が表示されれば成功．まだウィンドウを閉じない！

出力結果の見方

- この場合，標本 1 は男性，標本 2 は女性.
- 検定統計量：t 検定における検定統計量の実現値（t 値）
  - t (90) のカッコ内は，H_0 が真という仮定の下で検定統計量が従う t 分布の自由度（この場合は 90）
- この場合の両側検定の対立仮説は「男性と女性の消費支出額は異なる」
- 「両側 p 値」の下の「片側」は，右片側検定の p 値
  - t 値がマイナスの場合は，「両側 p 値」の下の「片側」は左片側検定の p 値
- この場合の右片側検定の対立仮説は「男性は女性よりも消費支出額が高い」

検定結果

- t 値は 1.33382.
- 両側検定
  - p 値は 0.1856 で，0.1 を超えている．
    ➡ 「男性と女性の消費支出額は同じ」という帰無仮説は棄却できない（採択される）．
    ➡ 男性と女性の消費支出額には，有意水準 10%で統計的に有意な差はない．
- 右片側検定
  - p 値は 0.09281 で，0.1 を下回り，0.05 を超えている．
    ➡ 「男性と女性の消費支出額は同じ」という帰無仮説は，有意水準 10%で棄却される（5%や 1%では採択）．
    ➡ 男性と女性の消費支出額には，有意水準 10%で統計的に有意な差があり，男性は女性よりも消費支出額が有意に高い．

$\Downarrow$

両側検定と右片側検定で異なる判断

二標本検定（母分散が異なると仮定する場合）

- 仮定
  - $(X_{M1}, X_{M2}, \cdots, X_{Mn_M})$ と $(X_{F1}, X_{F2}, \cdots, X_{Fn_F})$ は互いに独立.
  - 平均消費支出額は正規分布に従っており，母分散は男女で異なる.

$$
\begin{aligned}
X_{Mi} &\sim N\left(\mu_M, \sigma_M^2\right) \quad \text{for } i = 1, 2, \cdots, n_M, \\
X_{Fi} &\sim N\left(\mu_F, \sigma_F^2\right) \quad \text{for } i = 1, 2, \cdots, n_F.
\end{aligned}
$$

  - σ_M^2 ：男性の消費支出額の母分散
  - σ_F^2 ：女性の消費支出額の母分散

- この場合，検定統計量は，

$$t = \frac{\overline{X_M} - \overline{X_F}}{\sqrt{\frac{S_M^2}{n_M} + \frac{S_F^2}{n_F}}},$$

  - $$S_M^2 = \frac{1}{n_M - 1} \sum_{i=1}^{n_M} \left(X_{Mi} - \overline{X_M} \right)^2,$$
    $$S_F^2 = \frac{1}{n_F - 1} \sum_{i=1}^{n_F} \left(X_{Fi} - \overline{X_F} \right)^2.$$

  とするのが適切．
- この検定統計量は，t 分布には従わない（Behrens-Fisher 問題）．

- H_0 が真という仮定の下で，検定統計量が，近似的に自由度が

$$\frac{\left(S_M^2/n_M + S_F^2/n_F\right)^2}{\frac{\left(S_M^2/n_M\right)^2}{n_M-1} + \frac{\left(S_F^2/n_F\right)^2}{n_F-1}}$$

の t 分布に従うとして，検定する．

- ただし，gretl では母分散が等しいと仮定する場合と同様に，

$$t = \frac{\overline{X_M} - \overline{X_F}}{S\sqrt{\frac{1}{n_M} + \frac{1}{n_F}}},$$

が検定統計量として用いられ，自由度は前スライドに記載の，修正されたものが用いられる．

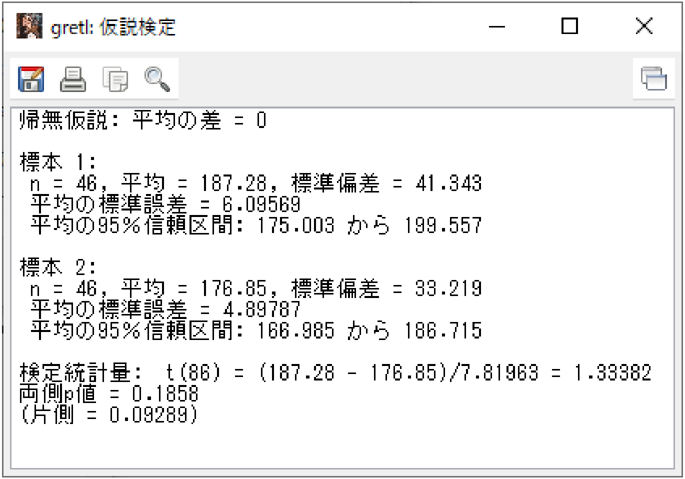
gretl での二標本検定（近似）

gretl のメニューバーから「ツール」→「検定統計量計算機」と操作し，「2 標本の平均」タブの，「共通の母標準偏差を仮定する」のチェックを外すと，両グループで母分散が異なると仮定した二標本検定が行われる．

実習 4

「男性と女性の消費支出額は同じ」という帰無仮説を，母分散が男女間で異なるという仮定の下で，Welch の検定を行ってみる.

1. gretl のメニューバーから「ツール」→「検定統計量計算機」と操作．前回の選択内容がそのまま記録されている.
2. 「共通の母標準偏差を仮定する」をクリックしてチェックを外す.
3. OK をクリック.



このような画面が表示されれば成功．まだウィンドウを閉じない！

検定結果

- t 値は男女間で母分散が等しいと仮定した場合と全く同じ値，p 値は男女間で母分散が等しいと仮定した場合とほぼ同じ値.
- 帰無仮説の採択・棄却の判断も変わらない.

$\Rightarrow$ このデータでは，男女間で母分散が等しいと仮定しても異なると仮定しても，「男性と女性の消費支出額は同じ」という帰無仮説の検定結果は本質的には変わらない.

その他の検定手法

- ２つの母集団の間で母分散が異なると仮定して二標本検定を行う他の手法として，ブートストラップ法（Bootstrap）を用いた検定がある．
  - 実際のデータからシミュレーションによって分布を生成．
    $\Rightarrow$ それを H_0 が真と仮定したときの検定統計量の分布とする．
    $\Rightarrow$ その分布に基づいて仮説検定を行う．
  - 詳細は，学部レベルを超えるので省略．

本日の作業はここまで．

gretl を終了しようとすると，「この gretl セッションを保存しますか？」というメッセージが表示されるが，今回はデータセットに変更を加えていないため，「いいえ」をクリックすればよい．上書き保存は不要．