・[課題 5]は、印刷体で（メール添付は不可）7 月 19 日の講義時間に提出すること。20 日は他の講義、21～24 日は不在のため受け取ることができない。25 日以降の提出では 26 日の講義までの添削・返却を保証できない。
・ホチキス止め不要、ただし各ページに名前（学生番号は 1 ページ目だけでよい）とページ番号を記すこと。

2011 年度　森林統計学　第 13 回課題［課題 5］平均値の差の検定

目的:二つの標本から、それらの母集団の平均値に差があるか否かを、統計的検定の方法により判断する（”第 8 章　仮説の検定”の復習）。

課題:（自主課題も含めて指示を記しているので注意;②-1 と⑤は自主課題）

① 比較対象が可能な（「2 つの平均値の差の検定」を適用するのが適切な）2 組の標本データを用意する。

- データは連番も付けて表にする。どのようなデータであるか、また比較の目的（検定により何がわかるか）を明記する。

注）各自比較してみたいデータを用いることを薦める。データ数（n）に特に基準や制限は設けないが、10～30 程度あれば検定手法に慣れる目的には足りる。ただし、統計学の理論によると、それぞれの標本の n はできるだけ等しくした方が検出力は高くなる（検定結果の信頼性が高い）ことがわかっている。

② 検定に先立ち、2 つの標本集団の分布状況を比較して考察する（[課題 1], [課題 2], [課題課題 4], “第 7 章　推定”の復習; 第 7 章までの知識を用いて「2 つの平均値の差」についてできるだけの考察を行なうことが目的)。

②-1. **[自主課題]** ヒストグラムによる比較を行なう（裏面の例を参照)。度数分布集計の境界値は二つの標本で共通とする。階級区分の数と境界値の決め方は［課題 1］の方法に順ずるが、比較のしやすさを考慮して各自の判断で適宜調整してよい。

②-2. 基礎情報として、データ数（n_1, n_2)、平均値（$\bar{x}_1, \bar{x}_2$）、標準偏差（s_1, s_2）を算出する。標準誤差（$s_{\bar{x}.1}, s_{\bar{x}.2}$）ともあわせて一覧表にしておくとよい。

②-3. 「平均値±標準誤差」の図による比較を行なう（裏面の例を参照)。

②-4. 以上の結果を用いて考察（2 つの標本集団の母集団の平均値に有意差があると考えられるかどうか）を行なう。（この後③で行なう検定結果と合っていなくてもよい)

注）実際の実験などでデータ数が少ない場合、ヒストグラムは省略せざるを得ない（ヒストグラムによる比較が適当でない）場合もあるが、「平均値±標準誤差」による比較考察は必要。

③ 帰無仮説を H_0：$\mu_1 = \mu_2$（$\mu_1 - \mu_2 = 0$)、対立仮説を H_1：$\mu_1 \neq \mu_2$（$\mu_1 - \mu_2 \neq 0$)、α（危険率・第 1 種の過誤の確率）は 0.05 として「2 つの平均値の差の検定」を行なう。

③-1. 適切な検定方法（[1] 正規分布に基づく検定・[2] t 検定・[3] Welch の t 検定）を選択する（選択の経緯を文章で説明すること)。検定方法の選択の仕方は配布資料を参照。

③-2. 選んだ検定方法により「2 つの平均値の差の検定」を行なう。

③-3. 検定結果について考察を加える（検定の結果から何が言えるか)。

④ 用いなかった他の 2 種類の検定方法による検定を行ない、③の結果と比較・考察する。

注）④での検定は、適切な検定手法ではないため、正しい結果ではない。実際の検定では、得られたデータに対して②-2, ②-3, ③-1～③-3 を行なうことで必要十分である。課題では、検定方法に慣れるため、および 2 種類の過誤に対する理解を深めるために、あえて適切でない検定方法を使って④を行なう（検定方法による検出力［1-β ：β は第 2 種の過誤の確率］を比較することになる）ということに注意。

⑤ **[自主課題]**: 別のデータを用意し①～④を行なう（反復練習のため；②-1 は省略してよい)。

注）データ数は「課題」が 30 程度なら「自主課題」は 10 程度（あるいは逆）のように変えてみるとよい。

ヒストグラムと「平均±標準誤差」の図の例:

(1) 教科書 p.183, 例 6 のデータを使った例

ヒストグラムからは 2 つの標本集団の分布の相違がみてとれる。「平均値±標準誤差」の図からは、95%信頼区間が「平均値±標準誤差×1.96」(正規分布の場合) であることから、誤差棒に重なりがあるかどうかでおよその有意性が判断できる。

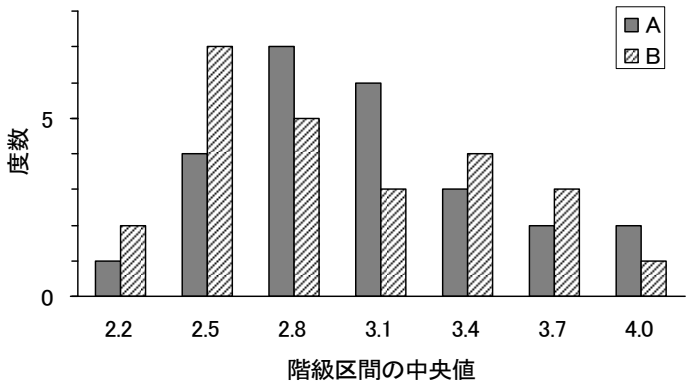

図-1. 教科書 p.183, 例 6 のデータのヒストグラム

※横軸の見出しには階級区間の中央値を示した (「階級区間」として"2.05～2.35"等のようにしてもよい)。また、この例題には単位は示されていないが、通常は階級区間の数値とともに単位 (cm, kg, など) を明記する。

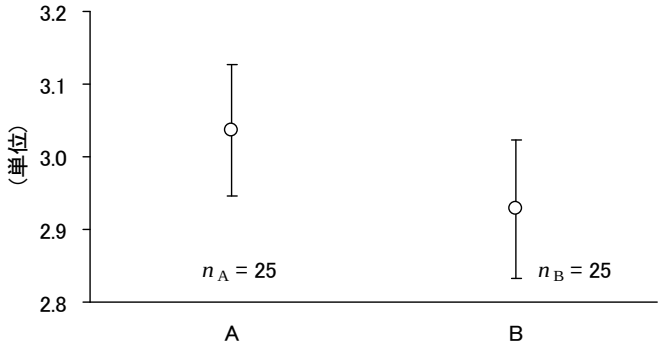

図-2. 教科書 p.183, 例 6 のデータの平均値と標準誤差

注) 縦棒は平均値±標準誤差。$n_A = n_B = 25$。

※n が共通の場合、注に記してあれば図中の書込みは省略してもよい。

(2) 教科書 p.179, 例 2 のデータを使った例

左が正しい例、右は誤って「平均値±標準偏差」としてしまった場合。標準偏差からは有意差を見極めることができない。ただし n が明記されていれば標準誤差を概算できるので、n を示しておくことは重要。

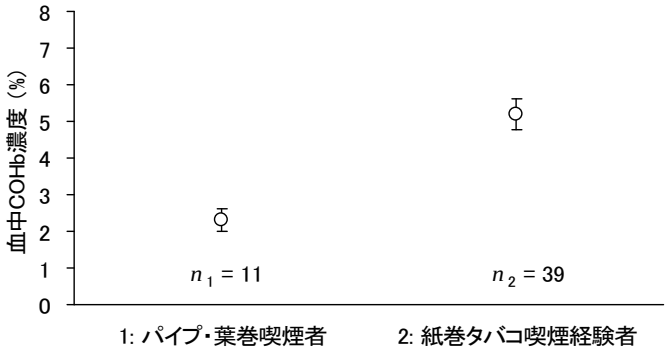

図-3a. 教科書 p.179, 例 2 のデータの平均値と標準誤差

注) 縦棒は平均値±標準誤差。$n_1 = 11$, $n_2 = 39$。

※n が異なる場合は図中に書き込む方がよい。その場合、注での n の明記は省略できる。

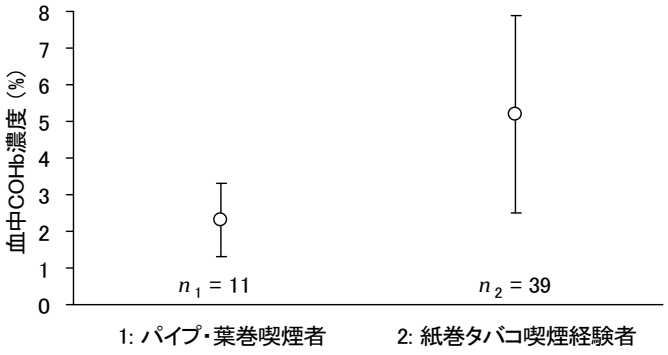

図-3b. 教科書 p.179, 例 2 のデータの平均値と標準偏差

注) 縦棒は平均値±標準偏差。$n_1 = 11$, $n_2 = 39$。

※間違った例。誤差棒を標準誤差でなく標準偏差であわらしてしまうと、有意差の判定に補助的に役立てるという目的には適さない。

注) この例では元データが示されていないため、ヒストグラムは作成できない。