月３「確率・統計」(第 10 回) 練習問題

- 計算結果は小数点以下１桁に丸めてよい。また、数値は既約分数で表してもよい。(小数点 2 桁の場合も示しておきます．)
- u =1.96 に対する標準正規分布の上側確率を $Q(1.96) = 0.025$ とする。
- 確率変数 X の期待値を$E[X]$ 、分散を$V[X]$ で表す。

以下の [　　　　]を埋めよ。【　　　　】の欄は、いずれかを○で囲むこと。

問１：

(1) 調査観察研究のうち，ある心疾患の発症者と非発症者のそれぞれの群について、過去にさかのぼって喫煙の有無を調査するような研究を [① ケース・コントロール]研究と呼び、一方で、喫煙者と非喫煙者を追跡して将来の心疾患の発生割合を比較するような研究を [② コホート]研究と呼ぶ。また，実験参加者をランダムに複数の群に割り付けて行う実験研究を[③ ランダム化比較]試験と呼ぶ。この中でエビデンスレベルが最も高いのは[④ ランダム化比較試験]である。

(2) 母集団からの標本抽出を何段階かに分けて行う抽出方法を[⑤ 多段]抽出、母集団の名簿から一定間隔で抽出する方法を[⑥ 系統]抽出、母集団の状況や特徴（たとえば人口比率や産業的特色など）に合わせて層を決め、それぞれの層から抽出を行う方法を[⑦ 層化（層別）]抽出、あらかじめ母集団をグループ分けしておき、グループを無作為抽出した上でグループ内の全員を調査する抽出方法を[⑧ 集落（クラスタ）]抽出と呼ぶ。

(3) 相関関係は因果関係を【⑨含意する or 含意しない 】。一見するとある二つの変数に因果関係があるように見えるが、第三の要因がそれぞれの変数に影響を与えるようなバイアスを[⑩交絡(バイアス)]と呼ぶ。

(4) 標本が母集団を代表していないことで生じるバイアスを[⑪ 選択バイアス]と呼ぶ。

(5) フィッシャーの三原則（反復、無作為化、局所管理）のうち、実験を行う場所や時間をブロックに区切ったうえで，各ブロックで全条件（全水準）の実験を行うことを[⑫ 局所管理]という。場所や時間で環境が異なってくるが、ブロックに区切ることで、各ブロック内では一定の環境で各条件（水準）の実験を比較することができる。また、これら三原則を満たす割り付け方法を[⑬ 乱塊法]という。

(6) 統計的仮説検定における第一種の過誤とは、帰無仮説が【⑭ 正しい or 正しくない 】にもかかわらず【⑮ 棄却してしまう or 棄却しない 】 誤りである。第二種の過誤とは、帰無仮説が 【⑯ 正しい or 正しくない 】 にもかかわらず 【⑰ 棄却してしまう or 棄却しない 】 誤りである。母集団に有意差がある場合、標本サイズを大きくすることによって第【⑱ 一 or 二 】種の過誤は減少する（ここは第 12 回で説明予定)。すなわち[⑲ 検定]力が大きくなる。第二種の過誤が起こる確率をβとすると，⑲は[⑳ $1-\beta$]で表される．

問 2: 以下では、表が出る確率が p 、裏が出る確率が $1-p$ のコインを考える。

(1) コイン投げのように、毎回 2 種類のいずれかの結果をとり、かつそれらの起こる確率がどの回も同じであるような独立試行を[① ベルヌーイ]試行と呼ぶ。表が出たときに 1、裏が出たときに 0 をとる確率変数 X を考えると、X は [② ベルヌーイ]分布に従う。期待値と分散の計算式から $E[X]=$ [③ p]、$V[X]=$ [④ $p(1-p)$]である。

(2) それぞれのコイン投げが①の試行として考えられるとき、 n 回コインを投げた時に表が出る回数は[⑤ 二項]分布に従う。i 回目のコイン投げの結果を、(1)の確率変数 X に添え字を加えた X_i で表すものとする。すると、表の出る回数は $Y=X_1+X_2+\cdots+X_n$ で表すことができる。①の試行が独立試行であることに注意すれば、Y の平均（期待値）$E[Y]=$ [⑥ np]、分散 $V[Y]=$[⑦ $np(1-p)$]であることが分かる。

(3) [⑧ 中心極限]定理を適用すると、n 回コインを投げた時に表が出る割合（比率）$R=Y/n$ は、n が大きければ近似的に[⑨ 正規]分布に従う。n, p を用いれば、割合 R の平均（期待値）$E[R]=$ [⑩ p]、分散 $V[R]=$[⑪ $\frac{p(1-p)}{n}$]と表すことができる。

(4) 母比率の差（たとえばリスク差）の検定では、コイン投げと同様の試行をそれぞれ n_A, n_B 回行うことで 2 群の標本が得られたと考える。2 つの母集団で母比率が共に p であれば、2 群の標本比率の差は、平均 [⑫ 0]、分散 [⑬ $p(1-p)\left(\frac{1}{n_A}+\frac{1}{n_B}\right)$]の正規分布に従うため、これを用いて母比率の差を検定できる。

(5) ある薬が疾病に効果があるかを検証するため、実験参加者 225 人に対する試験を行った。
その結果は以下の二元分割表の通りである。

	回復した	回復しなかった	計
新薬投与	83	52	135 (n_A)
プラセボ投与	42	48	90 (n_B)
計	125	100	225

薬投与群（処置群）とプラセボ投与群（対照群）の母比率（回復率）の両側検定を行う。帰無仮説でこれら 2 群の母比率がともに等しく p であるとする。p は未知であるが、 [⑭ 大数]の法則により $p=$ [⑮ 125/225 (≒ 0.556 or 0.56 or 0.6)]で近似できると考える。一方、標本比率の差は[⑯ $\frac{83}{135}-\frac{42}{90}$ (≒ 0.148 or 0.15 or 0.1)]であるため、検定統計量 $Z=$ [⑰ 2.19 or 2.2]であり、帰無仮説は有意水準 5%で棄却【⑱ される or されない 】。(Z = ⑯/sqrt(⑬) ただし p に⑮を代入)
(Z = 2.19 ＞1.96 より棄却)

(6) 対立仮説「新薬投与群とプラセボ投与群の回復率に差がある」を確かめる場合には、【⑲ 両側 or 片側 】検定を用いることが多い。両側検定は片側検定よりも有意差が【⑳ 認められやすく or 認められにくく 】なる。