Regression Analysis Case Study
Choosing the Best Fitting Model
Predicting SAT Scores from GPA and Income

Name Key

There has been a lot of discussion regarding the relationship between Scholastic Aptitude Test (SAT) scores and test-takers' family income (*New York Times*, August 27, 2009). It is generally believed that the wealthier a student's family, the higher the SAT score. Another commonly used predictor for SAT scores is the student's grade point average (GPA). Consider the data collected on 24 students.

1. Estimate 3 models and write the regression equations **(Round your answers to 4 decimal places.)**

MODEL 1	$SAT = \beta_0 + \beta_1 Income + \varepsilon$	$\widehat{SAT} = 1616.363 + 0.0015\,Income$
MODEL 2	$SAT = \beta_0 + \beta_1 GPA + \varepsilon$	$\widehat{SAT} = 1259.638 + 141.468\,GPA$
MODEL 3	$SAT = \beta_0 + \beta_1 Income + \beta_2 GPA + \varepsilon$	$\widehat{SAT} = 1104.26 + 0.0017\,Income + 150.99\,GPA$

2. Conduct individual hypothesis tests to determine if the coefficients differ from zero.

	Hypotheses	Coefficient Estimate (include sign)	P-value	Is the corresponding variable significant in explaining the variation in y?
MODEL 1 **Variable:** Income	$H_0: \beta_1 = 0$ $H_A: \beta_1 \neq 0$	0.0015	0.0204	Yes
MODEL 2 **Variable:** GPA	$H_0: \beta_1 = 0$ $H_A: \beta_1 \neq 0$	141.468	0.00002	Yes
MODEL 3 **Variable:** Income	$H_0: \beta_1 = 0$ $H_A: \beta_1 \neq 0$	0.0017	0.000001	Yes
MODEL 3 **Variable:** GPA	$H_0: \beta_2 = 0$ $H_A: \beta_2 \neq 0$	150.992	0.00000002	Yes

Regression Analysis Case Study
Choosing the Best Fitting Model
Predicting SAT Scores from GPA and Income

3. Summarize the 3 models in the table below

	MODEL 1	MODEL 2	MODEL 3
Multiple R	0.4702	0.7551	0.9300
R^2	0.2211	0.5701	0.8649
Adjusted R^2	0.1857	0.5506	0.8520
Standard Error of the Estimate	76.2217	56.6262	32.4902

4. Which model is more appropriate for making predictions? Explain the comparisons you made to determine the best-fitting model.

Model 3 is the best fitting model because it has
1) the highest Adj. R^2 at 0.8520
2) the lowest standard error of the estimate, S_e, at 32.49

5. In terms of the data, interpret the Coefficient of Determination for the model you chose.

Approximately 85.2% of the variation in SAT scores is explained by income and GPA.

6. In terms of the data, interpret the slope(s) of the line of best fit.

As income increase by $1000, SAT scores increase by 1.7 points, holding GPA constant.

As GPA increases by 1 point, SAT scores increase by 150.992 points, holding income constant.

7. Use the best-fitting model to predict SAT given the mean value of the explanatory variable(s). The mean income is $7,2833.33 and the mean GPA is 3.2783. **(Round intermediate calculations to 4 decimal places and final answer to 2 decimal places.)**

$$\hat{SAT} = 1104.258 + 123.817 + 494.997 = 1723$$

Regression Analysis Case Study
Choosing the Best Fitting Model
Predicting Sales from Advertising Costs and Unemployment

Name Key

The manager of a small sushi restaurant has noticed that the weak economic environment has hampered foot traffic in the area and caused a decline in sales. To thwart this trend, the manager designs an aggressive advertising campaign, which includes two-for-one coupons and early bird specials. Despite the fact that advertising increases overall costs, the manager believes the campaign has positively affected sales at the restaurant. The manager looks at monthly sales and advertising costs for the restaurant as well as the monthly unemployment rate from the county. Consider the data collected for 17 months. sales are in thousands of dollars

1. Estimate 3 models and write the regression equations **(Round your answers to 4 decimal places.)**

MODEL 1	$Sales = \beta_0 + \beta_1 AdsCost + \varepsilon$	$\widehat{Sales} = 15.26 + 0.023\ AdsCost$
MODEL 2	$Sales = \beta_0 + \beta_1 Unemp + \varepsilon$	$\widehat{Sales} = 31.28 - 0.418\ Unemp$
MODEL 3	$Sales = \beta_0 + \beta_1 AdsCost + \beta_2 Unemp + \varepsilon$	$\widehat{Sales} = 17.51 + 0.027\ AdsCost - 0.688\ Unemp$

2. Conduct individual hypothesis tests to determine if the coefficients differ from zero.

	Hypotheses	Coefficient Estimate (include sign)	P-value	Is the corresponding variable significant in explaining the variation in y?
MODEL 1 **Variable:** AdsCost	$H_0: \beta_1 = 0$ $H_A: \beta_1 \neq 0$	0.023	0.0075	Yes
MODEL 2 **Variable:** Unemp	$H_0: \beta_1 = 0$ $H_A: \beta_1 \neq 0$	−0.418	0.322	No
MODEL 3 **Variable:** AdsCost	$H_0: \beta_1 = 0$ $H_A: \beta_1 \neq 0$	0.027	0.0015	Yes
MODEL 3 **Variable:** Unemp	$H_0: \beta_2 = 0$ $H_A: \beta_2 \neq 0$	−0.688	0.038	Yes

Regression Analysis Case Study
Choosing the Best Fitting Model
Predicting Sales from Advertising Costs and Unemployment

3. Summarize the 3 models in the table below

	MODEL 1	MODEL 2	MODEL 3
Multiple R	0.6234	0.2555	0.7455
R^2	0.3887	0.0653	0.5558
Adjusted R^2	0.3479	0.003	0.4924
Standard Error of the Estimate	2.3078	2.8537	2.0362

4. Which model is more appropriate for making predictions? Explain the comparisons you made to determine the best-fitting model.

Model 3 is the best fitting model because it has
1) the highest Adj. R^2 of 0.4924
2) the smallest standard error of the estimate at 2.0362 S_e

5. In terms of the data, interpret the Coefficient of Determination for the model you chose.

Approximately 49% of the variation in sales is explained by advertising costs and the unemployment rate.

6. In terms of the data, interpret the slope(s) of the line of best fit.

- For every #1 increase in advertising costs, sales increase by 0.027 thousand dollars, holding unemployment constant.
- For every 1% increase in the unemployment rate, sales decreases by 0.688 thousand dollars, holding advertising costs constant.

7. Use the best-fitting model to predict Sales given the mean value of the explanatory variable(s). The mean Advertising cost is 582.35 and the mean Unemployment rate is 6.26. **(Round intermediate calculations to 4 decimal places and final answer to 2 decimal places.)**

$\hat{Sales} = 17.51 + 15.723 - 4.307 = 28.922$

$28,922