ASSIGNMENT–2
INJURY SEVERITY MODELLING

NAME: ANSHUMAN KUMAR YADAV
ROLL NO.: 241030018

Ques 1: Identify at least 3 additional explanatory variables present in the database that were not included in the initial model that in your opinion can influence the injury severity of motorcyclists. Discuss your a priori expectations.

Ans 1: Three additional explanatory variables present in the database that were not included in the initial model are as follows:

i. **X_LIGHT – Lighting Condition**
ii. **X_SUR_COND – Surface Condition**
iii. **X_CRASH_TYPE – Collision Type**

X_LIGHT: Represents the lighting condition at the time of the crash (e.g., daylight, dark with/without street lights).
Priori Expectation:
→ Crashes occurring in poor lighting conditions (e.g., dark with no street lights) are expected to result in higher injury severity due to reduced visibility.
→ Therefore, the coefficient for poor lighting conditions to be **positive**, indicating a greater likelihood of more severe injuries.

X_SUR_COND: Indicates the road surface condition at the crash site (e.g., dry, wet, icy, snowy).
Priori Expectation:
→ Slippery surfaces such as **wet or icy roads** increase the risk of losing control, which can lead to **more severe injuries**.
→ Hence, the coefficient for adverse surface conditions to be **positive**, showing a higher probability of severe outcomes.

X_CRASH_TYPE: Represents type of collision involved in the crash (e.g., rear-end, angle, head-on, sideswipe).
Priori Expectation:
→ Crashes like **head-on** or **angle** are typically more dangerous due to the high impact forces and the direction of collision. **Sideswipes** or **rear-end** crashes generally involve lower impact and may result in less severe injuries.
→ Severe crash types (e.g., head-on, angle) are expected to be **positively associated** with higher injury severity.

Ques 2: Include the variables in the ordered response model to assess their influence on crash outcomes and evaluate the statistical significance of the coefficients and assess whether the estimates match your a priori expectations.

Ans 2:

Code for including 3 variables is as follows:

```
### Set working directory (adjust to your location)
setwd("/Users/anshukryadav/Downloads/CE687_Lab_Ordered_Response_52bc21f3-b49e-4afc-9c77-09d21dac5835")
library(tidyverse)
library(MASS)
### Load the dataset
dat <- read.csv("Michigan_Motorcycle_Non_Intersection_Data_Subset.csv", stringsAsFactors = TRUE)
### Create ordered response variable (Injury Severity)
dat$Injury_Severity <- 0
dat$Injury_Severity[dat$Worst.Injury.in.Crash == "Possible Injury (C)"] <- 1
dat$Injury_Severity[dat$Worst.Injury.in.Crash == "Suspected Minor Injury (B)"] <- 2
dat$Injury_Severity[dat$Worst.Injury.in.Crash == "Suspected Serious Injury (A)"] <- 3
dat$Injury_Severity[dat$Worst.Injury.in.Crash == "Fatal Injury (K)"] <- 4
dat$Injury_Severity <- as.factor(dat$Injury_Severity)
### Explanatory variables
dat$Urban <- as.numeric(dat$Rural.Urban.Area == "Urban")
dat$Pedestrian <- as.numeric(dat$Crash..Pedestrian == "Pedestrian Involved")
dat$Late_Night <- as.numeric(dat$Time.of.Day %in% c("12:00 Midnight - 12:59 AM",
                                                    "1:00 AM - 1:59 AM",
                                                    "2:00 AM - 2:59 AM",
                                                    "3:00 AM - 3:59 AM",
                                                    "4:00 AM - 4:59 AM"))
dat$Parked_Vehicle <- as.numeric(dat$Crash..Lane.Departure == "Parked Vehicle")
### Additional categorical variables
dat$X_LIGHT <- factor(dat$Lighting.Conditions)
dat$X_SUR_COND <- factor(dat$Road.Conditions)
dat$X_CRASH_TYPE <- factor(dat$Crash.Type)
### ----- Ordered Probit Model -----
m1_probit_updated <- polr(Injury_Severity ~ Speed.Limit.at.Crash.Site + Urban +
                            Pedestrian + Parked_Vehicle + Late_Night +
                            X_LIGHT + X_SUR_COND + X_CRASH_TYPE,
                          data = dat, method = "probit")
cat("----- Ordered Probit Model Summary -----\n")
summary(m1_probit_updated)
# Log-Likelihood, AIC, BIC
ll_probit <- logLik(m1_probit_updated)
aic_probit <- AIC(m1_probit_updated)
bic_probit <- BIC(m1_probit_updated)
# Null model for pseudo-R²
null_model_probit <- polr(Injury_Severity ~ 1, data = dat, method = "probit")
ll_null_probit <- logLik(null_model_probit)

# McFadden's Pseudo-R²
pseudoR2_probit <- 1 - as.numeric(ll_probit) / as.numeric(ll_null_probit)
```

```r
cat("Log-Likelihood:", ll_probit, "\n")
cat("AIC:", aic_probit, "\n")
cat("BIC:", bic_probit, "\n")
cat("McFadden's Pseudo-R²:", pseudoR2_probit, "\n\n")
### ----- Ordered Logit Model -----
m1_logit_updated <- polr(Injury_Severity ~ Speed.Limit.at.Crash.Site + Urban +
                           Pedestrian + Parked_Vehicle + Late_Night +
                           X_LIGHT + X_SUR_COND + X_CRASH_TYPE,
                         data = dat, method = "logistic")
cat("----- Ordered Logit Model Summary -----\n")
summary(m1_logit_updated)
# Log-Likelihood, AIC, BIC
ll_logit <- logLik(m1_logit_updated)
aic_logit <- AIC(m1_logit_updated)
bic_logit <- BIC(m1_logit_updated)
# Null model for pseudo-R²
null_model_logit <- polr(Injury_Severity ~ 1, data = dat, method = "logistic")
ll_null_logit <- logLik(null_model_logit)
# McFadden's Pseudo-R²
pseudoR2_logit <- 1 - as.numeric(ll_logit) / as.numeric(ll_null_logit)
cat("Log-Likelihood:", ll_logit, "\n")
cat("AIC:", aic_logit, "\n")
cat("BIC:", bic_logit, "\n")
cat("McFadden's Pseudo-R²:", pseudoR2_logit, "\n")
```

Output:

```
> summary(m1_probit_updated)

Re-fitting to get Hessian

Call:
polr(formula = Injury_Severity ~ Speed.Limit.at.Crash.Site +
    Urban + Pedestrian + Parked_Vehicle + Late_Night + X_LIGHT +
    X_SUR_COND + X_CRASH_TYPE, data = dat, method = "probit")

Coefficients:
                                                    Value Std. Error  t value
Speed.Limit.at.Crash.Site                        0.003416   0.001037  3.29345
Urban                                           -0.118851   0.028437 -4.17939
Pedestrian                                       1.069433   0.199284  5.36638
Parked_Vehicle                                  -0.817899   0.166915 -4.90010
Late_Night                                       0.199084   0.053529  3.71917
X_LIGHTDark - Unlighted                         -0.133482   0.053038 -2.51670
X_LIGHTDawn                                     -0.343404   0.098915 -3.47170
X_LIGHTDaylight                                 -0.120735   0.046100 -2.61897
X_LIGHTDusk                                     -0.058394   0.078139 -0.74731
X_LIGHTOther                                    -0.429296   0.718013 -0.59789
X_LIGHTUnknown                                  -0.998031   0.369969 -2.69761
X_SUR_CONDDry                                    0.246609   0.268572  0.91822
X_SUR_CONDIce                                   -0.434737   0.454639 -0.95622
X_SUR_CONDMud, Dirt, Gravel                      0.144776   0.295514  0.48991
X_SUR_CONDOily                                  -0.171234   0.351402 -0.48729
X_SUR_CONDOther                                 -0.061104   0.356266 -0.17151
X_SUR_CONDSand                                  -0.363793   0.559456 -0.65026
X_SUR_CONDSlush                                  0.241764   1.070657  0.22581
X_SUR_CONDSnow                                   0.727725   0.447426  1.62647
X_SUR_CONDUnknown                               -0.131519   0.346667 -0.37938
X_SUR_CONDWater (standing/moving)               -0.868386   0.665426 -1.30501
X_SUR_CONDWet                                    0.010874   0.274538  0.03961
X_CRASH_TYPEBacking                             -1.669293   0.241599 -6.90936
X_CRASH_TYPEHead-On                              0.537244   0.114584  4.68864
X_CRASH_TYPEHead-On - Left Turn                  0.360932   0.096132  3.75453
X_CRASH_TYPEOther                               -0.273269   0.074948 -3.64610
X_CRASH_TYPERear-End                            -0.336844   0.058463 -5.76161
X_CRASH_TYPERear-End - Left Turn                -0.164764   0.131367 -1.25422
X_CRASH_TYPERear-End - Right Turn               -0.236325   0.226849 -1.04177
X_CRASH_TYPESideswipe - Opposite Directions     -0.387167   0.122412 -3.16282
X_CRASH_TYPESideswipe - Same Direction          -0.501122   0.065590 -7.64025
X_CRASH_TYPESingle Motor Vehicle                -0.092360   0.051986 -1.77661
X_CRASH_TYPEUnknown                             -0.321315   0.207892 -1.54559
```

```
Intercepts:
    Value   Std. Error t value
0|1 -0.7412  0.2812    -2.6356
1|2 -0.2449  0.2812    -0.8709
2|3  0.6784  0.2812     2.4127
3|4  1.8142  0.2819     6.4355

Residual Deviance: 20788.86
AIC: 20862.86
> cat("Log-Likelihood:", ll_probit, "\n")
Log-Likelihood: -10394.43
> cat("AIC:", aic_probit, "\n")
AIC: 20862.86
> cat("BIC:", bic_probit, "\n")
BIC: 21117.57
> cat("McFadden's Pseudo-R²:", pseudoR2_probit, "\n\n")
McFadden's Pseudo-R²: 0.02225251
> summary(m1_logit_updated)

Re-fitting to get Hessian

Call:
polr(formula = Injury_Severity ~ Speed.Limit.at.Crash.Site +
    Urban + Pedestrian + Parked_Vehicle + Late_Night + X_LIGHT +
    X_SUR_COND + X_CRASH_TYPE, data = dat, method = "logistic")

Coefficients:
                                                    Value Std. Error t value
Speed.Limit.at.Crash.Site                        0.005557   0.001755  3.1673
Urban                                           -0.201761   0.048297 -4.1776
Pedestrian                                       1.849635   0.331593  5.5780
Parked_Vehicle                                  -1.515431   0.301888 -5.0198
Late_Night                                       0.344105   0.093956  3.6624
X_LIGHTDark - Unlighted                         -0.233206   0.092327 -2.5259
X_LIGHTDawn                                     -0.576179   0.167584 -3.4381
X_LIGHTDaylight                                 -0.178533   0.079498 -2.2457
X_LIGHTDusk                                     -0.065754   0.134233 -0.4898
X_LIGHTOther                                    -0.686293   1.029187 -0.6668
X_LIGHTUnknown                                  -1.718335   0.628429 -2.7343
X_SUR_CONDDry                                    0.306658   0.451734  0.6788
X_SUR_CONDIce                                   -0.928384   0.780995 -1.1887
X_SUR_CONDMud, Dirt, Gravel                      0.134641   0.496036  0.2714
X_SUR_CONDOily                                  -0.344823   0.581354 -0.5931
X_SUR_CONDOther                                 -0.168206   0.598711 -0.2809
X_SUR_CONDSand                                  -0.607853   0.911308 -0.6670
X_SUR_CONDSlush                                  0.280679   1.581179  0.1775
X_SUR_CONDSnow                                   1.012976   0.844661  1.1993
X_SUR_CONDUnknown                               -0.324062   0.582496 -0.5563
X_SUR_CONDWater (standing/moving)               -1.987160   1.283717 -1.5480
X_SUR_CONDWet                                   -0.097424   0.462065 -0.2108
X_CRASH_TYPEBacking                             -2.741179   0.431299 -6.3556
X_CRASH_TYPEHead-On                              0.977720   0.213282  4.5842
X_CRASH_TYPEHead-On - Left Turn                  0.638854   0.167759  3.8082
X_CRASH_TYPEOther                               -0.474701   0.129585 -3.6632
X_CRASH_TYPERear-End                            -0.591941   0.101576 -5.8276
X_CRASH_TYPERear-End - Left Turn                -0.313752   0.220953 -1.4200
X_CRASH_TYPERear-End - Right Turn               -0.397353   0.385993 -1.0294
X_CRASH_TYPESideswipe - Opposite Directions     -0.711984   0.215556 -3.3030
X_CRASH_TYPESideswipe - Same Direction          -0.842500   0.113976 -7.3919
X_CRASH_TYPESingle Motor Vehicle                -0.144553   0.090433 -1.5984
X_CRASH_TYPEUnknown                             -0.611190   0.360087 -1.6973
Intercepts:
    Value   Std. Error t value
0|1 -1.3213  0.4745    -2.7848
1|2 -0.4891  0.4742    -1.0314
2|3  1.0137  0.4742     2.1376
3|4  3.1522  0.4768     6.6107

Residual Deviance: 20787.16
AIC: 20861.16
> cat("Log-Likelihood:", ll_logit, "\n")
Log-Likelihood: -10393.58
> cat("AIC:", aic_logit, "\n")
AIC: 20861.16
> cat("BIC:", bic_logit, "\n")
BIC: 21115.86
> cat("McFadden's Pseudo-R²:", pseudoR2_logit, "\n")
McFadden's Pseudo-R²: 0.02233283
```

X_LIGHT – Lighting Condition

Coefficient Interpretation (Probit vs Logit)

Lighting Condition	Probit Estimate	t-value	Logit Estimate	t-value	Significant?
Dark - Unlighted	-0.1335	-2.52	-0.2332	-2.53	Significant
Dawn	-0.3434	-3.47	-0.5762	-3.44	Significant
Daylight	-0.1207	-2.62	-0.1785	-2.25	Significant
Dusk	-0.0584	-0.75	-0.0658	-0.49	Not Significant
Other	-0.4293	-0.60	-0.6863	-0.67	Not Significant
Unknown	-0.9980	-2.70	-1.7183	-2.73	Significant

Interpretation:
• **Significant negative coefficients** for most lighting conditions, especially *Dawn*, *Dark-Unlighted*, and *Unknown*, suggest **worse lighting increases injury severity**.
• These match **a priori expectations**: poor visibility conditions like darkness or dawn are expected to correlate with more severe outcomes.

X_SUR_COND – Surface Condition

Coefficient Interpretation (Probit vs Logit)

Surface Condition	Probit Estimate	t-value	Logit Estimate	t-value	Significant?
Dry	+0.2466	0.92	+0.3067	0.68	Not Significant
Ice	-0.4347	-0.96	-0.9284	-1.19	Not Significant
Mud, Dirt, Gravel	+0.1448	0.49	+0.1346	0.27	Not Significant
Oily	-0.1712	-0.49	-0.3448	-0.59	Not Significant
Other	-0.0611	-0.17	-0.1682	-0.28	Not Significant
Sand	-0.3638	-0.65	-0.6079	-0.67	Not Significant
Slush	+0.2418	0.23	+0.2807	0.18	Not Significant
Snow	+0.7277	1.63	+1.0130	1.20	Not Significant
Unknown	-0.1315	-0.38	-0.3241	-0.56	Not Significant
Water (standing/moving)	-0.8684	-1.31	-1.9872	-1.55	Not Significant
Wet	+0.0109	0.04	-0.0974	-0.21	Not Significant

Interpretation:
• **None of the surface condition variables are statistically significant**, though **snow and water** show relatively large (but still insignificant) coefficients.
• **A priori expectation**: Slippery or uncertain surfaces (like **ice, snow, water**) would likely increase severity — the signs match that, but lack statistical strength (probably due to sample size or data imbalance).

X_CRASH_TYPE – Collision Type

Coefficient Interpretation (Probit vs Logit)

Crash Type	Probit Estimate	t-value	Logit Estimate	t-value	Significant?
Backing	-1.6693	-6.91	-2.7412	-6.36	Significant
Head-On	+0.5372	4.69	+0.9777	4.58	Significant
Head-On - Left Turn	+0.3609	3.75	+0.6389	3.81	Significant
Other	-0.2733	-3.65	-0.4747	-3.66	Significant
Rear-End	-0.3368	-5.76	-0.5919	-5.83	Significant
Rear-End - Left Turn	-0.1648	-1.25	-0.3138	-1.42	Not Significant
Rear-End - Right Turn	-0.2363	-1.04	-0.3974	-1.03	Not Significant
Sideswipe - Opposite Directions	-0.3872	-3.16	-0.7120	-3.30	Significant
Sideswipe - Same Direction	-0.5011	-7.64	-0.8425	-7.39	Significant
Single Motor Vehicle	-0.0924	-1.78	-0.1446	-1.60	Not Significant
Unknown	-0.3213	-1.55	-0.6112	-1.70	Not Significant

Interpretation:
• **Backing, Head-On, Head-On Left Turn, Sideswipes, Rear-End (main)** are **statistically significant**, and show strong effects.
• **Negative values (like for Backing)** → associated with **lower injury severity** (as expected, since low-speed).
• **Positive values (Head-On)** → associated with **higher injury severity**, which aligns well with **a priori expectations**.
• **Sideswipes and Rear-End** have negative signs, matching the idea they are typically lower-severity crashes.

<u>Ques 3</u>: Compare the model performance when compared to the original model using pseudo-R^2, Likelihood, AIC and BIC.
<u>Ans 3</u>:
Updated code for original model discussed in lab so that we can get pseudo-R^2, Likelihood, AIC and BIC in output itself:

```
### Set working directory
setwd("/Users/anshukryadav/Downloads/CE687_Lab_Ordered_Response_52bc21f3-b49e-4afc-9c77-09d21dac5835")
library(tidyverse)
library(MASS)
###### Load crash data #########
dat       <-       read.csv("Michigan_Motorcycle_Non_Intersection_Data_Subset.csv", stringsAsFactors = TRUE)
### Create ordered response variable: Injury Severity
dat$Injury_Severity <- 0
dat$Injury_Severity[dat$Worst.Injury.in.Crash == "Possible Injury (C)"] <- 1
dat$Injury_Severity[dat$Worst.Injury.in.Crash == "Suspected Minor Injury (B)"] <- 2
```

```r
dat$Injury_Severity[dat$Worst.Injury.in.Crash == "Suspected Serious Injury (A)"] <-
3
dat$Injury_Severity[dat$Worst.Injury.in.Crash == "Fatal Injury (K)"] <- 4
dat$Injury_Severity <- as.factor(dat$Injury_Severity)
### Create explanatory variables
dat$Urban <- as.numeric(dat$Rural.Urban.Area == "Urban")
dat$Pedestrian <- as.numeric(dat$Crash..Pedestrian == "Pedestrian Involved")
dat$Late_Night <- as.numeric(dat$Time.of.Day %in% c("12:00 Midnight - 12:59 AM",
                                                    "1:00 AM - 1:59 AM",
                                                    "2:00 AM - 2:59 AM",
                                                    "3:00 AM - 3:59 AM",
                                                    "4:00 AM - 4:59 AM"))
dat$Parked_Vehicle <- as.numeric(dat$Crash..Lane.Departure == "Parked Vehicle")
### ----- Ordered Probit Model -----
m1_probit <- polr(Injury_Severity ~ Speed.Limit.at.Crash.Site + Urban +
                    Pedestrian + Parked_Vehicle + Late_Night,
                  data = dat, method = "probit")
cat("----- Ordered Probit Model Summary -----\n")
summary(m1_probit)
# Metrics for Probit
ll_probit <- logLik(m1_probit)
aic_probit <- AIC(m1_probit)
bic_probit <- BIC(m1_probit)
null_probit <- polr(Injury_Severity ~ 1, data = dat, method = "probit")
ll_null_probit <- logLik(null_probit)
pseudoR2_probit <- 1 - as.numeric(ll_probit) / as.numeric(ll_null_probit)
cat("Probit Log-Likelihood:", ll_probit, "\n")
cat("Probit AIC:", aic_probit, "\n")
cat("Probit BIC:", bic_probit, "\n")
cat("Probit McFadden's Pseudo-R²:", pseudoR2_probit, "\n\n")
### ----- Ordered Logit Model -----
m1_logit <- polr(Injury_Severity ~ Speed.Limit.at.Crash.Site + Urban +
                   Pedestrian + Parked_Vehicle + Late_Night,
                 data = dat, method = "logistic")
cat("----- Ordered Logit Model Summary -----\n")
summary(m1_logit)
# Metrics for Logit
ll_logit <- logLik(m1_logit)
aic_logit <- AIC(m1_logit)
bic_logit <- BIC(m1_logit)
null_logit <- polr(Injury_Severity ~ 1, data = dat, method = "logistic")
ll_null_logit <- logLik(null_logit)
pseudoR2_logit <- 1 - as.numeric(ll_logit) / as.numeric(ll_null_logit)
cat("Logit Log-Likelihood:", ll_logit, "\n")
cat("Logit AIC:", aic_logit, "\n")
cat("Logit BIC:", bic_logit, "\n")
cat("Logit McFadden's Pseudo-R²:", pseudoR2_logit, "\n\n")
```

Output:

```
> summary(m1_probit)

Re-fitting to get Hessian

Call:
polr(formula = Injury_Severity ~ Speed.Limit.at.Crash.Site +
    Urban + Pedestrian + Parked_Vehicle + Late_Night, data = dat,
    method = "probit")

Coefficients:
                            Value Std. Error t value
Speed.Limit.at.Crash.Site  0.002253    0.00101   2.230
Urban                     -0.151719    0.02662  -5.698
Pedestrian                 1.045187    0.19743   5.294
Parked_Vehicle            -1.126514    0.15691  -7.180
Late_Night                 0.247619    0.04658   5.316

Intercepts:
    Value    Std. Error t value
0|1  -0.7554   0.0583   -12.9499
1|2  -0.2726   0.0580    -4.7033
2|3   0.6323   0.0582    10.8693
3|4   1.7462   0.0617    28.2927

Residual Deviance: 21099.42
AIC: 21117.42
> cat("Probit Log-Likelihood:", ll_probit, "\n")
Probit Log-Likelihood: -10549.71
> cat("Probit AIC:", aic_probit, "\n")
Probit AIC: 21117.42
> cat("Probit BIC:", bic_probit, "\n")
Probit BIC: 21179.38
> cat("Probit McFadden's Pseudo-R²:", pseudoR2_probit, "\n\n")
Probit McFadden's Pseudo-R²: 0.007646145
> summary(m1_logit)

Re-fitting to get Hessian

Call:
polr(formula = Injury_Severity ~ Speed.Limit.at.Crash.Site +
    Urban + Pedestrian + Parked_Vehicle + Late_Night, data = dat,
    method = "logistic")

Coefficients:
                            Value Std. Error t value
Speed.Limit.at.Crash.Site  0.003878   0.001713   2.264
Urban                     -0.262576   0.045271  -5.800
Pedestrian                 1.821210   0.328971   5.536
Parked_Vehicle            -2.082792   0.287090  -7.255
Late_Night                 0.403286   0.081456   4.951

Intercepts:
    Value    Std. Error t value
0|1  -1.2435   0.0989   -12.5727
1|2  -0.4378   0.0978    -4.4748
2|3   1.0277   0.0985    10.4387
3|4   3.1380   0.1107    28.3558

Residual Deviance: 21094.08
AIC: 21112.08
```

```
> cat("Logit Log-Likelihood:", ll_logit, "\n")
Logit Log-Likelihood: -10547.04
> cat("Logit AIC:", aic_logit, "\n")
Logit AIC: 21112.08
> cat("Logit BIC:", bic_logit, "\n")
Logit BIC: 21174.04
> cat("Logit McFadden's Pseudo-R²:", pseudoR2_logit, "\n\n")
Logit McFadden's Pseudo-R²: 0.007897511
```

Now comparing Model performance of Original model and Model including 3 new Variables (Refer code and output from Ques 2) :

Model Type	Model Version	Log-Likelihood	AIC	BIC	McFadden's Pseudo-R^2
Probit	**Original**	-10549.71	21117.42	21179.38	0.00765
	Updated	-10394.43	20862.86	21117.57	0.02225
Logit	**Original**	-10547.04	21112.08	21174.04	0.00790
	Updated	-10393.58	20861.16	21115.86	0.02233

Log-Likelihood
• Improved in both Probit and Logit models after including the new predictors (closer to 0 = better fit).
• Logit: -10547.04 → -10393.58
• Probit: -10549.71 → -10394.43

AIC (Akaike Information Criterion)
• Lower AIC indicates a better model with optimal complexity.
• Probit: 21117.42 → 20862.86 (↓254.56)
• Logit: 21112.08 → 20861.16 (↓250.92)

BIC (Bayesian Information Criterion)
• Like AIC but with a stronger penalty on model complexity.
• Probit: 21179.38 → 21117.57
• Logit: 21174.04 → 21115.86
• Drop in BIC suggests the added variables improved fit enough to justify their inclusion.

McFadden's Pseudo-R^2
• Shows substantial improvement, nearly tripled:
• Probit: 0.00765 → 0.02225
• Logit: 0.00790 → 0.02233

Hence:
• The updated models with the 3 additional predictors significantly outperform the original models across all metrics.

• Logit still slightly edges out Probit in terms of fit, but both benefit equally from the richer variable set.
• The new predictors (lighting, surface condition, crash type) contribute meaningfully to explaining injury severity.

Ques 4: Propose 2 additional variables that you would like to collect data on, and include in your model, and motivate their influence on injury severity. Also discuss whether the proposed explanatory variables will be added as numeric, count or binary variables.

Ans 4:

Two additional variables that I would like to collect data on, and include in your model, and motivate their influence on injury severity:

i. **Alcohol or Drug Use**
ii. **Helmet Use Type or Quality**

Alcohol or Drug Use:

• **Type:** Binary (Yes = 1, No = 0)
• **Motivation:**
→ Riders under the influence of alcohol or drugs have impaired judgment, slower reaction times, and reduced coordination.
→ This significantly increases the likelihood of crashes and worsens injury severity due to higher impact speeds or lack of evasive manoeuvres.
→ This variable directly captures risk-enhancing behaviour and has a well-established link to road injuries.

Helmet Use Type or Quality:

• **Type:** Categorical (e.g., "None", "Half Helmet", "Full-Face Helmet", "Unknown")
• **Motivation:**
→Not just helmet usage, but the type or quality of helmet matters significantly in protecting vital areas like the head and face.
→A full-face helmet provides more comprehensive protection than a half helmet or skull cap.
→Including this variable allows more accurate modelling of how protective gear influences injury outcomes.

Hence, incorporating additional variables like Alcohol or Drug Use (binary) and Helmet Use Type or Quality (categorical) can significantly improve the explanatory power of the injury severity model. These factors directly impact rider safety and crash outcomes. By capturing both **risky behaviour and protective measures**, the model would provide a more comprehensive understanding of the determinants of motorcycle crash injuries.