

第6章

問1 (a) のみ示す.

① $H_0: \mu_x = \mu_y$ vs. $H_1: \mu_x \neq \mu_y$ の両側 t 検定を行う. $n = 12, S = 6, \bar{X} = 21, \bar{Y} = 27$ より, 検定統計

$$\text{量 } T = \bar{X} - \bar{Y} / \sqrt{S^2 \left(\frac{1}{n} + \frac{1}{n} \right)} = -2.45 \text{ が得られる. 両側 P 値} = 0.023 \text{ であり, 5\% 有意水準で,}$$

H_0 を棄却できる. $\bar{X} < \bar{Y}$ より, $\mu_x < \mu_y$ と結論できる.

②

$$\cdot \mu_x \text{ の 95\% 信頼区間 : } \bar{X} \pm t_{11}(0.975)S/\sqrt{n} = (17.2, 24.8)$$

$$\cdot \mu_y \text{ の 95\% 信頼区間 : } \bar{Y} \pm t_{11}(0.975)S/\sqrt{n} = (23.2, 30.8)$$

$$\cdot \mu_x - \mu_y \text{ の 95\% 信頼区間 : } \bar{X} - \bar{Y} \pm t_{22}(0.975)S\sqrt{\left(\frac{1}{n} + \frac{1}{n}\right)} = (-11.1, -0.92)$$

③ $\bar{X}, \bar{Y}, \bar{X} - \bar{Y}$ の標準誤差はそれぞれ, $S/\sqrt{n}, S/\sqrt{n}, S\sqrt{2/n}$ であるため, S が同じ場合, n が大きいほど, あるいは, n が同じ場合, S が小さい場合ほど, 標準誤差は小さくなる. それに伴って, 検定統計量 T は 0 から離れ, P 値は小さくなり, 信頼区間の幅は小さくなる. つまり, 検定の精度が上がる.

```
> xbar <- 21; ybar <- 27
> n <- 12; s <- 6
> ( T <- (xbar-ybar)/(s*sqrt(2/n)) ) # (a) #
[1] -2.44949
> ( pt( T, df=n+n-2) ) *2 # p- 値 = 下側 p- 値 *2
[1] 0.02273287
> xbar + c(-1,1)*qt(.975,df=n-1)*s/sqrt(n) # mu_x の信頼区間
[1] 17.18778 24.81222
> ybar + c(-1,1)*qt(.975,df=n-1)*s/sqrt(n) # mu_y の信頼区間
[1] 23.18778 30.81222
> xbar-ybar + c(-1,1)*qt(.975,df=n+n-2)*s*sqrt(2/n) # mu_x-mu_y の信
頼区間
[1] -11.0799308 -0.9200692
> n <- 12; s <- 10
> ( T <- (xbar-ybar)/(s*sqrt(2/n)) )
[1] -1.469694
> ( pt( T, df=n+n-1) ) *2 # p- 値 = 下側 p- 値 * 2
[1] 0.155192
> xbar + c(-1,1)*qt(.975,df=n-1)*s/sqrt(n) # mu_x の信頼区間
[1] 14.6463 27.3537
> ybar + c(-1,1)*qt(.975,df=n-1)*s/sqrt(n) # mu_y の信頼区間
```

```

[1] 20.6463 33.3537
> xbar-ybar + c(-1,1)*qt(.975,df=n+n-2)*s*sqrt(2/n) # mu_x-mu_y の信
頼区間
[1] -14.466551 2.466551
> n <- 20; s <- 6 # (c) #
> ( T <- (xbar-ybar)/(s*sqrt(2/n)) )
[1] -3.162278
> ( pt( T, df=n+n-1) ) *2 # p- 値 = 下側 p- 値 * 2
[1] 0.003027422
> xbar + c(-1,1)*qt(.975,df=n-1)*s/sqrt(n) # mu_x の信頼区間
[1] 18.19191 23.80809
> ybar + c(-1,1)*qt(.975,df=n-1)*s/sqrt(n) # mu_y の信頼区間
[1] 24.19191 29.80809
> xbar-ybar + c(-1,1)*qt(.975,df=n+n-2)*s*sqrt(2/n) # mu_x-mu_y の信
頼区間
[1] -9.841018 -2.158982
> n <- 20; s <- 10 # (d) #
> ( T <- (xbar-ybar)/(s*sqrt(2/n)) )
[1] -1.897367
> ( pt( T, df=n+n-2) ) *2 # p- 値 = 下側 p- 値 * 2
[1] 0.06539761
> xbar + c(-1,1)*qt(.975,df=n-1)*s/sqrt(n) # mu_x の信頼区間
[1] 16.31986 25.68014
> ybar + c(-1,1)*qt(.975,df=n-1)*s/sqrt(n) # mu_y の信頼区間
[1] 22.31986 31.68014
> xbar-ybar + c(-1,1)*qt(.975,df=n+n-2)*s*sqrt(2/n) # mu_x-mu_y の信
頼区間
[1] -12.4016964 0.4016964

```

問 2

$n_x=20, \bar{X}=24.1, S_x^2=3.5, n_y=20, \bar{Y}=22.8, S_y^2=3.0$ とおく。

- ① 患者群の母平均を μ_x として, $H_0: \mu_x=25$ vs. $H_1: \mu_x<25$ の下側一標本 t 検定を行う。検定統計量 $T=(\bar{X}-25)/\sqrt{S_x^2/n_x}=-2.15$ が得られ, 下側 P 値 = 0.022 となる。したがって, 5% 有意水準で H_0 は棄却され, 患者群の検査値の平均値は 25 よりも小さいといえる。
- ②
 - ・患者群平均値 μ_x の 95% 信頼区間 = $\bar{X} \mp t_{19}(0.975)\sqrt{S_x^2/n_x} = (23.22, 24.98)$
 - ・健康者群平均値 μ_y の 95% 信頼区間 = $\bar{Y} \mp t_{19}(0.975)\sqrt{S_y^2/n_y} = (22.00, 23.61)$

③ $H_0: \mu_x = \mu_Y$ vs. $H_1: \mu_x \neq \mu_Y$ の両側 t 検定を行う。標本分散をプールすると $S^2 = \{(n_x - 1)S_x^2 + (n_y - 1)S_y^2\} / (n_x + n_y - 2) = 3.25$ となり、検定統計量 $\bar{X} - \bar{Y} \pm t_{n_x+n_y-2}(0.975)S \sqrt{\left(\frac{1}{n_x} + \frac{1}{n_y}\right)} = 3.25$

となる。両側 P 値 = 0.028 より、5%有意水準で、 H_0 を棄却できる。すなわち、二群の検査値の母平均に差があり、患者群のほうが健康者群よりも平均値が有意に高いと結論できる。

④ $\mu_x - \mu_Y$ の 95% 信頼区間 : $\bar{X} - \bar{Y} \pm t_{38}(0.975) \sqrt{S^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)} = (0.53, 2.67)$

⑤ 「2つの 95% 信頼区間が重なる場合は 5% 有意水準で異なるとはいえない」という主張は誤りである。なぜなら、個々の 95% 信頼区間が重なっていても、差の検定では有意な差が検出されることがある。これは、2つの推定値の差の標準誤差は、各群の標準誤差だけでなく両者の和に依存するため、個別の信頼区間の重なりと有意差の有無は一致しないことによる。一方で、「2つの 95% 信頼区間が重ならない場合は、5% 有意水準で異なるといえる」という主張は正しい。2つの真の母数が等しいためには、少なくとも一方の真の母数が対応する 95% 信頼区間の外に位置しなければならず、その確率は 5% 以下であるためである。ただし、この基準は非常に保守的であり、実際の有意水準は 5% よりかなり小さくなる点に注意が必要である。

```
> mean_samp <- c(24.1, 22.8)
> var_samp   <- c(3.5, 3.0)
> n <- c(20, 20)
> ( t <- (mean_samp[1] - 25) / sqrt(var_samp[1]/n[1]) ) # ①
[1] -2.151411
> pt(t, df=n[1]-1)      # 下側 P
[1] 0.02225943
>
> mean_samp[1] + c(-1, 1) * qt(0.975, df=n[1]-1) * sqrt(var_samp[1]/
n[1])    # ②
[1] 23.22443 24.97557
> mean_samp[2] + c(-1, 1) * qt(0.975, df=n[2]-1) * sqrt(var_samp[2]/
n[2])
[1] 21.98938 23.61062
> ( S2 <- ( (n[1]-1)*var_samp[1] + (n[2]-1)*var_samp[2] ) /
((n[1]+n[2]-2)) ) # ③
[1] 3.25
> ( t <- (mean_samp[1] - mean_samp[2]) / sqrt(S2*(1/n[1]+1/n[2])) )
[1] 2.280351
> ( 1 - pt(t, df=n[1]+n[2]-2) ) * 2    # p 値
[1] 0.02828438
```

```

> # ④
> ( mean_samp[1]-mean_samp[2] ) +
+      c(-1,1)*qt(0.975,df=n[1]+n[2]-2) * sqrt(s2) * sqrt(1/
n[1]+1/n[2])
[1] 0.1459178 2.4540822

```

問3

- ① データは血中濃度(単位： $\mu\text{g/mL}$)であり、両群ともヒストグラムは非対称で右裾が重く、ボックスプロットでも外れ値が多数確認できる。データは正規分布に従っているとはいがたい。
- ② 対照群と患者群の母平均値をそれぞれ μ_x, μ_y とし、 $H_0 : \mu_x = \mu_y$ vs. $H_1 : \mu_x \neq \mu_y$ の仮説に対し、Studentのt検定を行う。 $\bar{X} = 2.23, \bar{Y} = 5.03$ となり、t検定統計量は、 $T = -1.93$ （自由度38）と計算される。両側P値 = 0.061より、有意水準5%で帰無仮説を棄却できない。つまり、二群で血中濃度の変化に差があるとはいえない。
- ③ 対数変換後のデータは、分布が左右対称に近づき、ヒストグラムは正規分布状になった。
- ④ 対数変換後のt検定統計量は、 $T = -2.13$ （自由度38）と計算される。両側P値 = 0.04より、有意水準5%で帰無仮説を棄却できる。つまり、血中濃度の平均値は、患者群のほうが対照群よりも有意に高いといえる。②の結果と比較して、P値はわずかに小さくなり、有意性を示す結果となった。正規性に関する懸念の少ない対数変換後の解析結果のほうが②の結果よりも、信頼性が高いと考えられる。
- ⑤ 年齢や性別は血中物質量に影響を与える可能性があるため、これらの因子の分布が患者群と対照群で偏っていると、疾患の有無による比較に交絡（バイアス）が生じる可能性がある。そこで、年齢・性別でマッチングを行うことで、両群をこれらの因子について均一にし、不要なバイアスを減らして、純粋に疾患の影響を比較できるようにすることが目的である。

```

> x <- c(2.61, 2.90, 0.67, 0.05, 1.48, 1.83, 0.83, 0.56, 1.22, 4.58,
       1.13, 0.37, 0.64, 1.91, 14.38, 0.82, 1.05, 1.20, 3.01, 3.26)
> y <- c(0.23, 0.72, 3.37, 6.37, 7.54, 8.98, 3.36, 0.94, 23.74,
       3.99,
       1.93, 7.24, 3.68, 14.93, 0.69, 4.27, 0.65, 1.40, 1.10, 5.41)
> # ① ヒストグラム・ボックスプロット
> par(mfrow = c(2,3))
> hist(x, main="対照群", xlab="血中濃度 ( $\mu\text{g/mL}$ )")
> hist(y, main="患者群", xlab="血中濃度 ( $\mu\text{g/mL}$ )")
> boxplot(x, y )
> # ② 非変換データの t 検定
> t.test(x, y, var.equal=TRUE )

```

Two Sample t-test

```

data: x and y
t = -1.9296, df = 38, p-value = 0.06114
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-5.7415917  0.1375917
sample estimates:
mean of x mean of y
2.225      5.027
>                                         # ③ 対数変換
> logx <- log(x)
> logy <- log(y)
>
> hist(logx, main="対照群", xlab="log 血中濃度 (μg/mL)")
> hist(logy, main="患者群", xlab="log 血中濃度 (μg/mL)")
>
> boxplot(logx, logy)
>                                         # ④ 対数変換後の t 検定
> t.test(logx, logy, var.equal=TRUE)

Two Sample t-test

```

```

data: logx and logy
t = -2.1296, df = 38, p-value = 0.03974
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-1.53203415 -0.03878974
sample estimates:
mean of x mean of y
0.2444779 1.0298899

```

問 4

- ① 患者群と対照群の真の曝露率をそれぞれ π_1, π_2 とする。患者群のアスベスト曝露割合の点推定値は、 $\hat{\pi}_1 = 19/40 = 0.475$, 95% 信頼区間は $\hat{\pi}_1 \pm 1.96 \sqrt{\hat{\pi}_1(1-\hat{\pi}_1)/40} = (0.32, 0.63)$ 。同様に、対照群の点推定値は $\hat{\pi}_2 = 8/40 = 0.2$, 95% 信頼区間は $\hat{\pi}_2 \pm 1.96 \sqrt{\hat{\pi}_2(1-\hat{\pi}_2)/40} = (0.076, 0.32)$ となる。
- ② $H_0 : \pi_1 = \pi_2$ vs. $H_1 : \pi_1 \neq \pi_2$ の検定を行う。3つの方法のいずれも 5% 有意水準で H_0 を棄却で

き、患者群のほうが対照群よりも曝露率が高いことを示す。

- ① カイ二乗検定（連続修正）： $X^2 = 6.76$, (df=1), P 値 = 0.0093
- ② フィッシャーの検定：P 値 = 0.0172
- ③ 正規近似による検定： $Z = 2.60$, P 値 = 0.0093
- ④ $\hat{\pi}_1 - \hat{\pi}_2 \pm 1.96\sqrt{\hat{\pi}_1(1-\hat{\pi}_1)/40 + \hat{\pi}_2(1-\hat{\pi}_2)/40} = (0.077, 0.47)$ この信頼区間は 0 を含んでおらず、曝露割合は二群で異なる（患者群 > 対照群）ことを示す。
- ⑤ 一般に、このような観察研究（症例対照研究）では、曝露と疾病との間に有意な関連が認められたとしても、観測されていない因子（交絡因子）の影響を完全には排除できない。そのため、曝露が病気を「引き起こした」とする因果関係を直接結論することはできない。因果推論には、複数の研究結果や交絡調整など、さらなる検証が必要である。

```
> # ①  
> n1 <- 40; x1 <- 19 # 患者群  
> (pihat1 <- x1/n1) # 推定値  
[1] 0.475  
> pihat1 + c(-1, 1)*1.96*sqrt(pihat1*(1-pihat1)/n1) # 正規近似による  
95%信頼区間  
[1] 0.3202422 0.6297578  
> n2 <- 40; x2 <- 8 # 対照群  
> (pihat2 <- x2/n2)  
[1] 0.2  
> pihat2 + c(-1, 1)*1.96*sqrt(pihat2*(1-pihat2)/n2)  
[1] 0.07603872 0.32396128  
> # ②  
> A <- matrix(c(19, 8, 21, 32), 2, 2)  
> chisq.test(A, correct=F) # カイ二乗検定  
  
Pearson's Chi-squared test  
  
data: A  
X-squared = 6.7645, df = 1, p-value = 0.009299  
  
> pihat <- (x1+x2)/(n1+n2) # 正規分布に基づく検定  
> (z <- (pihat1-pihat2) / sqrt(pihat1*(1-pihat1)*(1/n1+1/n2)))  
[1] 2.600865  
> (1-pnorm(z))*2  
[1] 0.009298896  
> fisher.test(A) # フィッシャー検定
```

Fisher's Exact Test for Count Data

```
data: A
p-value = 0.01724
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
1.219033 11.252060
sample estimates:
odds ratio
3.559026
> ## ③
> pihat1-pihat2 + c(-1,1)*1.96*sqrt(pihat1*(1-pihat1)/n1+pihat2*(1-
pihat2)/n2)
[1] 0.07671643 0.47328357
```

問 5

McNemar 検定を行う。 π_1 を通常体重群, π_2 を肥満群の静脈瘤有率とし, $H_0: \pi_1 = \pi_2$ に対して, 両側検定を行う。検定統計量 = 2.12 で, カイ二乗分布 ($df=1$) に基づく P 値は, 0.14 で, 5%有意水準で H_0 を棄却できない。すなわち, このデータからは静脈瘤の発症と肥満に関連があるとはいえない。 $\pi_1 - \pi_2$ の 95% 信頼区間は (-0.01, 0.124) と計算される。

```
> mcnemar.test( matrix( c(19,5,12,86), 2, 2 ) )

McNemar's Chi-squared test with continuity correction

data: matrix(c(19, 5, 12, 86), 2, 2)
McNemar's chi-squared = 2.1176, df = 1, p-value = 0.1456

> (12-5)/122 + c(-1,1)*1.96*sqrt(12+5)/122
[1] -0.008863008 0.123617107
```

(Web Appendix)