Problem Set 1: Identification & Limited Dependent Variables

Identification

Exercise 1 (10 Marks)**.** Let Y, X and U be random variables where the unobservable U comes from a standard Normal distribution, i.e. $U \sim N(0,1)$, where

$$Y = \alpha + \beta X + U \tag{1}$$

Suppose we know the distribution of X (it is independent of α and β) and we know that $X \perp\!\!\!\perp U$. Are α and β identified by (1)? If yes, then prove it. If no, then display two or more values of the parameters for which the distribution of Y is the same.

Solution 1 (Identifiability)**.**

Claim 1. $\theta = (\alpha, \beta)$ is identified by (1).

Proof. If θ is not identified, then for any θ, there exists $\theta' \neq \theta$ such that

$$P(Y \leq t|X;\theta) = P(Y \leq t|X;\theta')$$

Since we know the distribution of X, we can restrict ourselves only to conditional distributions, since we know the distribution of x.

$$\begin{aligned} P(Y \leq t|X;\theta) &= P(\alpha + \beta X + U \leq t|X;\theta) \\ &= P(U \leq t - \alpha - \beta X|X;\theta) \\ &= \Phi(t - \alpha - \beta X) \end{aligned}$$

where Φ is the cumulative Normal distribution function. Suppose that for all X

$$\Phi(t - (\alpha + \beta X)) = \Phi(t - (\alpha' + \beta' X))$$

Since Φ is one-to-one, it follows that for all X

$$\begin{aligned} t - (\alpha + \beta X) &= t - (\alpha' + \beta' X) \\ \implies (\beta' - \beta)X &= \alpha - \alpha' \\ \implies \alpha = \alpha' &\text{ and } \beta = \beta' \end{aligned}$$

□

Exercise 2 (20 Marks)**.** `INPUTM12.txt` is a data file that contains 869 observations of American white male respondents in the National Longitudinal Study of Youth (NLSY). Each record consists of values for the variables (y, z, f, m), which are defined by:

$$\begin{aligned} y &= \text{indicator of high school completion (1 = yes, 0 = no)} \\ z &= \text{indicator of family status at age 14 (1 = intact, 0 = non-intact family)} \\ f &= \text{father's years of schooling} \\ m &= \text{mother's years of schooling} \end{aligned}$$

Suppose that the mother of an American white male has 12 years of schooling and you are asked to predict high school graduation. Assume that the 869 observations are a random sample of American white males. Use Stata software to do the following:

1. Estimate the best linear predictor of y given ($m = 12$) under square loss, by ordinary least squares.

2. Compute kernel estimates of $E(y|m = 12)$ using uniform and Gaussian kernels and bandwidths 0.5, 1.5 and 4.5; hence, there are six estimates in total.

3. Discuss the estimates computed under 1 and 2.

Solution 2 (Stata – kernreg)**.**

1. The best linear predictor of $(Y|M = 12)$, under square loss is given by the mean. Using an OLS regression to estimate the coefficient of the constant, the following Stata code 1 produces the estimate.

```
infile y z f m using INPUTM12.txt, clear

regress y m
predict yhat
list yhat if m==12
```

Listing 1: BLP of y given $m = 12$ under square loss by OLS.

This yielded the expectation of $Y|M = 12$

$$\underline{E(Y|M = 12)} = \underline{.8434434}$$

which is the best linear predictor of $Y|M = 12$, under square loss.
Note that the OLS estimation of $E(Y|M = 12)$ reduces to the non-parametric estimator of $E(Y|M = 12)$ when the covariates have positive probability (true since M is discrete and the data contains a subset where $M = 12$):

$$\begin{aligned}\hat{E}(Y|M = 12)_{OLS} &= \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2} \\ &= \frac{\sum_{i=1}^{N'} x_i y_i}{\sum_{i=1}^{N'} x_i^2} \\ &= \frac{\sum_{i=1}^{N} y_i \cdot 1[M_i = 12]}{\sum_{i=1}^{N} 1[M_i = 12]} \\ &= E_N(Y|M = 12)\end{aligned}$$

where N is the sample size (869), $N' < N$ is the number of observations for which $M = 12$ and $x_i = 1 \; \forall i : M_i = 12$.

2. The kernel estimates of $E(Y|M = 12)$ computed with the `kernreg` function, using the code shown in listing 2, are recorded in table 1. The Uniform Kernel (local average) and Gaussian Kernel (local weighted average) estimates were calculated as

$$\theta_N(M = 12, d_N) = \frac{\sum_{i=1}^{N} y_i \cdot 1[\rho(M_i, M = 12) < d_N]}{\sum_{i=1}^{N} 1[\rho(M_i, M = 12) < d_N]} \quad \text{(Uniform)}$$

$$E_N(Y|M = 12) = \frac{\sum_{i=1}^{N} y_i \cdot K\left[\frac{\rho(M_i, M=12)}{d_N}\right]}{\sum_{i=1}^{N} K\left[\frac{\rho(M_i, M=12)}{d_N}\right]} \quad \text{(Gaussian)}$$

where $d_N \in \{0.5, 1.5, 4.5\}$ is the bandwidth.

```
kernreg y m, g(yubar1) at(12) w1(0.5) rec

kernreg y m, g(yubar2) at(12) w1(1.5) rec

kernreg y m, g(yubar3) at(12) w1(4.5) rec

kernreg y m, g(ygbar1) at(12) w1(0.5) gau

kernreg y m, g(ygbar2) at(12) w1(1.5) gau

kernreg y m, g(ygbar3) at(12) w1(4.5) gau

list yubar1 yubar2 yubar3 ygbar1 ygbar2 ygbar3 in 1/1
```

Listing 2: Uniform & Gaussian kernel estimates of $E(y|m = 12)$ with different bandwidths.

Table 1: Uniform & Gaussian kernel estimates of $E(Y|M = 12)$ with bandwidths: 0.5, 1.5 and 4.5

Bandwidth	Uniform Kernel	Gaussian Kernel
0.5	0.888172	0.8865797
1.5	0.8781362	0.8720608
4.5	0.8543689	0.8533347

3. The point here is to compare OLS and nonparametric estimators. Essentially, by expanding the bandwidth, we use more data and so we get closer to the OLS estimate. This would receive full credit. Extra discussion follows.
   For $d_N < 1$, since $m \in \mathbb{N}$, only i such that $m_i = m = 12$ will be given a value of one (instead of zero) by the indicator function in the estimation.
   The Gaussian Kernel will put more weight on values close to $m = 12$, so $K\left[\frac{\rho(M_i,M=12)}{d_N}\right] = K\left[\frac{\rho(M=12,M=12)}{d_N}\right]$ for i such that $m_i = m = 12$ (this will be the only i where the indicator function will be non-zero in the estimation) but unlike the Uniform Kernel estimate, the indicator function $K[\cdot]$ will take a value that will not be exactly one; thus the Gaussian Kernel estimate in the case $d_N = 0.5$ will be different to the Uniform Kernel estimate.Once the bandwidths are increased above 1, e.g. $d_N \in \{1.5, 4.5\}$, the kernel estimators allow values further from $m = 12$ to be taken into account ($m \in \{11, 12, 13\}$ for $d_N = 1.5$ and $m \in \mathbb{N} : 8 \leq m \leq 16$ for $d_N = 4.5$). In fact, irrespective of the shape of the distribution (Uniform or Gaussian), the estimates seem to be decreasing with the size of the bandwidth. This is because these kernel estimators give more weight to values further away from $M = 12$ as the bandwidth increases.
   Increasing the bandwidth will allow for more observations to lie within the bandwidth and so the variance will come down. However, increasing the bandwidth will lead to a bias of the estimate (the values of M no longer lying within a bandwidth that can be made arbitrarily small and still contain the values of M). The curse of dimensionality problem arises in trying to chose the bandwidth to minimise the mean square error (variance and bias) as the sample size increases, especially as the dimensions increase (our case only looks at M, so there is only one dimension, hence only d_N, rather than d_N^k.

Exercise 3 (10 Marks)**.** A survey firm conducting an election poll can contact voters by any of three modes: internet, telephone or home interview. Let x denote the mode that the firm uses to contact a voter. Suppose that the firm contacts 1200 voters; 400 by internet ($x = 0$), 400 by telephone ($x = 1$) and 400 by home interview ($x = 2$). These voters are asked if they want Fianna Fáil to return to government in the year 2016. The possible responses are no ($y = 0$), indifferent ($y = \frac{1}{2}$) and yes ($y = 1$). Suppose that all voters have a

	Contact Mode		
Response to Survey Question	Internet ($x = 0$)	Telephone ($x = 1$)	Home Interview ($z = 2$)
$y = 0$	100	100	100
$y = \frac{1}{2}$	100	100	100
$y = 1$	100	100	100
no response	100	100	100

Table 2: A 2016 return for Fianna Fáil.

value of y, but some of them choose not to respond to the survey. Here are the data obtained: When answer the questions below, consider these 1200 voters to be the population of interest, not a sample drawn from a larger population. You are asked to predict y conditional on the event $(x = 1)$. Given the available data, what can you deduce about:

1. the best predictor of y conditional on $(x = 1)$, under square loss?
2. the best predictor of y conditional on $(x = 1)$, under absolute loss?

Solution 3 (Conditional Prediction with Incomplete Data)**.**
Let $z = 1$ denote response and $z = 0$ denote non-response in the following solutions.

1. The best predictor of y conditional on $(x = 1)$, under square loss is the conditional mean, $E(y|x = 1)$. By the law of iterated expectations:

$$E(y|x = 1) = E(y|z = 1, x = 1)P(z = 1|x = 1) + E(y|z = 0, x = 1)P(z = 0|x = 1)$$

We know that $P(z = 0|x = 1) = \frac{1}{4}$ and $P(z = 1|x = 1) = \frac{3}{4}$. The quantity $E(y|z = 0, x = 1)$ is unknown – we only know that $E(y|z = 0, x = 1) \in [0, 1]$. We must calculate $E(y|z = 1, x = 1)$:

$$\begin{aligned} E(y|z = 1, x = 1) &= 0 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{\cdot}\frac{1}{3} \\ &= \frac{1}{2} \end{aligned}$$

$$\begin{aligned} \therefore E(y|x = 1) &= \frac{1}{2}\frac{3}{4} + [0, 1]\frac{1}{4} \\ \therefore H[E(y|x = 1)] &= [\frac{3}{8}, \frac{5}{8}] \end{aligned}$$

2. The best predictor of y conditional on $(z = 1)$, under absolute loss is the conditional median, $E(y|z = 1) = \inf\{t : P(y \leq t|z = 1) \geq \frac{1}{2}\}$. We know that $P(z = 1|x = 1) = \frac{3}{4}$ and that $P(y = i|x = 1) = \frac{1}{3}$ for any $i \in \{0, \frac{1}{2}, 1\}$, i.e. all modes are equally likely. We do not know $P(y|z = 0, x = 1)$. We can choose one of two extremes, either all voters who were contacted by telephone $(x = 1)$ who did not respond $(z = 0)$ would have not wanted Fianna Fáil to return to government in 2016 $(y = 0)$ or – at the other side – all such voters would want Fianna Fáil to return to government $(y = 1)$. In the first case by the law of total probability:

$$P(y \leq 0|x = 1) = \frac{1}{3} \cdot \frac{3}{4} + 1 \cdot \frac{1}{4} = \frac{1}{2}$$

so $M(y|x = 1) = 0$. In the second case:

$$P(y \leq 0|x = 1) = \frac{1}{3} \cdot \frac{3}{4} = \frac{1}{4}$$

so $M(y|x=1) \neq 0$. In this case we will check if $M(y|x=1) = \frac{1}{2}$.

$$\begin{aligned} P(y = \frac{1}{2}|x=1) &= \frac{1}{3} \cdot \frac{3}{4} \\ &= \frac{1}{4} \\ \implies P(y \leq \frac{1}{2}|x=1) &= \frac{1}{4} + \frac{1}{4} \\ &= \frac{1}{2} \end{aligned}$$

So, in the second case $M(y|x=1) = \frac{1}{2}$. However, observe that $M(y|x=1) \neq 1$ for any distribution. We have assumed the extreme case that all voters who were contacted by telephone and did not respond would have wanted Fianna Fáil to return to government ($y = 1$). Yet, we did not get that $M(y|x=1) = 1$. Therefore, we conclude that $M(y|x=1) \in \{0, \frac{1}{2}\}$.
This proof relies on placing all the weight for the unobserved outcomes on either end of the distribution ($y = 0$ and $y = 1$). We found that even looking at these extreme cases, the median, middle voter cannot be at the top end ($y = 1$) in this example.

Exercise 4 (20 Marks)**.** Consider the problem of how sentencing juvenile offenders may affect their future criminality. Suppose we have available data on the sentencing and recidivism of males in Ireland who were born from 1980 through 1985 and who were convicted of offenses before they reached age 16. Let $t = b$ denote confinement in residential facilities and $t = a$ denote sentences that do not involve residential confinement. The outcome of interest is y defined by:

$$y = \begin{cases} 1 & \text{offender is not convicted of a subsequent crime the in five-year period following sentencing} \\ 0 & \text{offender is convicted of a subsequent crime in the five-year period following sentencing} \end{cases}$$

We have data for the study population as follows:

$$\begin{aligned} P(t = b) &= 0.10 \\ P(y = 0) &= 0.65 \\ P(y = 0|t = b) &= 0.75 \\ P(y = 0|t = a) &= 0.6 \end{aligned}$$

Consider two alternative policies: one mandating residential treatment for all offenders and the other mandating nonresidential treatment. The recidivism probabilities under these policies are $P[y(b) = 0]$ and $P[y(a) = 0]$, respectively.

1. If you assumed that judges in Ireland either purposefully or effectively sentence offenders at random to residential and nonresidential treatments, what could you conclude regarding $P[y(b) = 0]$ and $P[y(a) = 0]$?
2. What would be the identification regions for these potential recidivism probabilities using the empirical evidence alone?
3. What are the widths of the two intervals you calculated in 2? If they differ, why do they differ? If they do not differ, why do they not differ?
4. The average treatment effect in this setting is the difference in recidivism probabilities under the two alternative sentencing policies, i.e. $P[y(b) = 0] - P[y(a) = 0]$. Calculate the identification region for average treatment effect using the data alone. What is the width of this interval? Does it contain zero? Explain.

5. Calculate the average treatment effect in 2 under the assumption of treatment at random, i.e. under $P[y(a)|t = a] = P[y(a)|t = b]$ and $P[y(b)|t = a] = P[y(b)|t = b]$.

6. Finally, suppose that a legal researcher wants to use this data to support the abolition of sentences confining juvenile offenders to residences. In particular, she states the following:

   > Data indicate that juvenile offenders who are not sentenced to residential confinement have a lower probability of committing future crimes. The effect of nonresidential treatment is to lower the probability of juvenile offenders committing future crimes from 0.77 to 0.59.

   Does this statement accurately describe the empirical findings? Explain.

Solution 4 (Treatment Response)**.**

1. Random treatment assignment (random) implies that for $t' \in \{a, b\}$:

$$P(y(t')|t = t') = P(y(t')|t \neq t')$$

   So, by the law of total probability:

$$\begin{aligned} P(y(t')) &\overset{\text{LTP}}{=} P(y(t')|t = t')P(t = t') + P(y(t')|t \neq t')P(t \neq t') \\ &\overset{\text{random}}{=} \underbrace{P(y(t')|t = t')}_{P(y|t=t')}[\underbrace{P(t = t') + P(t \neq t')}_{1}] \end{aligned}$$

   which holds for all $t' \in \{a, b\}$. Therefore

$$\begin{aligned} P[y(b) = 0] &= P(y = 0|t = b) = 0.75 \\ P[y(a) = 0] &= P(y = 0|t = a) = 0.6 \end{aligned}$$

2. Without using any assumptions, again using the law of total probability:

$$\begin{aligned} P(y(b) = 0) &\overset{\text{LTP}}{=} \underbrace{P(y(b) = 0|t = b)P(t = b)}_{P(y=0|t=b)} P(t = b) + \underbrace{P(y(b) = 0|t = a)}_{\in[0,1]} \underbrace{P(t = a)}_{1-P(t=b)} \\ &= 0.75 \cdot 0.1 + [0, 1] \cdot 0.9 \\ &\in [0.075, 0.975] \\ P(y(a) = 0) &\overset{\text{LTP}}{=} \underbrace{P(y(a) = 0|t = a)}_{P(y=0|t=a)} P(t = a) + \underbrace{P(y(a) = 0|t = b)}_{\in[0,1]} P(t = b) \\ &= 0.6 \cdot 0.9 + [0, 1] \cdot 0.1 \\ &\in [0.54, 0.64] \end{aligned}$$

3. The width of $H[P(y(b) = 0)]$ is 0.9 and the width of $H[P(y(a) = 0)]$ is 0.1. In the first case, this is because this is the fraction of the study population who received treatment b and who, therefore, have unobservable outcomes under treatment b. Symmetrically, the region for $P(y(a) = 0)$ has width 0.1.

4. Let $ATE = P[y(b) = 0] - P[y(a) = 0]$. From part 2:

$$\begin{aligned} P[y(b) = 0] &\overset{2}{=} [\overbrace{0.075}^{L_b}, \overbrace{0.975}^{U_b}] \\ P[y(a) = 0] &\overset{2}{=} [\underbrace{0.54}_{}^{L_a}, \underbrace{0.64}_{}^{U_a}] \end{aligned}$$

The lower and upper bounds for ATE are given by

$$L_b - U_a = -0.565$$
$$U_b - L_a = 0.435$$

$$\therefore H[ATE] = [-0.565, 0.435]$$

$H[ATE]$ necessarily has a width of one and includes zero. To explain this, note that using data alone, the hypothesis of $ATE = 0$ is not refutable. Given that counterfactual outcomes are unobserved, it is possible that $y(a) = y(b)$ for every person in the population; technically, we say that $y_j(a) = y_j(b)$ for all persons j in the population. Therefore, the hypothesis of zero average treatment effects is not refutable using empirical evidence alone. The hypothesis may be made refutable only if the evidence is combined with sufficiently strong distributional assumptions.

5. We assumed random treatment assignment in part 1 and found that $P[y(a) = 0] = 0.6$ and $P[y(b) = 0] = 0.75$ so $ATE = 0.15$, indicating that nonresidential treatment is much better than residential treatment if the objective is to minimise recidivism.

6. No, this statement does not accurately describe the empirical finding.
We can only say that juvenile offenders who were sentenced to residential confinement had on average a higher probability of recidivism. We cannot say that the treatment of being sentenced to residential confinement increased the probability of recidivism.
The researcher has confused correlation with causation and has used a counterfactual (expressing what has not happened but what might or would happen if circumstances, i.e. data were different). The researcher is in effect extrapolating using the assumption of external validity, which is undermined by the fact that we are only looking at juvenile offenders *in Ireland* who were born during 1980-1985. Furthermore, changing the very structure of the sentences may change how people respond; this is related to the *Lucas critique*.
However, if the juvenile offenders were randomly sentenced to residential confinement as in part 5, then the researcher would be correct in saying that the effect of nonresidential treatment is to lower the probability of juvenile offenders committing future crimes.

LPM, Logit & Probit

Exercise 5 (20 Marks)**.** For this exercise you will need the dataset GRE.dta and the problems MUST be implemented in STATA where indicated. For this you will need to provide your STATA program and regression output. This data set has a binary response (outcome, dependent) variable called *admit*. There are three predictor variables: *gre, gpa and rank*. We will treat the variables gre and gpa as continuous. The variable rank takes on the values 1 through 4. Institutions (Colleges that students attended) with a rank of 1 have the highest prestige, while those with a rank of 4 have the lowest. Conduct a logit, probit and ols regression and interpret the coefficients. Compare the results from the 3 regressions and explain which model you prefer. Justify your answers.

Solution 5 (LPM, Logit & Probit)**.**
OLS: regress admit gre gpa i.rank
For every one unit change in GRE, the log odds of admission (versus non-admission) increases by 0.0004. For a one unit increase in GPA, the log odds of being admitted to graduate school increases by 0.155. The indicator variables for rank have a slightly different interpretation. For example, having attended an undergraduate institution with rank of 2, versus an institution with a rank of 1, decreases the log odds of admission by 0.16.

Logit: logit admit gre gpa i.rank
For every one unit change in GRE, the log odds of admission (versus non-admission) increases by 0.002. For

a one unit increase in GPA, the log odds of being admitted to graduate school increases by 0.804. The indicator variables for rank have a slightly different interpretation. For example, having attended an undergraduate institution with rank of 2, versus an institution with a rank of 1, decreases the log odds of admission by 0.675.

Probit: probit admit gre gpa i.rank
For a one unit increase in GRE, the z-score increases by 0.001. For each one unit increase in GPA, the z-score increases by 0.478. The indicator variables for rank have a slightly different interpretation. For example, having attended an undergraduate institution of rank of 2, versus an institution with a rank of 1 (the reference group), decreases the z-score by 0.415.

```
1 . use "/Users/ccs/Desktop/Econ Under/GRE.dta"

2 . summarize gre gpa
```

Variable	Obs	Mean	Std. Dev.	Min	Max
gre	400	587.7	115.5165	220	800
gpa	400	3.3899	.3805668	2.26	4

```
3 . tab rank
```

rank	Freq.	Percent	Cum.
1	61	15.25	15.25
2	151	37.75	53.00
3	121	30.25	83.25
4	67	16.75	100.00
Total	400	100.00	

```
4 . tab admit
```

admit	Freq.	Percent	Cum.
0	273	68.25	68.25
1	127	31.75	100.00
Total	400	100.00	

```
5 . tab admit rank
```

	rank				
admit	1	2	3	4	Total
0	28	97	93	55	273
1	33	54	28	12	127
Total	61	151	121	67	400

```
6 .
7 .
8 . regress admit gre gpa i.rank
```

Source	SS	df	MS
Model	8.70247579	5	1.74049516
Residual	77.9750242	394	.197906153
Total	86.6775	399	.217236842

Number of obs	=	400
F(5, 394)	=	8.79
Prob > F	=	0.0000
R-squared	=	0.1004
Adj R-squared	=	0.0890
Root MSE	=	.44487

admit	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gre	.0004296	.0002107	2.04	0.042	.0000153	.0008439
gpa	.155535	.0639618	2.43	0.015	.0297859	.2812842
rank						

2	-.1623653	.0677145	-2.40	0.017	-.2954922	-.0292385
3	-.2905705	.0702453	-4.14	0.000	-.428673	-.1524679
4	-.3230264	.0793164	-4.07	0.000	-.4789626	-.1670902
_cons	-.2589103	.2159904	-1.20	0.231	-.6835481	.1657275

```
 9 . logit admit gre gpa i.rank

Iteration 0:   log likelihood = -249.98826
Iteration 1:   log likelihood = -229.66446
Iteration 2:   log likelihood = -229.25955
Iteration 3:   log likelihood = -229.25875
Iteration 4:   log likelihood = -229.25875

Logistic regression                               Number of obs   =        400
                                                  LR chi2(5)      =      41.46
                                                  Prob > chi2     =     0.0000
Log likelihood = -229.25875                       Pseudo R2       =     0.0829
```

admit	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
gre	.0022644	.001094	2.07	0.038	.0001202	.0044086
gpa	.8040377	.3318193	2.42	0.015	.1536838	1.454392
rank						
2	-.6754429	.3164897	-2.13	0.033	-1.295751	-.0551346
3	-1.340204	.3453064	-3.88	0.000	-2.016992	-.6634158
4	-1.551464	.4178316	-3.71	0.000	-2.370399	-.7325287
_cons	-3.989979	1.139951	-3.50	0.000	-6.224242	-1.755717

```
10 . probit admit gre gpa i.rank

Iteration 0:   log likelihood = -249.98826
Iteration 1:   log likelihood = -229.29667
Iteration 2:   log likelihood = -229.20659
Iteration 3:   log likelihood = -229.20658

Probit regression                                 Number of obs   =        400
                                                  LR chi2(5)      =      41.56
                                                  Prob > chi2     =     0.0000
Log likelihood = -229.20658                       Pseudo R2       =     0.0831
```

admit	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
gre	.0013756	.0006489	2.12	0.034	.0001038	.0026473
gpa	.4777302	.1954625	2.44	0.015	.0946308	.8608297
rank						
2	-.4153992	.1953769	-2.13	0.033	-.7983308	-.0324675
3	-.812138	.2085956	-3.89	0.000	-1.220978	-.4032981
4	-.935899	.2456339	-3.81	0.000	-1.417333	-.4544654
_cons	-2.386838	.6740879	-3.54	0.000	-3.708026	-1.065649

Tobit and Censoring

Exercise 6 (10 Marks)**.** For this exercise you will need the dataset Honors2.dta and the problems MUST be implemented in STATA where indicated. For this you will need to provide your STATA program and regression output. The academic aptitude variable is apt which we will use as the dependent variable. Reading and math test scores are read and math respectively. The variable prog is the type of program the student is in, it is a categorical (nominal) variable that takes on three values, academic (prog = 1), general (prog = 2), and vocational (prog = 3). Using a histogram show the censoring of the variable apt. Conduct a tobit and ols regression and interpret the coefficients. Compare the results from the 2 regressions and explain which model you prefer. Justify your answers.

Solution 6 (Tobit & Censoring)**.**
OLS: regress apt read math i.prog
For a one unit increase in read, there is a 2.55 point increase in the predicted value of apt. A one unit increase in math is associated with a 5.38 unit increase in the predicted value of apt. The terms for prog have a slightly different interpretation. The predicted value of apt is 48.8 points lower for students in a vocational program (prog=3) than for students in an academic program (prog=1).

Tobit: tobit apt read math i.prog, ul(800)
For a one unit increase in read, there is a 2.7 point increase in the predicted value of apt. A one unit increase in math is associated with a 5.91 unit increase in the predicted value of apt. The terms for prog have a slightly different interpretation. The predicted value of apt is 46.14 points lower for students in a vocational program (prog=3) than for students in an academic program (prog=1).

```
1 . regress apt read math i.prog

      Source |       SS           df       MS      Number of obs   =       200
-------------+----------------------------------   F(  4,    195)  =     77.13
       Model |  1200325.87         4  300081.468   Prob > F        =    0.0000
    Residual |  758712.883       195   3890.8353   R-squared       =    0.6127
-------------+----------------------------------   Adj R-squared   =    0.6048
       Total |  1959038.76       199  9844.41585   Root MSE        =    62.377

------------------------------------------------------------------------------
         apt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        read |   2.552671   .5829987     4.38   0.000     1.402879    3.702463
        math |   5.383153   .6589949     8.17   0.000     4.083481    6.682825
             |
        prog |
          2  |  -13.74056    11.7439    -1.17   0.243    -36.90193    9.420818
          3  |  -48.83475    12.9815    -3.76   0.000    -74.43692   -23.23258
             |
       _cons |   242.7354   30.13952     8.05   0.000     183.2941    302.1767
------------------------------------------------------------------------------

2 .
3 . tobit apt read math i.prog, ul(800)

Tobit regression                                  Number of obs   =        200
                                                  LR chi2(4)      =     188.97
                                                  Prob > chi2     =     0.0000
Log likelihood = -1041.0629                       Pseudo R2       =     0.0832

------------------------------------------------------------------------------
         apt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        read |   2.697939    .618798     4.36   0.000     1.477582    3.918296
        math |   5.914485   .7098063     8.33   0.000     4.514647    7.314323
             |
        prog |
          2  |  -12.71476   12.40629    -1.02   0.307    -37.18173     11.7522
          3  |   -46.1439   13.72401    -3.36   0.001     -73.2096   -19.07821
             |
       _cons |    209.566   32.77154     6.39   0.000     144.9359    274.1961
-------------+----------------------------------------------------------------
      /sigma |   65.67672   3.481272                      58.81116    72.54228
------------------------------------------------------------------------------
  Obs. summary:          0  left-censored observations
                       183     uncensored observations
                        17 right-censored observations at apt>=800
```

Count data

Exercise 7 (10 Marks)**.** For this exercise you will need the dataset Honors3.dta and the problems MUST be implemented in STATA where indicated. For this you will need to provide your STATA program and regression output. num_awards is the dependent variable and indicates the number of awards earned by students at a high school in a year, math is a continuous independent variable and represents students' scores on their math final exam, and prog is a categorical independent variable with three levels indicating the type of program in which the students were enrolled. Conduct a poisson and ols regression and interpret the coefficients. Compare the results from the 2 regressions and explain which model you prefer. Justify your answers.

Solution 7 (Poisson)**.**
OLS: regress num_awards i.prog math, vce(robust)
The coefficient for math is .047. This means that the expected increase in log count for a one-unit increase in math is .047. The indicator variable 2.prog is the expected difference in log count between group 2 (prog=2) and the reference group (prog=1). Compared to level 1 of prog, the expected log count for level 2 of prog increases by about 1.1. The indicator variable 3.prog is the expected difference in log count between group 3 (prog=3) and the reference group (prog=1). Compared to level 1 of prog, the expected log count for level 3 of prog increases by about .21.

Poisson: poisson num_awards i.prog math, vce(robust)
The coefficient for math is .07. This means that the expected increase in log count for a one-unit increase in math is .07. The indicator variable 2.prog is the expected difference in log count between group 2 (prog=2) and the reference group (prog=1). Compared to level 1 of prog, the expected log count for level 2 of prog increases by about 1.1. The indicator variable 3.prog is the expected difference in log count between group 3 (prog=3) and the reference group (prog=1). Compared to level 1 of prog, the expected log count for level 3 of prog increases by about .37.

```
1 . poisson num_awards i.prog math, vce(robust)

  Iteration 0:   log pseudolikelihood = -182.75759
  Iteration 1:   log pseudolikelihood = -182.75225
  Iteration 2:   log pseudolikelihood = -182.75225

  Poisson regression                                Number of obs   =        200
                                                    Wald chi2(3)    =      80.15
                                                    Prob > chi2     =     0.0000
  Log pseudolikelihood = -182.75225                 Pseudo R2       =     0.2118
```

num_awards	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
prog						
2	1.083859	.3218538	3.37	0.001	.4530373	1.714681
3	.3698092	.4014221	0.92	0.357	-.4169637	1.156582
math	.0701524	.0104614	6.71	0.000	.0496485	.0906563
_cons	-5.247124	.6476195	-8.10	0.000	-6.516435	-3.977814

```
2 .
3 . regress num_awards i.prog math
```

Source	SS	df	MS
Model	61.1767822	3	20.3922607
Residual	159.443218	196	.813485805
Total	220.62	199	1.10864322

```
Number of obs =      200
F(  3,    196) =   25.07
Prob > F      =   0.0000
R-squared     =   0.2773
Adj R-squared =   0.2662
Root MSE      =   .90193
```

num_awards	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
prog						
2	.4786129	.1689563	2.83	0.005	.1454072	.8118187
3	.2125061	.1874332	1.13	0.258	-.1571386	.5821508
math	.0478888	.0077731	6.16	0.000	.0325592	.0632184
_cons	-2.195504	.4114165	-5.34	0.000	-3.006876	-1.384133