学籍番号: 公衆衛生学

疫学演習 (2019) 回答用紙と採点基準

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.

公式(1)により，共通分散/標準偏差を推定する：

$$\begin{aligned}
\because\quad S &= \sqrt{\frac{(n_A-1)S_A^2+(n_B-1)S_B^2}{n_A+n_B-2}} \\
\therefore\quad S &= \sqrt{\frac{(50-1)9^2+(150-1)9^2}{50+150-2}} = 9 \\
\Rightarrow\quad T &= \frac{\bar{X}_A-\bar{X}_B}{S\sqrt{1/n_A+1/n_B}} \\
&= \frac{69.2-60.2}{9\times\sqrt{1/50+1/150}} \\
&= \frac{9}{9\times 0.1633} = 6.1237
\end{aligned}$$

- 計算の過程と結果得点＝2

- 共通分散とT統計量両方の計算過程がないと無得点
- 片方だけ計算過程がある場合，得点＝1
- 過程があるけど，結果だけ不正解1点

```
2*pt(6.1237, 198, lower=FALSE)
```

```
## [1] 4.831141e-09
```

- P値の計算得点＝2
- P < 0.0001でも可

この検定結果は「両群のCOG平均値が等しい」という帰無仮説を棄却するために非常に強い証拠を提供したと言える.

- 結果の説明得点＝2
- 「両群のCOG平均値が(統計学的に)有意に異なる」でも可
- 「遺伝子変異あり群のCOG平均値は(統計学的に)有意に高い」でも可
- 「遺伝子変異なし群のCOG平均値は(統計学的に)有意に低い」でも可

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%)

テーブルから両群の標準偏差はそれぞれ 0.21, 0.10 だと推定され，分散 (variance) が等しいという前提が満たされていない． 1． の検定方法を使う時には，両群の分散が等しいという前提条件が必須だから，同じ Student t 検定を行うことができない（理由 1 点）．「両群の分散」が等しいという帰無仮説を検定するには F 検定を利用する：

$$
\begin{aligned}
F &= \frac{S_A^2}{S_B^2} \\
&= \frac{0.21^2}{0.10^2} \\
&= 4.41
\end{aligned}
$$

- F検定方法の計算過程を示すと結果 2 点
- 過程がない，結果だけ正しい無得点
- 過程があるけど，結果だけ不正解 1 点

自由度 (degree of freedom)はそれぞれ $n_A - 1 = 49; n_B - 1 = 149$，P値の計算は EZR を利用する：

```
2*pf(4.41, 49, 149, lower.tail = FALSE)
```

```
## [1] 2.971758e-12
```

- P 値の計算得点＝ 1
- P < 0.0001でも可

この検定結果は「両群の脳萎縮度の分散が等しい」という帰無仮説を棄却するために非常に強い証拠を提供したと言える．

- 結果の説明得点＝ 2
- 「両群の脳萎縮度の分散/標準偏差が(統計学的に)有意に異なる」でも可
- 「遺伝子変異あり群の脳萎縮度の分散/標準偏差が(統計学的に)有意に高い」でも可
- 「遺伝子変異なし群の脳萎縮度の分散/標準偏差が(統計学的に)有意に低い」でも可

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.の検定結果から，「両群の脳萎縮度の分散が等しい」という帰無仮説を棄却されたため，Welch の t 検定を採用する．

$$\begin{aligned}\Rightarrow\quad T &= \frac{\bar{X}_A - \bar{X}_B}{\sqrt{S_A^2/n_A + S_B^2/n_B}} \\ &= \frac{0.67-0.23}{\sqrt{0.21^2/50 + 0.10^2/150}} \\ &= \frac{0.44}{\sqrt{0.0009486667}} = 14.28551\end{aligned}$$

- T統計量の計算過程と結果２点
- 計算過程がない場合，結果あっても無得点
- 計算過程がある場合，結果間違っても１点

自由度は公式(2)により計算できる：

$$\begin{aligned}\mathbf{df} &= \frac{(S_A^2/n_A + S_B^2/n_B)^2}{(S_A^2/n_A)^2/(n_A-1) + (S_B^2/n_B)^2/(n_B-1)} \\ &= \frac{(0.21^2/50 + 0.10^2/150)^2}{(0.21^2/50)^2/(50-1) + (0.10^2/150)^2/(150-1)} \\ &= 58.58105\end{aligned}$$

- 自由度の計算過程と結果２点
- 計算過程がない場合，結果あっても無得点
- 計算過程がある場合，結果間違っても１点

P値の計算は EZR を利用する：

```
2*pt(14.28551, 58.58105, lower=FALSE)
```

```
## [1] 9.601543e-21
```

- P 値の計算得点＝2
- P < 0.0001でも可

この検定結果は「両群の脳萎縮度の平均値が等しい」という帰無仮説を棄却するために非常に強い証拠を提供したと言える.

- 結果の説明得点＝2
- 「両群の脳萎縮度の平均値が(統計学的に)有意に異なる」でも可
- 「遺伝子変異あり群の脳萎縮度の平均値は(統計学的に)有意に高い」でも可
- 「遺伝子変異なし群の脳萎縮度の平均値は(統計学的に)有意に低い」でも可

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%)

- それぞれ１点，四捨五入可

```
# 年齢の平均値
mean(Dataset$age)
```

[1] 16.97895

```
# 年齢の分散
var(Dataset$age)
```

[1] 69.5022

```
# 体重の平均値
mean(Dataset$wt)
```

[1] 9.644737

```
# 体重の分散
var(Dataset$wt)
```

[1] 3.513068

```
# 体重と年齢の共分散 covariance
cov(Dataset$wt, Dataset$age)
```

[1] 11.49089

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%)

- それぞれ 1 点，四捨五入可

```
# 傾き (slope)
beta <- cov(Dataset$wt, Dataset$age) / var(Dataset$age)
beta
```

```
## [1] 0.1653314
```

```
# 切片 (intercept)
alpha <- mean(Dataset$wt) - mean(Dataset$age)*beta
alpha
```

```
## [1] 6.837584
```

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%)

- 計算式1点

$$Y = 6.838 + 0.165X$$

- Y : 体重 (kg)； 0.5点，計算式の中に書いても可
- X : 年齢 (months)；0.5点，計算式の中に書いても可
- 0.165 ：子供の年齢が1ヶ月上がると，体重が平均的 0.165 kg (165 g) 高くなる；1点，単位がない場合0.5点
- 6.838 : 子供の年齢が0ヶ月の時に，体重の平均値は 6.838 kg． 1点，単位がない場合0.5点

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%)

- 計算式2点

$$Y = 7.152 + 0.164X_1 - 0.519X_2$$

- Y : 体重(kg)；1点，計算式の中に書いても可
- X_1 : 年齢 (months)；1点，計算式の中に書いても可
- $X_2 = 1$: 女性；2点
- $X_2 = 0$: 男性；2点
- 7.152 : 男の子が年齢 0 ヶ月の時の平均体重(kg)；2点，単位がない場合1点
- 0.164 : 同じ性別の子供の年齢が1ヶ月上がることによって，体重が平均的に 0.164 kg増える；2点，単位がない場合1点
- -0.519 : 子供年齢が同じ時に，女の子は男の子と比べ，体重が平均的に 0.519 kg 低い．2点，単位がない場合1点

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%)

- $X_1 = 34$；
- $X_2 = 1$;

$$\begin{aligned} Y &= 7.152 + 0.164 \times 34 - 0.519 \times 1 \\ &= 12.209\ (\mathbf{kg}) \end{aligned}$$

- 計算過程と結果5点，単位がない場合3点
- 計算過程がない場合，結果あっても1点
- 計算過程がある場合，結果間違っても3点

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%)

- それぞれ１点，四捨五入可

死亡	視力正常	視覚障害	合計
0	3971×4161/4298 = 3844.4232	327×4161/4298 = 316.5768	4161 (96.81%)
1	3971×137/4298 = 126.5768	327×173/4298 = 10.4232	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%)

$$\chi^2 = \frac{(3874 - 3844.4232)^2}{3844.4232} + \frac{(287 - 316.5768)^2}{316.5768} + \frac{(97 - 126.5768)^2}{126.5768} + \frac{(40 - 10.4232)^2}{10.4232}$$
$$= 93.829$$

- 計算過程と結果４点
- 計算過程がない場合，結果あっても1点
- 計算過程がある場合，結果間違っても3点

3.1.4 2×2 の分割表では，自由度は 1 (2%　= 2点)

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%)

- それぞれ 1 点，四捨五入可

	視力正常	視覚障害	合計
リスク (risk)	0.0244	0.1223	0.0319
オッズ (odds)	0.0250	0.1394	0.0329
対数オッズ (log-odds)	-3.689	-1.9704	-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% = 2点，四捨五入可)

$$\mathbf{OR} = 0.1394 \div 0.025 = 5.576$$

このオッズ比の対数を取った値 $\mathbf{log}(\mathbf{OR})$ は：(2% = 2点，四捨五入可)

$$\mathbf{log}(\mathbf{OR}) = 1.717$$

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%)

- それぞれ 1 点，四捨五入可

	視力正常	視覚障害	合計
リスク (risk)	0.0244	0.1223	0.0319
オッズ (odds)	0.0250	0.1394	0.0329
対数オッズ (log-odds)	-3.689	-1.9704	-3.414

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%)

各年齢層では，視覚障害と死亡の関係を評価するオッズ比は大きく異なることがわかり，年齢と共に減っていく傾向があるかもしれない(1点)．年齢を考慮せずに視覚障害と死亡との関係を評価することは妥当ではないことが示唆される．(1点)

- 「年齢を考慮せずに視覚障害と死亡との関係を評価することは実際の関連を過大評価する」でも可

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%)

- 年齢を考慮していない場合，視覚障害者は視力正常者と比べ，3年の間に死亡するオッズが5.57倍であり(1点)，95%信頼区間が 3.78 - 8.20 と推定される(1点).
- 多変量ロジスティクス回帰モデルを用いて，視覚障害と死亡の関係に年齢の交絡考慮した後，オッズ比が 2.20 になり(1点)，95%信頼区間が 1.41 - 3.44 と推定される(1点).
- このオッズ比が大きく変化した（小さくなった）ことは，年齢がこの関連の強い交絡因子であることを示唆している．(2点)
- また，年齢調整したオッズ比の95%信頼区間は 1 を跨いでいない．(2点)

以上の結果を踏まえて，「視覚障害者が視力正常者と比べ，観察期間中に死亡するオッズ比は1と等しい」を棄却するために，非常に強い証拠を提供した．(2点)

- 「視覚障害者が視力正常者と比べ，観察期間中に死亡するオッズ比は(統計学的に)有意に1より高い」でも可

4　問題４：生存分析 (10%)

4.1　単変量ハザード比，及び信頼区間の意味を説明せよ．(5%)

- 三年間の追跡期間における，視覚障害者の死亡ハザードは視力正常者の 5.197 倍，このハザード比の95% 信頼区間は 3.56 - 7.51 と推定される．(1点)
- 粗ハザード比の95%信頼区間は 1 を跨いでいない．(1点)

以上の結果から，「視覚障害者が視力正常者と比べ，観察期間中に死亡するハザード比は1と等しい」を棄却するために，非常に強い証拠を提供した．(3点)

- 「視覚障害者が視力正常者と比べ，観察期間中に死亡するハザード比は(統計学的に)有意に1より高い」でも可

4.2　年齢調整ハザード比，及び信頼区間の意味を説明せよ．(5%)

- 三年間の追跡期間における，視覚障害者の年齢調整死亡ハザードは視力正常者の 2.10 倍，このハザード比の 95% 信頼区間は 1.37 - 3.21 と推定される．(1点)
- 年齢調整ハザード比の95%信頼区間は 1 を跨いでいない．(1点)
- このハザード比が大きく変化した（小さくなった）ことは，年齢がこの関連の強い交絡因子であることを示唆している．(1点)

以上の結果から，「視覚障害者が視力正常者と比べ，観察期間中に死亡する年齢調整ハザード比は1と等しい」を棄却するために，非常に強い証拠を提供した．(2点)

- 「視覚障害者が視力正常者と比べ，観察期間中に死亡する年齢調整ハザード比は(統計学的に)有意に1より高い」でも可