Chapter 6: Homework

711378912 蔡宜誼

2025/04/28

作業：Seat-Belt Problem

實驗

Following is a sample of automobile accidents. The response categories are (1) not injured, (2) injured but not transported by emergency medical services, (3) injured and transported by emergency medical services but not hospitalized, (4) injured and hospitalized but did not die, (5) injured and died. Analyzed the data. Prepare a short report, summarizing your finding. Let SOI be the shorthand of severity of injury.

Gender	Location	Seat-Belt	SOI1	SOI2	SOI3	SOI4	SOI5
Female	Urban	No	7287	175	720	91	10
		Yes	11587	126	577	48	8
	Rural	No	3246	73	710	159	31
		Yes	6134	94	564	82	17
Male	Urban	No	10381	136	566	96	14
		Yes	10969	83	259	37	1
	Rural	No	6123	141	710	188	45
		Yes	6693	74	353	74	12

核心問題

1. 我們關心的主要問題是Seat-Belt到底有沒有起到降低車禍傷害程度的作用？
2. 除了有無使用安全帶，是否還有其它影響結果的變量？Gender & Location

Note: 探索性實驗，Seat-Belt 必須留在最後的模型裡，而 Gender 及 Location 則不一定

資料

```
gender <- rep(c("Female", "male"), each=4)
Location <- rep(rep(c("Urban", "Rural"), 2), each=2)
SB <- rep(c("No", "Yes"), 4)
SOI1 <- c(7287,11587,3246,6134,10381,10969,6123,6693)
SOI2 <- c(175,126,73,94,136,83,141,74)
SOI3 <- c(720,577,710,564,566,259,710,353)
SOI4 <- c(91,48,159,82,96,37,188,74)
SOI5 <- c(10,8,31,17,14,1,45,12)
DATA <- data.frame(gender, Location, SB, SOI1, SOI2, SOI3, SOI4, SOI5)
DATA
```

```
##   gender Location  SB  SOI1 SOI2 SOI3 SOI4 SOI5
## 1 Female    Urban  No  7287  175  720   91   10
## 2 Female    Urban Yes 11587  126  577   48    8
## 3 Female    Rural  No  3246   73  710  159   31
## 4 Female    Rural Yes  6134   94  564   82   17
## 5   male    Urban  No 10381  136  566   96   14
## 6   male    Urban Yes 10969   83  259   37    1
## 7   male    Rural  No  6123  141  710  188   45
## 8   male    Rural Yes  6693   74  353   74   12
```

```r
# 載入需要的庫
library(ggplot2)
library(tidyr)
library(dplyr)
```
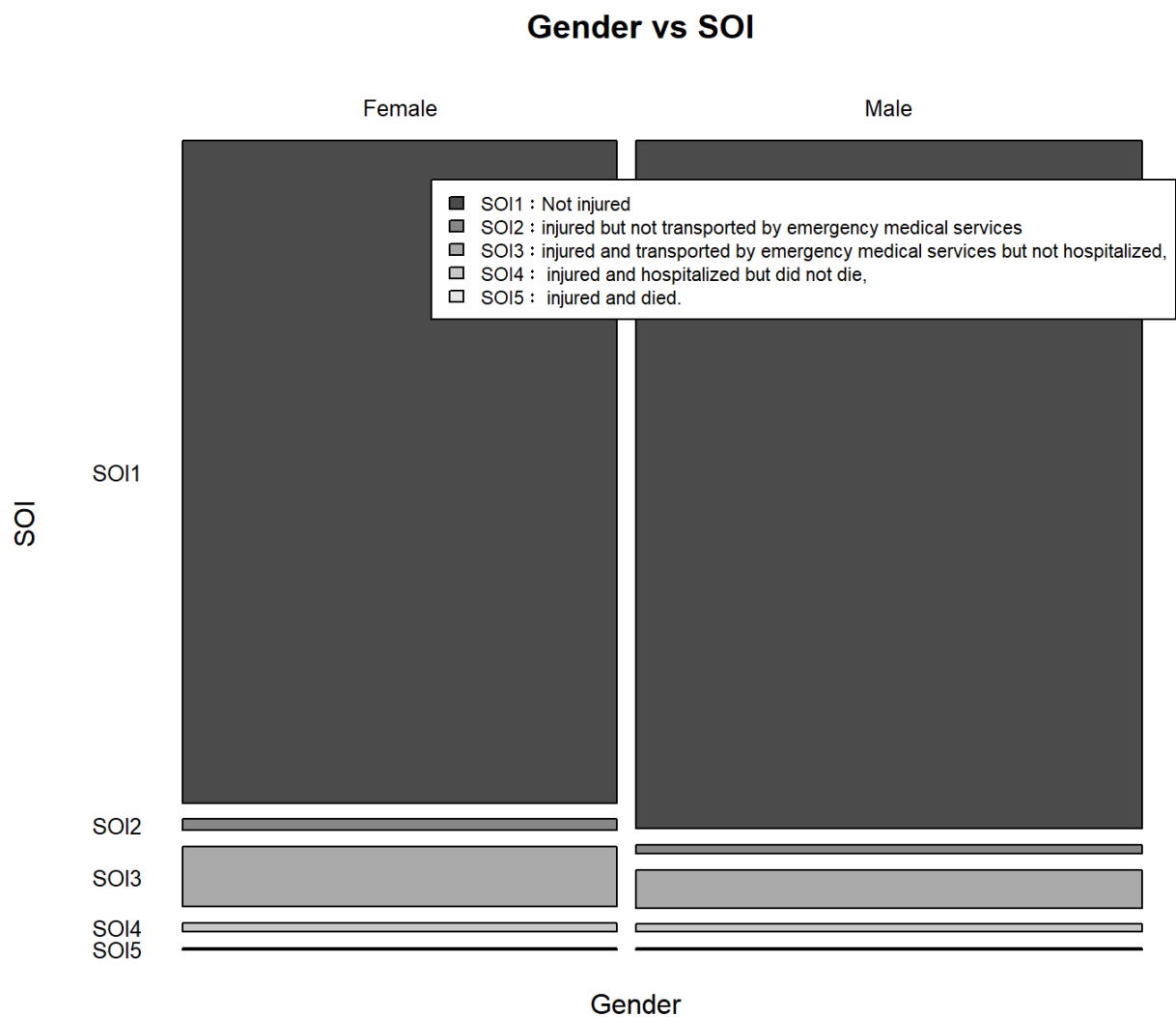
資料分析

1. 繪圖呈現資料

Gender vs SOI

```r
# Gender vs SOI
Gender_SOI <- rbind(
  Female = colSums(DATA[DATA$gender == "Female", c("SOI1","SOI2","SOI3","SOI4","SOI5")]),
  Male   = colSums(DATA[DATA$gender == "male", c("SOI1","SOI2","SOI3","SOI4","SOI5")])
)
head(Gender_SOI)
```

```
##         SOI1 SOI2 SOI3 SOI4 SOI5
## Female 28254  468 2571  380   66
## Male   34166  434 1888  395   72
```

```r
# 繪製馬賽克圖
mosaicplot(Gender_SOI,
           main = " Gender vs SOI",
           xlab = "Gender",
           ylab = "SOI",
           color = TRUE,
           las = 1,
           cex.axis = 0.8)
# 手動加入圖示 (Legend)
legend("topright", inset=c(0.02, 0.1), fill=gray.colors(5),
       legend=c("SOI1：Not injured",
                "SOI2：injured but not transported by emergency medical services",
                "SOI3：injured and transported by emergency medical services but not hospital
ized,",
                "SOI4： injured and hospitalized but did not die,",
                "SOI5： injured and died. "),
       cex=0.7, bg="white", box.col="black")
```

Gender vs SOI

這張馬賽克圖看起來稍微可以水平的切開，有機會是顯著的。

```
chisq_result <- chisq.test(Gender_SOI)
chisq_result
```

```
##
##  Pearson's Chi-squared test
##
## data:  Gender_SOI
## X-squared = 271.91, df = 4, p-value < 2.2e-16
```

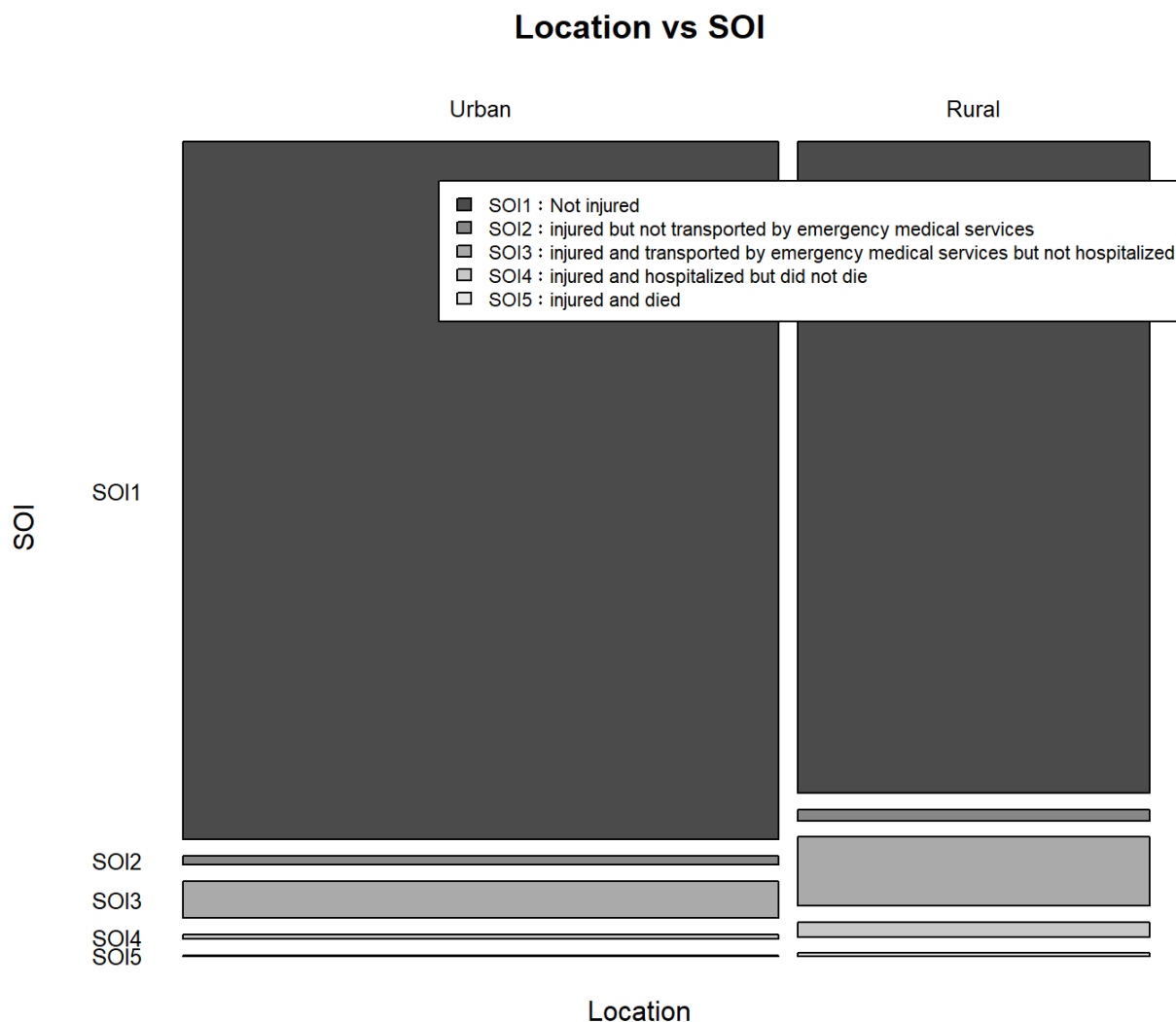這個結果顯示沒有足夠證據證明 性別（Gender） 和 傷害程度（SOI） 之間不存在有顯著的統計關聯，因為 p-value 小於0.05。

Location vs SOI

```
# Location vs SOI
Location_SOI <- rbind(
  Urban = colSums(DATA[DATA$Location == "Urban", c("SOI1", "SOI2", "SOI3", "SOI4", "SOI5")]),
  Rural = colSums(DATA[DATA$Location == "Rural", c("SOI1", "SOI2", "SOI3", "SOI4", "SOI5")])
)
head(Location_SOI)
```

```
##        SOI1 SOI2 SOI3 SOI4 SOI5
## Urban 40224  520 2122  272   33
## Rural 22196  382 2337  503  105
```

```
# 繪製馬賽克圖
mosaicplot(Location_SOI,
           main = "Location vs SOI",
           xlab = "Location",
           ylab = "SOI",
           color = TRUE,
           las = 1,
           cex.axis = 0.8)

# 手動加入圖示 (Legend)
legend("topright", inset=c(0.02, 0.1), fill=gray.colors(5),
       legend=c("SOI1：Not injured",
                "SOI2：injured but not transported by emergency medical services",
                "SOI3：injured and transported by emergency medical services but not hospital
ized",
                "SOI4：injured and hospitalized but did not die",
                "SOI5：injured and died"),
       cex=0.7, bg="white", box.col="black")
```

Location vs SOI

這張馬賽克圖看起來稍微可以水平的切開，有機會是顯著的。

```
# 進行卡方檢定
chisq_result <- chisq.test(Location_SOI)
chisq_result
```

```
##
##   Pearson's Chi-squared test
##
## data:  Location_SOI
## X-squared = 868.1, df = 4, p-value < 2.2e-16
```

這個結果顯示沒有足夠證據證明 地點（Location） 和 傷害程度（SOI） 之間不存在有顯著的統計關聯，因為 p-value 小於0.05。

- -
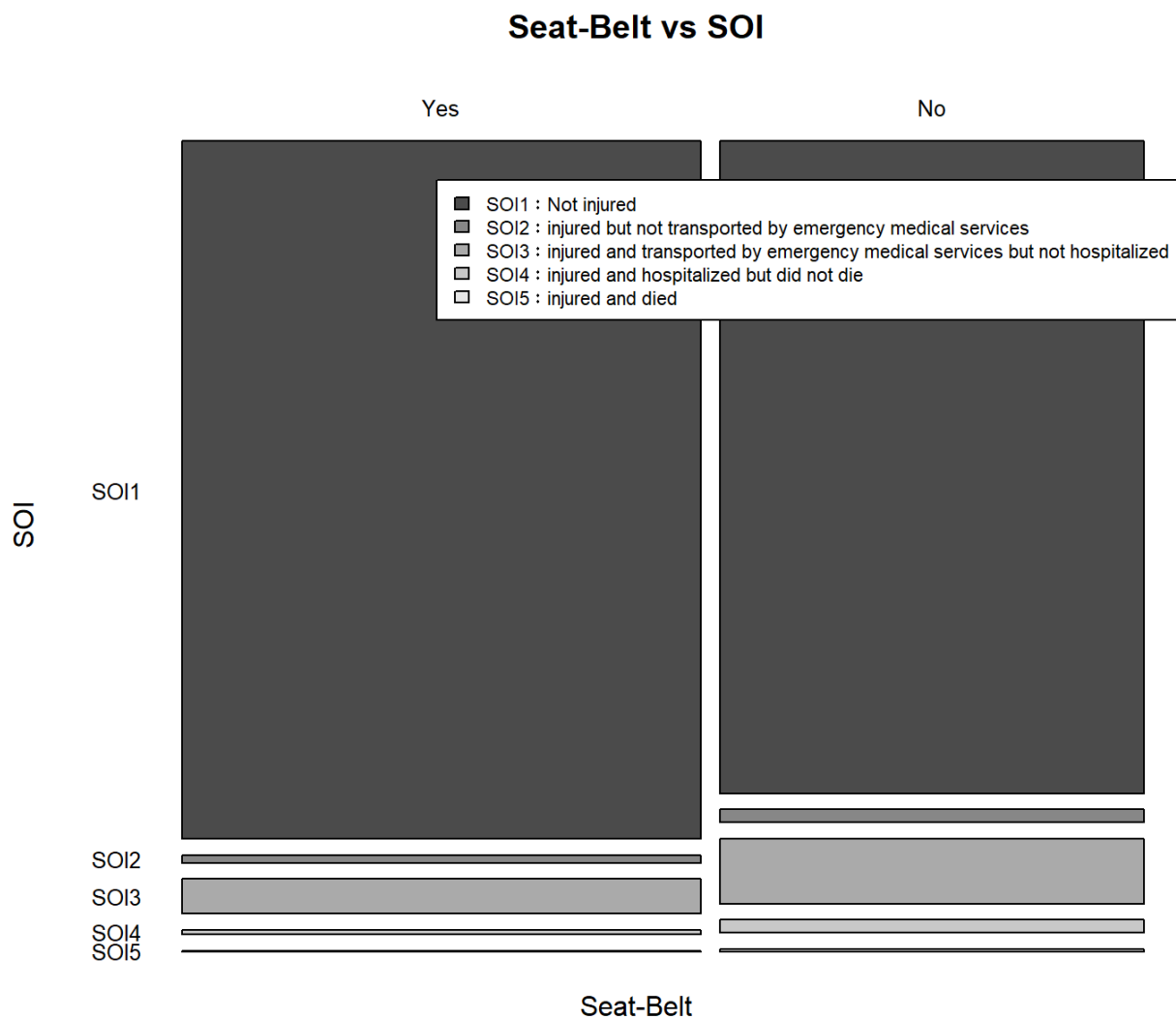
Seat-Belt vs SOI

```
# Seat-Belt vs SOI
SB_SOI <- rbind(
  Yes = colSums(DATA[DATA$SB == "Yes", c("SOI1", "SOI2", "SOI3", "SOI4", "SOI5")]),
  No = colSums(DATA[DATA$SB == "No", c("SOI1", "SOI2", "SOI3", "SOI4", "SOI5")])
)
head(SB_SOI)
```

```
##      SOI1 SOI2 SOI3 SOI4 SOI5
## Yes 35383  377 1753  241   38
## No  27037  525 2706  534  100
```

```
# 繪製馬賽克圖
mosaicplot(SB_SOI,
           main = "Seat-Belt vs SOI",
           xlab = "Seat-Belt",
           ylab = "SOI",
           color = TRUE,
           las = 1,
           cex.axis = 0.8)

# 手動加入圖示 (Legend)
legend("topright", inset=c(0.02, 0.1), fill=gray.colors(5),
       legend=c("SOI1：Not injured",
                "SOI2：injured but not transported by emergency medical services",
                "SOI3：injured and transported by emergency medical services but not hospital
ized",
                "SOI4：injured and hospitalized but did not die",
                "SOI5：injured and died"),
       cex=0.7, bg="white", box.col="black")
```

Seat-Belt vs SOI

這張馬賽克圖看起來稍微可以水平的切開，有機會是顯著的。

```
chisq_result <- chisq.test(SB_SOI)
chisq_result
```

```
##
##  Pearson's Chi-squared test
##
## data:  SB_SOI
## X-squared = 799.49, df = 4, p-value < 2.2e-16
```

這個結果顯示沒有足夠證據證明 是否繫安全帶（Seat-Belt） 和 傷害程度（SOI） 之間不存在有顯著的統計關聯，因為 p-value 小於0.05。

2. 模型建立與統計分析

```r
library(VGAM)
# 建立多類別邏輯回歸模型
# 使用 Parallel Logit Model
fit_full_par <- vglm(cbind(SOI1, SOI2, SOI3, SOI4, SOI5) ~ gender * Location * SB,   # 星星表
示把交互作用一起放進去
                      family = cumulative(parallel = TRUE),
                      data = DATA)
summary(fit_full_par)
```

```
##
## Call:
## vglm(formula = cbind(SOI1, SOI2, SOI3, SOI4, SOI5) ~ gender *
##     Location * SB, family = cumulative(parallel = TRUE), data = DATA)
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept):1                 1.18293    0.03610  32.764   <2e-16 ***
## (Intercept):2                 1.35734    0.03637  37.323   <2e-16 ***
## (Intercept):3                 3.22477    0.04674  68.994   <2e-16 ***
## (Intercept):4                 5.13170    0.09125  56.240   <2e-16 ***
## gendermale                    0.53683    0.04870  11.023   <2e-16 ***
## LocationUrban                 0.81606    0.04946  16.499   <2e-16 ***
## SBYes                         0.90840    0.05273  17.228   <2e-16 ***
## gendermale:LocationUrban      0.01381    0.06960   0.198   0.8427
## gendermale:SBYes             -0.06192    0.07709  -0.803   0.4219
## LocationUrban:SBYes          -0.17743    0.07302  -2.430   0.0151 *
## gendermale:LocationUrban:SBYes 0.14690   0.11204   1.311   0.1898
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2]),
## logitlink(P[Y<=3]), logitlink(P[Y<=4])
##
## Residual deviance: 151.0362 on 21 degrees of freedom
##
## Log-likelihood: -175.9734 on 21 degrees of freedom
##
## Number of Fisher scoring iterations: 4
##
## Warning: Hauck-Donner effect detected in the following estimate(s):
## '(Intercept):4'
##
##
## Exponentiated coefficients:
##                 gendermale              LocationUrban
##                  1.7105807                  2.2615660
##                      SBYes   gendermale:LocationUrban
##                  2.4803574                  1.0139083
##            gendermale:SBYes        LocationUrban:SBYes
##                  0.9399615                  0.8374172
## gendermale:LocationUrban:SBYes
##                  1.1582335
```

含交互作用模型 [性別、地點與是否佩戴安全帶對傷害程度的影響]：

1. 性別、地點 和 是否佩戴安全帶 都顯示對 傷害程度（SOI） 有顯著影響。

2. 性別與地點的交互作用 顯示顯著（p-value = 0.0151），應該保留在模型中。

3. 其他的交互作用項（如 gender:LocationUrban:SByes）則不顯著，可以考慮移除。

4. Hauck-Donner effect 顯示模型中某些類別的估計可能存在問題，需進一步檢查。

總體來說，這個模型顯示了性別、地點和是否佩戴安全帶對傷害程度的顯著影響，並且性別與地點的交互作用也應納入分析。**(這裡回答了第二個問題~)**

```
# 建立多類別邏輯回歸模型
# 非平行邏輯回歸模型
fit_full_nonpar <- vglm(cbind(SOI1, SOI2, SOI3, SOI4, SOI5) ~ gender * Location * SB,    # 星
星表示把交互作用一起放進去
                        family = cumulative(parallel = FALSE),    # Parallel = true 斜率共用
                        data = DATA)
```

```
## Warning in vglm.fitter(x = x, y = y, w = w, offset = offset, Xm2 = Xm2, :
## iterations terminated because half-step sizes are very small
```

```
## Warning in vglm.fitter(x = x, y = y, w = w, offset = offset, Xm2 = Xm2, : some
## quantities such as z, residuals, SEs may be inaccurate due to convergence at a
## half-step
```

```
summary(fit_full_nonpar)
```

```
## 
## Call:
## vglm(formula = cbind(SOI1, SOI2, SOI3, SOI4, SOI5) ~ gender *
##     Location * SB, family = cumulative(parallel = FALSE), data = DATA)
## 
## Coefficients:
##                                     Estimate Std. Error z value Pr(>|z|)
## (Intercept):1                        1.20479    0.03655  32.964  < 2e-16 ***
## (Intercept):2                        1.30502    0.03758  34.725  < 2e-16 ***
## (Intercept):3                        3.05425    0.07424  41.141  < 2e-16 ***
## (Intercept):4                        4.90599    0.18027  27.215  < 2e-16 ***
## gendermale:1                         0.52660    0.04921  10.701  < 2e-16 ***
## gendermale:2                         0.58848    0.05131  11.470  < 2e-16 ***
## gendermale:3                         0.34466    0.09973   3.456 0.000549 ***
## gendermale:4                         0.16389    0.23422   0.700 0.484095
## LocationUrban:1                      0.78531    0.04977  15.779  < 2e-16 ***
## LocationUrban:2                      0.90203    0.05258  17.156  < 2e-16 ***
## LocationUrban:3                      1.34032    0.12464  10.754  < 2e-16 ***
## LocationUrban:4                      1.81218    0.36417   4.976 6.48e-07 ***
## SBYes:1                              0.88744    0.05310  16.712  < 2e-16 ***
## SBYes:2                              0.93501    0.05551  16.844  < 2e-16 ***
## SBYes:3                              1.17413    0.12554   9.353  < 2e-16 ***
## SBYes:4                              1.09630    0.30243   3.625 0.000289 ***
## gendermale:LocationUrban:1           0.03153    0.06993   0.451 0.652068
## gendermale:LocationUrban:2          -0.05098    0.07456  -0.684 0.494088
## gendermale:LocationUrban:3          -0.12654    0.17074  -0.741 0.458606
## gendermale:LocationUrban:4          -0.19932    0.47592  -0.419 0.675349
## gendermale:SBYes:1                  -0.05030    0.07749  -0.649 0.516278
## gendermale:SBYes:2                  -0.09321    0.08202  -1.136 0.255790
## gendermale:SBYes:3                  -0.15672    0.17878  -0.877 0.380702
## gendermale:SBYes:4                   0.22992    0.44418   0.518 0.604726
## LocationUrban:SBYes:1               -0.15191    0.07325  -2.074 0.038085 *
## LocationUrban:SBYes:2               -0.22408    0.07809  -2.869 0.004112 **
## LocationUrban:SBYes:3               -0.17751    0.20910  -0.849 0.395907
## LocationUrban:SBYes:4               -0.47347    0.56273  -0.841 0.400140
## gendermale:LocationUrban:SBYes:1     0.12918    0.11225   1.151 0.249788
## gendermale:LocationUrban:SBYes:2     0.25436    0.12156   2.092 0.036395 *
## gendermale:LocationUrban:SBYes:3     0.24336    0.30905   0.787 0.431015
## gendermale:LocationUrban:SBYes:4     1.80131    1.22234   1.474 0.140572
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2]),
## logitlink(P[Y<=3]), logitlink(P[Y<=4])
## 
## Residual deviance: -1.742e-13 on 0 degrees of freedom
## 
## Log-likelihood: -100.4553 on 0 degrees of freedom
## 
## Number of Fisher scoring iterations: 5
## 
## Warning: Hauck-Donner effect detected in the following estimate(s):
## '(Intercept):3', '(Intercept):4'
## 
## 
```

```
## Exponentiated coefficients:
##                   gendermale:1                       gendermale:2
##                      1.6931651                          1.8012555
##                   gendermale:3                       gendermale:4
##                      1.4115048                          1.1780855
##                  LocationUrban:1                     LocationUrban:2
##                      2.1930764                          2.4646051
##                  LocationUrban:3                     LocationUrban:4
##                      3.8202733                          6.1237584
##                         SBYes:1                            SBYes:2
##                      2.4289144                          2.5472427
##                         SBYes:3                            SBYes:4
##                      3.2353317                          2.9930614
##      gendermale:LocationUrban:1        gendermale:LocationUrban:2
##                      1.0320350                          0.9502935
##      gendermale:LocationUrban:3        gendermale:LocationUrban:4
##                      0.8811385                          0.8192851
##             gendermale:SBYes:1               gendermale:SBYes:2
##                      0.9509469                          0.9110061
##             gendermale:SBYes:3               gendermale:SBYes:4
##                      0.8549421                          1.2584958
##          LocationUrban:SBYes:1            LocationUrban:SBYes:2
##                      0.8590681                          0.7992488
##          LocationUrban:SBYes:3            LocationUrban:SBYes:4
##                      0.8373498                          0.6228395
## gendermale:LocationUrban:SBYes:1 gendermale:LocationUrban:SBYes:2
##                      1.1379003                          1.2896315
## gendermale:LocationUrban:SBYes:3 gendermale:LocationUrban:SBYes:4
##                      1.2755289                          6.0576042
```

```
# 建立不包含交互作用項的模型
fit_red_par <- vglm(cbind(SOI1, SOI2, SOI3, SOI4, SOI5) ~ gender + Location + SB,
                family = cumulative(parallel = TRUE),
                data = DATA)
summary(fit_red_par)
```

```
## 
## Call:
## vglm(formula = cbind(SOI1, SOI2, SOI3, SOI4, SOI5) ~ gender +
##     Location + SB, family = cumulative(parallel = TRUE), data = DATA)
## 
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept):1  1.20337    0.02636   45.66   <2e-16 ***
## (Intercept):2  1.37773    0.02675   51.50   <2e-16 ***
## (Intercept):3  3.24477    0.03990   81.32   <2e-16 ***
## (Intercept):4  5.15156    0.08796   58.56   <2e-16 ***
## gendermale     0.54489    0.02719   20.04   <2e-16 ***
## LocationUrban  0.77321    0.02691   28.74   <2e-16 ***
## SBYes          0.82405    0.02760   29.86   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2]),
## logitlink(P[Y<=3]), logitlink(P[Y<=4])
## 
## Residual deviance: 159.6062 on 25 degrees of freedom
## 
## Log-likelihood: -180.2584 on 25 degrees of freedom
## 
## Number of Fisher scoring iterations: 4
## 
## Warning: Hauck-Donner effect detected in the following estimate(s):
## '(Intercept):3', '(Intercept):4'
## 
## 
## Exponentiated coefficients:
##    gendermale LocationUrban         SBYes
##      1.724423      2.166704      2.279721
```

無交互作用模型 [性別、地點（Urban/Rural）和 是否佩戴安全帶（SB）對 傷害程度（SOI） 的影響]：

1. 性別、地點 和 是否佩戴安全帶 對 傷害程度（SOI）都有顯著影響（p-value < 0.05）。

   ○ 性別（gendermale）顯示出顯著性，顯示男性與女性在傷害程度上的分佈有差異。

   ○ 地點（LocationUrban）顯示出顯著性，顯示城市與鄉村的傷害程度有差異。

   ○ 是否佩戴安全帶（SBYes）也顯示出顯著性，表明佩戴安全帶對傷害程度有顯著影響。

2. 交互作用：結果中未顯示交互作用項（如 gender:LocationUrban）有顯著性，因此可以認為這些變數的影響是獨立的。

3. Hauck-Donner效應：模型中警告顯示 (Intercept):3 和 (Intercept):4 存在 Hauck-Donner效應，這通常指的是這些參數的估計可能不穩定，可能需要進一步檢查。

Discuss that which model should be applied (use parallel model?)

```
# 進行偏差檢驗
LRT <- deviance(fit_full_nonpar) - deviance(fit_full_par)
pvalue <- 1 - pchisq(LRT, df = 1)
cat("比較是否要使用Parallel Logit Model: LRT =", LRT, ", p-value =", pvalue, "\n")
```

```
## 比較是否要使用Parallel Logit Model: LRT = -151.0362 , p-value = 1
```

從這個結果來看，LRT = -151.0362 和 p-value = 1 表示 parallel = TRUE 和 parallel = FALSE 兩個模型之間沒有顯著區別。這意味著 Proportional Odds 成立，因此 Parallel Logit Model 是合適的。

- -

Determine whether gender and location should be involved in the model

是的，性別 和 地點 應該納入模型中，因為這些變數對 SOI 顯示出顯著影響（p-value 小於 0.05）。

- -

Any interaction?

```
# 進行卡方檢定，檢查交互作用項是否顯著
LRT <- deviance(fit_red_par) - deviance(fit_full_par)
pvalue <- 1 - pchisq(LRT, df = 1)

cat("比較是否使用有交互作用模型: LRT =", LRT, ", p-value =", pvalue, "\n")
```

```
## 比較是否使用有交互作用模型: LRT = 8.569985 , p-value = 0.003417499
```

這表示在比較 有交互作用項模型 和 無交互作用項模型 時，卡方檢定的 p-value 小於 0.05，這意味著 交互作用項是顯著的。 結論： 由於 p-value = 0.0034 小於 0.05，我們可以 拒絕零假設（即交互作用項無顯著性），並得出結論：交互作用項對模型有顯著影響，因此應該保留交互作用項。

3. 結論

Report your final model and explain the meaning of parameters.

最終選擇的模型是 Cumulative Logit Model，並使用 Parallel Logit Model 假設，因為 SOI 是順序型資料，且 Proportional Odds成立。

1. 模型解釋：

   - 性別（gender）：顯示出顯著性（p-value < 0.05），表示性別對 傷害程度（SOI） 有顯著影響，男性和女性在不同傷害程度的分佈上有所區別。

   - 地點（Location）：也顯示顯著性（p-value < 0.05），顯示城市與鄉村的傷害程度有顯著差異。

   - 是否佩戴安全帶（SB）：顯示顯著性（p-value < 0.05），表明佩戴安全帶與傷害程度之間有顯著關聯。

2. 交互作用：

性別與地點的交互作用（gender:Location） 顯示顯著（p-value = 0.0151），表明性別和地點之間的交互作用對 SOI 產生影響。這意味著不同性別在不同地點的傷害程度有所不同。

3. 模型的參數解釋：

- 截距項（Intercept）：對每一類別（SOI1、SOI2、SOI3、SOI4）的基準概率進行估計。

- 性別（gender）：男性相對於女性的傷害程度變化。gendermale 的正值意味著男性的某些傷害程度的概率較高。

- 地點（Location）：城市地區與鄉村地區在傷害程度上的差異。LocationUrban 的正值表示城市地區的傷害程度較高。

- 是否佩戴安全帶（SB）：佩戴安全帶對傷害程度的影響。SBYes 的正值表示佩戴安全帶有助於減少傷害。

4. 模型檢驗：

卡方檢定結果顯示 交互作用項 顯示顯著（p-value = 0.0034），因此 可以拒絕H0（即交互作用項無顯著性）。這進一步支持了保留交互作用項的結論，也證明了 Parallel Logit Model 是合適的。