Introduction to Statistics: Homework 4

Model Answers

1.

a. The coefficient on # of days per week exercised indicates that for each additional day that an individual exercises, we expect them to lose 6.6 pounds. The T-statistic (6.0) has an absolute value greater than 2, so we can reject the null hypothesis that there is no relationship between days of the week exercised and weight loss. In other words, the relationship is statistically significant.

b. The coefficient on the constant means that the model predicts that individuals who exercise 0 days per week are expected to gain 8 pounds. The t-statistic indicates that this value is statistically distinguishable from zero.

c. The expected change in weight for someone who exercises 3 days per week is -8.0+3*6.6 = 11.8 pounds.

d. The estimate of the relationship between exercise and weight loss is probably biased by our failure to account for the fact that the number of days per week a person exercises is likely to be correlated with other explanations for weight loss. For example, people who exercise a lot are likely to eat more healthily. If we included a measure of the amount of junk food an individual eats as an independent variable (controlled for that variable) we would probably expect the estimate of the relationship between exercise and weight loss to be smaller. This is because some of the relationship between exercise and weight loss can also be explained by eating habits.

2.

a. The coefficient on number of days per week exercise is the estimated relationship between this variable and weight loss among men (for whom Gender=0). Among men, each additional day of exercise is associated with losing 9.6 pounds. The t-statistic is greater than 2, indicating that this relationship is distinguishable from zero.
   The coefficient on gender is the estimated effect of gender (being female rather than male) on weight loss among those who exercise 0 days per week. This coefficient means that among this group, females are expected to gain 2.3 pounds more than males. This relationship is also statistically distinguishable from zero.

b. The statistical significance of the coefficient on the interaction term is statistically significant. This tells us that the relationship between exercise and weight loss is significantly different across genders. Symmetrically this means that the effects of gender on weight loss depend on how many days per week an individual exercises.

c. The estimated slope of the relationship between days per week exercise and weight loss among men is 9.6. The estimated slope among women is 9.6-2.4 = 7.2.

3. Most Americans celebrate the winter holiday season in some way, regardless of their religious affiliation. Let's say we are interested in estimating the relationship between religious affiliation and spending on gifts during the winter holiday season. We predict the number of dollars an individual spends based on indicators for religious affiliation (leaving "atheist or agnostic" as the omitted category) and a measure of household income (in tens of thousands of dollars; for example, $40,000 per year = 4).

	Coefficient	Standard Error	T
Protestant	150.2	40.2	3.7
Roman Catholic	115.5	38.7	3.0
Jewish	95.2	35.4	2.7
Other Religion	45.6	34.4	1.3
Income ($10,000s)	256.4	30.8	8.3
Constant	85.0	34.2	2.5

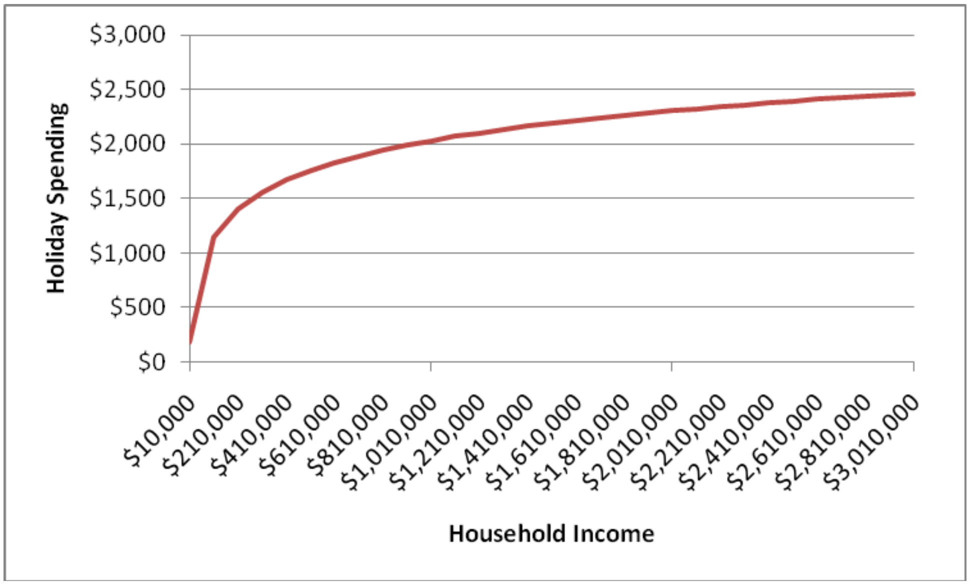
a. The coefficient on Other Religion means that, after controlling for the effects of income, individuals in this religious category are expected to spend $45.60 more than those who are atheist or agnostic (the reference category). The t-statistic is 1.3 indicating that this difference is not statistically distinguishable from zero.

b. The coefficient on Income means that, holding religious affiliation constant, each one unit increase in income (i.e., each increase of $10,000 in household income) is associated with a $256.40 increase in spending on holiday gifts. This slope is statistically different from zero.

c. The predicted amount of holiday season spending for a Roman Catholic with a household income of $200,000 per year is 85.0+115.5+256.4*20 = $5328.50.

d. The predicted amount of holiday season spending for a Protestant with a household income of $2,500,000 per year is 85.0+150.2+256.4*250 = $64335.20

e. The relationship might look something like the graph below. While we expect holiday spending to increase as income increases, we expect the increase in holiday spending associated with each unit increase in income to diminish as income gets very high. This transformation may fit the data better because we do not expect a change in income from $2,500,000 to $2,550,000 per year to increase holiday spending as much as an increase in income from $50,000 to $100,000.



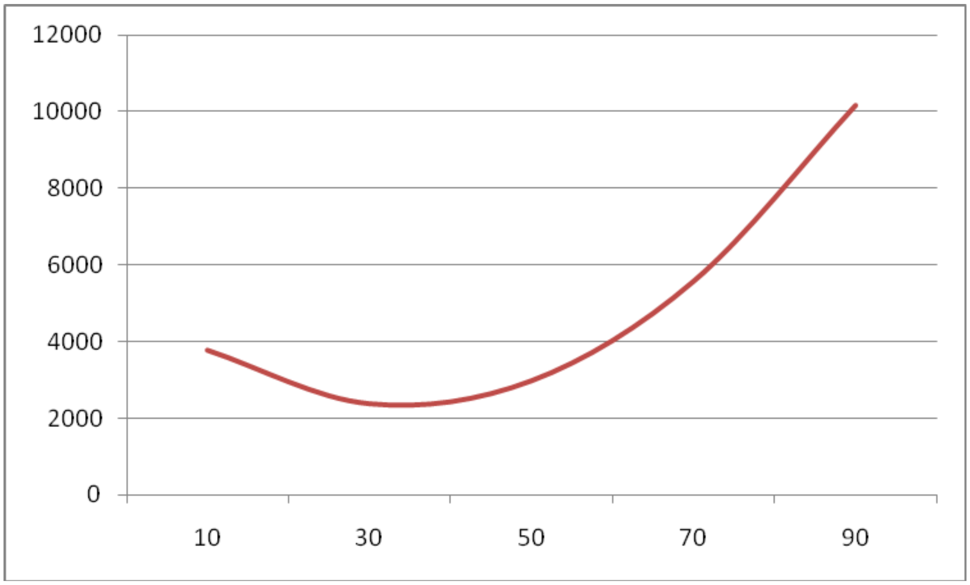
4.

a. The coefficient on age-squared is statistically significant (absolute value of the t-statistic is greater than 2). This means that allowing the relationship between age and health care costs to curve significantly improves the fit of the model.

b.

Age	Predicted value (health care cost)
10	$10*-170+10^2*2.5+220+5000=3770$
30	2370
50	2970
70	5570
90	10170

c. The relationship between age and health spending looks like the graph below. Health costs are fairly high among children, but go down among young adults. The model estimates that costs begin to rise again once a person reaches the age of about 40.



5.

a. Simply comparing achievement test scores across public and private schools is not a particularly persuasive way to estimate the effects of private schooling. Students at private schools are likely to differ systematically from those who attend public schools in ways that are likely to confound our estimate of the effects of private schooling. For example, students at private schools may be more likely to have wealthier or better-educated parents. These factors may explain the fact that students in private school perform better than those in public schools. In other words, students in private schools may perform better because their parents are (on average) better able to support their education rather than because private schools provide a better education.

b. The experiment described would provide a more defensible estimate of the effect of private education. Because students would be randomly assigned to attend public or private school we would not be faced with the problem of students in the two types of schools differing in systematic ways. The only difference between the two groups should be whether they were randomly selected to be "treated" with a public or private education. If we estimated a regression model predicting test scores with an indicator variable (1=selected to attend private school; 0=not selected), we would not need to control for other variables because in an experimental design like this there should not be any confounding explanations for the relationship between being selected to attend private school and test scores.