Business Analytics Yue Summer 2025

Assignment #3, Due 7/14/2025 or 7/15/2025

1. The following data show the results of the first job performance measured (in scores) of 10 employees and their grade point average.

First Job Performance Score (y)	GPA (x)
26	1.8
31	2.3
28	2.6
30	2.4
34	2.8
38	3.0
41	3.4
44	3.2
40	3.6
43	3.8

   (a) Develop a least squares estimated regression equation.
   (b) Compute the coefficient of determination and comment on the strength of the regression relationship.
   (c) Is the slope significant? Use a t test and let $\alpha = .05$.
   (d) At the 5% level of significance, test to determine if the model is significant (i.e., perform an F test).

2. Go to https://dasl.datadescription.com/ and download the heights and weights of students in a statistics class ("Heights_and_weights"). Then use the variable "Weight" as the dependent variable and "Height" as the independent variable.
   (a) Use Excel (or other statistical software) to finish the regression analysis. Then compare your outputs from t-test and ANOVA table, and comment on what you find.
   (b) Construct a 99% confidence interval for the estimate of slope.
   (c) Suppose we want to predict the weight of student of heights 60, 65, and 70 inches. Give your prediction values and their prediction intervals.

3. **Used Car Mileage and Price.** The Toyota Camry is one of the best-selling cars in North America. The cost of a previously owned Camry depends upon many factors, including the model year, mileage, and condition. To investigate the relationship between the car's mileage and the sales price for a 2007 model year Camry, the following data show the mileage and sale price for 19 sales.

(a) Develop a scatter diagram with the car mileage on the horizontal axis and the price on the vertical axis.

(b) Develop the estimated regression equation that could be used to predict the price ($1000s) given the kilometers (1000s).

(c) Test for a significant relationship at the .05 level of significance.

(d) Did the estimated regression equation provide a good fit? Provide an interpretation for the slope of the estimated regression equation.

(e) Suppose that you are considering purchasing a previously owned 2007 Camry that has been driven 97,000 kilometers. Using the estimated regression equation developed in part (c), predict the price for this car. Is this the price you would offer the seller?

Kilometers (1000s)	Price ($1000s)
35	16.2
47	16.0
58	13.8
76	11.5
101	12.5
124	12.9
117	11.2
140	13.0
148	11.8
163	10.8
177	8.3
45	12.5
95	11.1
109	15.0
109	12.2
146	13.0
68	15.6
105	12.7
177	8.3

4. **Buy Versus Rent.** Occasionally, it has been the case that home prices and mortgage rates dropped so low that in a number of cities the monthly cost of owning a home was less expensive than renting. The following data show the average asking rent for 10 markets and the monthly mortgage on the median priced home (including taxes and insurance) for 10 cities where the average monthly mortgage payment was less than the average asking rent (The Wall Street Journal).

(a) Develop the estimated regression equation that can be used to predict the monthly mortgage given the average asking rent.

(b) Construct a residual plot against the independent variable.

(c) Do the assumptions about the error term and model form seem reasonable in light of the residual plot?

City	Rent ($)	Mortgage ($)
Atlanta	840	539
Chicago	1062	1002
Detroit	823	626
Jacksonville, Fla.	779	711
Las Vegas	796	655
Miami	1071	977
Minneapolis	953	776
Orlando, Fla.	851	695
Phoenix	762	651
St. Louis	723	654

5. Use the variable selection commands of regression analysis in Minitab for the following data, e.g. stepwise and best subset regression. (Note: See the following outputs for reference.) Compare their differences and give your comments.

---

y: Total heat flux (kwatts)
x1: Insolation(watts/m2)
x2: Position of focal point in east direction (inches)
x3: Position of focal point in south direction (inches)
x4: Position of focal point in north direction (inches)
x5: Time of day

---

y	x1	x2	x3	x4	x5
271.8	783.35	33.53	40.55	16.66	13.2
264	748.45	36.5	36.19	16.46	14.11
238.8	684.45	34.66	37.34	17.66	15.68
230.7	827.8	33.13	32.52	17.5	10.53
254.6	860.45	35.75	33.71	16.4	11
257.9	875.15	34.46	34.14	16.28	11.31
263.9	909.45	34.6	34.85	16.06	11.96
266.5	905.55	35.38	35.89	15.93	12.58
229.1	756	35.85	33.53	16.6	10.66
239.3	769.35	35.68	33.79	16.41	10.85
258	793.5	35.35	34.72	16.17	11.41
257.6	801.65	35.04	35.22	15.92	11.91
267.3	819.65	34.07	36.5	16.04	12.85
267	808.55	32.2	37.6	16.19	13.58
259.6	774.95	34.32	37.89	16.62	14.21
240.4	711.85	31.08	37.71	17.37	15.56
227.2	694.85	35.73	37	18.12	15.83
196	638.1	34.11	36.76	18.53	16.41
278.7	774.55	34.78	34.62	15.54	13.1
272.3	757.9	35.77	35.4	15.7	13.63

268.4	753.35	36.44	35.96	16.45	14.51
254.5	704.7	37.82	36.26	17.62	15.38
224.7	666.8	35.07	36.34	18.12	16.1
181.5	568.55	35.26	35.9	19.05	16.73
227.5	653.1	35.56	31.84	16.51	10.58
253.6	704.05	35.73	33.16	16.02	11.28
263	709.6	36.46	33.83	15.89	11.91
265.8	726.9	36.26	34.89	15.83	12.65
263.8	697.15	37.2	36.27	16.71	14.06

- Output of Stepwise Regression:

Response is HeatFlux on 5 predictors, with N = 29

Step	1	2	3	4
Constant	607.1	483.7	389.2	270.2
North	-21.4	-24.2	-24.1	-21.1
T-Value	-8.34	-12.48	-12.92	-8.91
P-Value	0.000	0.000	0.000	0.000
South		4.80	5.32	5.34
T-Value		5.04	5.52	5.83
P-Value		0.000	0.000	0.000
East			2.1	3.0
T-Value			1.75	2.40
P-Value			0.092	0.025
Insolation				0.052
T-Value				1.92
P-Value				0.067
S	12.3	8.93	8.60	8.17
R-Sq	72.05	85.87	87.41	89.09
R-Sq(adj)	71.02	84.78	85.90	87.27
Mallows Cp	38.5	9.1	7.6	5.8

- Output of Best Subset Regression:

Best Subsets Regression: HeatFlux versus Insolation, East, ...

Response is HeatFlux

Vars	R-Sq	R-Sq(adj)	Mallows Cp	S	Insolation	East	South	North	Time
1	72.1	71.0	38.5	12.328				X	
1	39.4	37.1	112.7	18.154	X				
2	85.9	84.8	9.1	8.9321			X	X	
2	82.0	80.6	17.8	10.076				X	X
3	87.4	85.9	7.6	8.5978		X	X	X	
3	86.5	84.9	9.7	8.9110	X		X	X	
4	89.1	87.3	5.8	8.1698	X	X	X	X	
4	88.0	86.0	8.2	8.5550	X		X	X	X
5	89.9	87.7	6.0	8.0390	X	X	X	X	X