Pavitchaya
6004641251

Assignment 3

1.

```
. probit y x1 x2 x3 x4

Iteration 0:   log likelihood = -248.43455
Iteration 1:   log likelihood = -150.03919
Iteration 2:   log likelihood = -147.48531
Iteration 3:   log likelihood = -147.46882
Iteration 4:   log likelihood = -147.46881

Probit regression                                 Number of obs     =        400
                                                  LR chi2(4)        =     201.93
                                                  Prob > chi2       =     0.0000
Log likelihood = -147.46881                       Pseudo R2         =     0.4064
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	.3590739	.0371539	9.66	0.000	.2862536	.4318941
x2	-.8525746	.144481	-5.90	0.000	-1.135752	-.569397
x3	-.5735764	.2202882	-2.60	0.009	-1.005333	-.1418195
x4	-1.248569	.226762	-5.51	0.000	-1.693014	-.8041238
_cons	1.45664	.2037279	7.15	0.000	1.057341	1.85594

```
. fitstat

Measures of Fit for probit of y

Log-Lik Intercept Only:     -248.435     Log-Lik Full Model:         -147.469
D(395):                      294.938     LR(4):                       201.931
                                         Prob > LR:                     0.000
McFadden's R2:                 0.406     McFadden's Adj R2:             0.386
Maximum Likelihood R2:         0.396     Cragg & Uhler's R2:            0.557
McKelvey and Zavoina's R2:     0.640     Efron's R2:                    0.446
Variance of y*:                2.775     Variance of error:             1.000
Count R2:                      0.818     Adj Count R2:                  0.416
AIC:                           0.762     AIC*n:                       304.938
BIC:                       -2071.691     BIC':                        -177.966
```

Pavitchaya
6004641251

For Probit model, the overall test is indicated from the value of LR Chi-square Test. From the table, the value of LR Chi-square Test (LR chi2(4)) equals to 201.93. This is the Likelihood Ratio (LR) Chi-Square test that at least one of the predictors' regression coefficient is not equal to zero in the model. Moreover, the model is statistically significant because the p-value is less than 0.000. When we consider individual test, we look at the value of Z-test. And from the table, P-value of all independent variables are lower than 0.05. It means that all independent variables are statistically significant or can explain the dependent variables. For Pseudo R-square of this Probit model, the estimated result is 0.4064. However, pseudo R-squared only has meaning when compared to pseudo R-squared of other model with the same type, the same data, the same outcome. It indicates how well does the estimated result contribute to the likelihood of the model. The model that has higher pseudo R-squared is the better one. Lastly, Counted R-square of the Probit model is 0.818 or 81.8%. This means that the model can correctly predict the outcome by 81.8% or actual result is the same as predicted result by 81.8%.

```
. logit y x1 x2 x3 x4

Iteration 0:   log likelihood = -248.43455
Iteration 1:   log likelihood = -154.06753
Iteration 2:   log likelihood = -148.00091
Iteration 3:   log likelihood = -147.90887
Iteration 4:   log likelihood = -147.90869
Iteration 5:   log likelihood = -147.90869

Logistic regression                             Number of obs     =        400
                                                LR chi2(4)        =     201.05
                                                Prob > chi2       =     0.0000
Log likelihood = -147.90869                     Pseudo R2         =     0.4046

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .6299401   .0708979     8.89   0.000     .4909828    .7688974
          x2 |  -1.488248   .2597744    -5.73   0.000    -1.997396   -.9790992
          x3 |  -.9562902   .3882611    -2.46   0.014    -1.717268   -.1953124
          x4 |  -2.155321   .4055058    -5.32   0.000    -2.950097   -1.360544
       _cons |     2.5165   .3714373     6.78   0.000     1.788496    3.244503
------------------------------------------------------------------------------
```

Pavitchaya
6004641251

```
. fitstat

Measures of Fit for logit of y

Log-Lik Intercept Only:     -248.435     Log-Lik Full Model:         -147.909
D(395):                      295.817     LR(4):                       201.052
                                         Prob > LR:                     0.000
McFadden's R2:                 0.405     McFadden's Adj R2:             0.385
Maximum Likelihood R2:         0.395     Cragg & Uhler's R2:            0.555
McKelvey and Zavoina's R2:     0.622     Efron's R2:                    0.445
Variance of y*:                8.707     Variance of error:             3.290
Count R2:                      0.818     Adj Count R2:                  0.416
AIC:                           0.765     AIC*n:                       305.817
BIC:                       -2070.811     BIC':                        -177.086
```

For Logit model, the overall test is indicated from the value of LR Chi-square Test. From the table, the value of LR Chi-square Test (LR chi2(4)) equals to 201.05. This is the Likelihood Ratio (LR) Chi-Square test that at least one of the predictors' regression coefficient is not equal to zero in the model. Moreover, the model is statistically significant because the p-value is less than 0.000. When we consider individual test, we look at the value of Z-test. And from the table, P-value of all independent variables are lower than 0.05. It means that all independent variables are statistically significant or can explain the dependent variables. For Pseudo R-square of this Probit model, the estimated result is 0.4046. However, pseudo R-squared only has meaning when compared to pseudo R-squared of other model with the same type, the same data, the same outcome. It indicates how well does the estimated result contribute to the likelihood of the model. The model that has higher pseudo R-squared is the better one. Lastly, Counted R-square of the Probit model is 0.818 or 81.8%. This means that the model can correctly predict the outcome by 81.8% or actual result is the same as predicted result by 81.8%.

2.

To compare the goodness of fit of the two models, we consider Pseudo (McFadden) R2 of the two models. We can see from the tables that Pseudo R2 of Probit model is 0.4064, while Logit model's Pseudo R2 is only 0.4046. So, we can conclude that Probit model better fits the data than the Logit model.

Pavitchaya
6004641251

3.

```
. probit y, nolog

Probit regression                                 Number of obs     =        400
                                                  LR chi2(0)        =       0.00
                                                  Prob > chi2       =          .
Log likelihood = -248.43455                       Pseudo R2         =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .4887764   .0654634     7.47   0.000     .3604706    .6170822
------------------------------------------------------------------------------
```

Overall LR-test = 2 (log likelihood of unrestricted model – log likelihood of restricted model)
= 2(-147.46881+248.43455)
= 201.93

Pavitchaya
6004641251

4.

```
. logit y x1 x2 x3 x4, nolog

Logistic regression                             Number of obs     =        400
                                                LR chi2(4)        =     201.05
                                                Prob > chi2       =     0.0000
Log likelihood = -147.90869                     Pseudo R2         =     0.4046
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	.6299401	.0708979	8.89	0.000	.4909828	.7688974
x2	-1.488248	.2597744	-5.73	0.000	-1.997396	-.9790992
x3	-.9562902	.3882611	-2.46	0.014	-1.717268	-.1953124
x4	-2.155321	.4055058	-5.32	0.000	-2.950097	-1.360544
_cons	2.5165	.3714373	6.78	0.000	1.788496	3.244503

```
. mfx, predict(xb)

Marginal effects after logit
      y  = Linear prediction (log odds) (predict, xb)
         =    1.32418
```

variable	dy/dx	Std. Err.	z	P>|z|	[	95% C.I.	]	X
x1	.6299401	.0709	8.89	0.000	.490983	.768897		.454973
x2	-1.488248	.25977	-5.73	0.000	-1.9974	-.979099		.809344
x3	-.9562902	.38826	-2.46	0.014	-1.71727	-.195312		.556712
x4	-2.155321	.40551	-5.32	0.000	-2.9501	-1.36054		-.119684

I_hat = 2.52+0.62(0.45)-1.49(0.81)-0.96(0.56)+2.16(0.12)
= 1.32

Pavitchaya
6004641251

5.

```
. mfx

Marginal effects after logit
      y  = Pr(y) (predict)
         =   .7898763
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
x1	.1045522	.01146	9.12	0.000	.082083	.127022	.454973
x2	-.247007	.04388	-5.63	0.000	-.333011	-.161003	.809344
x3	-.1587171	.06397	-2.48	0.013	-.2841	-.033334	.556712
x4	-.3577223	.06679	-5.36	0.000	-.488633	-.226812	-.119684

```
. mfx, at(median)

Marginal effects after logit
      y  = Pr(y) (predict)
         =  .84127022
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
x1	.0841188	.00961	8.76	0.000	.065292	.102946	.655749
x2	-.1987326	.03349	-5.93	0.000	-.264373	-.133093	.692745
x3	-.1276979	.04944	-2.58	0.010	-.224597	-.030799	.488768
x4	-.28781	.05616	-5.12	0.000	-.397881	-.177739	-.109732

Pavitchaya
6004641251

6.

```
. mfx, at(0.5 1 0.5 0)

Marginal effects after probit
      y  = Pr(y) (predict)
         =     .69034
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
x1	.1266183	.01307	9.69	0.000	.101001	.152235	.5
x2	-.3006389	.05452	-5.51	0.000	-.4075	-.193778	1
x3	-.2022572	.07644	-2.65	0.008	-.352085	-.05243	.5
x4	-.4402764	.08419	-5.23	0.000	-.605293	-.27526	0

7. **Counted R2** = 81.75%

```
. estat clas

Logistic model for y
```

	True		
Classified	D	~D	Total
+	251	49	300
-	24	76	100
Total	275	125	400

```
Classified + if predicted Pr(D) >= .5
True D defined as y != 0
```

Sensitivity	Pr(+| D)	91.27%
Specificity	Pr(-|~D)	60.80%
Positive predictive value	Pr(D| +)	83.67%
Negative predictive value	Pr(~D| -)	76.00%
False + rate for true ~D	Pr(+|~D)	39.20%
False - rate for true D	Pr(-| D)	8.73%
False + rate for classified +	Pr(~D| +)	16.33%
False - rate for classified -	Pr(D| -)	24.00%
Correctly classified		81.75%

Pavitchaya
6004641251

8. **Counted R2** = 79.50%

```
. estat clas, cut(0.7)

Logistic model for y
```

Classified	True D	True ~D	Total
+	**217**	**24**	**241**
−	**58**	**101**	**159**
Total	**275**	**125**	**400**

Classified + if predicted Pr(D) >= .7
True D defined as y != 0

Sensitivity	Pr(+| D)	**78.91%**
Specificity	Pr(−|~D)	**80.80%**
Positive predictive value	Pr(D| +)	**90.04%**
Negative predictive value	Pr(~D| −)	**63.52%**
False + rate for true ~D	Pr(+|~D)	**19.20%**
False − rate for true D	Pr(−| D)	**21.09%**
False + rate for classified +	Pr(~D| +)	**9.96%**
False − rate for classified −	Pr(D| −)	**36.48%**
Correctly classified		**79.50%**