ØAMET2200 · Fall 2019
Mini Practice Final Exam

Instructor: Fenella Carpena

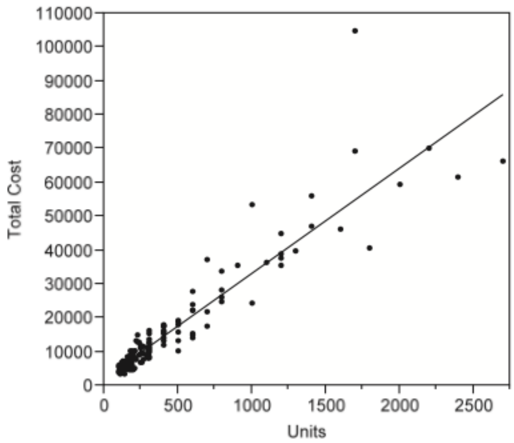
November 22, 2019

The questions here are designed to be completed within 1.5 hours. This mini practice final exam is therefore shorter than the actual final exam, and these practice questions are not meant to be representative of all topics that will be covered in the final exam.

Instructions: Show all relevant calculations and/or reasoning for each problem. For questions resulting in a final number or a short statement such as "we can reject the null at a 5% significance level," please draw a box around your final answer. Write clearly and legibly using normal-size text, and avoid excessively long or unfocused answers. Critical values are provided in the appendix tables attached.

Question 1 For each of the statements below, explain whether the statement is true, false, or uncertain. Provide a brief explanation, either mathematically or in words, for each of your answers. When providing an explanation in words, do not write more than three sentences.

(a) An analyst at a production facility for GPS units is interested in the relationship between weekly production (variable name: *Units*) and total cost of that production (variable name: *Total Cost*). A scatter plot of these two variables is shown below, together with the sample regression line. In this figure, we see evidence that the SRM condition of equal variances of the error term is violated.



(b) If the correlation between y and x is larger than 0.7, then a linear equation is appropriate to describe the association between the two variables.

(c) The presence of collinearity violates the multiple regression model (MRM) assumptions.

(d) Consider the simple regression model (SRM) $y = \beta_0 + \beta_1 x_1 + \epsilon$. Suppose we have an additional variable x_2 which is positively correlated with x_1. If we add x_2 in the regression, then the estimated coefficient on x_1 will decrease.

(e) Consider again the SRM in part (d). Because we did not include x_2 in the model, we are certain that x_2 is a confounding variable in the regression $y = \beta_0 + \beta_1 x_1 + \epsilon$.

(f) Composite panels used in the assembly of airplanes must meet certain specifications for thickness. When the process used by a manufacturer to produce these panels is working as designed, the mean thickness is 25 inches with a standard deviation of 0.4 inches. Every week, the quality control manager takes a simple random sample of panels. The manager then calculates the average thickness in the sample, to see if it is close to 25 inches.

Suppose that during the first week of a particular month, the quality control manager uses a sample of 25 panels and sets the control limits to 25 ± 0.24. During the second week of that month, the manager uses a sample of 36 panels and sets the control limits to 25 ± 0.13. In this case, the probability of a Type I error is higher during the second week.

(g) A software distributor has opened a new customer call center to assist customers with the installation and use of their software. The manager is interested in learning more about the amount of time required to help a customer (referred to as "service time"). Assume service time is a normally distributed random variable. The manager selected a random sample of 16 calls and found the average service time is 24 minutes with a standard deviation of 4.5 minutes. The manager would like to take another sample to determine a 95% confidence interval with margin of error of 1.5 minutes. To achieve this goal, the manager would need a sample size of 36 calls.

Question 2 The city of Tharbad (a fictional city in Norway) is considering providing all elementary schools with a computer lab (*datalab*). The basic idea is that by having a computer lab, students and teachers will have an additional tool for learning. Before implementing this project, however, the city government wants to understand if having a computer lab increases students' performance in national exams.

The city government has hired you as a consultant to determine whether schools with computer labs are more effective than schools without a computer lab. They give you data from a simple random sample of 755 students: 250 students from schools with a computer lab, and 505 students from schools without. The dataset contains variables for the student's scores in national exams (0 to 100 points) as well as parents' characteristics. Summary statistics for the exam scores are shown below.

	Exam Scores	
	With Computer Lab	Without Computer Lab
Mean	43.35	39.07
SD	15.61	15.76
Minimum	7.55	5.09
Maximum	95.86	100
No. of Students	250	505

(a) You decide to conduct a two-sample t-test to investigate if there is any difference in average exam scores between students in schools with a computer lab versus students in schools without a computer lab.

(i) Write down the null and alternative hypothesis.

(ii) What is the t-static and p-value for this test? Assume that the sample size conditions hold.

(b) You also estimate the following MRM

$$\text{ExamScore} = \beta_0 + \beta_1\text{ComputerLab} + \beta_2\text{ParentEducation} + \beta_3\text{LowIncome} + \epsilon$$

where ComputerLab is a dummy equal to 1 if the student is in a school has a computer lab and 0 otherwise; ParentEducation is the average education (in years) of the student's parents, and LowIncome is a dummy variable equal to 1 if the student is from a low-income family, and 0 otherwise. The results of the regression are shown below.

Regression Statistics	
R-squared	0.758
F-statistic	783.97
F-statistic (p-value)	0.000
s_e	7.805
Response Variable	ExamScore
Observations	755

	Coefficient	Standard Error	t-stat	$P > \|t\|$
Constant	13.856	0.644	21.51	0.000
ComputerLab	1.507	0.609	2.48	0.011
ParentEducation	1.478	0.056	26.10	0.000
LowIncome	−0.388	0.009	40.60	0.000

Interpret the coefficients for "Constant" and "ComputerLab" in the above regression.

(c) What is the 60% confidence interval for the coefficient on ParentEducation? Round your final answer to two decimal places. Interpret this interval in three or fewer sentences.

(d) Suppose that you wanted to test if after controlling for parent education, having a computer lab at school is more beneficial for students from high-income families, because these families can also afford to buy their own computer to use at home. Can you use the regression shown in part (b) to test this? If yes, explain the hypothesis test that you would carry out. If not, write down the regression and the hypothesis test you would need. Limit your answer to three or fewer sentences.

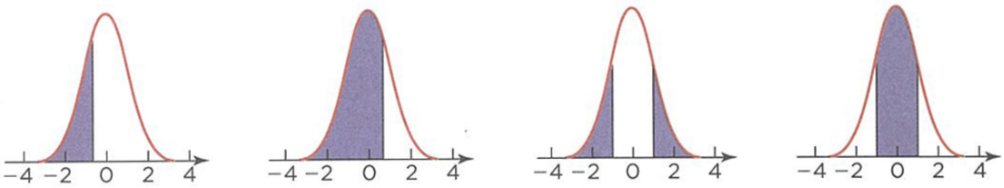
(e) Your colleague tells you that in schools with a computer lab, the number of applications for admission (opptakssøknad) always exceeds the number of available places. Therefore, a lottery is conducted. The lottery assigns each application a random number between 1 and 100. Students with a lottery number 51 or above receive admission (and attend the school with a computer lab). Students with a lottery number 50 and below do not receive admission (and attend a regular school without a computer lab). The city government gives you the lottery number for all students who applied for admission at schools with computer labs.

Suppose you compare the exam scores between students who "won" the lottery (and were admitted to the school with a computer lab) and students who "lost" the lottery (and were not admitted to the school with a computer lab). What types of confounding variables might we be concerned about in this comparison? Explain in three or fewer sentences. Assume that all students who "won" the lottery are attending a school with a computer lab, and students who "lost" the lottery are attending a school without a computer lab.

(f) You have data on the exam scores of lottery winner and lottery losers, as well as their lottery number. Write a regression model that would allow you to test whether exam scores were higher for lottery winners relative to lottery losers. Precisely define all the variables in your regression and the null and alternative hypothesis.

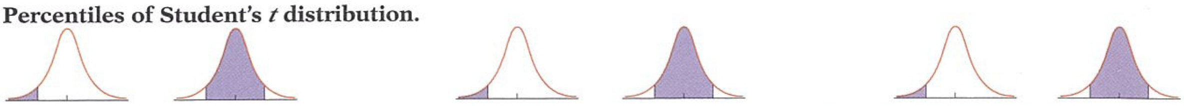
(g) Let's go back to the original question in this exercise: the city government wants to understand if having a computer lab increases students' performance in national exams. What does the approach in part (f) tell us about this question? Explain in three sentences or fewer.

	Reject $H_0: \rho_\varepsilon = 0$ if	
n	**D is less than**	**D is greater than**
15	1.36	2.64
20	1.41	2.59
30	1.49	2.51
40	1.54	2.46
50	1.59	2.41
75	1.65	2.35
100	1.69	2.31



z	$P(Z \leq -z)$	$P(Z \leq z)$	$P(\|Z\| > z)$	$P(\|Z\| \leq z)$
0	0.50	0.50	1	0
0.0502	0.48	0.52	0.96	0.04
0.1004	0.46	0.54	0.92	0.08
0.1510	0.44	0.56	0.88	0.12
0.2019	0.42	0.58	0.84	0.16
0.2533	0.40	0.60	0.80	0.20
0.3055	0.38	0.62	0.76	0.24
0.3585	0.36	0.64	0.72	0.28
0.4125	0.34	0.66	0.68	0.32
0.4677	0.32	0.68	0.64	0.36
0.4959	0.31	0.69	0.62	0.38
0.5244	0.30	0.70	0.60	0.40
0.5828	0.28	0.72	0.56	0.44
0.6433	0.26	0.74	0.52	0.48
0.6745	0.25	0.75	0.50	0.50
0.7063	0.24	0.76	0.48	0.52
0.7388	0.23	0.77	0.46	0.54
0.7722	0.22	0.78	0.44	0.56
0.8064	0.21	0.79	0.42	0.58
0.8416	0.20	0.80	0.40	0.60
0.8779	0.19	0.81	0.38	0.62
0.9154	0.18	0.82	0.36	0.64
0.9542	0.17	0.83	0.34	0.66
0.9945	0.16	0.84	0.32	0.68
1.0364	0.15	0.85	0.30	0.70
1.0803	0.14	0.86	0.28	0.72
1.1264	0.13	0.87	0.26	0.74
1.1750	0.12	0.88	0.24	0.76
1.2265	0.11	0.89	0.22	0.78
1.2816	0.10	0.90	0.20	0.80
1.3408	0.09	0.91	0.18	0.82
1.4051	0.08	0.92	0.16	0.84
1.4758	0.07	0.93	0.14	0.86
1.5548	0.06	0.94	0.12	0.88
1.6449	0.05	0.95	0.10	0.90
1.7507	0.04	0.96	0.08	0.92
1.8808	0.03	0.97	0.06	0.94
1.9600	0.025	0.975	0.05	0.95
2.0537	0.02	0.98	0.04	0.96
2.3263	0.01	0.99	0.02	0.98
2.5758	0.005	0.995	0.01	0.99
2.8070	0.0025	0.9975	0.005	0.995
3.0902	0.001	0.999	0.002	0.998
3.2905	0.0005	0.9995	0.001	0.999
3.7190	0.0001	0.9999	0.0002	0.9998
3.8906	0.00005	0.99995	0.0001	0.9999
4.2649	0.00001	0.99999	0.00002	0.99998
4.4172	0.000005	0.999995	0.00001	0.99999

T-TABLE **Percentiles of Student's *t* distribution.**



$df = 1$			$df = 2$			$df = 3$		
t	$P(T_1 \le -t)$	$P(-t \le T_1 \le t)$	t	$P(T_2 \le -t)$	$P(-t \le T_2 \le t)$	t	$P(T_3 \le -t)$	$P(-t \le T_3 \le t)$
3.078	0.1	0.8	1.886	0.1	0.8	1.638	0.1	0.8
6.314	0.05	0.9	2.920	0.05	0.9	2.353	0.05	0.9
12.71	0.025	0.95	4.303	0.025	0.95	3.182	0.025	0.95
31.82	0.01	0.98	6.965	0.01	0.98	4.541	0.01	0.98
63.66	0.005	0.99	9.925	0.005	0.99	5.841	0.005	0.99
318.3	0.001	0.998	22.33	0.001	0.998	10.21	0.001	0.998
636.6	0.0005	0.999	31.60	0.0005	0.999	12.92	0.0005	0.999
6366	0.00005	0.9999	99.99	0.00005	0.9999	28.00	0.00005	0.9999

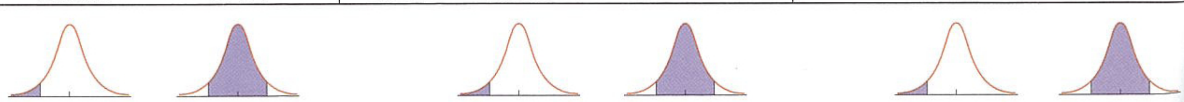
$df = 4$			$df = 5$			$df = 6$		
t	$P(T_4 \le -t)$	$P(-t \le T_4 \le t)$	t	$P(T_5 \le -t)$	$P(-t \le T_5 \le t)$	t	$P(T_6 \le -t)$	$P(-t \le T_6 \le t)$
1.533	0.1	0.8	1.476	0.1	0.8	1.440	0.1	0.8
2.132	0.05	0.9	2.015	0.05	0.9	1.943	0.05	0.9
2.776	0.025	0.95	2.571	0.025	0.95	2.447	0.025	0.95
3.747	0.01	0.98	3.365	0.01	0.98	3.143	0.01	0.98
4.604	0.005	0.99	4.032	0.005	0.99	3.707	0.005	0.99
7.173	0.001	0.998	5.893	0.001	0.998	5.208	0.001	0.998
8.610	0.0005	0.999	6.869	0.0005	0.999	5.959	0.0005	0.999
15.54	0.00005	0.9999	11.18	0.00005	0.9999	9.082	0.00005	0.9999

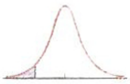
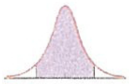
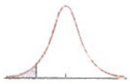
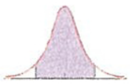
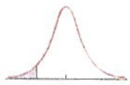
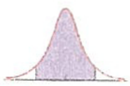
$df = 7$			$df = 8$			$df = 9$		
t	$P(T_7 \le -t)$	$P(-t \le T_7 \le t)$	t	$P(T_8 \le -t)$	$P(-t \le T_8 \le t)$	t	$P(T_9 \le -t)$	$P(-t \le T_9 \le t)$
1.415	0.1	0.8	1.397	0.1	0.8	1.383	0.1	0.8
1.895	0.05	0.9	1.860	0.05	0.9	1.833	0.05	0.9
2.365	0.025	0.95	2.306	0.025	0.95	2.262	0.025	0.95
2.998	0.01	0.98	2.896	0.01	0.98	2.821	0.01	0.98
3.499	0.005	0.99	3.355	0.005	0.99	3.250	0.005	0.99
4.785	0.001	0.998	4.501	0.001	0.998	4.297	0.001	0.998
5.408	0.0005	0.999	5.041	0.0005	0.999	4.781	0.0005	0.999
7.885	0.00005	0.9999	7.120	0.00005	0.9999	6.594	0.00005	0.9999

$df = 10$			$df = 11$			$df = 12$		
t	$P(T_{10} \le -t)$	$P(-t \le T_{10} \le t)$	t	$P(T_{11} \le -t)$	$P(-t \le T_{11} \le t)$	t	$P(T_{12} \le -t)$	$P(-t \le T_{12} \le t)$
1.415	0.1	0.8	1.397	0.1	0.8	1.383	0.1	0.8
1.895	0.05	0.9	1.860	0.05	0.9	1.833	0.05	0.9
2.365	0.025	0.95	2.306	0.025	0.95	2.262	0.025	0.95
2.998	0.01	0.98	2.896	0.01	0.98	2.821	0.01	0.98
3.499	0.005	0.99	3.355	0.005	0.99	3.250	0.005	0.99
4.785	0.001	0.998	4.501	0.001	0.998	4.297	0.001	0.998
5.408	0.0005	0.999	5.041	0.0005	0.999	4.781	0.0005	0.999
7.885	0.00005	0.9999	7.120	0.00005	0.9999	6.594	0.00005	0.9999

$df = 13$			$df = 14$			$df = 15$		
t	$P(T_{13} \le -t)$	$P(-t \le T_{13} \le t)$	t	$P(T_{14} \le -t)$	$P(-t \le T_{14} \le t)$	t	$P(T_{15} \le -t)$	$P(-t \le T_{15} \le t)$
1.350	0.1	0.8	1.345	0.1	0.8	1.341	0.1	0.8
1.771	0.05	0.9	1.761	0.05	0.9	1.753	0.05	0.9
2.160	0.025	0.95	2.145	0.025	0.95	2.131	0.025	0.95
2.650	0.01	0.98	2.624	0.01	0.98	2.602	0.01	0.98
3.012	0.005	0.99	2.977	0.005	0.99	2.947	0.005	0.99
3.852	0.001	0.998	3.787	0.001	0.998	3.733	0.001	0.998
4.221	0.0005	0.999	4.140	0.0005	0.999	4.073	0.0005	0.999
5.513	0.00005	0.9999	5.363	0.00005	0.9999	5.239	0.00005	0.9999

$df = 16$			$df = 17$			$df = 18$		
t	$P(T_{16} \le -t)$	$P(-t \le T_{16} \le t)$	t	$P(T_{17} \le -t)$	$P(-t \le T_{17} \le t)$	t	$P(T_{18} \le -t)$	$P(-t \le T_{18} \le t)$
1.337	0.1	0.8	1.333	0.1	0.8	1.33	0.1	0.8
1.746	0.05	0.9	1.740	0.05	0.9	1.734	0.05	0.9
2.120	0.025	0.95	2.110	0.025	0.95	2.101	0.025	0.95
2.583	0.01	0.98	2.567	0.01	0.98	2.552	0.01	0.98
2.921	0.005	0.99	2.898	0.005	0.99	2.878	0.005	0.99
3.686	0.001	0.998	3.646	0.001	0.998	3.610	0.001	0.998
4.015	0.0005	0.999	3.965	0.0005	0.999	3.922	0.0005	0.999
5.134	0.00005	0.9999	5.044	0.00005	0.9999	4.966	0.00005	0.9999



$df = 19$		
t	$P(T_{19} \le -t)$	$P(-t \le T_{19} \le t)$
1.328	0.1	0.8
1.729	0.05	0.9
2.093	0.025	0.95
2.539	0.01	0.98
2.861	0.005	0.99
3.579	0.001	0.998
3.883	0.0005	0.999
4.897	0.00005	0.9999

$df = 20$		
t	$P(T_{20} \le -t)$	$P(-t \le T_{20} \le t)$
1.325	0.1	0.8
1.725	0.05	0.9
2.086	0.025	0.95
2.528	0.01	0.98
2.845	0.005	0.99
3.552	0.001	0.998
3.850	0.0005	0.999
4.837	0.00005	0.9999

$df = 22$		
t	$P(T_{22} \le -t)$	$P(-t \le T_{22} \le t)$
1.321	0.1	0.8
1.717	0.05	0.9
2.074	0.025	0.95
2.508	0.01	0.98
2.819	0.005	0.99
3.505	0.001	0.998
3.792	0.0005	0.999
4.736	0.00005	0.9999

$df = 24$		
t	$P(T_{24} \le -t)$	$P(-t \le T_{24} \le t)$
1.318	0.1	0.8
1.711	0.05	0.9
2.064	0.025	0.95
2.492	0.01	0.98
2.797	0.005	0.99
3.467	0.001	0.998
3.745	0.0005	0.999
4.654	0.00005	0.9999

$df = 26$		
t	$P(T_{26} \le -t)$	$P(-t \le T_{26} \le t)$
1.315	0.1	0.8
1.706	0.05	0.9
2.056	0.025	0.95
2.479	0.01	0.98
2.779	0.005	0.99
3.435	0.001	0.998
3.707	0.0005	0.999
4.587	0.00005	0.9999

$df = 28$		
t	$P(T_{28} \le -t)$	$P(-t \le T_{28} \le t)$
1.313	0.1	0.8
1.701	0.05	0.9
2.048	0.025	0.95
2.467	0.01	0.98
2.763	0.005	0.99
3.408	0.001	0.998
3.674	0.0005	0.999
4.530	0.00005	0.9999

$df = 30$		
t	$P(T_{30} \le -t)$	$P(-t \le T_{30} \le t)$
1.31	0.1	0.8
1.697	0.05	0.9
2.042	0.025	0.95
2.457	0.01	0.98
2.75	0.005	0.99
3.385	0.001	0.998
3.646	0.0005	0.999
4.482	0.00005	0.9999

$df = 32$		
t	$P(T_{32} \le -t)$	$P(-t \le T_{32} \le t)$
1.309	0.1	0.8
1.694	0.05	0.9
2.037	0.025	0.95
2.449	0.01	0.98
2.738	0.005	0.99
3.365	0.001	0.998
3.622	0.0005	0.999
4.441	0.00005	0.9999

$df = 34$		
t	$P(T_{34} \le -t)$	$P(-t \le T_{34} \le t)$
1.307	0.1	0.8
1.691	0.05	0.9
2.032	0.025	0.95
2.441	0.01	0.98
2.728	0.005	0.99
3.348	0.001	0.998
3.601	0.0005	0.999
4.405	0.00005	0.9999

$df = 36$		
t	$P(T_{36} \le -t)$	$P(-t \le T_{36} \le t)$
1.306	0.1	0.8
1.688	0.05	0.9
2.028	0.025	0.95
2.434	0.01	0.98
2.719	0.005	0.99
3.333	0.001	0.998
3.582	0.0005	0.999
4.374	0.00005	0.9999

$df = 40$		
t	$P(T_{40} \le -t)$	$P(-t \le T_{40} \le t)$
1.303	0.1	0.8
1.684	0.05	0.9
2.021	0.025	0.95
2.423	0.01	0.98
2.704	0.005	0.99
3.307	0.001	0.998
3.551	0.0005	0.999
4.321	0.00005	0.9999

$df = 50$		
t	$P(T_{50} \le -t)$	$P(-t \le T_{50} \le t)$
1.299	0.1	0.8
1.676	0.05	0.9
2.009	0.025	0.95
2.403	0.01	0.98
2.678	0.005	0.99
3.261	0.001	0.998
3.496	0.0005	0.999
4.228	0.00005	0.9999

$df = 60$		
t	$P(T_{60} \le -t)$	$P(-t \le T_{60} \le t)$
1.296	0.1	0.8
1.671	0.05	0.9
2.000	0.025	0.95
2.390	0.01	0.98
2.660	0.005	0.99
3.232	0.001	0.998
3.460	0.0005	0.999
4.169	0.00005	0.9999

$df = 75$		
t	$P(T_{75} \le -t)$	$P(-t \le T_{75} \le t)$
1.293	0.1	0.8
1.665	0.05	0.9
1.992	0.025	0.95
2.377	0.01	0.98
2.643	0.005	0.99
3.202	0.001	0.998
3.425	0.0005	0.999
4.110	0.00005	0.9999

$df = 100$		
t	$P(T_{100} \le -t)$	$P(-t \le T_{100} \le t)$
1.290	0.1	0.8
1.660	0.05	0.9
1.984	0.025	0.95
2.364	0.01	0.98
2.626	0.005	0.99
3.174	0.001	0.998
3.390	0.0005	0.999
4.053	0.00005	0.9999

$df = 125$		
t	$P(T_{125} \le -t)$	$P(-t \le T_{125} \le t)$
1.288	0.1	0.8
1.657	0.05	0.9
1.979	0.025	0.95
2.357	0.01	0.98
2.616	0.005	0.99
3.157	0.001	0.998
3.370	0.0005	0.999
4.020	0.00005	0.9999

$df = 150$		
t	$P(T_{150} \le -t)$	$P(-t \le T_{150} \le t)$
1.287	0.1	0.8
1.655	0.05	0.9
1.976	0.025	0.95
2.351	0.01	0.98
2.609	0.005	0.99
3.145	0.001	0.998
3.357	0.0005	0.999
3.998	0.00005	0.9999

$df = \infty$		
t	$P(Z \le -t)$	$P(-t \le Z \le t)$
1.282	0.1	0.8
1.645	0.05	0.9
1.960	0.025	0.95
2.326	0.01	0.98
2.576	0.005	0.99
3.090	0.001	0.998
3.291	0.0005	0.999
3.891	0.00005	0.9999