1. Which of these equations is it possible to estimate using regression techniques?

   a.) \[ Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon \]

   This equation is already in a form interpreted to be capable of estimating through standard regression techniques. See equation (1.12) in the Studenmund book – this is the definition of a multivariate regression model with two explanatory variables.

   b.) \[ Y_i = e^{\beta_0 + \beta_1 X_{1i} + \varepsilon_i} \]

   This equation is capable of being transformed into a something a little more familiar:

   First, take the log of both sides:

   \[ \ln Y_i = \ln e^{\beta_0 + \beta_1 X_{1i} + \varepsilon_i} \]

   Using the properties of logs given in Chapter 7 of Studenmund:

   \[ \ln Y_i = (\beta_0 + \beta_1 X_{1i} + \varepsilon_i) \ln e \]

   Since the \( \ln e = 1 \), we then have:

   \[ \ln Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i \]

   which is just the standard semilog equation from Chapter 7.

   c.) \[ Y_i = \beta_0 X_{1i}^{\beta_1} X_{2i}^{\beta_2} e^{\varepsilon_i} \]

   Again, this equation can be transformed into something more useful by taking the log of both sides:

   \[ \ln Y_i = \ln (\beta_0 X_{1i}^{\beta_1} X_{2i}^{\beta_2} e^{\varepsilon_i}) \]

   Using the properties of logs given on the bottom of page 205, we can rewrite the right-side as:

   \[ \ln Y_i = \ln \beta_0 + \ln X_{1i}^{\beta_1} + \ln X_{2i}^{\beta_2} + \ln e^{\varepsilon_i} \]

   Using the exponent rule of logs also listed on page 205, we can rewrite it as:

   \[ \ln Y_i = \ln \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + \varepsilon_i \ln e \]

   And since \( \ln e = 1 \), we have:
\[ \ln Y_i = \ln \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + \varepsilon_i, \]

which is just the double-log form listed in Chapter 7 (\( \ln \beta_0 \) is still just a constant).

d.) \[ Y_i = \beta_0 + X_{1i}^{\beta_1} + X_{2i}^{\beta_2} + \varepsilon_i \]

This model is not able to be estimated by any regression techniques, as the coefficients in the exponent of addition terms prohibit it from being estimated. Even if one tries to take the log of both sides of the model, it would get them no where in advancing this model as something capable of being estimated.

e.) \[ Y_i = \beta_0 + \beta_1 (1/X_{1i}) + \beta_2 X_{2i} + \varepsilon_i \]

If you look at equation (7.13) from Studenmund, this model is the standard Inverse Form equation given. It is quite capable of being estimated, with the values of the coefficients described in pages 211-212.

2) We are interested in the effect of the presence of various benefit programs on employee views of their boss….

A) Interpret the impact of each factor on job satisfaction and test the null.

Critical t-values:

<table>
<thead>
<tr>
<th></th>
<th>Two tailed</th>
<th>one tailed</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>1.645</td>
<td>1.282</td>
</tr>
<tr>
<td>5%</td>
<td>1.960</td>
<td>1.645</td>
</tr>
<tr>
<td>1%</td>
<td>2.576</td>
<td>2.32</td>
</tr>
</tbody>
</table>

Note that all the hypothesis tests except the test for the coefficient on female have a positive alternative hypothesis and a null of less than or equal to zero. The test for female is two tailed.

*Education*

i.) An additional year of schooling would bring about an expected increase in job satisfaction of .14 points, all else equal.

ii.) Hypothesis test: \( t = .14 / .76 = 0.184 \) \( \Rightarrow \) Can not reject at any confidence level.

*Female*

i.) We would expect females to be .10 points higher on the job satisfaction scale, all else equal.

ii.) Hypothesis test: \( t = |.10 / .04| = 2.5 \) \( \Rightarrow \) Can reject the null at 10% and 5%, but not at the 1% confidence level.
Health
i.) We would expect employees working for an employer who provides health insurance to be 1.1 points higher on the job satisfaction scale, all else equal.
ii.) Hypothesis test: $t = \frac{1.1}{.85} = 1.29 \Rightarrow$ reject in a 10% test, but no other

Pension
i.) We would expect employees working for an employer who provides a pension plan to be .05 points lower on the job satisfaction scale, all else equal.
ii.) The coefficient has the wrong sign to reject the null, its negative, so we do not even have to construct the t-statistic

Party
i.) We would expect employees working for an employer who throws an annual seasonal party to rate 1.5 points higher on the job satisfaction scale, all else equal.
ii.) Hypothesis test: $t = \frac{1.5}{.67} = 2.24 \Rightarrow$ Can reject null at 10% and 5% levels, but not at 1%.

B.) Why can the distinction between significance and magnitude be important?

While a variable can have a significant coefficient, that does not necessarily mean that the variable has any particular impact on the dependent variable due to the fact that the size of the coefficient is extremely small. Just because a variable is deemed significant does not mean it plays any major role in influencing the dependent variable due to the miniscule effect the coefficient. Similarly, a large coefficient that is deemed statistically insignificant at a high degree may, because it is not significant, play no role, as we wouldn’t even be sure that there was an effect in the population equation. In this model, we see that being female raises satisfaction by .10, and is significant. However, compared to the other dummy variables, .10 is a rather small effect.

C.) The interpretation of the intercept.

The intercept means that with all variables equal to zero, it would be the value we would expect of job satisfaction, namely a negative 8.12. A naïve person may be concerned about the value of the intercept because it is negative, and they might argue that a person can’t have a negative value of job satisfaction. Usually this claim would be worthless, as we would not expect to have all zeroes in the explanatory variables (for example, do we really expect to survey someone with zero education?).

D.) Adding in of age variable.

There are three things to consider here. First, the coefficient on age is significant at the 10% confidence level, leading us to believe that age might need to be included. Second, the r-bar-squared value jumps by .02 – this isn’t a significant jump, but it rises nonetheless. Finally, and maybe most importantly, the sign of the pension variable goes from negative to positive, which was the expected theoretical sign of the pension variable all along, suggesting to us that age was mistakenly omitted in the first model since the pension variable may have been picking up some of the age impact.
3. You hear two colleagues discussing the use of indicator variables in regression models…

A.) Explain how occupational indicators should be entered into a regression model and which of their approaches is correct.

If we have four occupational dummy variables, one should enter *three* of those occupational dummies into the model, leaving a fourth dummy as the base category. The first researcher’s method is incorrect, as entering all four occupational variables does not allow for a comparison category; similarly the second researcher’s claim that the use of managerial/professional/sales, service and manual occupations as the *only* way to enter occupational variables is flawed too – one could have used service, manual and clerical variables and left managerial/professional/sales out and used it as the base category to get the same results.

B.) Interpret each coefficient and perform a hypothesis test.

<table>
<thead>
<tr>
<th>Critical t-values:</th>
<th>Two tailed</th>
<th>one tailed</th>
</tr>
</thead>
<tbody>
<tr>
<td>10% -</td>
<td>1.645</td>
<td>1.282</td>
</tr>
<tr>
<td>5% -</td>
<td>1.960</td>
<td>1.645</td>
</tr>
<tr>
<td>1% -</td>
<td>2.576</td>
<td>2.32</td>
</tr>
</tbody>
</table>

Note that these are all one tailed tests. The null for age and tenure is a value less than or equal to zero, the null for parttime is greater than or equal to zero.

**Age**

i.) We expect that adding a year to a person’s age would lead to an increase in their benefit level by $127, all else equal.

ii.) Hypothesis test: $t=127/94.1 = 1.35 \Rightarrow$ We reject the null in a 10% test, but no other.

**Tenure**

i.) An additional year of tenure would be expected to bring about an additional $517 in benefits, all else equal.

ii.) Hypothesis test: $t=517/256 = 2.02 \Rightarrow$ We can reject the null at the 10% and 5% confidence levels, but not at the 1% level.

**Parttime**

i.) A part-time worker would be expected to have $214 more in benefits that a full-time worker, all else equal.

ii.) We expected a negative effect but found a positive effect. We cannot reject the null (that B is greater than or equal to zero) because our coefficient is consistent with that null.
C.) Predict the amount paid in benefits for a person who is 52 years old, has 5 years with their current employer and works part-time.

Benefit = 1024 + 127 * age + 517 * tenure + 214 * parttime

Benefit = 1024 + 127 (52) + 517 () + 214 (1)

Benefit = $38,862

I gave full credit for those of you who treated 5 years as 5 months.

4.) How can you improve on Hay?

As we saw in the final Minitab assignment, it is better to regress your dependent variable (in this case salary) on each of the individual evaluation categories instead of regressing it on a summation of the evaluation points. When you use a summation of job evaluation points, you are imply that each of the categories has an equal impact on salary, a restriction that rarely bears fruit. A simple way to make your pitch would be to check the F statistic in a regression on the individual variables – it is likely that you’ll find that the statistics suggest that they in fact have different effects. Thus, a regression on their summation, thus imposing a restriction of equality, is incorrect and a misspecification.

We also saw that non-linear forms, such as taking the log the dependent variable, or use of a double log model, and allowing for a no-linear relationship between earnings and JEP with a polynominal form had advantages as did adding dummy variables for industry.
5.) For each of the functional forms, explain the relationship between a one unit increase in the explanatory variable and the change in the value of lost property.

A.)
SAFEXP: A one thousand dollar increase (one unit) in expenditures on public safety is expected to lead to a 25,000 dollar decrease in value lost in crimes.
POPULATION: A one person increase (one unit) in population would be expected to lead to a 500 dollar increase in the value lost in crimes.

B.)
SAFEXP: A one thousand dollar increase (one unit) in expenditures on public safety is expected to lead to a 125% decrease in value lost in crimes.
POPULATION: A one person increase (one unit) in population would be expected to lead to a 13% increase in the value lost in crimes.

C.)
SAFEXP: A one percent increase in expenditures on public safety would be expected to decrease the value lost in crimes by 3.63 percent.
POPULATION: A one percent increase in population would be expected to lead to a .05 increase in the value lost in crimes.

D.)
SAFEXP: We would expect a one-unit change safexp to lead to a marginal effect of: 4.05+2*(-.0051)*SAFEXP in the change of percent in value lost in crimes.
POPULATION: We would expect a one unit increase in population to lead to a 10 percent increase in value lost in crimes.