Chapter 9 – Multiple Regression

Case 9.1.1. Effects of Light on Meadowfoam Flowering – A Randomized Experiment *R&S p.238-239*.

Step 1: Copy the data into a Minitab Worksheet: use these steps:

File \rightarrow Open Worksheet \rightarrow Browse your local directory and upload the csv file case0901.csv. The data will appear as columns in Minitab with titles FLOWERS, TIME, and INTENS. Note that both TIME and INTENS are categorical predictor variables.

The categorical predictor TIME has 2 levels: level 1 indicates timing level or day 0, i.e., Late (at PFI), and level 2 indicates timing level or day 24, i.e., Early (24 days before PFI).

The categorical predictor INTENS denotes light intensity (µmol/m²/sec) at 6 levels: 150, 300, 450, 600, 750, and 900. *See R&S Display 9.1*. The data is shown in tabular form in *R&S Display 9.2*.

Step 2: Scatterplots of FLOWER versus INTENS and FLOWER versus TIME are shown below, as well as summary statistics for FLOWER by TIME and INTENS.





Descriptive Statistics: FLOWERS

Results for TIME = 1.00

Variable	INTENS	Ν	N*	Mean	SI	E Mean	StDev	Minimum	Q1	Median	Q3
FLOWERS	150.00	2	0	69.85		7.55	10.68	62.30	*	69.85	*
	300.00	2	0	54.750		0.550	0.778	54.200	*	54.750	*
	450.00	2	0	55.75		6.15	8.70	49.60	*	55.75	*
	600.00	2	0	42.55		3.15	4.45	39.40	*	42.55	*
	900.00	2	0	39.35		2.55	3.61	36.80	*	39.35	*
Variable	INTENS	Ма	ximu	m							
FLOWERS	150.00		77.4	0							
	300.00	5	5.30	0							
	450.00		61.9	0							
	600.00		45.7	0							
	750.00		44.9	0							
	900.00		41.9	0							
Results f	or TIME	= 2	.00								
Variable	INTENS	Ν	N*	Mean	SE	Mean	StDev	Minimum	Q1	Median	Q3
FLOWERS	150.00	2	0	76.70		1.10	1.56	75.60	*	76.70	*
	300.00	2	0	73.55		4.45	6.29	69.10	*	73.55	*
	450.00	2	0	64.05		7.05	9.97	57.00	*	64.05	*
	600.00	2	0	57.55		5.35	7.57	52.20	*	57.55	*
	750.00	2	0	52.95		7.35	10.39	45.60	*	52.95	*
	900.00	2	0	48.50		4.10	5.80	44.40	*	48.50	*

!]]		
Variable	INTENS	Maxımum
FLOWERS	150.00	77.80
	300.00	78.00
	450.00	71.10
	600.00	62.90
	750.00	60.30
	900.00	52.60

Step 3: Fit a Multiple Linear Regression of FLOWER on two predictor variables, INTENS and TIME. To do this, Go to Stat \rightarrow Regression; select FLOWERS into Response Variable and select INTENS and TIME into Predictors.

a,

Regre	ession		×
C1 C2 C3	Flowers Time Intensity	R <u>e</u> sponses: Flowers	×
		<u>C</u> ontinuous predictors: Intensity Time	<u> </u>
		Categorical predictors:	_
			<u>^</u>
		Model Optio <u>n</u> s Co <u>d</u> ing <u>S</u> ta	epwise
	Select	<u>G</u> raphs <u>R</u> esults St	orage
	Help	<u>K</u>	Cancel

Regression Analysis: Flowers versus Intensity, Time

 Analysis of Variance

 Source
 DF
 Adj SS
 Adj MS
 F-Value
 P-Value

 Regression
 2
 3466.7
 1733.35
 41.78
 0.000

 Intensity
 1
 2579.8
 2579.75
 62.18
 0.000

 Time
 1
 887.0
 886.95
 21.38
 0.000

 Error
 21
 871.2
 41.49
 41.49

 Lack-of-Fit
 9
 215.3
 23.92
 0.44
 0.889

 Pure Error
 12
 655.9
 54.66
 54.66

Model Summary

	S	R-sq	R-s	q(adj)	R-sq(pred)		
6.441	.07	79.92%		78.00%	7	3.84%		
Coeff	icie	nts						
Term Const	ant	5	Coef 9.15	SE Coe 4.9	f T-V 5 1	alue 1.94	P-Value 0.000	VIF
Inter	nsity	-0.0	4047	0.0051	3 –	7.89	0.000	1.00
Time		1	2.16	2.6	3	4.62	0.000	1.00
Regre	essio	n Equa	tion					
Flowe	ers =	59.15	- 0.	04047 I	ntensi	ty +	12.16 Tin	ie
Fits	and 1	Diagno	stics	for Un	usual	Obser	vations	
Obs 2	Flowe	ers .40 6	Fit 5.24	Resid 12.16	Std Resid 2.08	R		
R La	arge :	residua	al					

Step 4: Scatterplot with Fitted Linear Regressions for the two groups, viz., Early Timing and Late Timing. Go to Graphs \rightarrow Scatterplot \rightarrow With Regression and Groups; select FLOWERS into Y Variables and INTENS into X Variables; select TIME into Categorical Variables for Grouping, and click Ok to produce this graph. Also produce Residual plots.

	Scatterplot - With Re	ression and Groups	×
:::: R	C1 FLOWERS C2 TIME C3 INTENS C4 day0 C5 day24 C6 L150 C7 L300 C8 L450 C9 L600 C10 L750 C11 L900 C12 INTENS*day24 C13 FIT51 C14 FIT52 C15 FIT53 C16 TIME_1.00		
	Select	Scale Labels Data Vie Multiple Graphs Data Options	∍w 5.2 9.1
	Help		8.1 el 8.1



Step 5: Scatterplot with Fitted Linear Regressions for the two groups, viz., Early Timing and Late Timing. Go to Graphs \rightarrow Scatterplot \rightarrow With Regression and Groups; select FLOWERS into Y Variables and TIME into X Variables; select INTENS into Categorical Variables for Grouping, and click Ok to produce this graph.





Step 6: **More Extensive way to carry out Multiple Regression,** with Categorical Predictors. Create Indicator Variables and Interaction Variables.

Create Indicator variables corresponding to the categorical variable TIME. Go to Calc \rightarrow Make Indicator Variables, and select TIME. This creates two new columns C4 and C5, we rename them day0 and day24. Similarly, we may also construct six indicator variables corresponding to INTENS, see *R&S Display 9.7* for the indicator variables L150 – L900.

Also, create an Interaction Variable between INTENS and day24 by going to Calc and calculating into Column C12 INTENS*day24.

\$	2 2 5	70 8 32 28	0 000			
٢	1ake:	Indicator ¥ariab	oles			×
	C1 C2 C3	FLOWERS TIME INTENS	Indicator variables for:	TIME	j	
	C4 C5	day0 day24	Store indicator variables in	columns:		
	C6	L150	Distinct Value		Column	
	68	1450	1.00	'TIME_1.00'		
i I	C9	L600	2.00	'TIME_2.00'		
	C10 L750 C11 L900 C12 INTENS*day24 C13 FIT51 C14 FIT52 C15 FIT53 C16 TIME_1.00 C17 TIME_2.00					
	L	Select				
		Help			<u>O</u> K	Cancel

Step 7: To run a regression of FLOWER on INTENS, day24 and INTENS*day24, go to STAT \rightarrow Regression and select variables for analysis as shown below, and results follow. The fitted values are stored in FITS1 in column C13. To do this, click on Storage and click on Fits.

Regression		R	egression: Graphs			×
C1 Flowers C2 Time C3 Intensity C4 time#intens	Responses: Flowers Continuous predictors: Intensity Time 'time *intens' Categorical predictors:			Residuals for plots:	Regular esiduals plity plot of residuals sus fits sus order ariables:	< x
Select	<u>M</u> odel	Optio <u>n</u> s Coding Graphs <u>R</u> esults	Select Help		<u>o</u> k	Cancel
Help		<u>O</u> K	Cancel			
Regression					×	1 1
C1 Flowers C2 Time C3 Inten: Reg C4 time*i	Responses: ression: Storage Fits Residuals Standardized residuals Deleted residuals Leverages Cook's distance DFITS	☐ Co <u>e</u> ffi ☐ Design	cients n <u>m</u> atrix	X		
Seli Help	Help		<u>O</u> K	Cancel OK	tepwise Storage Cancel	

Regression Analysis: Flowers versus Intensity, Time, time*intens

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	3467.28	1155.76	26.55	0.000

Intensity	1	281 62	281 62	6 47	0 019
incensie	-	201.02	201.02	0.17	0.010
Time	1	153.22	153.22	3.52	0.075
time*intens	1	0.58	0.58	0.01	0.910
Error	20	870.66	43.53		
Lack-of-Fit	8	214.73	26.84	0.49	0.841
Pure Error	12	655.92	54.66		
Total	23	4337.94			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
6.59795	79.93%	76.92%	70.95%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	60.10	9.71	6.19	0.000	
Intensity	-0.0423	0.0166	-2.54	0.019	10.00
Time	11.52	6.14	1.88	0.075	5.20
time*intens	0.0012	0.0105	0.12	0.910	14.20

Regression Equation

Flowers = 60.10 - 0.0423 Intensity + 11.52 Time + 0.0012 time*intens

Fits and Diagnostics for Unusual Observations

				Std	
Obs	Flowers	Fit	Resid	Resid	
2	77.40	65.46	11.94	2.11	R

R Large residual



Next, carry out a regression of FLOWER on INTENS and day24, but no interaction term. Output is shown below.

Regression: Graphs	×
Residuals for plots: Regular	Responses:
	Flowers
Residuals plots	
C Individual plots	Continuous predictors:
☐ <u>H</u> istogram of residuals	Intensity
☐ Normal probability plot of residuals	
☐ Residuals <u>v</u> ersus fits	
☐ Re <u>s</u> iduals versus order	⊥
Four in one	Categorical predictors:
Deciduals versus the variables:	Time
Select	Model Options Coding Stepwise
Help <u>O</u> K Cancel	<u>G</u> raphs <u>R</u> esults <u>St</u> orage
49.6 1 450]
619 1 450 Help	<u>Q</u> K Cancel

Regre	ssion							×
C1 C2	Flowe	rs Regression:	Responses:				×	
C3	Inten:	Fits	storage	Co <u>e</u> fficients			~	-
		Residual	ls	Design <u>m</u> atrix				
		I <u>S</u> tandar	dized residuals residuals					
		Leverag	jes					<u>_</u>
			distance					
		, <u> </u>						
								<u>_</u>
								tepwise
	Sel	Help		<u></u>	к	Cancel		Storage
	Help					<u>0</u> K		Cancel

Regression Analysis: Flowers versus Intensity, Time

Method							
Categorical predictor coding (1, 0)							
Analysis of Va	ırian	ce					
Source	DF	Adj SS	Adj MS	F-Value	P-Value		
Regression	2	3466.7	1733.35	41.78	0.000		
Intensity	1	2579.8	2579.75	62.18	0.000		
Time	1	887.0	886.95	21.38	0.000		
Error	21	871.2	41.49				
Lack-of-Fit	9	215.3	23.92	0.44	0.889		
Pure Error	12	655.9	54.66				
Total	23	4337.9					
Model Summary							
S R-s	q R	-sq(adj)	R-sq(pr	ed)			
6.44107 79.92	18	78.00%	73.	84%			

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	71.31	3.27	21.78	0.000	
Intensity	-0.04047	0.00513	-7.89	0.000	1.00
Time					
2	12.16	2.63	4.62	0.000	1.00

Regression Equation

Time 1 Flowers = 71.31 - 0.04047 Intensity 2 Flowers = 83.46 - 0.04047 Intensity

Fits and Diagnostics for Unusual Observations

				Sta	
0bs	Flowers	Fit	Resid	Resid	
2	77.40	65.24	12.16	2.08	R



Finally, carry out a regression of FLOWER on INTENS only, output is shown below.

Regress	ion		×
C1 1 C2 7 C3 1 C4 7	Flowers Time Intensity time*intens FITS	R <u>e</u> sponses: Flowers	4
C5 FII C6 FII C7 FII	FITS_1 FITS_2	<u>Continuous predictors:</u> Intensity	<u></u>
		C <u>a</u> tegorical predictors:	×
		Model Optio <u>n</u> s Co <u>d</u> ing	Stepwise
_	Select	<u>G</u> raphs <u>R</u> esults	S <u>t</u> orage
He	elp	<u>K</u>	Cancel

Regression Analysis: Flowers versus Intensity

Analysis of Var	rianc	e					
Source Regression Intensity Error Lack-of-Fit Pure Error Total	DF 1 22 4 18 23	Adj SS 2579.8 2579.8 1758.2 103.8 1654.4 4337.9	Adj MS 2579.75 2579.75 79.92 25.94 91.91	F-Value 32.28 32.28 0.28	P-Value 0.000 0.000 0.886		
Model Summary							
S R-so 8.93966 59.479	S R-sq R-sq(adj) R-sq(pred) 8.93966 59.47% 57.63% 52.42%						
Coefficients							
Term Constant 7 Intensity -0.0	Coef 77.39)4047	E SE Coe 9 4.1 7 0.0071	ef T-Val 16 18. 12 -5.	lue P-Va .60 0.0 .68 0.0	lue VIF 000 000 1.00		
Regression Equation							
Flowers = 77.39 - 0.04047 Intensity							



Case 9.1.2. Why do some Mammals have Large Brains for their Size – An Observational Study. *R&S p.239-242*.

Step 1: Copy the data into a Minitab Worksheet: use these steps:

File \rightarrow Open Worksheet \rightarrow Browse your local directory and upload the csv file case0902.csv. To display the data in Minitab, go to Data \rightarrow Display Data, and copy the columns C1-C5 in the window on the right. The data will appear as five columns in Minitab. *See R&S Display 9.4* for data display on 96 different mammals.

Step 2: Go to Graph \rightarrow Matrix Plots \rightarrow Simple; copy in Columns C2-C5 into the window on the right and click OK to get the plot below.

	Matrix Plot - Matrix of Pl	ots, Simple			×
	C2 BRAIN C3 BODY C4 GESTATIO C5 LITTER	Graph variables: BRAIN-LITTER		Ā	
		Matrix Options	<u>S</u> cale D <u>a</u> ta Options	Labels	
ГІ: 5	Select				
-	Help		<u>o</u> k	Cancel	

i>	Matrix Display O Eull O Lower left O Upper right		
	Variable Label Placement		
	Diagonal Boundary		
H			
-	Help	<u>o</u> ĸ	Cancel



Step 3: Go to Calc \rightarrow Calculator; and save into Columns C6-C9, the LN transformations of C2-C5. Then, go to Graph \rightarrow Matrix Plots \rightarrow Simple; copy in Columns C6-C9 into the window on the right and click OK to get the plot below.

Matrix Plot - Matrix of Pl	ots, Simple		×
C2 BRAIN C3 BODY C4 GESTATIO C5 LITTER C6 log BRAIN C7 log BODY C8 log GESTATION C9 log LITTER	<u>G</u> raph variables: C6-C9 <u>Matrix Options</u> <u>D</u> ata View	<u>S</u> cale D <u>a</u> ta Options	Labels
Select			-
Help		<u>o</u> k	Cancel
I 2I 0.1310	13 _3.01594	3 93183 0 404	547



Step 4: Go to Stat \rightarrow Regression; select C6 log BRAIN into Response window and select C7- C9 into Predictors window; select Residuals Plots for graphs and click ok. See *R&S Display 9.15*.

Regre	ssion		×	
C2 Brain C3 Body C4 Gestation C5 Litter C6 LN(Brain) C7 LN(Body) C8 LN(Gestati C9 LN(Litter)	Brain Body Gestation Litter LN(Brain)	R <u>e</u> sponses: ['LN(Brain)'	*	
	LN(Body) LN(Gestation) LN(Litter)	<u>C</u> ontinuous predictors: 'LN(Body)' 'LN(Gestation)' 'LN(Litter)'	<u>_</u>	
		Ci	Categorical predictors:	
		Model Options Coding	Stepwise	
	Select	<u>G</u> raphs <u>R</u> esults	S <u>t</u> orage	
	Help	<u>O</u> K	Cancel	

Regression Analysis: LN(BRAIN) versus LN(Body), LN(Gestation), LN(Litter)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	427.076	142.359	631.60	0.000
LN(Body)	1	70.189	70.189	311.41	0.000
LN(Gestation)	1	1.986	1.986	8.81	0.004
LN(Litter)	1	1.612	1.612	7.15	0.009
Error	92	20.736	0.225		
Total	95	447.812			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.474755	95.37%	95.22%	95.03%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.855	0.662	1.29	0.200	
LN(Body)	0.5751	0.0326	17.65	0.000	3.79
LN(Gestation)	0.418	0.141	2.97	0.004	6.27
LN(Litter)	-0.310	0.116	-2.67	0.009	2.54

Regression Equation

LN(BRAIN) = 0.855 + 0.5751 LN(Body) + 0.418 LN(Gestation) - 0.310 LN(Litter)

Fits and Diagnostics for Unusual Observations

		Std Resid	Resid	Fit	LN(BRAIN)	0bs
	R	2.45	1.144	6.233	7.378	25
Х		-0.04	-0.018	5.211	5.193	28
	R	-2.03	-0.931	7.311	6.380	48
	R	3.36	1.575	5.595	7.170	53
Х		-0.18	-0.075	2.937	2.862	72
	R	-2.05	-0.954	6.476	5.521	86

R Large residual X Unusual X