

THE OPEN UNIVERSITY OF SRI LANKA

B. Sc. Degree Programme, Continuing Education Programme

APPLIED MATHEMATICS-LEVEL05

ADU5301- REGRESSION ANALYSIS I

FINAL EXAMINATION 2023/2024

Duration: Two Hours.



Date: 04.04.2024

Time: 9.30 a.m- 11.30 a.m

Answer FOUR questions only.

Instructions:

- This question paper consists of 06 questions. Answer only four questions.
- Statistical Tables are attached at the end of the question paper. When reading values, you may use the closest degrees of freedom given in the table.
- Where appropriate, method of least squares will be used to fit the regression model.
- Non-programmable calculators are permitted.

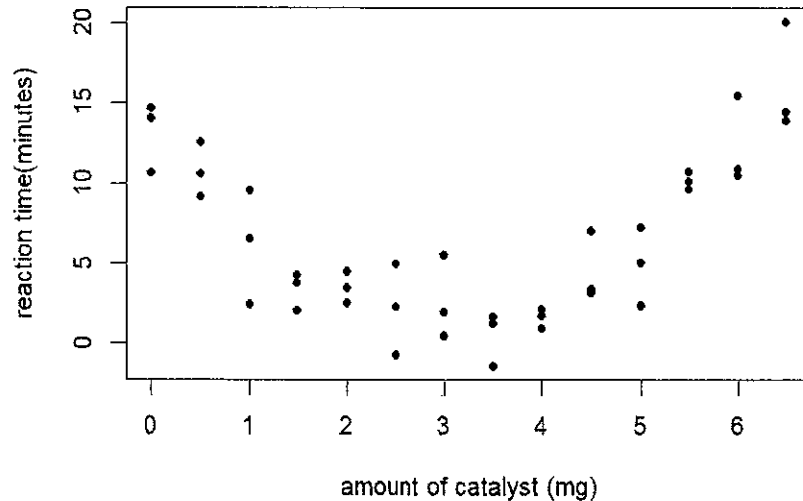
1. A manufacturer claims that addition of a certain chemical improves the durability of a wood polish. Following summary statistics were computed from durability (y) measured to the nearest month on 28 polished wood samples, after adding known amounts (x) of the said chemical, prior to polishing. The amount of chemical used for the experiment had varied from 0 (no chemical added) to 20 grams per litre of polish.

$$\sum x = 276, \sum y = 354.5, \sum x^2 = 3796, \sum y^2 = 5169.25, \sum xy = 4178.4$$

The experimenter wants to fit a simple linear regression model for the durability, using the amount of chemical added as the predictor variable.

- Clearly state the assumptions needed to fit the model, using the method of least squares.
- Compute least squares estimates for the slope and intercept parameters.
- Write down the equation of the fitted model.
- Based on the fitted model, estimate the durability, if 5 grams of the said chemical was added per litre.
- Estimate the improvement in durability associated with 4 grams of increase of the chemical, per litre.

2. i) Briefly describe two advantages of examining scatter plots of variables, in regression model building.
- ii) A plot of observed reaction times (in minutes), y , of a chemical process in 42 samples, after adding known amounts (mg) of a catalyst, x , to each sample is given below.



State whether you agree or disagree with each of the following statements, based on the above plot. In each case, give reasons for your answer.

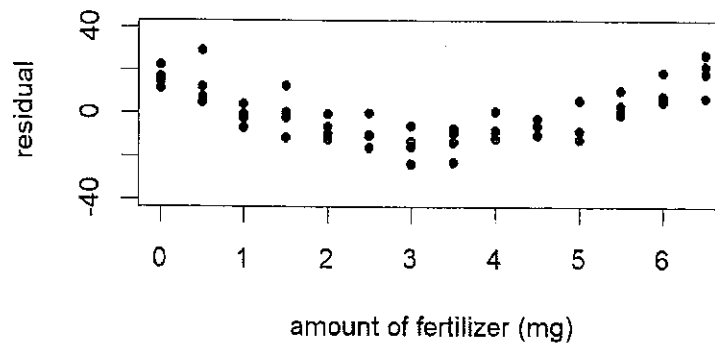
- a) Pearson correlation coefficient is appropriate to assess the strength of the association between the reaction time and the amount of catalyst added.
 - b) The relationship between reaction time and the amount of catalyst cannot be modelled using a simple linear regression model.
 - c) If a simple linear regression model is fitted using the amount of catalyst as the predictor variable, observations with identical reaction times will have identical fitted values.
 - d) The plot indicates violation of the constant variance assumption.
3. In a study to assess the effect of protein on the weight gain of rats, a researcher measured the weight gain (mg) of 34 rats, three months after giving a diet with known amounts (mg) of protein, daily. The protein contents in the daily diets were in the range from 0 mg to 5mg. The model $y = \beta_0 + \beta_1 x + \epsilon$ was fitted to the data using the method of least squares with weight gain as the response variable and the protein content in the diet as the predictor variable. The following table gives some parts of the output from the model fit.

Parameter	Estimate	Standard error
β_0	2.43	0.28
β_1	0.26	0.10

Residual sum of squares = 30.57

- i) Clearly explain what the value 0.10 in the table represents in relation to this study.
 - ii) Construct a 95% confidence interval for the slope parameter.
 - iii) Using the results in part (ii) or otherwise, test the validity of each of the following claims.
 - a) Protein in the diet has no significant effect on the weight gain of rats.
 - b) A change of 2 mg of protein content in the diet is associated with a change of 0.5 mg of weight gain in three months.
 - iv) Estimate the random error in the weight gain of rats.
4. i) State whether fitting a simple linear regression model, using the method of least squares is appropriate to meet the stated objectives. In each case, give reasons for your answer.
- a) A researcher wants to estimate the expected time a parachute would take to reach the ground when released from a known height, far from the ground. Data are available on the location and time, at each of 40 different locations immediately after release of the parachute. .
 - b) A researcher wants to predict the number of seeds that will germinate from a packet of 50 grams. Data were available on the number of seeds that germinated from each of 40 packets of similar seeds with weights in the range of 10 grams to 100 grams, but none of those packets was of 50 grams. Available data showed that a regression function of the form $n = w^{2/3}$ is appropriate, where n denote the expected number of seeds that will germinate from a packet of weight equal to w .
- ii) The following figure illustrates a plot of residuals against the predictor variable obtained from fitting the model $y = \beta_0 + \beta_1 x + \epsilon$ using the method of least squares, to the data

collected on the yield (y) of a tomato species with the amount of fertilizer added (x) as the predictor variable.



State whether each of the following statements is true or false according to the above plot. In each case, give reasons for your answer.

- a) The plot is adequate to conclude that there are no extreme values in the data set.
 - b) A plot of residuals against the fitted values will have a similar pattern as in the given plot.
 - c) The plot shows violation of the constant variance assumption.
 - d) Since there are similar number of positive and negative residuals, the random errors will have a normal distribution.
 - e) A plot of fitted values against the predictor variable will have a pattern similar to the given plot.
5. A researcher fitted the regression function $\beta_0 + \beta_1 x$, to the data collected in a study to assess the effect of ambient temperature ($^{\circ}\text{C}$), x , on the drying time (in minutes), y , of a certain paint. Least squares estimates of the slope and intercept parameters from this model fit, based on 44 observations were -0.65 and 45.62 respectively. The coefficient of determination R^2 and the mean squared error were 0.71 and 7.56 respectively.
- i) Clearly explain, what one can conclude from the value of R^2 , in relation to this study.
 - ii) Calculate the Pearson correlation coefficient and clearly explain what one can conclude from the value you computed
 - iii) Estimate the random error in the drying times of this paint,
 - a) Based on the fitted model
 - b) Assuming that the researcher had not recorded data on the ambient temperature
 - iv) Estimate the standard error of the least squares estimator for the slope parameter

6. The following descriptive statistics were computed from the lengths (cm) of the stem and the longest petal length of 32 lotus flowers.

Description	Sample mean	Standard deviation
Length of petal (y)	3.63	0.71
Length of stem (x)	11.5	2.33

Pearson correlation coefficient between the length of stem and length of petal = 0.69.

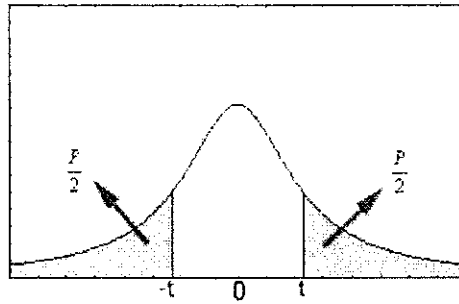
Part of the Analysis of variance (ANOVA) table obtained by fitting a simple linear regression model for y using x as the predictor variable is given below.

Source of variation	Sum of squares	Degrees of freedom
Regression	(a)	(d)
Residual	(b)	(e)
Total	(c)	(f)

- Compute the missing values indicated by (a), (b), (c), (d), (e) and (f) in the ANOVA table.
- Using a 5% significance level, test whether the predictor variable significantly contributes to predict the variation in the response variable. Clearly state your findings.

***** Copyrights reserved. *****

Table A2: Student's t - Distribution



P	50	20	10	5	2	1	0.2	0.1
Degrees of freedom								
1	1.00	3.08	6.31	12.7	31.8	63.7	318	637
2	0.82	1.89	2.92	4.30	6.96	9.92	22.3	31.6
3	0.76	1.64	2.35	3.18	4.54	5.84	10.2	12.9
4	0.74	1.53	2.13	2.78	3.75	4.60	7.17	8.61
5	0.73	1.48	2.02	2.57	3.36	4.03	5.89	6.87
6	0.72	1.44	1.94	2.45	3.14	3.71	5.21	5.96
7	0.71	1.42	1.89	2.36	3.00	3.50	4.79	5.41
8	0.71	1.40	1.86	2.31	2.90	3.36	4.50	5.04
9	0.70	1.38	1.83	2.26	2.82	3.25	4.30	4.78
10	0.70	1.37	1.81	2.23	2.76	3.17	4.14	4.59
12	0.70	1.36	1.78	2.18	2.68	3.05	3.93	4.32
15	0.69	1.34	1.75	2.13	2.60	2.95	3.73	4.07
20	0.69	1.32	1.72	2.09	2.53	2.85	3.55	3.85
24	0.68	1.32	1.71	2.06	2.49	2.80	3.47	3.75
30	0.68	1.31	1.70	2.04	2.46	2.75	3.39	3.65
40	0.68	1.30	1.68	2.02	2.42	2.70	3.31	3.55
60	0.68	1.30	1.67	2.00	2.39	2.66	3.23	3.46
∞	0.67	1.28	1.64	1.96	2.33	2.58	3.09	3.29

F ව්‍යාප්තිය
 F ප්‍රභවය
 The F Distribution

F ව්‍යාප්තියේ යම් P ප්‍රතිශතයක් F හි පිටතින් පිහිටීමට අවශ්‍ය F හි අගය සමඟ වෙනුවෙන් ලබාගත හැක.

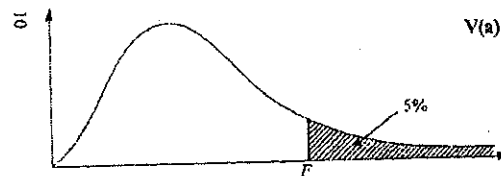
විචලිතයාගේ සඳහා වූ ස්වයංසන් ඇස්තමේන්තු කුඩාකළ අනුපාතයෙන් F ව්‍යාප්තිය ලැබේ. මෙම එක් එක් ඇස්තමේන්තුයාගේ අනර්ථය සුවිශේෂ අංක ඇති බැවින්, භාවිතා කර යුක්තේ කුඩා F ව්‍යාප්තිය දුළු ලෙස ලැබීමට, ලබියේ සුවිචලිත අංකය වන n_1 ද, හරයේ සුවිචලිත අංකය වන n_2 ද දී සිසිය යුතුය.

இங் அட்டவணையானது, F பரம்பலில் ஒரு தரப்பட்ட ஊர்முன்தக்திலும் கூடிய, F பெறுமதிகளைத் தருகின்றது

F பரம்பலானது, மாற்றற்றின் ஒன்றின் இரு சார மதிப்புகளை ஒன்றை மற்றதால் பிரிக் உண்டாகும். இச் சார மதிப்புகள் ஒவ்வொன்றின் சுவாரஸ்யம் இருப்பதால் எந்த F பரம்பலானது கருதப்படல் வேண்டுமெனக் குறிப்பிட்டு, தொகுதியின் சுவாரஸ்யம் σ , பகுதியின் சுவாரஸ்யம் σ_1 , இன்னும் தரப்படல் வேண்டும்.

These tables give the values of F for which a given percentage of the F -distribution is greater than F .

The F -distribution arises when two independent estimates of a variance are divided one by the other. Each of these estimates has its degrees of freedom associated with it, thus to specify which particular F -distribution is to be considered, the degrees of freedom of both the numerator n_1 , and the denominator n_2 , must be given.



5 பூமிக்கெனது துருவ F வகைநிலை
 F பரம்பலின் 5 சதவீத புள்ளிகள்
 5 percent points of the F distribution

		r											
		1	2	3	4	5	6	7	8	10	12	14	
$n_1 =$	$n_2 =$												
2	2	18.5	19.0	19.20	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.5	
3	3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.79	8.74	8.64	
4	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.96	5.91	5.77	
5	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.74	4.68	4.53	
6	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.06	4.00	3.84	
7	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.64	3.57	3.41	
8	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.35	3.28	3.12	
9	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.14	3.07	2.90	
10	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.98	2.91	2.74	
12	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.75	2.69	2.51	
15	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.54	2.48	2.29	
20	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.35	2.28	2.08	
24	24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.25	2.18	1.98	
30	30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.16	2.09	1.89	
40	40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.08	2.00	1.79	
60	60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	1.99	1.92	1.70	