Midterm 2

EC031-S25

Name: ________________________________

For this midterm, you will be asked several questions about what we've learned in class. There are five questions in this exam. **Choose and answer 4 out of 5 questions.** You will not receive extra credit for answering all questions.

The exam is out of 100 points. By default, all math-involved questions require showing your work, unless otherwise stated. Simply writing down the right answer is not enough for full credit.

For "explain your answer" questions, the length of the answer is not as important as being able to clearly explain your thinking. As a general rule, 3-5 sentences should suffice.

The amount of space that I give you is not indicative of how much is needed to answer the question. Clearly mark the questions that you choose.

Note: The last page of the exam is a t-distribution table.

Tip: Before starting the exam, look through all the questions and choose the ones you want to answer.

Question 1 [25 points]

Assume that there are two random variables X, Y and Z with expected values μ_x, μ_y, and μ_z and variances σ_x^2 and σ_y^2, and σ_z^2. Let the covariance between any two random variables be: σ_{ij} where i and j are the random variables, X, Y or Z. Note that the covariance between these variables need not be 0 (i.e. they aren't necessarily independent).

Any other letters (a, b, c etc. . .) can be considered non-random parameters. For each of the following problems, rewrite the expressions in terms of their means, variances and covariances when possible. If not possible, write what you think the answer is and explain why.

a. [7 points] $E(XY + bZ)$

Since $\text{Cov}(X,Y) = E[(X-\bar{X})(Y-\bar{Y})]$

$= E(XY) + \mu_x \mu_y$

So $E(XY) = \sigma_{xy} - \mu_x \mu_y$

$$\boxed{\sigma_{xy} - \mu_x \mu_y + b\mu_z}$$

b. [5 points] $\text{Var}(bX + Y)$

$V(bX + Y) = V(bX) + V(Y) + 2\,\text{Cov}(bX, Y)$

$$= \boxed{b^2 \sigma_x + \sigma_y + 2b\,\sigma_{xy}}$$

c. [7 points] $\text{Cov}(X, a)$

0

d. [6 points] $\text{Var}(X + Y + Z)$ if the covariances between all variables are 0 .

$\sigma_x + \sigma_y + \sigma_z$

Question 2 [25 points]

Suppose we have the following data with $N = 4$:

X_{1i}	Y_i	$\hat{Y}_i$	e_i
1	2	3	-1
0	4	5	-1
1	4	3	1
0	6	5	1

Where e_i is the residual for each observation and $\hat{Y}_i$ is the predicted value of Y_i.

a. [10 points] Calculate b_0 and b_1 for the regression line $Y = b_0 + b_1 X_1$.

$$\bar{X} = \frac{1 + 0 + 1 + 0}{4} = \frac{2}{4} = \frac{1}{2}$$

$$\bar{y} = \frac{2 + 4 + 4 + 6}{4} = \frac{16}{4} = 4$$

$X-\bar{X}$	$y-\bar{y}$	$(X-\bar{X})(y-\bar{y})$	$(X-\bar{X})(X-\bar{X})$
$\frac{1}{2}$	-2	-1	$\frac{1}{4}$
$-\frac{1}{2}$	0	0	$\frac{1}{4}$
$\frac{1}{2}$	0	0	$\frac{1}{4}$
$-\frac{1}{2}$	2	-1	$\frac{1}{4}$

$$b_1 = \frac{-1 + 0 + 0 + -1}{(\frac{1}{4})4} = -2$$

$$b_0 = \bar{Y} - b_1 \bar{X} = 4 - (-2)\frac{1}{2} = 4 + 1 = 5$$

b. [5 points] Using the regression line from part a, calculate the residuals and predicted values for each observation and fill them in the table above.

$\hat{Y}_i$
$5-2(1)=3$
$5-2(0)=5$
$5-2(1)=3$
$5-2(0)=5$

c. [5 points] Calculate the TSS (total sum of squares), and the sum of squared residuals.

TSS

$(y-\bar{y})$	$(y_i-\bar{y})^2$
$2-4=-2$	4
$4-4=0$	0
$4-4=0$	0
$6-4=2$	4

$TSS = 4+0+0+4 = \boxed{8}$

SSR

e_i^2
1
1
1
1

$\boxed{SSR = 4}$

d. [5 points] Calculate the R^2 value for the regression line.

$$R^2 = 1 - \frac{SSR}{TSS}$$

$$= 1 - \frac{4}{8} = \boxed{.5}$$

Question 3 [25 points]

In 2-3 sentences or bullet points, explain what Type I and Type II errors are and how they are linked to a researcher's choice of significance level (i.e., the value of α the researcher sets).

- Type I error is when you reject the null, but it's actually true

- Type II error is when you fail to reject the null even though it's false

- α is the probability of a Type I error.
- But low $\alpha \rightarrow$ CI $\uparrow$ & critical values larger, making it "easier" for the CI to include the null value & failing to reject the null. So Type II error rises.

Question 4 [25 points]

A market research firm used a sample of individuals to rate the purchase potential of a particular product before and after the individuals saw a new television commercial about the product. The purchase potential ratings were based on a 0 to 10 scale, with higher values indicating a higher purchase potential. The null hypothesis stated that the mean rating "after" would be less than or equal to the mean rating "before." Rejection of this hypothesis would show that the commercial improved the mean purchase potential rating.

Use $\alpha = .05$ and the following data to test the hypothesis and comment on the value of the commercial.

Make sure to show all your work:

- Write down your null and alternative hypotheses
- Calculate the test statistic
- Find the critical value
- Make a decision
- Explain your answer

Individual	After	Before
1	6	5
2	6	4
3	7	7
4	4	3
5	3	5
6	9	8
7	7	5
8	6	6

$H_0: \mu_{After} \leq \mu_{Befor}$

$H_a: \mu_{After} > \mu_{Befor}$

Let $\mu_D = \mu_{After} - \mu_{Befor}$

$\bar{d} = .625, \quad S_d = 1.30$

Question 4 (extra space)

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{.625 - 0}{1.30 / \sqrt{8}} = 1.36$$

$df = n - 1 = 7$

Critical value, $t_7 = 1.36$

Do not reject null

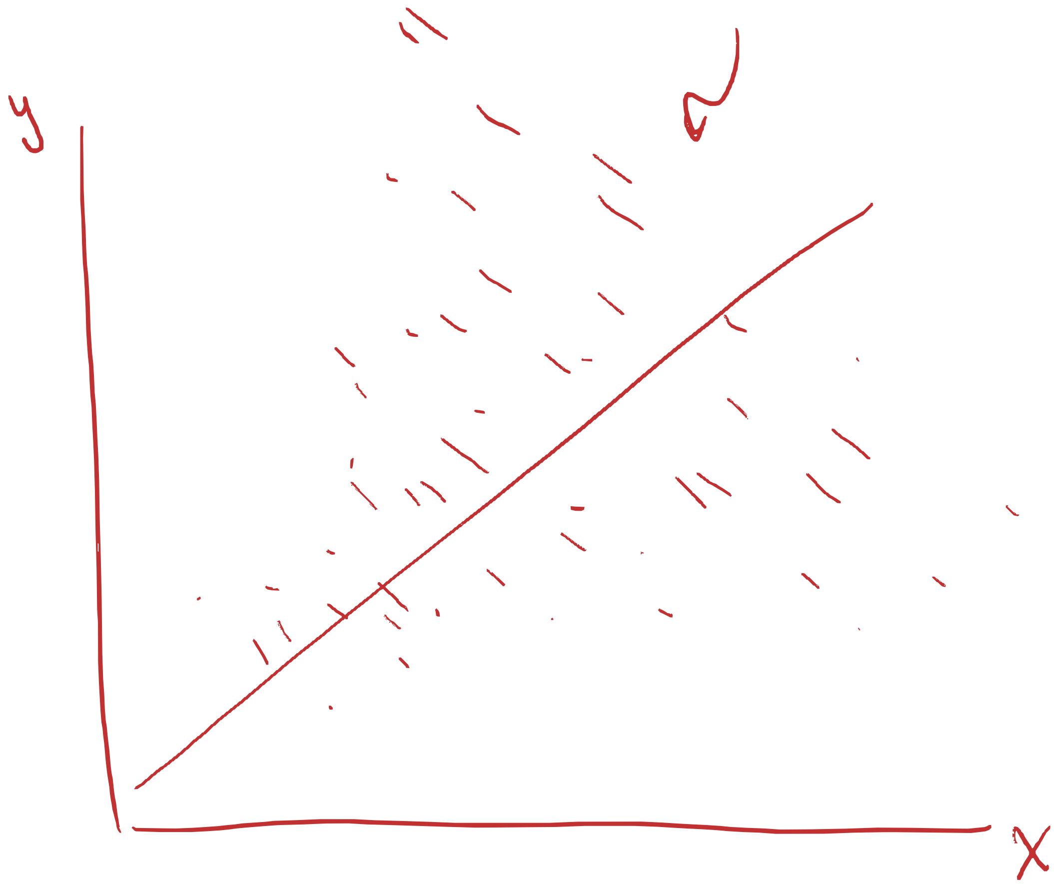
The comercial doesn't improve the mean potential to purchase.

Question 5 [25 points]

a. [8 points] State the Classical Linear Regression Model (CLRM) assumption for homoskedastic errors. Then explain or draw a diagram to show an example when this assumption is violated.

$$V(\varepsilon_i \mid X_i) = \sigma_\varepsilon^2$$

heteroskedasticity



b. [8 points] State the CLRM assumption about the linearity of the model. Then write down a model that violates this assumption and explain why.

You can write the model like so:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Example violating it (not the only one):

$$Y_i = \beta_0 + \beta_1^2 X_i + \varepsilon_i$$

c. [7 points] State the CLRM assumption about the independence of the errors. Then write down a model that violates this assumption and explain why.

$E(\varepsilon_i \varepsilon_j) = 0$

This would be violated in the case of time series or serial correlation.

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										