Statistics and Econometrics: Problem Set 2

Siow Meng Low

23 October 2016

Question 1

Question 1a: Stock Market Performance on CEO Salary

The equation is given as:

$$log(salary) = \beta_0 + \beta_1 log(sales) + \beta_2 roe + \beta_3 ros + u$$

The null hypothesis states that *ros* has no effect on CEO salary after controlling for *sales* and *roe*. This can be expressed as: H_0: $\beta_3 = 0$

The alternative hypothesis states that higher *ros* increases CEO salary (i.e. better stock market performance increases a CEO's salary). This can be expressed as: H_1: $\beta_3 > 0$

Question 1b: Linear Regression Results and Effects of *ros*

Table 1 displays the model information of regressing CEO salary (in natural log form) on the firm sales (in natural log form), return on equity, and return on firm's stock.

Table 1: Regression Model for Question 1b

	Log(CEO Salaries)
lsales	0.2803***
	(0.0353)
roe	0.0174***
	(0.0041)
ros	0.0002
	(0.0005)
Constant	4.3117***
	(0.3154)
N	209
R^2	0.2827
Adjusted R^2	0.2722
Residual Std. Error	0.4832 (df = 205)
F Statistic	26.9293*** (df = 3; 205)

Notes:	***Significant at the 1 percent level.
	**Significant at the 5 percent level.
	*Significant at the 10 percent level.

The question specifies that *ros* is in percentage form, hence an increase of 50 basis points refer to 0.50 increase in *ros* value. If *ros* increases by 50 basis points, CEO salary (*salary*) is predicted to increase by 0.01%. This effect on CEO salary is practically very small.

Question 1c: Hypothesis Testing of *ros* Parameter

The null hypothesis and alternative hypothesis are:

- H_0: $\beta_3 = 0$
- H_1: $\beta_3 > 0$

From the above, it is clear that we would do an one-tail test. The critical value at 10% significance level is thus 1.2816 (assume normal distribution since we have large degree of freedom, > 120).

The t-statistic is $\dfrac{\hat{\beta_3}}{se(\hat{\beta_3})} = 0.446$. Since this value is way smaller than the critical value 1.2816, we fail to reject the null hypothesis ($\beta_3 = 0$). We conclude that *ros* has no effect on CEO salary.

Question 1d: Inclusion of *ros* in Final Model

Since *ros* is both economically insignificant and statistically insignificant, I would not include it in the final model. Leaving *ros* out allows us to include other important independent variables that my be highly correlated with *ros*.

Question 2

Question 2a: Effects of Marijuana on Wage

The equation is in the following form:

$$log(wage) = \beta_0 + \delta_0 female + \beta_1 marijuana + \beta_2 educ + \beta_3 exper + u$$

The variables are described below:

- *wage*: Respondent's wage
- *female*: Respondent's gender, $= 1$ if the respondent is female
- *marijuana*: Number of times respondent smoked marijuana last month
- *educ*: Years of education
- *exper*: Years of experience
- *u*: Error term

From the above equation, smoking marijuana one more time per month is estimated to change wage by approximately $(100 \cdot \beta_1)\%$, holding other independent variables fixed. Smoking marijuana five more times per month is estimated to change wage by approximately $(100 \cdot 5\beta_1)\%$, holding other independent variables fixed.

Question 2b: Different Effects of Drug Usage for Men and Women

The equation for this model is to be expressed in this form:

$$log(wage) = \beta_0 + \delta_0 female + \beta_1 marijuana + \delta_1 female \cdot marijuana + \beta_2 educ + \beta_3 exper + u$$

female·marijuana is the interaction term between *female* and *marijuana*. To test that there are no differences in the effects of drug usage for men and women, we would need to test the null hypothesis $H_0 : \delta_1 = 0$

Question 2c: Drug Usage Categories

In this question, there is no mention of interactions between marijuana usage and gender, hence we leave out the interactive effects in this section. The equation for this model is to be expressed in this form (nonuser is used as base group):

$$log(wage) = \beta_0 + \delta_0 light + \delta_1 moderate + \delta_2 heavy + \delta_3 female + \beta_1 educ + \beta_2 exper + u$$

The three new independent variables are described below:

- *light*: $= 1$ if respondent is a light marijuana user (1 to 5 times per month)
- *moderate*: $= 1$ if respondent is a moderate marijuana user (6 to 10 times per month)
- *heavy*: $= 1$ if respondent is a heavy marijuana user (more than 10 times per month)

Question 2d: Null Hypothesis - Marijuana Usage Has No Effect on Wage

To test the null hypothesis that marijuana usage has no effect on wage, we would need to test if ($H_0 : \delta_0 = 0, \delta_1 = 0, \delta_2 = 0$) in the equation written in question 2c.

In other words, we would need to test if these three variables have a joint effect on wage. To perform this test, we would need two models:

- Unrestricted Model: Same model as in Question 2c
- Restricted Model: Exclude these three variables (related to marijuana usage) in the unrestricted model

Using the R^2 value from the two models, we shall calculate the F-statistic $F = \dfrac{(R_{ur}^2 - R_r^2)/q}{(1 - R_{ur}^2)/(n - k - 1)}$ and compare it with the $F_{q,n-k-1}$ critical value according to the desired signficance level. In this case, we know that $q = 3$ and $k = 6$.

If the F-statistic, F, is smaller than the critical value, we fail to reject the null hypothesis and we say that the three variables are jointly insignificant, which means marijuana usage has no effect on wage and it is justified to drop the three variables from the model.

Question 2e: Potential Problems with Causal Inference

To draw causal inference using any independent variable in the model, it must not be correlated with the error term, u (in other words, Zero-Conditional-Mean Assumption must hold). For instance, to capture a causal relationship between wages and drug usage, *marijuana* and u must be uncorrelated.

However, there may be some other uncaptured factors in u (e.g. family income, drug habits of family members, parents' educational level), which are correlated with individual's drug habit and can affect wages. To draw causal inference using *marijuana*, we will need to capture these data in order to control for these factors.

Question 3

Question 3a: Linear Regression Using Experience and Position

Table 2 displays the model information of regressing points per game (scored by player) on the player's experience (both *exper* and the quadratic form, *expersq*) and player's position.

Table 2: Regression Model for Question 3a

	Points per Game
exper	1.2807***
	(0.3285)
expersq	−0.0718***
	(0.0241)
guard	2.3147**
	(1.0004)
forward	1.5446
	(1.0023)
Constant	4.7608***
	(1.1786)
N	269
R^2	0.0910
Adjusted R^2	0.0772
Residual Std. Error	5.6683 (df = 264)
F Statistic	6.6055*** (df = 4; 264)

Notes:	***Significant at the 1 percent level.
	**Significant at the 5 percent level.
	*Significant at the 10 percent level.

Question 3b: Guard vs Center

Holding experience fixed, a guard is predicted to score 2.31 more points than a center per game.

To test whether this difference is statistically significant, we can test if $H_0 : \beta_{guard} = 0$. The corresponding p-Value is 0.0214, which is smaller than 0.05 (assuming we use 5% significance level). Therefore, we reject the null hypothesis H_0 and conclude that this difference is statistically significant.

Question 3c: Effect of Marriage

The right column of table 3 displays the model information of regressing points per game (scored by player) on the player's experience (both *exper* and the quadratic form, *expersq*), player's position, and player's marital status.

Table 3: Regression Model for Question 3c

| | Points per Game | |
| | Question 3a Model | Question 3c Model |
	(1)	(2)
exper	1.2807***	1.2333***
	(0.3285)	(0.3342)
expersq	−0.0718***	−0.0704***
	(0.0241)	(0.0242)
guard	2.3147**	2.2863**
	(1.0004)	(1.0017)
forward	1.5446	1.5409
	(1.0023)	(1.0030)
marr		0.5843
		(0.7404)
Constant	4.7608***	4.7029***
	(1.1786)	(1.1817)
N	269	269
R^2	0.0910	0.0931
Adjusted R^2	0.0772	0.0759
Residual Std. Error	5.6683 (df = 264)	5.6724 (df = 263)
F Statistic	6.6055*** (df = 4; 264)	5.4014*** (df = 5; 263)

Notes:	***Significant at the 1 percent level.
	**Significant at the 5 percent level.
	*Significant at the 10 percent level.

Holding position and experience fixed, married players are predicted to score 0.58 more points per game than unmarried players. However, the p-Value for $H_0 : \beta_{marr} = 0$ is equal to 0.4308. Since this value is way larger than 0.05 (assuming 5% significance level), we cannot reject the null hypothesis H_0, which states that the effect of *marr* is statistically insignificant. Therefore, we cannot conclude that the productivity (based on points per game) of married players is different from unmarried players.

Question 3d: Interactions of Marriage with Experience Variables

The right column of table 4 displays the model information of regressing points per game (scored by player) on the player's experience (both *exper* and the quadratic form, *expersq*), player's position, player's marital status, and the interaction between marital status and experience.

Table 4: Regression Model for Question 3d

| | Points per Game | |
	Question 3c Model	Question 3d Model
	(1)	(2)
exper	1.2333***	0.7026
	(0.3342)	(0.4340)
expersq	−0.0704***	−0.0295
	(0.0242)	(0.0327)
guard	2.2863**	2.2508**
	(1.0017)	(1.0000)
forward	1.5409	1.6291
	(1.0030)	(1.0020)
marr	0.5843	−2.5375
	(0.7404)	(2.0382)
exper:marr		1.2797*
		(0.6823)
expersq:marr		−0.0936*
		(0.0489)
Constant	4.7029***	5.8162***
	(1.1817)	(1.3488)
N	269	269
R^2	0.0931	0.1058
Adjusted R^2	0.0759	0.0818
Residual Std. Error	5.6724 (df = 263)	5.6541 (df = 261)
F Statistic	5.4014*** (df = 5; 263)	4.4126*** (df = 7; 261)

Notes:
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

From the above report, we can observe that *marr* is statistically insignificant while the two interaction terms are only significant at 10 percent level. If we assume 5% significance level, none of them is individually significant.

Next, we would need to test if these three variables are jointly significant. To test this, we perform F-test to test the difference between unrestricted model and restricted (without *marr*, *exper:marr*, and *expersq:marr*) model.

The p-Value for H_0 (which states that the three variables are jointly insignificant) is 0.2303. Assuming 5% significance level, this value is way higher than 0.05. Consequently, the null hypothesis H_0 cannot be rejected and we conclude that these three variables are jointly insignificant. In summary, there is no strong evidence that marital status affects points per game in this new model.

Question 3e: Adding Assists Per Game

The right column of table 5 displays the model information of regressing assists per game (by player) on the player's experience (both *exper* and the quadratic form, *expersq*), player's position, and player's marital status.

Table 5: Regression Model for Question 3e

	Points per Game Question 3c Model	Assists per Game Question 3e Model
	(1)	(2)
exper	1.2333***	0.4436***
	(0.3342)	(0.1004)
expersq	−0.0704***	−0.0267***
	(0.0242)	(0.0073)
guard	2.2863**	2.4917***
	(1.0017)	(0.3008)
forward	1.5409	0.4475
	(1.0030)	(0.3012)
marr	0.5843	0.3219
	(0.7404)	(0.2224)
Constant	4.7029***	−0.2258
	(1.1817)	(0.3549)
N	269	269
R^2	0.0931	0.3499
Adjusted R^2	0.0759	0.3375
Residual Std. Error (df = 263)	5.6724	1.7035
F Statistic (df = 5; 263)	5.4014***	28.3076***

Notes: ***Significant at the 1 percent level.
                        **Significant at the 5 percent level.
                        *Significant at the 10 percent level.

Similar to the results in Question 3c, both *forward* and *marr* are not statistically significant. However, the *guard* variable is more significant in this new model (from 5 percent level in Question 3c to 1 percent level in Question 3e). This is expected since a guard often leads the team in assists performance.

From this model, married players are predicted to have 0.32 more assists per game than unmarried players. However, the p-Value for for $H_0 : \beta_{marr} = 0$ is equal to 0.1489. This value (although smaller than the p-Value in 3c) is still larger than 0.05, assuming 5% significance level. As a result, we cannot reject the null hypothesis H_0 where the effect of *marr* (on assists per game) is statistically insignificant. Therefore, we cannot conclude that married players have more assists per game than unmarried players.