

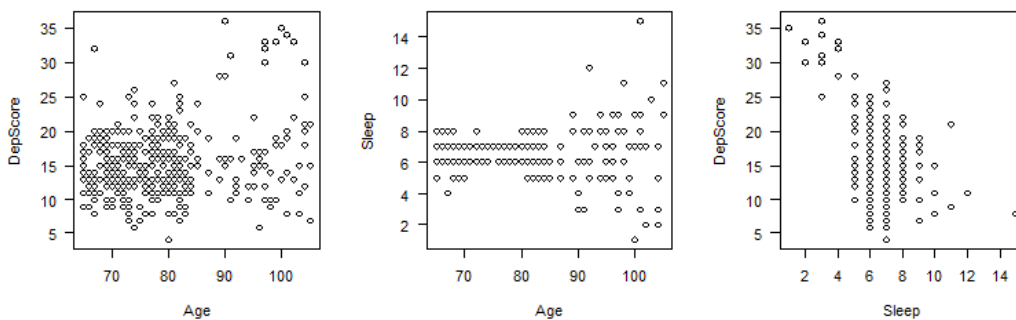
(解答例)

1. 以下の各文の下線部についての設問に答えよ。なお、すべて架空の事例であり、仮に実際に似た状況があったとしても偶然の一致である。

(1) 日本のある町(人口約5,000人、65歳以上人口2,000人)で、高齢者を対象にうつ状態と睡眠時間の関連を調べる研究を町の保健当局と共同で実施するため、(A)保健当局に依頼して住民基本台帳のデータから65-69歳、70-74歳、75-79歳、80-84歳、85歳以上の年齢層別にランダムサンプリングで80人ずつの対象者を決めて貰い、保健師経由で地元の保健指導員(一種の住民ボランティア)に質問紙の配布と回収を依頼したところ、有効回収率が100%であり、以下の結果が得られた(<http://minato.sip21c.org/ebhc/depsleep2019.txt> というタブ区切りテキスト形式でアップロード済。変数はPID(個人ID番号)、Age(年齢)、Sleep(睡眠時間)、DepScore(うつ得点)の4つ)。

年齢階級(歳)	N	Mean ± SD (うつ得点)	Mean ± SD (睡眠時間)
65~69	80	15.06 ± 3.91	6.45 ± 0.84
70~74	80	14.76 ± 4.03	6.36 ± 0.56
75~79	80	14.95 ± 3.63	6.60 ± 0.49
80~84	80	15.09 ± 4.35	6.45 ± 0.76
85~	80	17.51 ± 7.22	6.63 ± 2.40

このデータから年齢と睡眠時間とうつ得点の関係を分析するには、まず図示する。



このデータでは、年齢階級が上がると、うつ得点の平均値が上がっているように見える(Welchの方法による一元配置分散分析の結果、 $p=0.053$ で年齢階級の効果は有意でない)。睡眠時間の平均値は年齢と無関係にみえる(Welchの方法による一元配置分散分析の結果、 $p=0.073$ で年齢階級の効果は有意でない)。(B)うつ得点も睡眠時間も、ばらつきが年齢とともに大きくなっているように見える。うつ得点と睡眠時間には負の相関があるように見える。(C)ピアソンの積率相関係数を計算すると、 $-0.41$ (95%信頼区間は $-0.49 \sim -0.32$ )であり、 $p < 0.001$ ( $p = 1.88 \times 10^{-17}$ )となって帰無仮説が棄却されることから、うつ得点と睡眠時間には強い負の相関関係があるといえる。年齢と睡眠時間両方のうつ得点への影響を検討するため、(D)年齢と睡眠時間を説明変数、うつ得点を目的変数とする重回帰分析を行うと、偏回帰係数が年齢  $0.1068$  ( $p < 0.001$ )、睡眠時間  $-1.6910$  ( $p < 0.001$ )なので、うつ得点には年齢よりも睡眠時間の方が強く影響していると言える。

(A) このサンプリング方法について不適切な点があれば指摘し、改善案を示せ。適切ならば適切と記せ。

正しい。日本の市町村では、基本的に住民が住民基本台帳に登録されているので、市町村をカウンターパートとして調査をする場合はランダムサンプルしてもらうことは可能である。

(B) この場合、ばらつきが年齢階級間で有意に異なるかを示すのに適切な検定手法は何か？睡眠時間についての実行結果はどうなるか？

3群以上の間で分散に差が無いという帰無仮説の検定なので、バートレットの検定を用いる。

うつ得点については、 $Bartlett's K-squared = 56.851$ ,  $df = 4$ ,  $p-value = 1.329e-11$  であり、睡眠時間については、 $Bartlett's K-squared = 301.14$ ,  $df = 4$ ,  $p-value < 2.2e-16$  であり、ともに有意水準5%で分散に有意な差があるといえる。

(C) この検定手法、結果、解釈に不適切な点があれば指摘し、改善案を示せ。適切ならば適切と記せ。

検定方法と結果は正しいが、解釈が一部不適切である。相関係数 $-0.41$ は強い負の相関ではなく、中程度の負の相関である。

(D) この分析手法、結果、解釈に不適切な点があれば指摘し、改善案を示せ。適切ならば適切と記せ。

重回帰分析で説明変数間で目的変数への影響の相対的な大きさを調べる目的では、偏回帰係数は元の数値の大きさの影響を受けるので使えない。標準化偏回帰係数は年齢  $0.22$ 、睡眠時間 $-0.43$  であり、偏相関係数の二乗は年齢  $0.035$ 、睡眠時間  $0.180$  なので、睡眠時間の方が強く影響しているといえる。

(2) 以下の研究方法, 結果, 解釈に不適切な点があれば指摘し, 改善案を示せ。適切ならば適切と記せ。

「大規模なコホート研究の結果から玄米摂取には体重減少効果があり, 白米摂取には体重増加効果があることがわかったので, 5人の肥満者の主食を白米から玄米に変えて貰ったところ, 開始前と半年経過後の2時点間で体重(kg)が70→67, 140→130, 86→85, 95→90, 85→83と変化した, 対応のあるt検定をするとp値が0.05より大きかったので, 玄米食には統計的に有意な減量効果があるとはいえなかった。」

この方法で得られた結果では  $p=0.058$  であり有意ではないが, 差の平均値は 4.2 kg で, 95%信頼区間は  $-0.22 \sim 8.62$  なので, サンプルサイズが小さく検出力不足により統計的有意性が示せていないだけである可能性が高い。従って解釈が不適切である。差の標準偏差 3.56kg から 4.2kg の差を有意水準 5%で検出するためにサンプルサイズ 5 では, 検出力が 0.45 しかない。検出力 0.8 で検定するにはサンプルサイズは少なくとも 8 必要だった。その意味で研究方法も不適切であった。

(3) 下線部の説明に不適切な点があれば指摘し, 改善案を示せ。適切ならば適切と記せ。

マーカーの測定値がある閾値を超えると疾病 X を疑うべきとされる標準的な定量的検査方法 A に対して, より安価または迅速等の理由で, 同じマーカーの新しい検査方法 B を開発したとする。B が妥当であることを示すには, 十分な数の患者と健常者に対して A と B での測定を行い, 対応のある t 検定を行って有意差がなければ良い。

対応のある t 検定で有意な差が無いだけでは, A と B の差の平均値が 0 ということしか示せていないので, 例えば絶対値が小さいときは  $A > B$ , 大きいときは  $A < B$  のような系統的なズレがあるか, ばらつきが大きい可能性があり, B の妥当性を示すには不十分である。これに加えて相関が高く傾きが 1 と差が無いことを示す必要があり, そのためには Bland-Altman プロットを行う。

(4) 下線部の研究方法, 結果, 解釈に不適切な点があれば指摘し, 改善案を示せ。適切ならば適切と記せ。

「慢性肝炎患者をランダムに 22 人ずつの 2 群に分け, 片方はプレドニソロンを投与し, 残りは経過観察したところ, 観察終了時にプレドニソロン群では 11 人が生存しており, 経過観察群では 6 人が生存していたが, フィッシャーの直接確率検定の結果, p 値は 0.215 で観察期間終了後の生存確率に有意差はなかった。死亡または観察打ち切りまでの月数のデータ (<http://minato.sip21c.org/ebhc/hepatitis2019.txt> にアップロードしてあり, 変数は time (死亡または観察打ち切りまでの月数), flag (観察終了時死亡が 1, 生存が 0), group (1 がプレドニソロン群, 2 が経過観察群) の 3 つである)があるので, カプラン=マイヤ法による生存時間の中央値推定をした結果, 生存時間の中央値はプレドニソロン群が 146.0 ヶ月 (95%信頼区間は 96 ヶ月から無限大), 経過観察群が 40.5 ヶ月 (95%信頼区間は 29 ヶ月から無限大) で, 95%信頼区間が重なっているのでプレドニソロンには統計的に有意な延命効果があるとはいえなかった。」

統計的有意性を示すにはログランク検定か一般化ウィルコクソン検定を用いる。ログランク検定の結果, カイ二乗値は 4.7, 自由度 1,  $p=0.03$  であり, 有意水準 5%で統計的に有意な差があり, プレドニソロンには有意な延命効果があった。

2. 集団において疾病の有病割合(prevalence proportion)とは何か, またこの値を得るためにどのようなデザインの研究が必要かを簡潔に説明せよ。有病割合とは, 調査時点における集団全員のうち当該疾病である人の割合を意味する。横断研究(断面的研究)により得る。

3. ブドウ糖飴を食べると短期記憶が向上するかどうかを調べるため, 10 人の被験者に対して, 飴を食べる前後でフラッシュ記憶をしてもらって得点を比較するという実験研究を行った。結果が下表の通りだったとき, 飴を食べて短期記憶は向上したと言えるか? 有意水準 5%で検定せよ (EZR や関数電卓を使っても良い)。近似値として自由度 9 の t 分布の 97.5%点 2.262,  $\sqrt{2}=1.414$ ,  $\sqrt{3}=1.732$ ,  $\sqrt{5}=2.236$  を使って良い。

飴を食べる前の得点	8	6	3	7	6	7	4	8	5	6
飴を食べた後の得点	9	8	2	8	9	8	3	9	5	7

$t = 2.058$ ,  $df = 9$ ,  $p\text{-value} = 0.06971$  であり, 得点変化の平均値は 0.8 (95%信頼区間は  $-0.079 \sim 1.679$ ) なので, 有意水準 5%で統計的に有意な差はなく, 飴を食べて短期記憶が向上したとはいえない。

4. 軽度なヨガをしながら瞑想することがストレス軽減効果をもつかどうかを調べる目的で事前-事後デザインの実験研究を実施した。アウトカムとしては唾液中クロモグラニン A 濃度の変化をみた。唾液中クロモグラニン A 濃度は安静時で平均 3.5 (pmol/mg protein), 繰り返し測定値の標準偏差は 0.5 (pmol/mg protein) という先行研究がある。クロモグラニン A 濃度が 1 (pmol/mg protein) 低下したら臨床的にストレス軽減効果があったと判定するとしたとき, 有意水準 5%, 検出力 80%で対応のある片側 t 検定をして効果の有無を判定する予定である場合, 必要サンプルサイズは何人か。

`power.t.test(delta=1, sd=0.5, sig.level=0.05, power=0.8, type="paired", alternative="one.sided")`の結果,  $n=3.3$  なので 4 人必要。