



**UJIAN TENGAH SEMESTER**  
**FAKULTAS EKONOMI UNIVERSITAS INDONESIA**

Mata Kuliah : **Ekonometrika 2/CSPD (ECEU601304)-A**  
Program : S1 Reguler  
Tanggal Ujian : 23 Oktober 2019  
Waktu : 150 Menit (2 ½ Jam)  
Sifat Ujian : Closed Book/Closed Notes  
Tim Pengajar : **I Dewa Gede Karma Wisana**  
Aufa Doarest/Dhaniel Ilyas  
Djoni Hartono

**Petunjuk Umum:**

1. Perhatikan, lembar soal ini terdiri dari **9 Halaman** dan **4 Soal**. Semua soal wajib dijawab. Bobot penilaian untuk setiap soal tersedia di akhir pertanyaan.
2. **Boleh** menggunakan kalkulator. **Tidak boleh** menggunakan telepon seluler/komputer/laptop sebagai kalkulator.
3. Sifat ujian adalah tutup buku (*closed book*) dan tutup catatan (*closed notes*).

---

**Soal 1. (20 poin)**

1. Perhatikan Spesifikasi *Linear Probability Model* berikut:

$$Y_i = Pr(Y_i = 1 | X_i) + \epsilon_i = E(Y_i | X_i) + \epsilon_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_m X_{im} + \epsilon_i$$

Dengan asumsi  $E(\epsilon_i) = 0$

$Y_i = 1$  jika sebuah peristiwa terjadi

$Y_i = 0$  jika sebuah peristiwa tidak terjadi

$i$  adalah banyaknya sampel



Distribusi probabilitas  $Y$  adalah  $Pr(Y_i = 1) = \pi_i$  untuk nilai  $Y_i = 1$  dan  $Pr(Y_i = 0) = (1 - \pi_i)$  untuk nilai  $Y_i = 0$ , dimana  $0 \leq \pi_i \leq 1$ . Hal ini dikarenakan kemungkinan nilai  $Y$  hanya terdiri dari 2 nilai yaitu 0 atau 1.

Perhatikan tabel kemungkinan nilai  $Y_i$  berikut:

| Kemungkinan nilai $Y_i$ | Probabilita keluarnya kemungkinan nilai $Y_i$ tertentu |
|-------------------------|--|
| 0                       | $Pr(Y_i = 0) = 1 - \pi_i$                              |
| 1                       | $Pr(Y_i = 1) = \pi_i$                                  |

- a. Turunkan  $E(Y_i)$  dengan konsep *expected value*! Jelaskan dari hasil turunan tersebut! Mengapa kita bisa mengkonstruksikan model probabilitas? Jelaskan dengan menggunakan hasil turunan tersebut! **(5 poin)**

Misalkan seorang peneliti me-run sebuah spesifikasi sederhana dari *Linear Probability Model* sebagai berikut:

$$mwarm = \beta_0 + \beta_1 age + \beta_2 ed + \beta_3 prst + e_i$$

Dimana  $mwarm = 1$  jika seorang wanita bersikap hangat dan memperhatikan anaknya dan 0 jika wanita tersebut relatif tidak terlalu memperdulikan anaknya;  $age$  = Umur wanita (dalam tahun);  $ed$  = lama pendidikan formal yang dikenyam wanita tersebut (dalam tahun); dan  $prst$  adalah indeks tingkat prestise dari pekerjaan wanita tersebut (makin tinggi prestise makin tinggi jabatannya) dengan skala 0-100.

```
. reg mwarm age ed prst
```

| Source   | SS         | df    | MS         | Number of obs | = | 2,293  |
|----------|------------|-------|------------|---------------|---|--------|
| Model    | 34.6479441 | 3     | 11.5493147 | F(3, 2289)    | = | 49.73  |
| Residual | 531.623316 | 2,289 | .23225134  | Prob > F      | = | 0.0000 |
|          |            |       |            | R-squared     | = | 0.0612 |
|          |            |       |            | Adj R-squared | = | 0.0600 |



Total | 566.27126 2,292 .24706425 Root MSE = .48192

| ----- | -----     | -----     | ----- | ----- | -----                | -----     |
|-------|-----------|-----------|-------|-------|----------------------|-----------|
| mwarm | Coef.     | Std. Err. | t     | P> t  | [95% Conf. Interval] |           |
| ----- | -----     | -----     | ----- | ----- | -----                | -----     |
| age   | .0054174  | .0006398  | 8.47  | 0.000 | .0041628             | .006672   |
| ed    | -.0151384 | .0041201  | -3.67 | 0.000 | -.023218             | -.0070588 |
| prst  | -.0017211 | .0008594  | -2.00 | 0.045 | -.0034064            | -.0000357 |
| _cons | .4544887  | .0567739  | 8.01  | 0.000 | .3431551             | .5658223  |
| ----- | -----     | -----     | ----- | ----- | -----                | -----     |

- b. Interpretasikanlah arah tanda dan nilai koefisien pada model diatas ini! **(5 poin)**
- c. Jelaskan minimal dua kekurangan *Linear Probability Model* terkait dengan nilai *predicted probability*,  $Pr(Y_i = 1 | X_i)$ ! Berikan satu contoh untuk mendukung jawaban anda dengan menggunakan hasil run stata diatas! **(5 poin)**
- d. Jelaskan alternatif metode estimasi ekonometri lainnya yang dapat memperbaiki kekurangan dari *Linear Probability Model* tersebut! Jelaskan mengapa metode yang anda sarankan akan dapat memperbaiki kekurangannya? Jelaskan dengan cukup detail, ringkas dan padat! **(5 poin)**
2. Dengan menggunakan data *loanapp.dta*, Anda diminta untuk membahas peluang aplikasi pinjaman seseorang diterima oleh pihak bank yang dipengaruhi oleh status demografi dan sosio-ekonominya ( $Pr[approve|X]$ ). Variabel yang terdapat dalam *loanapp.dta* adalah sebagai berikut:

- approve*** : Loan approval from bank; 1=Approved, 0=Not Approved
- married*** : Marital status; 1=Married, 0=Otherwise
- appinc*** : Income of Loan applicant (USD '000)
- male*** : 1=Male, 0=Female
- white*** : Race; 1=White, 0=Otherwise

Adapun hasil estimasi model di atas dengan menggunakan model Logit adalah sebagai berikut:



```
. logit approve married appinc male white

Iteration 0:  log likelihood = -737.97933
Iteration 1:  log likelihood = -701.47807
Iteration 2:  log likelihood = -692.53653
Iteration 3:  log likelihood = -692.50501
Iteration 4:  log likelihood = -692.50501
```

```
Logistic regression              Number of obs   =    1971
                                LR chi2(4)       =    90.95
                                Prob > chi2        =    0.0000
Log likelihood = -692.50501      Pseudo R2      =    0.0616
```

|         | Coef.     | Std. Err. | z     | P> z  | [95% Conf. Interval] |
|---------|-----------|-----------|-------|-------|----------------------|
| married | .4478597  | .1532945  | 2.92  | 0.003 | .1474081 .7483113    |
| appinc  | -.001232  | .0006872  | -1.79 | 0.073 | -.0025788 .0001149   |
| male    | -.1375788 | .1880388  | -0.73 | 0.464 | -.506128 .2309703    |
| white   | 1.453855  | .1533245  | 9.48  | 0.000 | 1.153344 1.754365    |
| _cons   | .7757226  | .1939455  | 4.00  | 0.000 | .3955963 1.155849    |

```
Marginal effects after logistic
y = Pr(approve) (predict)
= .88943801
```

| variable | dy/dx     | Std. Err. | z     | P> z  | [ 95% C.I. ]     | X       |
|----------|-----------|-----------|-------|-------|------------------|---------|
| married* | .0466843  | .01679    | 2.78  | 0.005 | .013775 .079594  | .659564 |
| appinc   | -.0001211 | .00007    | -1.80 | 0.073 | -.000253 .000011 | 84.8767 |
| male*    | -.0130846 | .01728    | -0.76 | 0.449 | -.04695 .020781  | .813293 |
| white*   | .2080362  | .02752    | 7.56  | 0.000 | .154105 .261968  | .846271 |

(\*) dy/dx is for discrete change of dummy variable from 0 to 1

Logistic model for approve

| Classified | True |     | Total |
|------------|------|-----|-------|
|            | D    | ~D  |       |
| +          | 1725 | 243 | 1968  |
| -          | 2    | 1   | 3     |
| Total      | 1727 | 244 | 1971  |

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

|                               |            |        |
|-------------------------------|------------|--------|
| Sensitivity                   | Pr( +  D)  | 99.88% |
| Specificity                   | Pr( -  ~D) | 0.41%  |
| Positive predictive value     | Pr( D  +)  | 87.65% |
| Negative predictive value     | Pr( ~D  -) | 33.33% |
| False + rate for true ~D      | Pr( +  ~D) | 99.59% |
| False - rate for true D       | Pr( -  D)  | 0.12%  |
| False + rate for classified + | Pr( ~D  +) | 12.35% |
| False - rate for classified - | Pr( D  -)  | 66.67% |
| Correctly classified          |            | 87.57% |

Berdasarkan hasil estimasi di atas, Anda diminta untuk menjawab beberapa pertanyaan berikut:

- Interpretasikan makna koefisien dari variabel *married* dan *male*! (5 poin)
- Apakah Pseudo-R<sup>2</sup> merupakan indikator yang tepat untuk mengukur keakuratan model? Jelaskan! (5 poin)
- Apa makna dari *correctly classified*? Jelaskan (5 poin)



- d. Apabila ada 2 orang memiliki *income* yang sama dan berasal dari ras kulit putih, hitunglah perbedaan peluang antara kedua orang itu jika orang pertama sudah menikah dan berjenis kelamin wanita dan orang kedua belum menikah dan berjenis kelamin pria! **(5 poin)**

**Soal 3. (20 poin)**

Seorang peneliti ilmu politik menganalisis tentang voting terhadap seorang kandidat kepala daerah melakukan survey terhadap 1.825 pemilih di suatu daerah dan menanyakan pilihan mereka terhadap kandidat tersebut. Sebuah variabel sederhana dibuat  $support=1$  bagi mereka yang memilih kandidat tersebut dan  $support=0$  bagi mereka yang tidak memilih kandidat tersebut. Diduga bahwa para pendukung kandidat tersebut akan didominasi oleh para penduduk tua dan mereka yang memiliki tingkat pendidikan rendah namun berpendapatan tinggi. Untuk menguji dugaan tersebut, juga dikumpulkan data mengenai umur (*age*) para pemilih (dalam tahun), pendidikan atau *education* (dalam tahun lamanya sekolah), dan pendapatan atau *income* (dalam jutaan Rupiah per tahun). Ide awal yang dibangun oleh si peneliti tersebut adalah berupa model sebagai berikut:

$$\Pr(support = 1) = F(x'\beta), \text{ dimana } F(\bullet) \text{ adalah sebuah fungsi, dan}$$
$$x'\beta = \beta_1 + \beta_2 age + \beta_3 education + \beta_4 \log(income)$$

Pertanyaan :

- Jelaskan bagaimanakah menginterpretasikan koefisien *age*,  $\beta_2$ , jika kita hanya melakukan regresi dari variabel *support* terhadap variabel-variabel lainnya dengan menggunakan metode OLS? **(5 poin)**
- Apakah masalah-masalah yang mungkin muncul dari upaya melakukan regresi dengan menggunakan metode OLS terhadap model di atas? **(5 poin)**
- Apabila kemudian kita ingin melakukan estimasi dengan menggunakan model logit, sedemikian  $\Pr(support = 1) = \frac{\exp(x'\beta)}{1 + \exp(x'\beta)}$ . Jelaskan bagaimanakah menginterpretasikan koefisien *age*,  $\beta_2$ , dari pendekatan ini? **(5 poin)**

Jelaskan mengenai konstruksi/rumus kriteria goodness-of-fit (seperti  $R^2$  di dalam OLS) [Hint: jelaskan minimal 2 ukuran-ukuran yang Anda ketahui] **(5 poin)**

**Soal 4. (40 poin)**

Saat ini, Ehsan sedang mengerjakan proyek mengenai kondisi ketimpangan di Indonesia. Ia terinspirasi dari bulletin yang ditulis oleh staf peneliti di International Monetary Fund (IMF) yang



mengatakan bahwa salah satu penyebab meningkatnya ketimpangan di Amerika Serikat adalah menurunnya partisipasi tenaga kerja dalam mengikuti organisasi serikat pekerja. Hal itu menyebabkan tenaga kerja tidak memiliki *bargaining power* yang cukup kuat dihadapan para *employer* dalam meningkatkan taraf hidup mereka.

Namun, sebelum melanjutkan penelitiannya lebih dalam lagi, Ehsan ingin mengetahui faktor-faktor apa saja yang dapat membuat seorang pekerja bergabung dalam organisasi serikat pekerja. Model yang ia gunakan dalam penelitiannya adalah:

$$\Pr(\text{union} = 1) = f(\text{age}, \text{yeduc}, \text{urban}, \text{black})$$

Dengan: *union* : 1=anggota serikat pekerja, 0=bukan anggota serikat pekerja  
*age* : umur pekerja (dalam tahun)  
*yeduc* : lamanya sekolah (dalam tahun)  
*urban* : 1=bekerja di kota, 0=tidak bekerja di kota  
*black* : 1=ras *black*, 0=bukan ras *black*

Dalam penelitiannya, Ehsan mengestimasi model ekonometrikanya dengan *probit* dan *logit*.  
Pertanyaan:

- Apa saja persamaan dan perbedaan dalam *logit* dan *probit* dalam mengestimasi *binary dependent variable*? (10 poin)
- Dengan menggunakan *logit*, berapakah peluang seorang pekerja untuk bergabung ke dalam organisasi serikat pekerja ketika ia adalah tamatan SMA, berumur 26 tahun, tinggal di urban, dan berasal dari ras *black*? Bandingkanlah juga dengan hasil menggunakan *probit*! Apakah hasil estimasi *logit* dan *probit* berbeda? (10 poin)
- Ehsan memutuskan untuk menambah variabel baru pada modelnya. Model yang baru adalah:

$$\Pr(\text{union} = 1) = f(\text{age}, \text{yeduc}, \text{urban}, \text{black}, \text{senior})$$

Dengan *senior*: 1=umur pekerja  $\geq 40$  tahun, 0=sebaliknya. Ia menggunakan *logit* dalam estimasinya. Menurut anda, apakah apabila seorang pekerja sudah senior akan memengaruhinya untuk bergabung dalam anggota serikat pekerja? Lalu, adanya variabel *senior* menambah derajat ketepatan model secara signifikan? Berikan pendapat anda! (20 poin)





## 2) Hasil Regresi Logit (*old*)

```

Logistic regression                               Number of obs   =   26200
                                                  LR chi2(4)      =   507.36
                                                  Prob > chi2     =   0.0000
Log likelihood = -13610.55                    Pseudo R2       =   0.0183
  
```

| union | Coef.    | Std. Err. | z      | P> z  | [95% Conf. Interval] |
|-------|----------|-----------|--------|-------|----------------------|
| age   | .0097576 | .0023515  | 4.15   | 0.000 | .0051488 .0143664    |
| yeduc | .0786546 | .0064382  | 12.22  | 0.000 | .066036 .0912733     |
| urban | .274689  | .0352925  | 7.78   | 0.000 | .2055171 .343861     |
| black | .5547836 | .0326907  | 16.97  | 0.000 | .4907111 .6188562    |
| _cons | -2.93562 | .1079878  | -27.18 | 0.000 | -3.147272 -2.723968  |

Marginal effects after logit  
 $y = \text{Pr}(\text{union})$  (predict, p)  
 = .14832169

| variable | dy/dx    | Std. Err. | z     | P> z  | [ 95% C.I. ]    | X  |
|----------|----------|-----------|-------|-------|-----------------|----|
| age      | .0012326 | .00029    | 4.30  | 0.000 | .000671 .001794 | 25 |
| yeduc    | .0099358 | .00083    | 12.03 | 0.000 | .008317 .011555 | 12 |
| urban*   | .0381447 | .00474    | 8.05  | 0.000 | .028853 .047436 | 0  |
| black*   | .0843937 | .0056     | 15.06 | 0.000 | .073409 .095378 | 0  |

(\*) dy/dx is for discrete change of dummy variable from 0 to 1

Logistic model for union

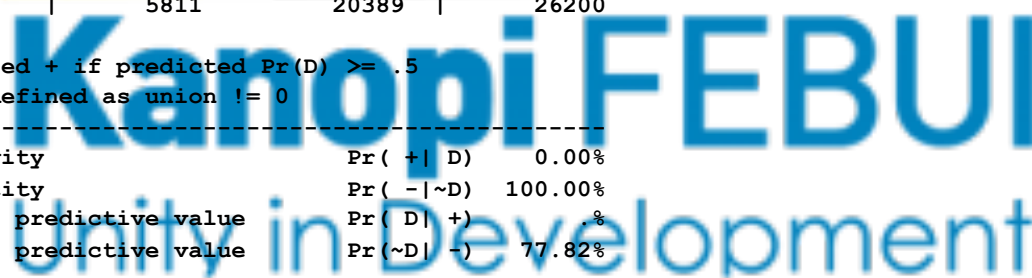
| Classified | True |       | Total |
|------------|------|-------|-------|
|            | D    | ~D    |       |
| +          | 0    | 0     | 0     |
| -          | 5811 | 20389 | 26200 |
| Total      | 5811 | 20389 | 26200 |

Classified + if predicted  $\text{Pr}(D) \geq .5$   
 True D defined as union != 0

|                           |             |         |
|---------------------------|-------------|---------|
| Sensitivity               | Pr( +   D)  | 0.00%   |
| Specificity               | Pr( -   ~D) | 100.00% |
| Positive predictive value | Pr( D   +)  | .%      |
| Negative predictive value | Pr( ~D   -) | 77.82%  |

|                               |             |         |
|-------------------------------|-------------|---------|
| False + rate for true ~D      | Pr( +   ~D) | 0.00%   |
| False - rate for true D       | Pr( -   D)  | 100.00% |
| False + rate for classified + | Pr( ~D   +) | .%      |
| False - rate for classified - | Pr( D   -)  | 22.18%  |

Correctly classified 77.82%







### 3) Hasil Regresi Logit (*new*)

```

Logistic regression
Log likelihood = -13608.865
Number of obs   =    26200
LR chi2(5)      =    510.73
Prob > chi2     =    0.0000
Pseudo R2      =    0.0184
  
```

| union  | Coef.     | Std. Err. | z      | P> z  | [95% Conf. Interval] |          |
|--------|-----------|-----------|--------|-------|----------------------|----------|
| age    | .0125014  | .0027888  | 4.48   | 0.000 | .0070354             | .0179674 |
| yeduc  | .0780103  | .0064452  | 12.10  | 0.000 | .065378              | .0906427 |
| urban  | .2763247  | .0353066  | 7.83   | 0.000 | .207125              | .3455244 |
| black  | .5520489  | .032724   | 16.87  | 0.000 | .487911              | .6161868 |
| senior | -.1145139 | .0626302  | -1.83  | 0.067 | -.2372668            | .008239  |
| _cons  | -3.001562 | .1139467  | -26.34 | 0.000 | -3.224893            | -2.77823 |

Logistic model for union

| Classified | True |       | Total |
|------------|------|-------|-------|
|            | D    | ~D    |       |
| +          | 0    | 0     | 0     |
| -          | 5811 | 20389 | 26200 |
| Total      | 5811 | 20389 | 26200 |

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

|                               |            |         |
|-------------------------------|------------|---------|
| Sensitivity                   | Pr( +  D)  | 0.00%   |
| Specificity                   | Pr( -  ~D) | 100.00% |
| Positive predictive value     | Pr( D  +)  | .%      |
| Negative predictive value     | Pr( ~D  -) | 77.82%  |
| False + rate for true ~D      | Pr( +  ~D) | 0.00%   |
| False - rate for true D       | Pr( -  D)  | 100.00% |
| False + rate for classified + | Pr( ~D  +) | .%      |
| False - rate for classified - | Pr( D  -)  | 22.18%  |
| Correctly classified          |            | 77.82%  |

**Kanopi FEBUI**  
Unity in Development

----- Selamat Mengerjakan, Semoga Sukses! -----