SIE 330R Homework, Spring 2023

Agustin Espinoza

HW 6 (Chapter 4)

3/14/2023

Homework must be readable! Do not just send in numbers or charts. You must explain the homework answers Preferred to receive homework in Word doc format with any excel or Minitab results pasted into word document. You may choose to use pdf which is also OK.

Put answers to all questions in one document NOT in separate documents

• 4.3. Blocking is a technique that can be used to control the variability transmitted by uncontrolled nuisance factors in an experiment.

True   False

When the source of nuisance variability is known and uncontrollable, the blocking technique can be implemented to reduce the variability of the dependent variable results and lead to a more accurate model.

4.14. An article in *Communications of the ACM* (Vol. 30, No. 5, 1987) studied different algorithms for estimating software development costs. Six algorithms were applied to several different software development projects and the percent error in estimating the development cost was observed. Some of the data from this experiment is show in the table below.

	Project					
Algorithm	1	2	3	4	5	6
1 (SLIM)	1244	21	82	2221	905	839
2 (COCOMO-A)	281	129	396	1306	336	910
3 (COCOMO-R)	220	84	458	543	300	794
4 (COCOMO-C)	225	83	425	552	291	826
5 (FUNCTION POINTS)	19	11	-34	121	15	103
6 (ESTIMALS)	-20	35	-53	170	104	199

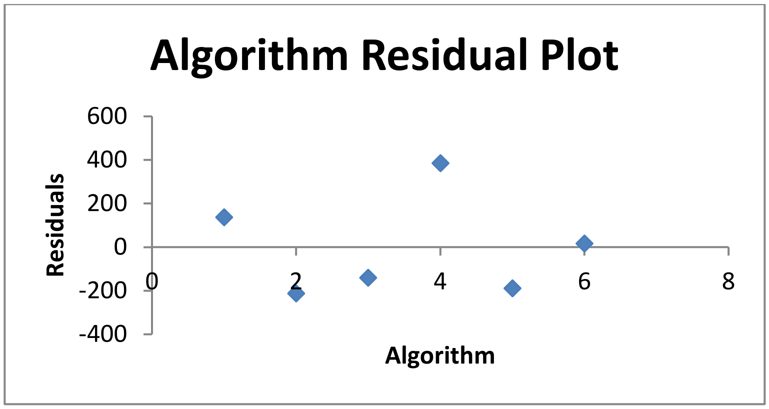
(a) Do the algorithms differ in their mean cost estimation accuracy?

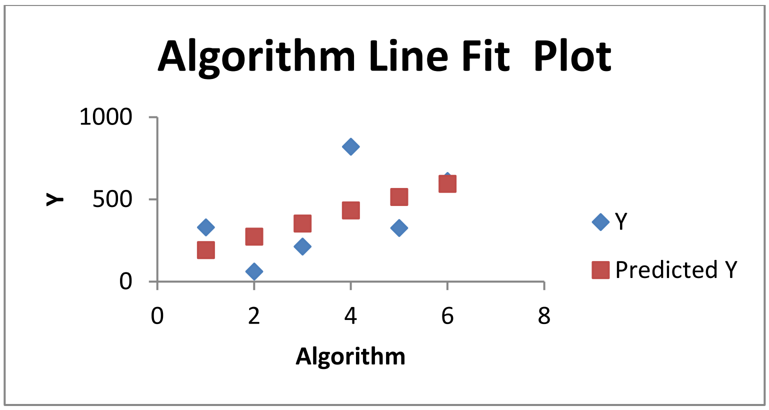
The ANOVA below provides the F-value, 5.377, which is greater than the F critical value. Therefore, the mean cost estimation accuracies differ significantly.

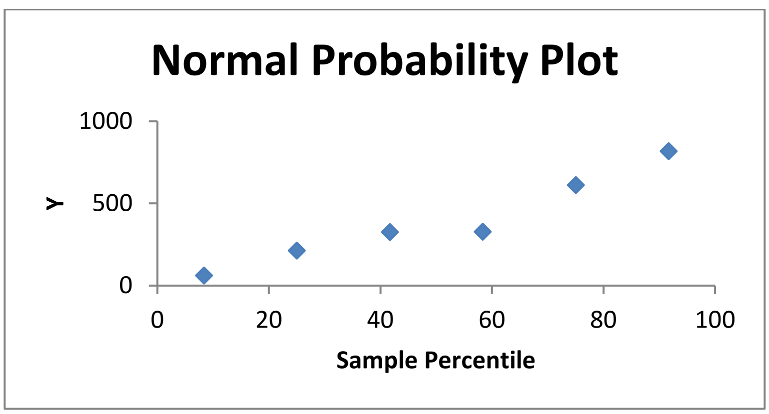
Anova: Two-Factor Without Replication						
SUMMARY	*Count*	*Sum*	*Average*	*Variance*		
1 Slim	6	5312	885.3333333	661519.4667		
2 COCOMO-A	6	3358	559.6666667	203937.8667		
3 COCOMO-R	6	2399	399.8333333	64260.96667		
4 COCOMO-C	6	2402	400.3333333	69639.86667		
5 FUNCTION POINTS	6	235	39.16666667	3581.766667		
6 ESTIMALS	6	435	72.5	10442.7		
1	6	1969	328.1666667	216024.5667		
2	6	363	60.5	2082.3		
3	6	1274	212.3333333	57476.26667		
4	6	4913	818.8333333	651728.5667		
5	6	1951	325.1666667	96648.56667		
6	6	3671	611.8333333	129780.5667		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	2989130.472	5	597826.0944	5.376958485	0.001720107	2.602987403
Columns	2287339.472	5	457467.8944	4.114550869	0.007295258	2.602987403
Error	2779573.694	25	111182.9478			
Total	8056043.639	35				

(b) Analyze the residuals from this experiment.

The residual analysis provided below is a result of the Blocking ANOVA technique.


Algorithm Residual Plot
Residuals
600
400
200
0
-200
-400
0
2
4
6
8
Algorithm


Algorithm Line Fit Plot
Y
1000
500
0
0
2
4
6
8
Algorithm
Y
Predicted Y


Normal Probability Plot
Y
1000
500
0
0
20
40
60
80
100
Sample Percentile

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.545670771
R Square	0.29775659
Adjusted R Square	0.122195738
Standard Error	258.7042286
Observations	6

ANOVA

	df	SS	MS	F	Significance F
Regression	1	113511.7337	113511.7337	1.696030672	0.262732378
Residual	4	267711.5116	66927.87791		
Total	5	381223.2454			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	110.9222222	240.8405299	0.460562939	0.669038199	-557.7582882	779.6027327	-557.7582882	779.6027327
X Variable 1	80.53809524	61.84213908	1.302317424	0.262732378	-91.16320911	252.2393996	-91.16320911	252.2393996

RESIDUAL OUTPUT

Observation	Predicted Y	Residuals	Standard Residuals
1	191.4603175	136.7063492	0.590799562
2	271.9984127	-211.4984127	-0.914026088
3	352.5365079	-140.2031746	-0.605911683
4	433.0746032	385.7587302	1.667121462
5	513.6126984	-188.4460317	-0.814401333
6	594.1507937	17.68253968	0.07641808

PROBABILITY OUTPUT

Percentile	Y
8.333333333	60.5
25	212.3333333
41.66666667	325.1666667
58.33333333	328.1666667
75	611.8333333
91.66666667	818.8333333

(c) Which algorithm would you recommend for use in practice?

Based on the analysis provided above, I would recommend the FUNCTIONAL POINTS algorithm for practical use as it has the lowest cost estimation error.

4.54. Physics graduate student Laura Van Ertia has conducted a complete randomized design with a single factor, hoping to solve the mystery of the unified theory and complete her dissertation. The results of this experiment are summarized in the following ANOVA display:

Source	*DF*	*SS*	*MS*	*F*
Factor	?	?	14.18	?
Error	?	37.75	?	
Total	23	108.63		

Answer the following questions about this experiment.

(a) The sum of squares for the factor is?

$SS_{Total} = SS_{Factor} + SS_{Error}$

$108.63 = SS_{Factor} + 37.75$

$SS_{Factor} = 108.63 – 37.75$

SS_{Factor} = 70.88

(b) The number of degrees of freedom for the single factor in the experiment is?

df_{Factor} = (sum of squares for the factor) / (MS for the factor)

df_{Factor} = 70.88 / 14.18

df_{Factor} = 5

(c) The number of degrees of freedom for the error is?

$df_{Total} = df_{Factor} + df_{Error}$

$23 = 5 + df_{Error}$

df_{Error} = 18

(d) The mean square for error is?

df_{Error} = (sum of squares for the error) / (MS for the error)

MS for the error = (sum of squares for the error) / (df_{Error})

MS for the error = 37.75 / 18

MS for the error = 2.10

(e) The value of the test statistic is?

$F = MS_{Factor} / MS_{Error}$

F = 14.18 / 2.10

F = 6.75

(f) If the significance level is 0.05, your conclusions are not to reject the null hypothesis.

Yes or No

The statement above is False. If the significance level is 0.05, and your statistical test produces a test statistic value greater than 0.05, then your conclusions would be to reject the null hypothesis.

(g) The P-value for the test statistic is?

P-value = 0.001

(i) Laura used how many levels of the factor in this experiment?

6 levels. This problem is solved by adding 1 to the factor DF value. In this case, the factor DF is 5. Therefore, there are 6 levels of the factor.

(j) Laura replicated this experiment how many times?

Laura replicated this experiment 4 times.

(k) Suppose that Laura had actually conducted this experiment as a random complete block design and the sum of squares for the blocks was 12. Reconstruct the ANOVA display above to reflect this new situation. How much has the blocking reduced the estimate of the experimental error?

Source	*DF*	*SS*	*MS*	*F*

Factor	?	?	14.18	?
Error	?	37.75	?	
Total	23	108.63		

The blocking reduced the SS_{Error} by 12 and the MS_{Error} *by* 0.67.

Categorical Question 1:

A one sample test of proportions can be evaluated with a Z score

True False

A one-sample test of proportions can be evaluated with a Z score if the sample size is sufficiently large and if the assumptions for using a normal approximation are met.

Categorical Question 2:

A Chi-Square contingency table has 4 rows and 6 columns. How many Degrees of Freedom will the test of independence have?

A. 24
B. 23
C. 15
D. 9

The formula for calculating Degrees of Freedom (DF) for a Chi-Square test of independence is:

$$DF = (\# \text{ of rows} - 1)(\# \text{ of cols} - 1)$$

In this case, we have 4 rows and 6 columns:

$$DF = (4 - 1)(6 - 1)$$
$$DF = 3(5)$$
$$DF = 15$$

Therefore, a Chi-Square contingency table with 4 rows and 6 columns will have 15 DF.