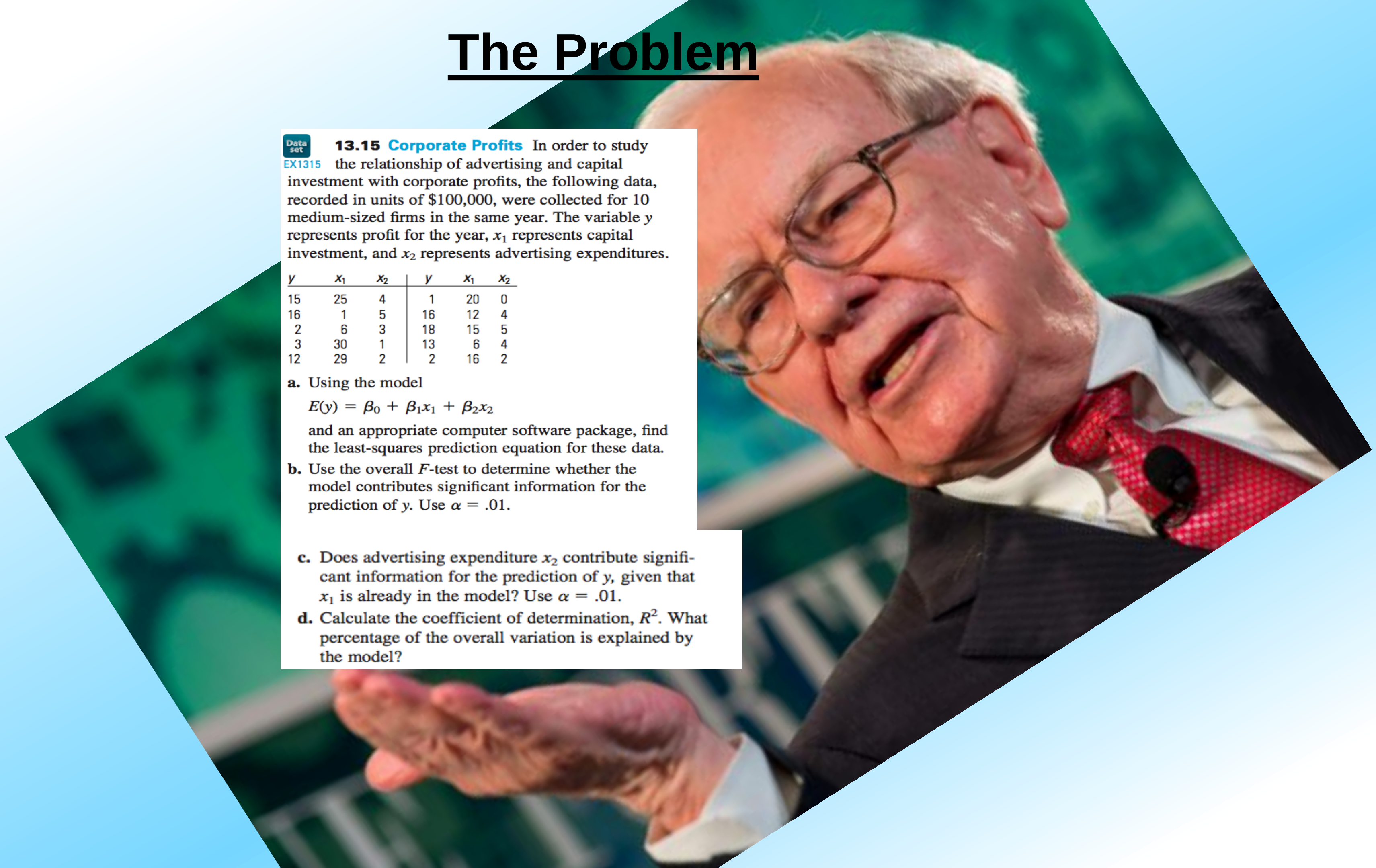
Group 4 Presentation (2)
MTH-404

By:Jeesun S.
Paul Ycay
Ali Syed

The Problem



13.15 Corporate Profits In order to study the relationship of advertising and capital investment with corporate profits, the following data, recorded in units of \$100,000, were collected for 10 medium-sized firms in the same year. The variable y represents profit for the year, x_1 represents capital investment, and x_2 represents advertising expenditures.

Data set EX1315

y	x_1	x_2	y	x_1	x_2
15	25	4	1	20	0
16	1	5	16	12	4
2	6	3	18	15	5
3	30	1	13	6	4
12	29	2	2	16	2

a. Using the model

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

and an appropriate computer software package, find the least-squares prediction equation for these data.

b. Use the overall F-test to determine whether the model contributes significant information for the prediction of y. Use $\alpha = .01$.

c. Does advertising expenditure x_2 contribute significant information for the prediction of y, given that x_1 is already in the model? Use $\alpha = .01$.

d. Calculate the coefficient of determination, R^2. What percentage of the overall variation is explained by the model?

Excel Data

y	x_1	x_2
15	25	4
16	1	5
2	6	3
3	30	1
12	29	2
1	20	0
16	12	4
18	15	5
13	6	4
2	16	2

Regression equation:

$$y = -8.18 + 0.29x_1 + 4.43x_2$$

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 99.0%*	*Upper 99.0%*
Intercept	-8.17702	4.205988	-1.94414	0.092967	-18.1226	1.768562	-22.8958	6.541765
x1	0.292132	0.135714	2.152556	0.068355	-0.02878	0.613044	-0.1828	0.767061
x2	4.434303	0.800243	5.541193	0.000868	2.542028	6.326578	1.633864	7.234741

Analysis of variance

Source	DF	SS	MS	F	P
Regression	2	355.22	177.61	16.28	0.002
Residual Error	7	76.38	10.91		
Total	9	431.60			

Source	DF	Seq SS
X1	1	20.16
X2	1	335.05

A) From the analysis of variance, we can conclude that the least square prediction equation is $\hat{y} = -8.18 + 0.29x_1 + 4.43x_2$

B) Use the overall F-test to determine whether the model contributes significant information for the prediction of y. Use α = 0.01

$h(o)$: β1 = 0

$h(a)$: β1 ≠ 0

f = MSR/MSE

f = 177.61 / 10.91

f= 16.28

TABLE 6 *(continued)*

df_2	α	df_1 = 1	2	3	4	5	6	7	8	9
10	0.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
	0.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
	0.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
	0.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
	0.005	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97
11	0.100	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
	0.050	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
	0.025	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
	0.010	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
	0.005	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54
12	0.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
	0.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
	0.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
	0.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
	0.005	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20
13	0.100	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
	0.050	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
	0.025	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
	0.010	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
	0.005	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94
14	0.100	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
	0.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
	0.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
	0.010	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
	0.005	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72
15	0.100	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
	0.050	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
	0.025	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
	0.010	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
	0.005	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54
16	0.100	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
	0.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
	0.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
	0.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
	0.005	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38
17	0.100	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
	0.050	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
	0.025	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
	0.010	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
	0.005	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25
18	0.100	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
	0.050	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
	0.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
	0.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
	0.005	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14
19	0.100	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
	0.050	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
	0.025	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
	0.010	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
	0.005	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04
20	0.100	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
	0.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
	0.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
	0.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
	0.005	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96

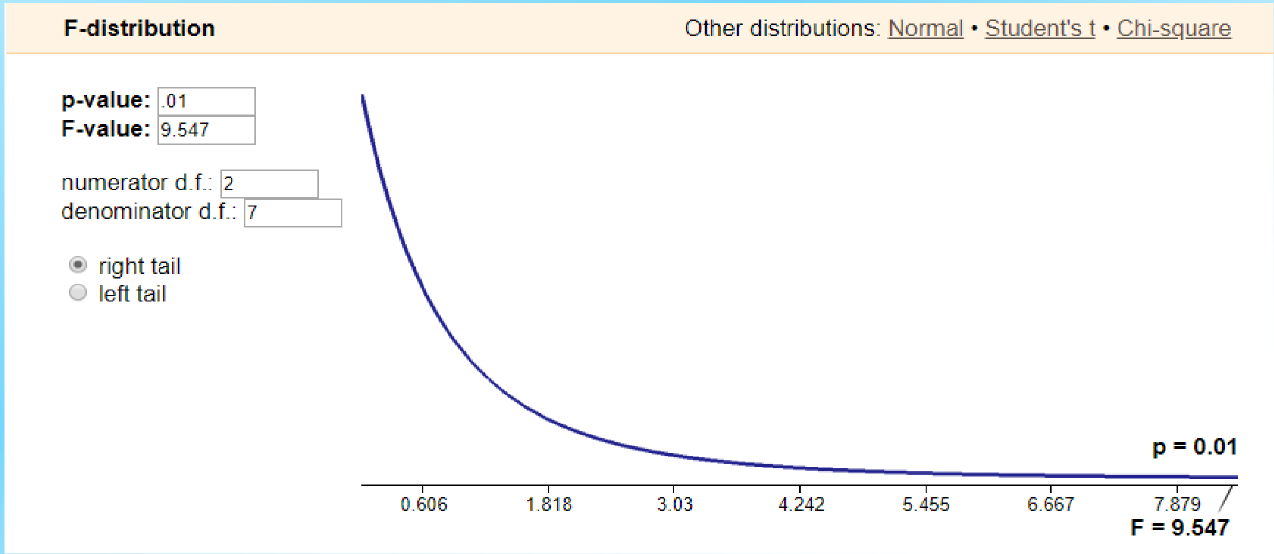
f_a = 0.01 with df_1=2 and df_2=7 is non existent in table 6, so we go to the p-value. Since the p-value, from ANOVA table was 0.002 < α = 0.01, **we reject the Null Hypothesis** and conclude that there is a significant evidence for the prediction of y.

F-Test

$F_{.01,2,7} = 9.55$

$\because$ our $f > f_a$

We Can conclude that the regression to be highly significant. At least, one of the predictor variables is contributing significant information for the prediction of the response variable y. That is, profits for the year.



C)

c. Does advertising expenditure x_2 contribute significant information for the prediction of y, given that x_1 is already in the model? Use $\alpha = 0.01$.

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 99.0%	Upper 99.0%
Intercept	-8.17702	4.205988	-1.94414	0.092967	-18.1226	1.768562	-22.8958	6.541765
x1	0.292132	0.135714	2.152556	0.068355	-0.02878	0.613044	-0.1828	0.767061
x2	4.434303	0.800243	5.541193	0.000868	2.542028	6.326578	1.633864	7.234741

$h(o)$: = x_2 is not significant $h(a)$: = x_2 is significant

From the table above, the t-stat associated with x_2 is 5.54. Looking in the table, the p-value associated with x_2 is 0.0008, which is less than α = 0.01. Since it is less than, we can reject the null hypothesis and claim the significance of x_2. Calculating the p-value by hand was difficult since df of 7 with t stat 5.54 exceeded the t-table, thus excel had to interpret the p-value.

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										

D)

d. Calculate the coefficient of determination, R^2. What percentage of the overall variation is explained by the model?

Regression Statistics	
Multiple R	0.907204
R Square	0.823019
Adjusted R Square	0.772453
Standard Error	3.30335
Observations	10

r^2=SSR/Total SS

=355.251/431.6

=0.823102409

Standard Error = 3.30335
R-Sq = 82.3%
R-Sq(adj) = 77.2%

∴ A 82% overall variation is explained by the model, meaning this model works well with the given data