מודלים סטטיסטיים ויישומיהם 52518 תשע"ח – פתרון תרגיל 8

להגשה עד 25.12.17 בשעה 23:55

1. קובץ הנתונים ex8dat.txt מציג ניסוי לבדיקת יעילותה של תרופה לשיפור תפקוד הכליות בקרב חולי סכרת. עמודות הקובץ הן (משמאל לימין במסד הנתונים) ID, Dose Group, Type of Diabetes ו-Y (מדד תפקוד הכליה).

א. הריצו ANOVA דו-כווני ללא אינטראקציות עם המשקלות

$$\pi_i = \frac{n_{i.}}{N}, \tau_j = \frac{n_{.j}}{N}$$

ב. מצאו ביטוי עבור α_I לפי $\alpha_1, \dots, \alpha_{I-1}, n_{1.}, \dots, n_{I.}$.

ג. במודל נטול האינטראקציות, מצאו את המטריצה C המקיימת $\Delta = C\begin{pmatrix}\alpha_1\\ \vdots \\ \alpha_{I-1}\end{pmatrix}$, כאשר Δ הוא אוסף כל הקונטרסטים הזוגיים (מהצורה $\alpha_i - \alpha_{i'}$).

ד. במודל נטול האינטראקציות, חשבו בעזרת השיטה המדויקת רווחי סמך בו-זמניים לכל הקונטרסטים הזוגיים (מהצורה $\alpha_i - \alpha_{i'}$).

ה. חזרו על הסעיפים הקודמים, הפעם במודל עם אינטראקציות ותוך שימוש במשקלות הבאים:

$$\pi_i = \frac{1}{I}, \tau_1 = 0.75, \tau_2 = 0.25$$

א. הניתוח המתקבל:

	DF	Sum Sq	MSE	F	$P(>F)$
A	3	47.4813	15.8271	22.1303	$2.0042 \cdot 10^{-12}$
B	1	0.4460	0.446	0.6236	0.4306
E	201	143.7509	0.7152		

ב.

$$I = 4 \rightarrow \alpha_4 = -\frac{n_{1.}}{n_{4.}}\alpha_1 - \frac{n_{2.}}{n_{4.}}\alpha_2 - \frac{n_{3.}}{n_{4.}}\alpha_3$$

ג.

$$\Delta = \begin{pmatrix}\alpha_1-\alpha_2\\ \alpha_1-\alpha_3\\ \alpha_2-\alpha_3\\ \alpha_1-\alpha_4\\ \alpha_2-\alpha_4\\ \alpha_3-\alpha_4\end{pmatrix} = \begin{pmatrix}\alpha_1-\alpha_2\\ \alpha_1-\alpha_3\\ \alpha_2-\alpha_3\\ \alpha_1-\left(-\frac{n_{1.}}{n_{4.}}\alpha_1-\frac{n_{2.}}{n_{4.}}\alpha_2-\frac{n_{3.}}{n_{4.}}\alpha_3\right)\\ \alpha_2-\left(-\frac{n_{1.}}{n_{4.}}\alpha_1-\frac{n_{2.}}{n_{4.}}\alpha_2-\frac{n_{3.}}{n_{4.}}\alpha_3\right)\\ \alpha_3-\left(-\frac{n_{1.}}{n_{4.}}\alpha_1-\frac{n_{2.}}{n_{4.}}\alpha_2-\frac{n_{3.}}{n_{4.}}\alpha_3\right)\end{pmatrix} = \underbrace{\begin{pmatrix}1 & -1 & 0\\ 1 & 0 & -1\\ 0 & 1 & -1\\ 1+\frac{n_{1.}}{n_{4.}} & \frac{n_{2.}}{n_{4.}} & \frac{n_{3.}}{n_{4.}}\\ \frac{n_{1.}}{n_{4.}} & 1+\frac{n_{2.}}{n_{4.}} & \frac{n_{3.}}{n_{4.}}\\ \frac{n_{1.}}{n_{4.}} & \frac{n_{2.}}{n_{4.}} & 1+\frac{n_{3.}}{n_{4.}}\end{pmatrix}}_{C}\begin{pmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\end{pmatrix}$$

ד. יש לנו את המטריצה $D = (Q_1^T Q_1)^{-1}$ המכילה את המידע הרלוונטי על α_i,מטריצת הקונטרסטים היא C לעיל. נגדיר Ω כאברי האלכסון של CDC בחזקת $-\frac{1}{2}$ ואז את $R = \Omega CDC^T \Omega$, עבור $\alpha = 0.05$ נקבל את רווחי הסמך הבאים:

$$CI = \begin{bmatrix}0.0350 & 0.8984\\ 0.3444 & 1.2115\\ -0.1243 & 0.7468\\ 0.8819 & 1.7365\\ 0.4132 & 1.2717\\ 0.0999 & 0.9626\end{bmatrix}$$

ה. הניתוח המתקבל:

	DF	Sum Sq	MSE	F	$P(> F)$
A	3	37.1992	12.3997	17.3348	$4.9209 \cdot 10^{-10}$
B	1	0.4292	0.4292	0.6000	0.4395
AB	3	2.1196	0.7065	0.9877	0.3996
E	198	141.6313	0.7153		

ורווחי הסמך:

$$CI = \begin{bmatrix} -0.1032 & 0.8278 \\ 0.2667 & 1.2100 \\ -0.0873 & 0.8395 \\ 0.7568 & 1.6774 \\ 0.4076 & 1.3021 \\ 0.0298 & 0.9277 \end{bmatrix}$$

קוד R לשאלה זו:

```
# READ DATA
indat = read.table('ex8dat.txt')
A = indat[,2]
B = indat[,3]
Y = indat[,4]
nA = table(A)
nB = table(B)
#In the code that follows, I use the fact that I know that the possible values of A are 1,2,3,4
#and the possible values of B are 1,2. This could be avoided with a bit of extra coding.

#Q1 Parts A and B - Set up variables to be used in regressions
N = length(Y)
XA1 = rep(0,N)
XA2 = rep(0,N)
XA3 = rep(0,N)
XB = rep(0,N)
nA14 = nA[1]/nA[4]
nA24 = nA[2]/nA[4]
nA34 = nA[3]/nA[4]
nB12 = nB[1]/nB[2]
for (r in 1:N) {
  if (A[r]==1) {XA1[r]=1}
  if (A[r]==2) {XA2[r]=1}
  if (A[r]==3) {XA3[r]=1}
  if (A[r]==4) {
    XA1[r] = -nA14
    XA2[r] = -nA24
    XA3[r] = -nA34
  }
  if (B[r]==1) {XB[r]=1}
  if (B[r]==2) {XB[r]=-nB12}
}

#Q1 Parts A and B - Run regressions
model.a.b = lm(Y ~ XA1 + XA2 + XA2 + XA3 + XB)
model.a = lm(Y ~ XA1 + XA2 + XA3)
model.b = lm(Y ~ XB)
sse.a.b = sum(model.a.b$residuals^2)
sse.a = sum(model.a$residuals^2)
sse.b = sum(model.b$residuals^2)

#Q1 Parts A and B - Compute SS, MS, F, and p values
sse.no.int = sse.a.b
mse.no.int = sse.no.int/(N-5)
```

```
ssa.no.int = sse.b - sse.a.b
msa.no.int = ssa.no.int/3
F.A = msa.no.int/mse.no.int
p.A = 1 - pf(F.A,3,N-5)
ssb.no.int = sse.a - sse.a.b
msb.no.int = ssb.no.int
F.B = msb.no.int/mse.no.int
p.B = 1 - pf(F.B,1,N-5)
df = c(3,1,N-5)
SS = c(ssa.no.int,ssb.no.int,sse.no.int)
MS = c(msa.no.int,msb.no.int,mse.no.int)
F = c(F.A,F.B,NA)
p.value = c(p.A,p.B,NA)
anova.no.int = cbind(df,SS,MS,F,p.value)
anova.no.int = data.frame(anova.no.int)
rownames(anova.no.int) = c('A','B','Error')
anova.no.int

#Q1 Part C
C = rbind(
  c(1,-1,0),
  c(1,0,-1),
  c(0,1,-1),
  c(1+nA14,nA24,nA34),
  c(nA14,1+nA24,nA34),
  c(nA14,nA24,1+nA34))
eta = coef(model.a.b)
alf = eta[2:4]
del.rs.hat = C %*% alf
xmat = cbind(rep(1,N),XA1,XA2,XA3,XB)
G = mse.no.int * solve(t(xmat)%*%xmat)
Ga = G[2:4,2:4]
cov.dif = C %*% Ga %*% t(C)
sd.dif = sqrt(diag(cov.dif))
sd.dif.mat.i = diag(1/sd.dif)
R = sd.dif.mat.i %*% cov.dif %*% sd.dif.mat.i
library(mvtnorm)
ts = qmvt(0.95,tail='both.tails',corr=R, df = 217)
tstar = ts$quantile
cihw = tstar * sd.dif
lower.conf.lim = del.rs.hat - cihw
upper.conf.lim = del.rs.hat + cihw
r.index = c(1,1,2,1,2,3)
s.index = c(2,3,3,4,4,4)
result.C = cbind(r.index,s.index,del.rs.hat,lower.conf.lim,upper.conf.lim)
result.C = data.frame(result.C)
names(result.C) = c('r','s','del.rs.hat','lower.conf.lim','upper.conf.lim')
result.C

#Q1 Part D
# we have to express alpha4 in terms of the other alphas
alf4coef = -c(nA14,nA24,nA34)
c = c(-1,0,0.5) + 0.5*alf4coef
psi.hat = sum(c*alf)
sd = sqrt(t(c) %*% Ga %*% c)
fcrit = qf(0.95,3,N-5)
xi = sqrt(3*fcrit)
cihw = xi * sd
psi.L = psi.hat - cihw
psi.U = psi.hat + cihw
result.D = cbind(psi.hat,psi.L,psi.U)
```

```
result.D = data.frame(result.D)
names(result.D) = c('psi.hat','psi.L','psi.U')
result.D

#ADDED 5778: REDO VARIABLES WITH NEW WEIGHTS
N = length(Y)
XA1 = rep(0,N)
XA2 = rep(0,N)
XA3 = rep(0,N)
XB = rep(0,N)
for (r in 1:N) {
  if (A[r]==1) {XA1[r]=1}
  if (A[r]==2) {XA2[r]=1}
  if (A[r]==3) {XA3[r]=1}
  if (A[r]==4) {
    XA1[r] = -1
    XA2[r] = -1
    XA3[r] = -1
  }
  if (B[r]==1) {XB[r]=1}
  if (B[r]==2) {XB[r]=-3}
}

#Q1 Part E - Set up variables and run regressions
XAB1 = XA1*XB
XAB2 = XA2*XB
XAB3 = XA3*XB
full = lm(Y ~ XA1 + XA2 + XA3 + XB + XAB1 +XAB2 + XAB3)
model.a.ab = lm(Y ~ XA1 + XA2 + XA3 + XAB1 +XAB2 + XAB3)
model.b.ab = lm(Y ~ XB + XAB1 + XAB2 + XAB3)
sse.full = sum(full$residuals^2)
sse.a.ab = sum(model.a.ab$residuals^2)
sse.b.ab = sum(model.b.ab$residuals^2)

#Q1 Part E - Compute SS, MS, F, and p values
sse = sse.full
mse = sse/(N-8)
ssa = sse.b.ab - sse
msa = ssa/3
F.A = msa/mse
p.A = 1 - pf(F.A,3,N-8)
ssb = sse.a.ab - sse
msb = ssb
F.B = msb/mse
p.B = 1 - pf(F.B,1,N-8)
ssab = sse.a.b - sse
msab = ssab/3
F.AB = msab/mse
p.AB = 1 - pf(F.AB,3,N-8)
df = c(3,1,3,N-8)
SS = c(ssa,ssb,ssab,sse)
MS = c(msa,msb,msab,mse)
F = c(F.A,F.B,F.AB,NA)
p.value = c(p.A,p.B,p.AB,NA)
anova = cbind(df,SS,MS,F,p.value)
anova = data.frame(anova)
rownames(anova) = c('A','B','AB','Error')
anova

#Q1 Part E - Pairwise comparisons
#Same C matrix as before
```

```
eta = coef(full)
alf = eta[2:4]
del.rs.hat = C %*% alf
xmat = cbind(rep(1,N),XA1,XA2,XA3,XB,XAB1,XAB2,XAB3)
G = mse * solve(t(xmat)%*%xmat)
Ga = G[2:4,2:4]
cov.dif = C %*% Ga %*% t(C)
sd.dif = sqrt(diag(cov.dif))
sd.dif.mat.i = diag(1/sd.dif)
R = sd.dif.mat.i %*% cov.dif %*% sd.dif.mat.i
ts = qmvt(0.95,tail='both.tails',corr=R, df = 198)
tstar = ts$quantile
cihw = tstar * sd.dif
lower.conf.lim = del.rs.hat - cihw
upper.conf.lim = del.rs.hat + cihw
r.index = c(1,1,2,1,2,3)
s.index = c(2,3,3,4,4,4)
result.E2 = cbind(r.index,s.index,del.rs.hat,lower.conf.lim,upper.conf.lim)
result.E2 = data.frame(result.E2)
names(result.E2) = c('r','s','del.rs.hat','lower.conf.lim','upper.conf.lim')
result.E2
```

2. יהיה מודל ANOVA דו-כווני (עם או בלי אינטראקציה). כמו בכתה, נסמן ב-H את הבלוק במטריצה $(X^TX)^{-1}$ המתייחס לאיברים $\alpha_1, \dots, \alpha_{I-1}$. נסמן ב-$\hat{Y}$ את וקטור התחזיות תחת המודל המלא וב-$\hat{Y}^{(0)}$ את וקטור התצפיות תחת ההשערה $H_A: \alpha_1 = \cdots = \alpha_I = 0$. ניתן להראות כי $\|\hat{Y} - \hat{Y}^{(0)}\|^2 = \hat{\alpha}^T H^{-1} \hat{\alpha}$ ואז F_A (סטטיסטי המבחן F המשמש לבדיקת ההשערה H_A) מקיים

$$F_A = \frac{\hat{\alpha}^T H^{-1} \hat{\alpha}/(I-1)}{s^2}$$

תחת H_A, $F_A \sim F_{I-1,p}$ כאשר p הוא מספר דרגות החופש ($N-IJ$ במודל עם אינטראקציות, $N-I-J+1$ במודל נטול אינטראקציות).

א. עבור נתוני הקובץ ex8dat.txt, ודאו כי מתקיימת הזהות $F_A = \frac{\hat{\alpha}^T H^{-1} \hat{\alpha}/(I-1)}{s^2}$.

ב. נסמן את הערך האמתי של הוקטור α באופן α^*. הוכיחו כי לכל α^* מתקיים

$$\frac{(\hat{\alpha} - \alpha^*)^T H^{-1} (\hat{\alpha} - \alpha^*)/(I-1)}{s^2} \sim F_{I-1,p}$$

ג. השתמשו בתוצאות הנ"ל בכדי לוודא את נכונות רווח הסמך שהוצג בכתה עבור רווח סמך בשיטת Scheffé. בססו את ההוכחה שלכם על ההוכחה לרווח סמך בשיטת Scheffé במקרה החד-כווני.

א. קוד R (המשך לשאלה 1):

```
#Q2 - Set up variables
XA1 = rep(0,N)
XA2 = rep(0,N)
XA3 = rep(0,N)
XB = rep(0,N)
for (r in 1:N) {
  if (A[r]==1) {XA1[r]=1}
  if (A[r]==2) {XA2[r]=1}
  if (A[r]==3) {XA3[r]=1}
  if (A[r]==4) {
    XA1[r] = -1
    XA2[r] = -1
    XA3[r] = -1
  }
  if (B[r]==1) {XB[r]=1}
  if (B[r]==2) {XB[r]=-0.15/0.85}
```

```
}
XAB1 = XA1*XB
XAB2 = XA2*XB
XAB3 = XA3*XB

#Q2 - Run regressions
full = lm(Y ~ XA1 + XA2 + XA3 + XB + XAB1 +XAB2 + XAB3)
model.b.ab = lm(Y ~ XB + XAB1 + XAB2 + XAB3)
sse.full = sum(full$residuals^2)
sse.b.ab = sum(model.b.ab$residuals^2)
sse = sse.full

#Q2 - Compute F.A value using first method
ssa = sse.b.ab - sse
msa = ssa/3
mse = sse/(N-8)
F.first = msa/mse
F.first

#Q2 - Compute F.A value using second method
eta = coef(full)
alf = eta[2:4]
xmat = cbind(rep(1,N),XA1,XA2,XA3,XB,XAB1,XAB2,XAB3)
G = mse * solve(t(xmat)%*%xmat)
Ga = G[2:4,2:4]
F.second = t(alf) %*% solve(Ga) %*% alf / 3
F.second
```

ב. במודל עם אינטראקציות מתקיים:

$$Y = X\eta + \epsilon \to \hat{\eta} = \begin{pmatrix} \hat{\mu} \\ \hat{\alpha} \\ \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = (X^T X)^{-1} X^T Y$$

אז עבור G שהיא השורות $2, \dots, I$ של המטריצה $(X^T X)^{-1}$ יתקיים $\hat{\alpha} = G X^T Y$. בעת, נסמן את הערך האמתי של η באופן η^*, מתקיים $\eta^* = (X^T X)^{-1} X^T X \eta^*$ ולכן $\alpha^* = G X^T X \eta^*$, נסמן:

$$\Delta = \hat{\alpha} - \alpha^* = G X^T Y - G X^T X \eta^* = G X^T (Y - X\eta^*) = G X^T \tilde{Y}$$

כלומר Δ זהה לאומד של α אותו היינו מקבלים עבור הנתונים $\tilde{Y}$, אבל עבור $\tilde{Y}$ הוקטור α^* מתאפס ולכן נקבל כי

$$\frac{(\hat{\alpha} - \alpha^*)^T H^{-1} (\hat{\alpha} - \alpha^*)/(I-1)}{\tilde{s}^2} \sim F_{I-1, N-IJ}$$

כאשר $\tilde{s}^2$ הוא הערך של s^2 עבור נתוני $\tilde{Y}$, כלומר כדי לסיים עלינו להוכיח כי $\tilde{s}^2 = s^2$:

$$P_X = X(X^T X)^{-1} X^T \to \tilde{e} = (I - P_X)\tilde{Y} = (I - P_X)(Y - X\eta^*) = Y - P_X Y - X\eta^* + P_X X \eta^*$$
$$= (Y - \hat{Y}) + (X\eta^* - X(X^T X)^{-1} X^T X \eta^*) = e + (X\eta^* - X\eta^*) = e \to \tilde{s}^2 = s^2$$

ההוכחה למודל ללא אינטראקציות זהה.

ג. המטריצה H היא חיובית וסימטרית, לכן קיימת מטריצה A עבורה $(A^{-1})^T A^{-1} = H^{-1}$, $AA^T = H$. אחת הדרכים למצוא את A היא דרך פירוק ספקטרלי: אם $H = U\Lambda U^T$ אז נוכל לקבל $A = U\Lambda^{\frac{1}{2}} U^T$. בעת:

$$\left|\hat{\psi}(d) - \psi(d)\right| = |d^T \hat{\alpha} - d^T \alpha^*| = |d^T(\hat{\alpha} - \alpha^*)| = |d^T A A^{-1}(\hat{\alpha} - \alpha^*)| = \left|(A^T d)^T \left(A^{-1}(\hat{\alpha} - \alpha^*)\right)\right|$$

מא"ש קושי-שוורץ נקבל:

$$\left|(A^T d)^T \left(A^{-1}(\hat{\alpha} - \alpha^*)\right)\right| \le \|A^T d\| \cdot \|A^{-1}(\hat{\alpha} - \alpha^*)\| = \left((d^T A A^T d) \left((\hat{\alpha} - \alpha^*) {A^{-1}}^T A^{-1} (\hat{\alpha} - \alpha^*) \right) \right)^{\frac{1}{2}}$$
$$= \left((d^T H d) \left((\hat{\alpha} - \alpha^*) H^{-1} (\hat{\alpha} - \alpha^*) \right) \right)^{\frac{1}{2}}$$

כלומר

$$\max_{d\neq 0}\frac{\left|\hat{\psi}(d)-\psi(d)\right|}{s\sqrt{V(d)}}=\left(\frac{(\hat{\alpha}-\alpha^*)^T H^{-1}(\hat{\alpha}-\alpha^*)}{\tilde{s}^2}\right)^{\frac{1}{2}}=(I-1)^{\frac{1}{2}}\left(\frac{(\hat{\alpha}-\alpha^*)^T H^{-1}(\hat{\alpha}-\alpha^*)/(I-1)}{\tilde{s}^2}\right)^{\frac{1}{2}}$$

ולכן

$$\max_{d\neq 0}\frac{\left|\hat{\psi}(d)-\psi(d)\right|}{s\sqrt{(I-1)V(d)}}=\left(\frac{(\hat{\alpha}-\alpha^*)^T H^{-1}(\hat{\alpha}-\alpha^*)/(I-1)}{\tilde{s}^2}\right)^{\frac{1}{2}}$$

הראינו בסעיף ב' כי $\frac{(\hat{\alpha}-\alpha^*)^T H^{-1}(\hat{\alpha}-\alpha^*)/(I-1)}{\tilde{s}^2}\sim F_{I-1,p}$ ולכן:

$$P\left(\left|\hat{\psi}(d)-\psi(d)\right|\leq s\sqrt{(I-1)F_{I-1,p}^{1-\omega}V(d)}\right)=P\left(\max_{d\neq 0}\frac{\left|\hat{\psi}(d)-\psi(d)\right|}{s\sqrt{(I-1)V(d)}}\leq\sqrt{F_{I-1,p}^{1-\omega}}\right)$$
$$=P\left(\left(\frac{(\hat{\alpha}-\alpha^*)^T H^{-1}(\hat{\alpha}-\alpha^*)/(I-1)}{\tilde{s}^2}\right)^{\frac{1}{2}}\leq\sqrt{F_{I-1,p}^{1-\omega}}\right)=1-\omega$$

3. יהי מודל רגרסיה לא-לינארי עבור משתנה תגובה רציף Y, המוגדר באופן:

$$Y_i=h(\beta^T X_i)+\epsilon_i$$

כאשר h גזירה ומונוטונית עולה וכן $\epsilon_i\underset{iid}{\sim}N(0,\sigma^2)$. מצאו את פונקצית לוג-הנראות וכן את הנגזרות החלקיות שלה מסדר ראשון.

$$Y_i=h(\beta^T X_i)+\epsilon_i\rightarrow Y_i\sim N(h(\beta^T X_i),\sigma^2)$$

$$L(Y;\beta,\sigma^2,X)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{\left(Y_i-h(\beta^T X_i)\right)^2}{2\sigma^2}\right\}=(2\pi\sigma^2)^{-\frac{n}{2}}\cdot\prod_{i=1}^{n}\exp\left\{-\frac{\left(Y_i-h(\beta^T X_i)\right)^2}{2\sigma^2}\right\}$$

$$l(Y;\beta,\sigma^2,X)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma^2)+\sum_{i=1}^{n}-\frac{\left(Y_i-h(\beta^T X_i)\right)^2}{2\sigma^2}$$
$$=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i-h(\beta^T X_i)\right)^2$$
$$=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i^2-2Y_i h(\beta^T X_i)+\left(h(\beta^T X_i)\right)^2\right)$$
$$=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma^2)-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}Y_i^2-2\sum_{i=1}^{n}Y_i h(\beta^T X_i)+\sum_{i=1}^{n}\left(h(\beta^T X_i)\right)^2\right)$$
$$=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma^2)-\frac{1}{2}(\sigma^2)^{-1}\left(\sum_{i=1}^{n}Y_i^2-2\sum_{i=1}^{n}Y_i h(\beta^T X_i)+\sum_{i=1}^{n}\left(h(\beta^T X_i)\right)^2\right)$$

$$\frac{\partial l}{\partial\sigma^2}=-\frac{n}{2\sigma^2}-\frac{1}{2}\cdot -1\cdot(\sigma^2)^{-2}\cdot\left(\sum_{i=1}^{n}Y_i^2-2\sum_{i=1}^{n}Y_i h(\beta^T X_i)+\sum_{i=1}^{n}\left(h(\beta^T X_i)\right)^2\right)$$
$$=\frac{1}{2\sigma^2}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(Y_i-h(\beta^T X_i)\right)^2-n\right)$$

$$\frac{\partial l}{\partial\beta}=-\frac{1}{2\sigma^2}\left(-2\sum_{i=1}^{n}Y_i h'(\beta^T X_i)+2\sum_{i=1}^{n}\left(h(\beta^T X_i)\right)h'(\beta^T X_i)\right)=\frac{1}{2\sigma^2}\sum_{i=1}^{n}h'(\beta^T X_i)\left[\left(h(\beta^T X_i)\right)-Y_i\right]$$

4. יהי Y משתנה מקרי בינארי המקיים $Y_i\in\{0,1\}$. נתייחס למודל הלוגיסטי

$$P(Y=1|X=x)=\frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

כאשר X משתנה מקרי בינארי המקיים $X_i\in\{0,1\}$. הראו כי במקרה זה קיים פתרון אלגברי סגור למשוואת הנראות ומצאו אותו.

$$L(\beta_0,\beta_1;x)=\prod_{i=1}^{N}\left(\frac{e^{\beta_0+\beta_1x_i}}{1+e^{\beta_0+\beta_1x_i}}\right)^{y_i}\left(1-\frac{e^{\beta_0+\beta_1x_i}}{1+e^{\beta_0+\beta_1x_i}}\right)^{1-y_i}=\prod_{i=1}^{N}\left(\frac{e^{\beta_0+\beta_1x_i}}{1+e^{\beta_0+\beta_1x_i}}\right)^{y_i}\left(\frac{1}{1+e^{\beta_0+\beta_1x_i}}\right)^{1-y_i}$$

$$=\prod_{i=1}^{N}\left(e^{\beta_0+\beta_1x_i}\right)^{y_i}\left(\frac{1}{1+e^{\beta_0+\beta_1x_i}}\right)^{1}=\prod_{i=1}^{N}\frac{\left(e^{\beta_0+\beta_1x_i}\right)^{y_i}}{1+e^{\beta_0+\beta_1x_i}}$$

$$l=\log L=\sum_{i=1}^{N}y_i\log\left(e^{\beta_0+\beta_1x_i}\right)-\sum_{i=1}^{N}\log\left(1+e^{\beta_0+\beta_1x_i}\right)=\sum_{i=1}^{N}y_i(\beta_0+\beta_1x_i)-\sum_{i=1}^{N}\log\left(1+e^{\beta_0+\beta_1x_i}\right)$$

$$\frac{\partial l}{\partial\beta_0}=\sum_{i=1}^{N}y_i-\sum_{i=1}^{N}\frac{e^{\beta_0+\beta_1x_i}}{1+e^{\beta_0+\beta_1x_i}}=0\to\sum_{i=1}^{N}y_i=\sum_{i=1}^{N}\frac{e^{\beta_0+\beta_1x_i}}{1+e^{\beta_0+\beta_1x_i}}$$

$$=\sum_{i=1}^{N}x_i\cdot\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+(N-\sum_{i=1}^{N}x_i)\frac{e^{\beta_0}}{1+e^{\beta_0}}$$

$$\frac{\partial l}{\partial\beta_1}=\sum_{i=1}^{N}y_ix_i-\sum_{i=1}^{N}\frac{x_ie^{\beta_0+\beta_1x_i}}{1+e^{\beta_0+\beta_1x_i}}=0\to\sum_{i=1}^{N}y_ix_i=\sum_{i=1}^{N}\frac{x_ie^{\beta_0+\beta_1x_i}}{1+e^{\beta_0+\beta_1x_i}}=\sum_{i=1}^{N}x_i\cdot\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$$

נציב בנגזרת $\frac{\partial l}{\partial\beta_0}$ ונקבל:

$$\sum_{i=1}^{N}y_i=\sum_{i=1}^{N}x_i\cdot\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}+\left(N-\sum_{i=1}^{N}x_i\right)\frac{e^{\beta_0}}{1+e^{\beta_0}}=\sum_{i=1}^{N}y_ix_i+\left(N-\sum_{i=1}^{N}x_i\right)\frac{e^{\beta_0}}{1+e^{\beta_0}}$$

$$\to\left(N-\sum_{i=1}^{N}x_i\right)\frac{e^{\beta_0}}{1+e^{\beta_0}}=\sum_{i=1}^{N}(y_i-y_ix_i)=\sum_{i=1}^{N}y_i(1-x_i)\to\frac{e^{\beta_0}}{1+e^{\beta_0}}=\frac{\sum_{i=1}^{N}y_i(1-x_i)}{\left(N-\sum_{i=1}^{N}x_i\right)}$$

$$\to logit\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)=\hat{\beta}_0=logit\left(\frac{\sum_{i=1}^{N}y_i(1-x_i)}{\left(N-\sum_{i=1}^{N}x_i\right)}\right)=\log\left(\frac{\sum_{i=1}^{N}y_i(1-x_i)}{N-\sum_{i=1}^{N}x_i-\sum_{i=1}^{N}y_i(1-x_i)}\right)$$

$$=\log\left(\frac{\sum_{i=1}^{N}y_i(1-x_i)}{\sum_{i=1}^{N}(1-x_i)-\sum_{i=1}^{N}y_i(1-x_i)}\right)\to\boxed{\hat{\beta}_0=\log\left(\frac{\sum_{i=1}^{N}y_i(1-x_i)}{\sum_{i=1}^{N}(1-y_i)(1-x_i)}\right)}$$

בעת נוכל לקבל גם

$$\sum_{i=1}^{N}y_ix_i=\sum_{i=1}^{N}x_i\cdot\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\to\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}=\frac{\sum_{i=1}^{N}y_ix_i}{\sum_{i=1}^{N}x_i}\to logit\left(\frac{\sum_{i=1}^{N}y_ix_i}{\sum_{i=1}^{N}x_i}\right)=logit\left(\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right)$$

$$=\hat{\beta}_0+\hat{\beta}_1\to\hat{\beta}_1=logit\left(\frac{\sum_{i=1}^{N}y_ix_i}{\sum_{i=1}^{N}x_i}\right)-logit\left(\frac{\sum_{i=1}^{N}y_i(1-x_i)}{\left(N-\sum_{i=1}^{N}x_i\right)}\right)$$

$$\to sigmoid\left(\hat{\beta}_1\right)=\frac{\sum_{i=1}^{N}y_ix_i}{\sum_{i=1}^{N}x_i}-\frac{\sum_{i=1}^{N}y_i(1-x_i)}{\left(N-\sum_{i=1}^{N}x_i\right)}=\frac{N\sum_{i=1}^{N}y_ix_i-\sum_{i=1}^{N}x_i\sum_{i=1}^{N}y_ix_i-\sum_{i=1}^{N}x_i\sum_{i=1}^{N}y_i(1-x_i)}{\sum_{i=1}^{N}x_i\left(N-\sum_{i=1}^{N}x_i\right)}$$

$$=\frac{N\sum_{i=1}^{N}y_ix_i-\sum_{i=1}^{N}x_i\left(\sum_{i=1}^{N}y_ix_i+\sum_{i=1}^{N}y_i(1-x_i)\right)}{\sum_{i=1}^{N}x_i\left(N-\sum_{i=1}^{N}x_i\right)}=\frac{N\sum_{i=1}^{N}y_ix_i-\sum_{i=1}^{N}x_i\sum_{i=1}^{N}y_i}{\sum_{i=1}^{N}x_i\left(N-\sum_{i=1}^{N}x_i\right)}$$

$$\to\boxed{\hat{\beta}_1=logit\left(\frac{N\sum_{i=1}^{N}y_ix_i-\sum_{i=1}^{N}x_i\sum_{i=1}^{N}y_i}{\sum_{i=1}^{N}x_i\left(N-\sum_{i=1}^{N}x_i\right)}\right)}$$

אם נציג את הנתונים בטבלה 2×2 עם $n_{ij}=\sum_k I\{x_k=i,y_k=j\},\ i,j\in\{0,1\}$ אז נקבל:

$$\hat{\beta}_0=\log\left(\frac{n_{01}}{n_{00}}\right)$$

ובאופן דומה

$$\hat{\beta}_1=\log\left(\frac{n_{00}n_{11}}{n_{01}n_{10}}\right)$$

וזהו הביטוי הקלאסי עבור לוג מנת יחס הסיכויים בטבלת 2×2.