ECO 307 Practice Exam 1

Fall 2015

Problem 1: A state legislator is interested in determining whether additional public funding for elementary education leads to improved student learning outcomes. A data set of most high schools in the state includes variables for the percentage of 4th grade students with a passing grade on a standardized math test (`math4`), the level of public expenditures per student at the school (`exppp`), and a measure for the percentage of students coming from low-income families (`lunch`) (equal to the percentage of students who qualify for financial assistance for school hot lunch).

A researcher ran the following regressions:

```
lmmath <- lm(scale(math4) ~ exppp + lunch, data=data)
summary(lmmath)
```

```
##
## Call:
## lm(formula = scale(math4) ~ exppp + lunch, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6595 -0.4553  0.0414  0.4675  2.7139
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.838e-01  9.085e-02   4.225 2.51e-05 ***
## exppp        1.044e-04  1.754e-05   5.951 3.20e-09 ***
## lunch       -2.360e-02  7.251e-04 -32.539  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7952 on 1820 degrees of freedom
## Multiple R-squared:  0.3684, Adjusted R-squared:  0.3677
## F-statistic: 530.8 on 2 and 1820 DF,  p-value: < 2.2e-16
```

```
confint(lmmath)
```

```
##                     2.5 %        97.5 %
## (Intercept)  2.056482e-01  0.5620246849
## exppp        6.998792e-05  0.0001388029
## lunch       -2.501737e-02 -0.0221730246
```

```r
slmmath <- lm(scale(math4) ~ scale(exppp) + scale(lunch), data=data)
summary(slmmath)
```

```
##
## Call:
## lm(formula = scale(math4) ~ scale(exppp) + scale(lunch), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6595 -0.4553  0.0414  0.4675  2.7139
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -7.933e-17  1.862e-02   0.000        1
## scale(exppp)  1.140e-01  1.916e-02   5.951  3.2e-09 ***
## scale(lunch) -6.233e-01  1.916e-02 -32.539  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7952 on 1820 degrees of freedom
## Multiple R-squared:  0.3684, Adjusted R-squared:  0.3677
## F-statistic: 530.8 on 2 and 1820 DF,  p-value: < 2.2e-16
```

```r
confint(slmmath)
```

```
##                    2.5 %      97.5 %
## (Intercept)  -0.03652585  0.03652585
## scale(exppp)  0.07641908  0.15155739
## scale(lunch) -0.66087682 -0.58573851
```

A. (7 points) Is there statistical evidence that more public expenditures per student lead to an improvement in student success, as measured by 4th grade standardized test math scores? Test the appropriate hypothesis.

B. (7 points) Is there statistical evidence that having more low-income students leads to lower success in math standardized test scores?

C. (7 points) How much does a one standard deviation increase in public expenditures per student influence math test scores? Give a point estimate and report and interpret a 95% confidence interval.

D. (7 points) Estimate the percentage of 4th grade children that pass the standardized math test at a school that has public expenditures per student equal to $4,000 and has 25% of its students from low-income familites.

E. (7 points) Consider the R output below. Another state legislator claims that public expenditures does **not** have any impact on student success, and he has statistical evidence to back it up this counter claim (see below). Which answer is more likely to be correct? Consider the correlation estimated below between public expenditures per student and the percentage of children from low income famililies.

```
lmmath <- lm(scale(math4) ~ exppp , data=data)
summary(lmmath)
```

```
##
## Call:
## lm(formula = scale(math4) ~ exppp, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6273 -0.5193  0.2239  0.7573  1.4942
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.484e-01  1.139e-01   1.303    0.193
## exppp       -2.857e-05  2.145e-05  -1.332    0.183
##
## Residual standard error: 0.9998 on 1821 degrees of freedom
## Multiple R-squared:  0.0009729,  Adjusted R-squared:  0.0004243
## F-statistic: 1.773 on 1 and 1821 DF,  p-value: 0.1831
```

```
cor.test(x=data$exppp, y=data$lunch)
```

```
##
##  Pearson's product-moment correlation
##
## data:  data$exppp and data$lunch
## t = 10.22, df = 1821, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1890287 0.2758773
## sample estimates:
##       cor
## 0.2329173
```

Problem 2: Consider a data set on employees that includes the following variables: 1. Annual income in dollars 2. Annual dollar amount contributed to retirement savings 3. (Annual savings contributions) / (Annual income) 3. Dummy variable for whether or not the employee is married

For each of the following research questions, state the appropriate univariate or bivariate hypothesis test and state the null and alternative hypotheses.

A. (7 points) Do married employees contribute a different amount of income to savings each year than non-married employees?

B. (7 points) Do contributions to savings increase as income increases?

C. (7 points) Do employees on average contribute at least 10% of their income to retirement savings?

Problem 3: Consider a data set on employees that includes the following variables:

1. `inc`: Annual income in thousands of dollars
2. `nettfa`: Total financial assets (i.e. total lifetime savings)
3. `marr`: Dummy variable for whether or not the person is married (`marr=1` if married)
4. `male`: Dummy variable for sex (`male=1` if male)
5. `age`: Age of employee in years

Consider the following regression:

```
lmfa <- lm(log(nettfa) ~ log(inc) + male + marr + age
           + male*marr + male*log(inc) + marr*log(inc),
           data=data)
summary(lmfa)
```

```
##
## Call:
## lm(formula = log(nettfa) ~ log(inc) + male + marr + age + male *
##     marr + male * log(inc) + marr * log(inc), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.2112 -0.9029  0.2182  1.1173  5.3560
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -7.546936   0.302099 -24.982  < 2e-16 ***
## log(inc)       2.169877   0.088527  24.511  < 2e-16 ***
## male           1.093030   0.373697   2.925  0.00346 **
## marr           0.686258   0.329949   2.080  0.03758 *
## age            0.048167   0.002045  23.559  < 2e-16 ***
## male:marr      0.127439   0.133087   0.958  0.33832
## log(inc):male -0.325226   0.110374  -2.947  0.00323 **
## log(inc):marr -0.300304   0.096788  -3.103  0.00193 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.631 on 6021 degrees of freedom
## Multiple R-squared:  0.3391, Adjusted R-squared:  0.3384
## F-statistic: 441.4 on 7 and 6021 DF,  p-value: < 2.2e-16
```

A. (6 points) How much does financial assets increase when income for married men increases by 1%?

B. (6 points) How much does financial assets increase when income for married women increases by 1%?

C. (6 points) How much does financial assets increase when income for non-married men increases by 1%?

D. (6 points) How much does financial assets increase when income for non-married women increases by 1%?

E. (7 points) Is there statistical evidence that married versus non-married people's financial assets grow by different amount as their income rises?

F. (6 points) What percentage of the variability in $ln(financial\ assets)$ is explained by the variables in the regression model?

G. (7 points) Consider removing the dummy variable for whether the employee married, which also involves removing all the interactions terms involving this variable. The code below estimates this new regression model and compares the models. Test the hypothesis that the married variable and its interaction terms contributed to explaining net financial assets.

```
lmfa_nomarried <- lm(log(nettfa) ~ log(inc) + male + age + male*log(inc), data=data)
summary(lmfa_nomarried)
```

```
##
## Call:
## lm(formula = log(nettfa) ~ log(inc) + male + age + male * log(inc),
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1993 -0.9272  0.2273  1.1250  5.2413
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.634062   0.176343 -37.620   <2e-16 ***
## log(inc)       1.827181   0.041747  43.768   <2e-16 ***
## male           0.717094   0.332909   2.154   0.0313 *
## age            0.048170   0.002051  23.483   <2e-16 ***
## log(inc):male -0.161837   0.092429  -1.751   0.0800 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.637 on 6024 degrees of freedom
## Multiple R-squared:  0.3342, Adjusted R-squared:  0.3337
## F-statistic: 755.8 on 4 and 6024 DF,  p-value: < 2.2e-16
```

```
anova(lmfa, lmfa_nomarried)
```

```
## Analysis of Variance Table
##
## Model 1: log(nettfa) ~ log(inc) + male + marr + age + male * marr + male *
##     log(inc) + marr * log(inc)
## Model 2: log(nettfa) ~ log(inc) + male + age + male * log(inc)
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1   6021 16018
## 2   6024 16138 -3   -120.65 15.118 8.444e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```