Continuous Assessment in Applied Econometrics – 2H

Anderson Vil
CY Cergy Paris Université
2026 – 2027
Master 1 – Mock exam 3

Answer the questions in order. All answers must be justified. Calculators are allowed. Throughout, log denotes the natural logarithm. Unless stated otherwise, the assumptions of the classical linear model, including normally distributed errors, are taken to hold. Critical values are read from the tables at the end of the paper. When the number of degrees of freedom is not printed, use the closest lower row, and use the row ∞ above 120. The marking scheme is indicative.

Exercise 1: An intern's report on umbrella sales ***(4 points)***

A company runs umbrella kiosks in Paris train stations. An intern studied $N = 20$ autumn days: x_i is the rainfall of the day (mm) and y_i the number of umbrellas sold. He kept the following sums:

$$\sum_i x_i = 60, \qquad \sum_i y_i = 400, \qquad \sum_i x_i^2 = 380, \qquad \sum_i y_i^2 = 8{,}900, \qquad \sum_i x_i y_i = 1{,}560.$$

His report on the OLS regression $y_i = \beta_0 + \beta_1 x_i + u_i$ contains the following statements:

(R1) $\hat{\beta}_0 = 14.6$, $\hat{\beta}_1 = 1.8$ and $R^2 = 0.72$.

(R2) "The residuals of my regression sum to 4.2."

(R3) "When sales are counted in dozens of umbrellas, the R^2 falls to 0.06: the model then explains almost nothing."

(R4) "Without rain, nobody should buy an umbrella, so I also estimated the model without constant. Its SSR is 200, lower than the SSR of the model with constant (252): the model without constant fits better."

1. Check statement (R1) from the sums, showing your calculations.
2. For each of the statements (R2), (R3) and (R4), say whether the reported result is possible. Justify your answer with a property of OLS, and give the correct value when the reported one is wrong.
3. The intern concludes: "each additional millimetre of rain causes 1.8 extra umbrella sales". In two or three lines, is a causal reading more or less credible here than in most regressions estimated on observational data?

Exercise 2: A new substrate for tomato plants ***(5 points)***

An agronomy laboratory tests a new growing substrate on $N = 6$ tomato plants. Plants 1 to 3 are grown in the standard substrate and plants 4 to 6 in the new one. The yields (kg per plant) are

$$y = (2,\ 3,\ 4,\ 5,\ 5,\ 8)'.$$

Let $d_i = 1$ if plant i is grown in the new substrate and 0 otherwise, and consider the model $y_i = \beta_0 + \beta_1 d_i + u_i$, with $X = [\iota,\ d]$ where ι is a vector of ones.

1. Compute $X'X$, $(X'X)^{-1}$ and $\hat{\beta}$. Interpret $\hat{\beta}_0$ and $\hat{\beta}_1$.
2. Show that the projection matrix $P = X(X'X)^{-1}X'$ is block diagonal, with two 3×3 blocks whose elements all equal $1/3$. Deduce $\hat{y} = Py$ without using $\hat{\beta}$, and check that $P^2 = P$.

3. Compute $\hat{u} = My$, check that $X'\hat{u} = 0$ and interpret its two equations. Compute $\hat{u}'\hat{u}$.

4. Compute the trace of P and deduce the trace of M. Use it to compute $\hat{\sigma}^2$, then the standard error of $\hat{\beta}_1$. Test at the 5% level the hypothesis that the new substrate increases the yield. Would a two-sided test lead to the same conclusion?

5. A technician wants to add the column $s = \iota - d$ (equal to 1 for the standard substrate) to X. In two or three lines, explain why this is impossible.

Exercise 3: Taxi fares, in euros and in dollars ***(5 points)***

A city's taxi regulator studies $N = 400$ rides. It estimates by OLS (standard errors in parentheses):

$$\widehat{fare} = \underset{(0.35)}{3.80} + \underset{(0.05)}{1.45}\, dist + \underset{(0.08)}{0.40}\, wait, \qquad R^2 = 0.88,$$

where *fare* is the fare paid (euros), *dist* the distance (km) and *wait* the number of minutes spent stopped in traffic. In the sample, the mean distance is 8 km and the mean waiting time is 5 minutes.

1. Interpret the coefficients on *dist* and *wait*. The official tariff is 1.50 euros per km: test, at the 5% level and against a two-sided alternative, whether the estimated rate per km is compatible with it.

2. A travel website re-estimates the equation with the fare in dollars (1 euro = 1.10 dollar) and the distance in miles (1 mile = 1.609 km), the waiting time being unchanged. Give the new coefficients and standard errors. What happens to the t statistics and to the R^2? Justify.

3. The regulator also estimates $\widehat{\log(fare)} = 1.10 + \underset{(0.02)}{0.78}\log(dist)$, $R^2 = 0.79$. Without re-estimating, give the coefficients and the standard error of the slope in the regression of $\log(fare/dist)$ on a constant and $\log(dist)$. Interpret this slope.

4. Test at the 5% level the hypothesis that fares are proportional to distance.

5. Using the level equation, compute the elasticity of the fare with respect to distance at the mean point of the sample. In two or three lines, explain why both this elasticity and the slope of question 3 are below one.

Exercise 4: The diffusion of electric vehicles across départements ***(6 points)***

A think tank studies the 96 départements of metropolitan France in 2025. The dependent variable is the share of electric vehicles in new private car registrations (in %). The regressors are *chargers*, the number of public charging points per 10,000 inhabitants, the log of the median disposable income (euros), *urban*, the share of the population living in urban areas (%), and *mount*, a dummy equal to 1 for mountain départements. Column (4) reports the regression of $\log(inc)$ on a constant and *chargers*.

Table 1 – Share of electric vehicles in registrations, OLS estimates

	(1)	(2)	(3)	(4)
Dependent variable	*EVshare*	*EVshare*	*EVshare*	$\log(inc)$
chargers	0.981	0.486	0.503	0.0662
	(0.075)	(0.112)	(0.111)	(0.0050)
$\log(inc)$		7.465	6.995	
		(1.366)	(1.405)	
urban			0.005	
			(0.009)	
mount			−0.517	
			(0.237)	
Constant	0.668	−72.381	−67.936	9.786
	(0.258)	(13.369)	(13.619)	(0.017)
N	96	96	96	96
R^2	0.644	0.731	0.746	0.654

Notes: standard errors in parentheses. *EVshare* is measured in % (e.g. 3.5 for 3.5%).

We denote by $\tilde{\beta}_1$ the coefficient on *chargers* in column (1), by $\hat{\beta}_1$ and $\hat{\beta}_2$ the coefficients on *chargers* and $\log(inc)$ in column (2), and by $\tilde{\delta}_1$ the coefficient on *chargers* in column (4).

1. Interpret the coefficients on *chargers* and on $\log(inc)$ in column (2); for the latter, compute the effect of a 10% higher median income.

2. Test, in column (2) and at the 5% level, the hypothesis that charging points raise the EV share. State H_0 and H_1.

3. Using only the R^2 of column (3), test at the 5% level the joint significance of its regressors.

4. (a) Test whether *urban* and *mount* can be jointly excluded from column (3), at the 5% and at the 10% level.

(b) Show that *mount* is individually significant at the 5% level (two-sided). In two lines, explain how this can be reconciled with your answer to (a).

5. (a) Interpret $\tilde{\delta}_1$.

(b) Verify numerically the relation $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2\tilde{\delta}_1$.

(c) Is the effect of charging points overestimated or underestimated in column (1)? By how much? Explain in one sentence.

6. In two or three lines each:

(a) Rural départements (low *urban*) register few new cars, so their EV share is measured from few registrations and varies much more, for given values of the other regressors. What does this imply for your answers to questions 2 to 4?

(b) Operators install charging points where electric vehicles are already numerous. What does this imply for the interpretation of the coefficient on *chargers* in column (3)?

TABLE G.2 Critical Values of the *t* Distribution

		Significance Level				
1-Tailed:		.10	.05	.025	.01	.005
2-Tailed:		.20	.10	.05	.02	.01
Degrees of Freedom	1	3.078	6.314	12.706	31.821	63.657
	2	1.886	2.920	4.303	6.965	9.925
	3	1.638	2.353	3.182	4.541	5.841
	4	1.533	2.132	2.776	3.747	4.604
	5	1.476	2.015	2.571	3.365	4.032
	6	1.440	1.943	2.447	3.143	3.707
	7	1.415	1.895	2.365	2.998	3.499
	8	1.397	1.860	2.306	2.896	3.355
	9	1.383	1.833	2.262	2.821	3.250
	10	1.372	1.812	2.228	2.764	3.169
	11	1.363	1.796	2.201	2.718	3.106
	12	1.356	1.782	2.179	2.681	3.055
	13	1.350	1.771	2.160	2.650	3.012
	14	1.345	1.761	2.145	2.624	2.977
	15	1.341	1.753	2.131	2.602	2.947
	16	1.337	1.746	2.120	2.583	2.921
	17	1.333	1.740	2.110	2.567	2.898
	18	1.330	1.734	2.101	2.552	2.878
	19	1.328	1.729	2.093	2.539	2.861
	20	1.325	1.725	2.086	2.528	2.845
	21	1.323	1.721	2.080	2.518	2.831
	22	1.321	1.717	2.074	2.508	2.819
	23	1.319	1.714	2.069	2.500	2.807
	24	1.318	1.711	2.064	2.492	2.797
	25	1.316	1.708	2.060	2.485	2.787
	26	1.315	1.706	2.056	2.479	2.779
	27	1.314	1.703	2.052	2.473	2.771
	28	1.313	1.701	2.048	2.467	2.763
	29	1.311	1.699	2.045	2.462	2.756
	30	1.310	1.697	2.042	2.457	2.750
	40	1.303	1.684	2.021	2.423	2.704
	60	1.296	1.671	2.000	2.390	2.660
	90	1.291	1.662	1.987	2.368	2.632
	120	1.289	1.658	1.980	2.358	2.617
	∞	1.282	1.645	1.960	2.326	2.576

Examples: The 1% critical value for a one-tailed test with 25 *df* is 2.485. The 5% critical value for a two-tailed test with large (> 120) *df* is 1.96.

Source: This table was generated using the Stata® function invttail.

Copyright 2016 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE G.3a 10% Critical Values of the *F* Distribution

		Numerator Degrees of Freedom									
		1	2	3	4	5	6	7	8	9	10
Denominator Degrees of Freedom	10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
	11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25
	12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19
	13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14
	14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10
	15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
	16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03
	17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00
	18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98
	19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96
	20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94
	21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92
	22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90
	23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89
	24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88
	25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87
	26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86
	27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85
	28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84
	29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83
	30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82
	40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76
	60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71
	90	2.76	2.36	2.15	2.01	1.91	1.84	1.78	1.74	1.70	1.67
	120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65
	∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60

Example: The 10% critical value for numerator $df = 2$ and denominator $df = 40$ is 2.44.
Source: This table was generated using the Stata® function invFtail.

Copyright 2016 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE G.3b 5% Critical Values of the *F* Distribution

		Numerator Degrees of Freedom									
		1	2	3	4	5	6	7	8	9	10
Denominator Degrees of Freedom	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
	11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
	13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
	14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
	17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
	18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
	19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
	21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
	22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
	23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
	24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
	25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
	26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
	27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
	28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
	29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
	30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
	40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
	60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
	90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
	120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
	∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Example: The 5% critical value for numerator $df = 4$ and large denominator $df(\infty)$ is 2.37.

Source: This table was generated using the Stata® function invFtail.

Copyright 2016 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE G.3c 1% Critical Values of the *F* Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom: 1	2	3	4	5	6	7	8	9	10
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
90	6.93	4.85	4.01	3.54	3.23	3.01	2.84	2.72	2.61	2.52
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

Example: The 1% critical value for numerator $df = 3$ and denominator $df = 60$ is 4.13.

Source: This table was generated using the Stata® function invFtail.

Copyright 2016 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.