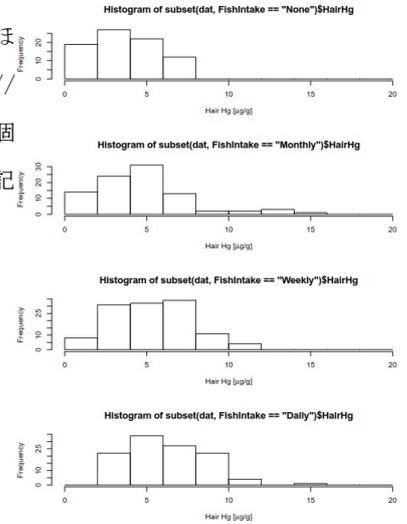


(解答例)

1. 以下の各文の下線部について解釈や解析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。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) 途上国の人口約 4,000 人の R 島には Google Earth で見ると同じような規模の村が 40 ほど散在している。近年、魚介類や鯨類を多食する住民で神経症状を示す者が現れ、水銀中毒の可能性が疑われたため、実態調査として、(A)40 の村からランダムに 4 つの村を選び、それぞれの村については全数を調査するという方法で選ばれた 400 人の住民を対象に、魚介類及び鯨類の摂取頻度について、[1]ほとんど食べない、[2]月 1 回以上週 1 回未満、[3]週 1 回以上毎日未満、[4]ほぼ毎日、から選んでもらうと同時に、毛髪水銀濃度を調べたところ、下表の結果が得られた(<http://minato.sip21c.org/fish-Hg-2018.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), HairHg (毛髪水銀濃度), HighHairHg (5µg/g 以上なら 1, 未満なら 0), FishIL (上記の魚介類及び鯨類摂取頻度カテゴリ番号) の 4 つ)。



Eating fish or whale frequency	N	Median (Hg 濃度)	Mean ± SD (Hg 濃度)	毛髪 Hg 濃度 5 µg/g 以上
1. None	80	3.44	3.55 ± 2.02	27
2. Monthly	90	4.34	4.72 ± 2.89	36
3. Weekly	120	5.45	5.36 ± 2.30	69
4. Daily	110	5.97	6.20 ± 2.37	72

このデータから魚介類及び鯨類の摂取頻度と水銀曝露の関係を分析するには、2 つのアプローチがある。1 つは、成人耐容摂取量の目安である 5 µg/g 以上の毛髪中水銀濃度を示すかどうか魚介類及び鯨類の摂取頻度と独立かを検定する方法である。(B)フィッシャーの直接確率検定をすると、 $p=1.34 \times 10^{-5}$ となって帰無仮説が棄却され、魚介類及び鯨類の摂取頻度と毛髪中水銀高値に統計的に有意な関連があるといえる。もう 1 つのアプローチは、魚介類及び鯨類の摂取頻度の違いが毛髪水銀濃度に影響しているかどうかを調べる方法である。Welch の方法による一元配置分散分析の結果は、F 値 24.273, 第一自由度 3, 第二自由度 210.39, $p < 0.001$ で帰無仮説は棄却される。(C)さらに魚介類及び鯨類の摂取頻度が異なる群間で毛髪水銀濃度に差があるのか総当たりで調べるには、Welch の t 検定を繰り返せば良い。

- (A) 適切。こういう場合はクラスターサンプリングが良い。
 - (B) 適切。ただしサンプルサイズもそれなりに大きく摂取頻度の異なる群間で毛髪水銀高値の人の割合に差がないことを帰無仮説にした場合に期待度数が小さくなるセルもないので、多群間の割合の差のカイ二乗検定でも良い。
 - (C) Welch の t 検定を繰り返すと検定の多重性の問題が生じるので不適切。Holm か FDR で多重性を調整するか Tukey の HSD を用いるべき。Holm でも FDR でも Monthly と Weekly の間を除きすべての組合せについて有意水準 5% で有意な差がある。
- (2) 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→65, 140→135, 95→85, 95→90, 85→80 と変化した。対応のある t 検定をすると $p=0.0008$ となるので、介入研究においても玄米食には統計的に有意な減量効果があると結論できた。

- (レベル 1 の解答) 検定方法と結論に矛盾はないが、 $p=0.003883$ が正しい。
 - (レベル 2 の解答) 主食を白米から玄米に変えるという条件設定のみで、米の量やおかず等について何も制御していないので、この介入の仕方では、玄米食の効果ではなく、それがこの対象者に引き起こした別の変化が体重減少につながった可能性を否定できない
- (3) 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、相関係数 0.8 以上の有意な正の相関があれば良い。

絶対値が異なっていたり系統的にずれている(例えば低濃度域では A>B, 高濃度域では A<B となりやすい)可能性があるので、Bland-Altman プロットが必要。

(4) 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/hepatitis-2018.txt> にアップロードしており、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、観察終了時生存か死亡かを無視して 2 群間で観察された生存月数の平均値を Welch の t 検定で比較したところ、プレドニソロン群が 109.5 ヶ月、経過観察群が 64.7 ヶ月で、p 値が 0.013 だったため、有意水準 5% で考えると、プレドニソロンには統計的に有意な延命効果があると結論できた。

(レベル 1 の解答) 適切(観察終了時生存か死亡かを無視して平均値の差の検定を行えば下線部の結論は正しい)。

(レベル 2 の解答) 生存者についてはその後どれくらい生きるかが不明なので、この分析は不適切であり、打ち切りを考慮して生存時間をログランク検定すると、 $p=0.0309$ となる。結論としては、有意水準 5% でプレドニソロンには有意な延命効果がある、は変わらない。

2. 集団において疾病の発生率(incidence rate)とは何か、またこの値を得るためにどういふデザインの研究が必要かを簡潔に説明せよ。

コホート研究を行う必要がある。疾病が発生した人数(incidence)を観察人時で割ることにより、単位時間当たりの疾病の発生率(速度)が得られる。

3. 1900 年頃に行われた、Sir Wright が開発したチフスワクチンの有効性を検証するための 6 つの研究結果について、Karl Pearson は単純に 6 つの研究で得られた四分相関係数(ワクチン接種の有無とチフス感染後治癒するか死亡するか)のクロス集計データの平均値をとって 0.193 という値を得たことから、ワクチンとして推奨するには効果が

が小さすぎると判定した。しかしアウトカムを死亡として

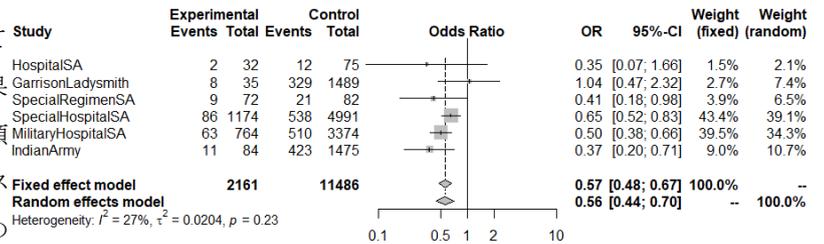
オッズ比を用いたメタアナリシスを行うと、6 つの研究結果

を統合したオッズ比は固定効果モデルで 0.57 (95% 信頼

区間が [0.48, 0.67]) となり、ワクチン接種は有意にチフス

感染による死亡を防ぐことができると示される。Pearson の

判定が不適切なものになった原因について考察せよ。



(難問) 6 つの研究のサンプルサイズの違いを無視して相関係数の平均値をとってしまったため、相対的に予防接種を打った群のサンプルサイズが小さい研究におけるランダムなバラツキの影響の大きさに起因すると思われる低い相関の研究 (GarrisonLadysmith) が結果を歪めた。

4. 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、飴を食べると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)。

手計算する際の近似値としては自由度 9 の t 分布の 97.5% 点 が 2.262 であることと $\sqrt{2}=1.414$, $\sqrt{3}=1.732$, $\sqrt{5}=2.236$ を使って良い。

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

差 1 1 2 1 1 1 1 1 1 1 1 差の平均 1.1, 差の標準誤差 $\sqrt{0.1/\sqrt{10}}=0.1$

したがって、 $t \text{ 値} = 1.1/0.1 = 11 > 2.262$ より有意水準 5% で統計的に有意に向上したと言える。

$t.test(c(9,7,5,8,7,8,6,9,8,7), c(8,6,3,7,6,7,5,8,7,6), paired=TRUE)$ でも同じ結果が得られる。ちなみに $p=1.61 \times 10^{-6}$

5. 新しく考案したストレッチが従来のストレッチよりも身体の柔軟性を上げるかどうか調べる目的で RCT を実施した。アウトカムとしては立位体前屈の変化をみた。従来のストレッチの後は、ストレッチ前に比べて平均 3cm (標準偏差 2cm) 深く前屈できるというデータがある。新しいストレッチによる前屈距離の伸びが 1cm 以上大きければ臨床的に意味があるとすれば、有意水準 5%、検出力 80% で両側 t 検定をして効果を比較する予定である場合、2 群同数として、必要サンプルサイズは何人ずつか。

$power.t.test(delta=1, sd=2, sig.level=0.05, power=0.8)$ の結果で $n=63.76...$ と出るので、各群 64 人ずつ。(EZR では 63 人ずつと出るので、それでも良い。PS では 64 人と出る) (レベル 2 の解答) なお、この場合は短期的な効果を見ているので、十分な washout 期間をおき同じ対象者で 2 種類のストレッチを比べて良い。PS で Paired を指定して計算すると 33 人 (EZR で対応のある 2 群間の平均値の差で計算すると 34 人) で済む。