学籍番号:　　　　　　　　　　　　　　　　　　　　　　　　　　　　公衆衛生学

疫学演習 (2019) 回答用紙

氏名:　　　　　　　　　　　　　　　　　　締め切り:2019年6月28日(金)

1　問題1:両群間計量データの平均値を比較する (20%)

1. 帰無仮説を「遺伝子変異ありと変異なし両群の間で,COGの平均値は等しい」とする.上記のデータ及び適切な方法を使って検定し,検定の結果を分かりやすく説明せよ.なお,分散が等しいと仮定できる場合,以下の式で両群の共通標準偏差が計算できる:(6%)

$$S = \sqrt{\frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2}} \tag{1}$$

- S_A : A群の標準偏差;
- n_A : A群の人数;
- S_B : B群の標準偏差;
- n_B : B群の人数;
- S : A群及びB群の共通標準偏差;
- $n_A + n_B - 2$: 分散が等しい時の自由度.

2. 遺伝子変異ありとなしの群の間の脳萎縮度 (atrophy) を比較する場合，1. と同じ検定方法を用いてよいか？ それを判断するにはどの検定方法を使えばよいかを説明し,実際にこの検定方法を実施せよ. (6%)

3. 2.の結果を踏まえて，帰無仮説「両群の脳萎縮度の平均値が等しい」を検定せよ．なお，両群の分散が等しいという前提が満たされていない時に，自由度(df)の計算式は以下となる：(8%)

$$\mathbf{df} = \frac{(S_A^2/n_A + S_B^2/n_B)^2}{(S_A^2/n_A)^2/(n_A - 1) + (S_B^2/n_B)^2/(n_B - 1)} \tag{2}$$

2　問題2:線形回帰モデル (30%)

2.3 年齢,体重それぞれの平均値,分散を求めよ;また,年齢と体重の共分散を算出せよ.なお,EZRで計量データの平均値を計算するには,コマンド `mean(変数名)` を使う;共分散を計算したい時には,コマンド `cov(変数1, 変数2)` を利用する.

以下のコードをRスクリプトに入力して,実行をクリックしてください.(結果を下の余白に記入すること) (5%)

```
# 年齢の平均値
mean(Dataset$age)
# 年齢の分散
var(Dataset$age)
# 体重の平均値
mean(Dataset$wt)
# 体重の分散
var(Dataset$wt)
# 体重と年齢の共分散 covariance
cov(Dataset$wt, Dataset$age)
```

2.4 年齢を説明変数,体重を目的変数とする場合,年齢の傾き(回帰係数),と切片を求めよ.なお,分散と共分散の定義は以下とする,$\bar{X}$ は X の平均値を示す:

- 分散 variance:

$$\begin{aligned}\mathbf{Var}(X) &= \frac{(X_1-\bar{X})^2+(X_2-\bar{X})^2+\cdots+(X_n-\bar{X})^2}{n-1}\\ &= \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}\end{aligned}$$

- 共分散 covariance:

$$\begin{aligned}\mathbf{Cov}(X,Y) &= \frac{(X_1-\bar{X})(Y_1-\bar{Y})+(X_2-\bar{X})(Y_2-\bar{Y})+\cdots+(X_n-\bar{X})(Y_n-\bar{Y})}{n-1}\\ &= \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{n-1}\end{aligned}$$

以下のコードをRスクリプトに入力して,実行をクリックしてください.(結果を下の余白に記入すること) (2%)

```
# 傾き (slope)
beta <- cov(Dataset$wt, Dataset$age) / var(Dataset$age)
beta
# 切片 (intercept)
alpha <- mean(Dataset$wt) - mean(Dataset$age)*beta
alpha
```

2.6 今まで計算した傾きと切片の数字を用いて,年齢と体重の関係を線形と考える場合の計算式を記入せよ.傾きと切片の計算結果の意味をそれぞれ記述せよ.(4%)

2.8 重回帰線形モデルの計算結果を用いて,体重の平均値を年齢と性別の線形モデルで表示せよ.各回帰係数の意味を説明せよ.(14%)

2.9 上記の重回帰線形モデルを用いて,年齢が34ヶ月の女の子の体重の予測値を計算せよ.(5%)

3　問題3：χ^2 検定，オッズ比，ロジスティック回帰モデル (40%)

3.1 もし，視覚障害と対象者の死亡リスクに関連がない場合，下の表（各セルの期待値の人数）を答えよ：(4%)

死亡	視力正常	視覚障害	合計
0			4161 (96.81%)
1			137 (3.19%)
合計	3971 (100%)	327 (100%)	4298 (100%)

3.1.2 上記の2つの表の数字を使って χ^2 統計量を計算せよ (4%)

3.1.4 2 × 2 の分割表では，自由度は ＿＿＿＿＿＿＿＿ (2%)

3.1.5 視覚障害と死亡の関係を示すテーブルのデータをもとに，下表を完成せよ：(6%)

	視力正常	視覚障害	合計
リスク (risk)			0.0319
オッズ (odds)			0.0329
対数オッズ (log-odds)			-3.414

視覚障害と死亡の関連を示すオッズ比を算出せよ：(2%)

$$\mathbf{OR} =$$

このオッズ比の対数を取った値 $\mathbf{log(OR)}$ は：(2%)

$$\mathbf{log(OR)} =$$

3.2 年齢の影響を考慮する

	視覚障害 (0 = no, 1 = yes)									
死亡	0	1	0	1	0	1	0	1	0	1
1 = yes	29	2	38	10	15	11	15	17	97	40
0 = no	2301	22	1271	124	212	69	90	72	3874	287
n										
年齢	15-34		35-54		55-64		65 +		Total	

上記のデータをよく見ると，視覚障害のオッズは年齢と共に上昇している (年齢が15-34歳群の$(2 + 22)/(29 + 2301) = 0.010$から年齢が65歳以上群の$(17 + 72)/(15 + 90) = 0.848$に上がっている)．しかし，年齢の上昇と共に，死亡のオッズも上がる．年齢はここで，交絡因子 (confounder) と定義される．

3.2.1 以上のデータと解説をよく理解した上で，下表を完成せよ：(8%)

	オッズ		
年齢	視力正常	視覚障害	オッズ比
15-34	29/2301 = 0.01260		
35-54	0.02990		
55-64	0.07075		
65+	0.16667		

各年齢層では視覚障害と死亡との関連はどう変化しているか？(2%)

3.2.2.3 単変量ロジスティック回帰モデルで評価した粗オッズ比 (crude odds ratio) と比べ,年齢調整オッズ比はどう変わったかを説明せよ.(10%)

4　問題4:生存分析 (10%)

- 単変量ハザード比,及び信頼区間の意味を説明せよ.(5%)

- 年齢調整ハザード比,及び信頼区間の意味を説明せよ.(5%)