



UJIAN TENGAH SEMESTER
FAKULTAS EKONOMI UNIVERSITAS INDONESIA

Mata Kuliah : Ekonometrika 2 (CSPD) - **Kelas A**
Program : S1 Reguler
Tanggal Ujian : 22 Oktober 2018
Waktu : 150 Menit (2 ½ Jam)
Sifat Ujian : Closed Book/Closed Notes
Pengajar : Dhaniel Ilyas

Petunjuk Umum:

1. Perhatikan, lembar soal ini terdiri dari **7 Halaman** dan **5 Soal**. Terdapat dua kelompok soal: soal wajib/parallel dan soal pilihan. Semua soal wajib/parallel harus dijawab. Sementara soal pilihan, cukup pilih dan kerjakan dua soal dari tiga soal yang tersedia. Bobot penilaian untuk setiap soal tersedia di akhir pertanyaan.
2. **Boleh** menggunakan kalkulator. **Tidak boleh** menggunakan telepon seluler/komputer/laptop sebagai kalkulator.
3. Sifat tutup buku (*closed book*) dan tutup catatan (*closed notes*)

SOAL 1 dan SOAL 2 adalah SOAL WAJIB

Soal 1. (25 poin)

1. Perhatikan Spesifikasi *Linear Probability Model* berikut:

$$Y_i = Pr(Y_i = 1|X_i) + \epsilon_i = E(Y_i|X_i) + \epsilon_i = Z_i + \epsilon_i$$

$$\text{Dimana } Z_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_m X_{im}$$

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.

- a. Jelaskan bahwa nilai ϵ_i hanya mempunyai dua kemungkinan. Spesifikasikan kedua kemungkinan ini dengan menggunakan variabel Z_i ! (**5 poin**)
- b. Buktikan nilai $var(\epsilon_i) = Z_i(1 - Z_i)$! (Hint: Gunakan rumus varians: $var(X) = E[X - E(X)]^2$ dan rumus Expected Value.) Misalkan seorang peneliti menggunakan metode OLS untuk mengestimasi nilai-nilai parameter pada



spesifikasi *Linear Probability Model* diatas. Validkah penggunaan metode OLS ini?
Jelaskan dengan singkat dan padat. (10 poin)

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

$$BELI = \beta_0 + \beta_1 TAMB + \epsilon_i$$

Dimana $BELI = 1$ jika seorang individu membeli mobil teknologi *hybrid* dan 0 jika tidak membelinya; dan $TAMB =$ Biaya tambahan harga mobil *hybrid* secara relatif dibandingkan mobil standar (dalam ribu dollar)

regress BELI TAMB

Source	SS	df	MS			
Model	40.2272555	1	40.2272555	Number of obs =	1000	
Residual	1271713.21	998	.193259263	F(3, 309) =	208.15	
				Prob > F	= 0.0000	
				R-squared	= 0.1726	
				Adj R-squared	= 0.1717	
				Root MSE	= .43961	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
TAMB	-.2810605	.0194809	-14.43	0.000	-.3192888	-.2428322
_cons	.9888035	.0450873	21.93	0.000	.9003267	1.07728

- Apakah secara umum menurut anda hasil run spesifikasi diatas ini sudah cukup baik? (Hint: Perhatikan R^2 , t-test dan nilai koefisien). Interpretasikan nilai koefisien pada model diatas ini! (5 poin)
- Jelaskan kekurangan *Linear Probability Model* yang paling penting terkait dengan nilai *predicted probability*, $Pr(Y_i = 1|X_i)$! Ilustrasikan kekurangan ini dengan hasil run model *Linear Probability Model* pada pertanyaan c.! (Hint: Gunakan sembarang nilai TAMB yang masuk akal) (5 poin)

Soal 2. (25 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 , β_2 , jika kita hanya melakukan regresi dari variabel $support$ terhadap variabel-variabel lainnya dengan menggunakan metode OLS? (5 poin)
- Apabila kemudian kita ingin melakukan estimasi dengan menggunakan model logit, sedemikian $Pr(support = 1) = \exp(x'\beta) / [1 + \exp(x'\beta)]$. Jelaskan bagaimanakah menginterpretasikan koefisien age , β_2 , dari pendekatan ini? (10 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)
- Misalkan, diusulkan untuk melakukan estimasi ulang terhadap model di atas menggunakan model Probit. Aspek-aspek apakah dari hasil estimasi yang akan berbeda dan aspek-aspek apakah yang tidak berbeda? (5 poin)

SOAL 3 sampai dengan SOAL 5 adalah SOAL PILIHAN. Pilihlah 2 diantara 3 soal dibawah. Pastikan nomor soal dan jawaban Anda sudah sesuai

Soal 3. (25 poin)

Vira dari Kementerian Kesehatan ingin meneliti faktor yang mempengaruhi seorang anak mengalami stunting (tinggi anak berada di bawah -2 standar deviasi dari tinggi median referensi WHO). Diketahui bahwa stunting dapat terjadi karena berbagai faktor seperti faktor lingkungan, nutrisi, dan genetik. Dengan mempertimbangkan seluruh faktor tersebut, Vira akhirnya merumuskan model dari stunting anak sebagai berikut:

$$Pr(stunting = 1) = f(\lnjarakair, Proteinpc, tinggiIbu, tinggiBpk)$$

Dengan: Stunting: 1 = Anak mengalami Stunting, 0 = Anak tidak mengalami stunting
Lnjarakair: Jarak sumber air bersih ke rumah (Ln)
Proteinpc: Konsumsi Protein per kapita (juta rupiah)
TinggiIbu: Tinggi ibu (cm)
TinggiBpk: Tinggi ayah (cm)

Dalam penelitiannya, peneliti menggunakan model ekonometrika LPM dan logit. Pertanyaan:

- Apakah anda ada masukan kepada peneliti mengenai penggunaan model LPM? Apa saja kelemahan dan kelebihan dari menggunakan LPM? Apakah LPM dapat menjadi aproksimasi awal dari hubungan antar variabel? Apakah sebaiknya peneliti tetap menggunakan model LPM atau adakah model alternatif? (5 poin)
- Dari hasil model logit, hitunglah Pseudo-R²! Apakah pseudo-R² merupakan goodness-of-fit yang baik untuk model Logit? Apakah ada alternatif dari pseudo-R²? (5 poin)
- Dari hasil klasifikasi oleh model Logit, apa makna dari *correctly classified*? Bagaimana mendapatkan angka 76.03%? (5 poin)
- Menggunakan hasil LPM dan logit, hitunglah probabilitas anak stunting ($Pr(Y = Stunting)$) apabila jarak air ke rumah adalah 1000m, pengeluaran protein percapita sebesar Rp 500.000, tinggi ibu 180 cm, dan tinggi ayah 203 cm! Apakah asumsi Kolgomorov terpenuhi? (5 poin)
- Berdasarkan hasil logit, berapakah peluang seorang anak mengalami stunting dengan karakteristik pada poin (d)! Interpretasikan! (5 poin)



Lampiran Soal 3.

Hasil regresi model LPM

Source	SS	df	MS	Number of obs =	2587
Model	20.5430924	4	5.13577309	F(4, 2582) =	29.86
Residual	444.033637	2582	.171972749	Prob > F =	0.0000
				R-squared =	0.0442
				Adj R-squared =	0.0427
Total	464.57673	2586	.179650708	Root MSE =	.4147

stunting	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lnjarakair	.0241926	.004591	5.27	0.000	.0151901 .0331951
Proteinpc	-.7860655	.1360355	-5.78	0.000	-1.052815 -.5193158
TinggiIbu	-.0035362	.0008538	-4.14	0.000	-.0052104 -.001862
TinggiBpk	-.0028057	.0005718	-4.91	0.000	-.003927 -.0016844
_cons	1.236205	.1561617	7.92	0.000	.9299906 1.54242

Hasil regresi model logit

Logistic regression	Number of obs =	2587
	LR chi2(4) =	120.31
	Prob > chi2 =	0.0000

Log likelihood = -1349.283

stunting	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lnjarakair	.1253008	.0247225	5.07	0.000	.0768455 .1737561
Proteinpc	-6.968696	1.151818	-6.05	0.000	-9.226218 -4.711175
TinggiIbu	-.0171841	.0047252	-3.64	0.000	-.0264454 -.0079228
TinggiBpk	-.0133475	.0029655	-4.50	0.000	-.0191597 -.0075352
_cons	3.739343	.8505596	4.40	0.000	2.072277 5.406409

Klasifikasi dari model Logit

Logistic model for stunting

Classified	True		Total
	D	~D	
+	12	25	37
-	595	1955	2550
Total	607	1980	2587

Classified + if predicted Pr(D) >= .5
True D defined as stunt14 != 0

Sensitivity	Pr(+ D)	1.98%
Specificity	Pr(- ~D)	98.74%
Positive predictive value	Pr(D +)	32.43%
Negative predictive value	Pr(~D -)	76.67%
False + rate for true ~D	Pr(+ ~D)	1.26%
False - rate for true D	Pr(- D)	98.02%
False + rate for classified +	Pr(~D +)	67.57%
False - rate for classified -	Pr(D -)	23.33%
Correctly classified		76.03%



Soal 4. (25 poin)

Cholil ingin meneliti mengenai salah satu program bantuan sosial dari pemerintah. Dikarenakan ketertarikannya pada dunia pendidikan, Cholil memutuskan untuk meneliti mengenai program BSM (Bantuan Siswa Miskin). BSM merupakan salah satu program bantuan sosial pemerintah yang diberikan langsung kepada siswa SD hingga SMA dengan besarnya bergantung pada jenjang pendidikan yang sedang diambil. Cholil ingin meneliti apakah dengan diberikannya BSM akan meningkatkan angka partisipasi sekolah murni pada level SMA dimana sampelnya adalah anak-anak yang berusia 15-18 tahun. Akhirnya Cholil membuat model sebagai berikut:

$$Pr(SMA = 1) = f(\text{penerimaBSMSMA}, \text{nilaiBSM}, \text{nikah}, \text{avyosortu}, \text{poor})$$

Dimana: SMA: 1 = Sedang Sekolah pada jenjang SMA, 0 = selain itu
 PenerimaBSMSMA: 1 = Penerima BSM SMA, 0 = non-penerima BSM SMA
 NilaiBSM = Nilai BSM (Rupiah)
 Nikah: 1 = Menikah, 0 = Belum Menikah/Cerai
 Avyosortu = Rata-rata tingkat pendidikan Orang Tua (Tahun)
 Poor: 1 = Miskin, 0 = Tidak miskin

Dalam penelitiannya, Cholil mengestimasi model ekonometrika dengan logit. Pertanyaan:

- Apakah terjadi *mistargeting* dari BSM? Berapa persentase *mistargeting* dari BSM? Berapakah peluang siswa miskin akan mendapatkan BSM? (5 poin)
- Apabila dilakukan perhitungan dari *odds ratio* dari regresi logit tidak terdapat iterasi seperti pada regresi logit biasa. Mengapa tidak terdapat iterasi tersebut? (5 poin)
- Hitunglah *odds ratio* dari variabel nilaiBSM dan nikah! Interpretasikan! (5 poin)
- Cholil mencoba untuk menggunakan model probit sebagai alternatif. Apakah perbedaan dan persamaan model probit dengan model logit? (5 poin)
- Interpretasikan *marginal effect* kedua tabel untuk variabel penerimaBSMSMA dan nilaiBSM! Mengapa terjadi perbedaan angka diantara kedua hasil tersebut? Apakah penerima BSM SMA akan bersekolah di level SMA? (5 poin)

Lampiran Soal 4

Tabulasi Silang Penerima BSM dengan Status Kemiskinan

poor	penerimaBSMSMA		Total
	0	1	
0	64,023	6,742	70,765
1	10,077	1,268	11,345
Total	74,100	8,010	82,110

Hasil Odds Ratio Regresi Logit

Logistic regression	Number of obs	=	82110
	LR chi2(5)	=	9015.08
	Prob > chi2	=	0.0000
Log likelihood = -50921.961	Pseudo R2	=	0.0813



SMA	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
penerimabsmSMA	4.695604	.4422582	16.42	0.000	3.904097	5.64758
nilaibsmSMA		1.04e-07	2.46	0.014		
nikah		.0151893	-23.75	0.000		
avyosortu	1.118688	.0023479	53.44	0.000	1.114096	1.123299
poor	.4676545	.0102432	-34.70	0.000	.448003	.4881681
_cons	.6724628	.0115249	-23.15	0.000	.6502496	.6954349

Hasil regresi Probit

Iteration 0: log likelihood = -55429.502
 Iteration 1: log likelihood = -50964.715
 Iteration 2: log likelihood = -50947.357
 Iteration 3: log likelihood = -50947.345
 Iteration 4: log likelihood = -50947.345

Probit regression

Number of obs = 82110
 LR chi2(5) = 8964.31
 Prob > chi2 = 0.0000
 Pseudo R2 = 0.0809

Log likelihood = -50947.345

SMA	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
penerimabsmSMA	.9037	.050109	18.03	0.000	.8054881	1.001912
nilaibsmSMA	1.39e-07	5.47e-08	2.54	0.011	3.17e-08	2.46e-07
nikah	-.7392376	.0301401	-24.53	0.000	-.7983112	-.6801641
avyosortu	.0679835	.0012578	54.05	0.000	.0655183	.0704487
poor	-.4658129	.0133729	-34.83	0.000	-.4920234	-.4396025
_cons	-.2364445	.0105001	-22.52	0.000	-.2570243	-.2158647

Hasil Average Marginal Effect Regresi Probit

Average marginal effects
 Model VCE : OIM

Number of obs = 82110

Expression : Pr(SMA), predict()
 dy/dx w.r.t. : penerimabsmSMA nilaibsmSMA nikah avyosortu poor

	dy/dx	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
penerimabsmSMA	.320434	.0176746	18.13	0.000	.2857923	.3550757
nilaibsmSMA	4.93e-08	1.94e-08	2.54	0.011	1.12e-08	8.73e-08
nikah	-.2621189	.0105706	-24.80	0.000	-.2828368	-.241401
avyosortu	.0241056	.0004222	57.10	0.000	.0232781	.0249331
poor	-.165168	.004635	-35.63	0.000	-.1742525	-.1560835

