PROBLEM SET 4
Due back by noon of Tuesday April 11.

Name: VALLICELLI LORENZO

Id: 197661

Instructions:

- Make sure you are working on your problem set as each problem set is different.
- **The answers to the questions of this problem set are to be given exclusively in the answer sheet**
- **The answers sheet MUST be printed and not photocopied. Photocopies will not be accepted.**
- Questions marked with the symbol ♣ admit more than one correct answer
- Please fill the boxes in the answer sheet completely using a **black pen** as follows

Question 1: ■ B C D E

- The answer sheet must not be creased or folded otherwise your problem set won't be graded.
- You can hand back your problem set after the class on Tuesday.

Question 1 What does it mean for an estimator to be consistent?

A The estimator is statistically significant.

B The estimator converges in probability to the parameter that it is estimating.

C The estimator is normally distributed.

D The expected value of the estimator is equal to the parameter that it is estimating.

Question 2 What does it mean for an estimator to be unbiased?

A The estimator converges in probability to the parameter to be estimated.

B The estimator is equal to the parameter to be estimated.

C As the sample size tends to infinity, the estimator gets closer and closer to the true value of the parameter.

D The expected value of the estimator is equal to the parameter to be estimated.

Question 3 We regress the final grade on the Applied Econometrics and Statistics final exam (from 0 to 30) on the number of hours the students spent studying. Our model is

$$finalgrade_i = \beta_0 + \beta_1 hours\_study_i + u_i$$

We find that $\hat{\beta}_1 = 0.2$ and $SE(\hat{\beta}_1) = 0.01$. Suppose we measure the $hours\_study_i$ in minutes instead of hours and re-run the regression. What is the standard error of the new estimator of β_1?

A We cannot provide an answer with the info provided.

B 0.00017.

C 0.6.

D 0.01.

Question 4 ♣ We want to study the effect of being married and being female on the hourly wage. We have four dummies: `marfem` if the person is married and female, `marmal` if the person is married and male, `singfem` if the person is single and female, and finally `singmal` if the person is single and male. In which of the following model the coefficient β_3 can be interpreted as the difference in the average hourly wage between a single woman and a single man?

A $y_i = \beta_0 + \beta_1 marfem_i + \beta_2 marmal_i + \beta_3 singfem_i + u_i$

B $y_i = \beta_0 + \beta_1 marfem_i + \beta_2 singmal_i + \beta_3 singfem_i + u_i$

C $y_i = \beta_0 marfem_i + \beta_1 marmal_i + \beta_3 singfem_i + \beta_4 singmal_i + u_i$

D $y_i = \beta_0 + \beta_1 marfem_i + \beta_2 marmal_i + \beta_3 singmal_i + u_i$

E *None of these answers are correct.*

Question 5 In a linear model $y_i = \beta_0 + \beta_1 x_i + u_i$, the error are conditional homoskedastic if

A $Var(y_i|x_i)$ is constant.

B $E(y_i|x_i)$ is constant.

C $Var(u_i)$ is constant.

D $E(u_i|x_i)$ is constant.

Question 1

You are asked to investigate what are the determinants of car fatalities in the US. You have traffic fatalities data on the "lower 48" US states (i.e. all states with the exception Alaska and Hawaii), for 1988. In particular, you have the following variables:

- `fatal` vehicle fatalities per ten thousand people
- `unemp` unemployment rate (measured from 0 to 100)
- `income` per capita personal income in 1987 (in thousands of dollars)
- `beertax` tax on case of beer in 1987 dollar
- `miles` average miles per driver
- `jail` =1 if state has mandatory jail for drunk driving
- `service` = 1 if state has mandatory community service
- `youngdrivers` percent of drivers aged 15-24 (measured from 0 to 1).

The descriptive statics of the sample are provided in Table 25.

Table 249: Vehicle fatalities: descriptive statistics.

Statistic	N	Mean	St. Dev.	Min	Max
fatalities	48	2.070	0.521	1.231	3.236
unemp	48	5.456	1.838	2.400	10.900
income	48	14.894	2.628	10.699	22.193
beertax	48	0.480	0.435	0.043	2.194
miles	48	8,615.830	1,115.104	5,789.922	11,812.110
youngdrivers	48	0.162	0.022	0.073	0.221

With this goal in mind, you estimate a linear model that relates fatalities and the other variable in the dataset. You obtain:

$$
\begin{aligned}
fatalities &= \underset{(0.9333)}{0.5708} + \underset{(0.0363)}{0.049}\, unemp - \underset{(0.0248)}{0.0573}\, income + \underset{(0.1317)}{0.1442}\, beertax \\
&\quad + \underset{(0.0001)}{0.0002}\, miles + \underset{(0.1622)}{0.1279}\, jail + \underset{(0.1856)}{0.0973}\, service + \underset{(2.6413)}{0.3288}\, youngdrivers
\end{aligned}
$$

Question 6 What is the interpretation of the coefficient on `youngdrivers`?

- [A] A one percentage point increase in the fraction of `youngdrivers` is associated with about 0.33 more vehicle fatalities per million of people, holding everything else equal.
- [B] A one percentage point increase in the fraction of `youngdrivers` is associated with about 0.33 more vehicle fatalities per ten thousand of people, holding everything else equal.
- [C] A one percentage point increase in the fraction of `youngdrivers` is associated with about 33 more vehicle fatalities per million of people, holding everything else equal.
- [D] A one percentage point increase in the fraction of `youngdrivers` is associated with about 33 more vehicle fatalities per hundred of people, holding everything else equal.

Question 7 What is the interpretation of the coefficient on `income`?

- [A] Holding everything else equal, States with higher per-capita personal income have less fatalities. In particular, an extra $2,000 is associated with a reduction of about 1.2 fatalities per hundred of people
- [B] Holding everything else equal, States with higher per-capita personal income have less fatalities. In particular, an extra $1,000 is associated with a reduction of about 6 fatalities per million of people.
- [C] Holding everything else equal, States with higher per-capita personal income have more fatalities. In particular, an extra $1,000 is associated with an increase reduction of about 0.6 fatalities per thousand of people.
- [D] Holding everything else equal, States with higher per-capita personal income have less fatalities. In particular, an extra $1,000 is associated with a reduction of about 0.06 fatalities per hundred of people.
- [E] Holding everything else equal, States with higher per-capita personal income have more fatalities. In particular, an extra $1,000 is associated with an increase reduction of about 6 fatalities per hundred of people.

Question 8 ♣ What is the interpretation of the coefficient on `beertax`?

- [A] Holding other variables constant, taxing a case of beer one extra dollar cent more is associated with an increase in fatalities that can be estimated to about 0.14 more death per million of people.
- [B] Holding other variables constant, taxing a case of beer 50 cent more is associated with an increase in fatalities that can be estimated to about 7 more death per hundred of people.
- [C] Holding other variables constant, taxing a case of beer 50 cent more is associated with a decrease in fatalities that can be estimated to about 7 more death per hundred of people.
- [D] Holding other variables constant, taxing a case of beer one extra dollar cent more is associated with an increase in fatalities that can be estimated to about 0.14 more death per thousand of people.
- [E] Holding other variables constant, taxing a case of beer 50 cent more is associated with an increase in fatalities that can be estimated to about 7 more death per million of people.
- [F] *None of these answers are correct.*

Question 9 ♣ The R^2 of this regression is 0.64. We run a second regression including one more variable, state investment in road maintenance per mile. The R^2 is now 0.69. What can we conclude?

- [A] We should prefer the new regression as it is more meaningful.
- [B] The variables in the second model capture about 69% of the variance in fatalities.
- [C] The variables in the first model capture about 64% of the variance in fatalities
- [D] We cannot conclude much, since the R^2 always increases when a new variable is added to a regression.
- [E] Since it increases the R^2, the investment variables is statistically significant.
- [F] *None of these answers are correct.*

Question 10 Is the coefficient on `beertax` statistically different from zero at the $\alpha = 0.1$ significance level?

- [A] Yes, because the t-statistic is not larger than the associated critical value of 1.96.
- [B] No, because the t-statistic is not larger than the associated critical value of 1.96.
- [C] No, because the t-statistic is larger than the associated critical value of 1.96.
- [D] Yes, because the t-statistic is larger than the associated critical value of 1.

Question 11 The 90% confidence interval for the effect of an increase in income of $\$3,000$ if given by

- [A] $(-0.217, -0.127)$
- [B] $(-0.1926, 0.078)$
- [C] $(-3.102, -3.012)$
- [D] $(-0.294, -0.049)$

Question 12 The p-value for a two-sided test on the coefficient on `income` is 0.0209. What can we conclude?

- [A] There is a 96.3% probability that the coefficient is statistically significant.
- [B] The coefficient is statistically significant at the $\alpha = 0.01$ level, but not at $\alpha \geq 0.05$ levels.
- [C] The coefficient is statistically significant at both the $\alpha = 0.05$ and $\alpha = 0.10$ level, but not at the $\alpha = 0.01$ level.
- [D] In order to conclude anything, I need to now the whether the errors of the regression are homoskedastic or not.

Question 2

Earnings functions are one of the most investigated relationships in economics. These typically relate the earnings to a series of explanatory variables such as education, work experience, gender, race, etc.

Consider using data on a sample of Italian employees aged between 15 and 64. The sample is from the Labor Force Survey from ISTAT which is the most comprehensive survey on labor related issues. The sample we use here is a small subsample (3,000 observations) of the one provided by ISTAT and used in class.

We estimate first a regression model linking `hourlywage` (monthly net wage divided the hour worked in that month) with years of education (`educ`) and a dummy for gender (`male`, which is = 1 if the individual is a male and zero otherwise). The model we consider is the following

$$\log hourlywage = \beta_0 + \beta_1 educ + \beta_2 male + u.$$

The regression results are shown below:

Table 250: Labor force survey: OLS regression

	Dependent variable:
educ	0.035
	(0.006)
male	0.052
	(0.027)
Constant	3.072
	(0.080)

Note: Note: Heteroskedastic robust standard errors in parenthesis.

Question 13 Given the estimation results, what is the interpretation you can give of β_1, the coefficient on `educ`?

- [A] Controlling for gender, an extra year of education is associated with 3.5 extra euro net per month
- [B] Controlling for gender, an extra 3 years of education are associated with a decrease in hourly net wage of about 10.5%
- [C] Controlling for gender, an extra 3 years of education are associated with an increase in hourly net wage of about 10.5%
- [D] Controlling for gender, β_2 is very close to zero and so not important

Question 14 What would happen to β_0 and β_1 if we were to substitute the `male` dummy with a `female` (which would equal to 1 if the individual is a female and 0 otherwise)?

- [A] The coefficients will not change.
- [B] The new coefficients would become $\beta_0 = 3.124$ and $\beta_1 = -0.052$.
- [C] I don't know: I would need to re-run the regression to give a sensible answer.
- [D] The new intercept will remain unchanged, while the new β_1 will be equal to -0.052.

Question 15 Given the estimation results, what is the interpretation you can give of β_2, the coefficient on `male`?

- [A] Controlling for education, females work more hours and ern more money.
- [B] Controlling for education, females tend to earn on average 5.2% higher hourly wage than males.
- [C] Controlling for education, females work more hours and earn more money.
- [D] Controlling for education, males tend to earn on average 5.2% higher hourly wage than female.

Question 16 Which of the following statement is false?

- [A] β_1 is statistically significant at 5% significance level.
- [B] β_1 is statistically significant at 10% significance level.
- [C] β_1 is not significant at 5% significance level.
- [D] β_1 is statistically significant at 0.1% significance level.

Question 3

Consider the following table taken from Stock and Watson's textbook.

	Dependent variable:						
	testscore						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
str	−0.998	−0.734	−0.968	−0.531	64.300	83.700	65.300
	(0.239)	(0.231)	(0.540)	(0.300)	(25.500)	(29.700)	(25.500)
english	−0.122	−0.176					−0.166
	(0.032)	(0.032)					(0.032)
str^2					−3.420	−4.380	−3.470
					(1.290)	(1.510)	(1.290)
str^3					0.059	0.075	0.060
					(0.022)	(0.025)	(0.022)
lunch	−0.547	−0.398		−0.411	−0.420	−0.418	−0.402
	(0.022)	(0.030)		(0.029)	(0.028)	(0.029)	(0.030)
$\log(income)$		11.600		12.100	11.700	11.800	11.500
		(1.740)		(1.760)	(1.730)	(1.750)	(1.730)
$english \geq 20$			5.640	5.500	−5.470	816.000	
			(16.700)	(9.140)	(1.030)	(435.000)	
$str \times (english \geq 20)$			−1.280	−0.578		−123.000	
			(0.844)	(0.465)		(66.400)	
$str^2 \times (english \geq 20)$						6.120	
						(3.350)	
$str^3 \times (english \geq 20)$						−0.101	
						(0.056)	
Constant	700.000	659.000	682.000	654.000	252.000	122.000	245.000
	(4.690)	(7.680)	(10.500)	(8.890)	(166.000)	(192.000)	(166.000)
Observations	420	420	420	420	420	420	420
R^2	0.775	0.796	0.310	0.797	0.801	0.803	0.801
Adjusted R^2	0.773	0.794	0.305	0.795	0.798	0.799	0.798

Question 17 Consider the specification of column (6). What is the interpretation of the coefficient on $\log(income)$?

- [A] An increase in income of 1% is associated with an increase in testscore of 1.2%
- [B] An increase in income of 1% is associated with an increase in testscore of 12 testcore points
- [C] An increase in income of 1% is associated with an increase in testscore of 12%
- [D] An increase in income of 1% is associated with an increase in testscore of 0.12%

Question 18 What is the fraction of variance explained by the model in column (2)?

- [A] 3.1%
- [B] 31%
- [C] 78%
- [D] 80%

Question 19 What is the fraction of variance explained by the model in column (4)?

- [A] 78%
- [B] 3.1%
- [C] 31%
- [D] 80%

Question 20 Consider the regression in column (3). An increase in *str* from 20 to 21 is associated with

- [A] a decrease in *testscore* of about 1
- [B] a decrease in *testscore* of about 1.3
- [C] a decrease in *testscore* of about 2.2
- [D] an increase in *testscore* of about 1.3

Question 21 Consider the regression in column (3). An increase in *str* from 20 to 22 is associated with

- [A] a decrease in *testscore* of about 2
- [B] a decrease in *testscore* of about 2.6
- [C] a decrease in *testscore* of about 4.4
- [D] an increase in *testscore* of about 2.6

+125/11/6+

Name: **VALLICELLI LORENZO** Id: **197661**

- Answers must be given exclusively on this sheet: answers given on the other sheets will be ignored.
- **This sheet MUST be printed out and not photocopied. Photocopies will not be accepted.**
- Please fill the boxes below completely using a **black pen**.
- Do not crease or fold.
- You can hand back your problem set by putting it into my mailbox on the fifth floor of the viale Romania campus by noon of Friday, March 25 at noon.

Question 1: A B C D
Question 2: A B C D
Question 3: A B C D
Question 4: A B C D E
Question 5: A B C D
Question 6: A B C D
Question 7: A B C D E
Question 8: A B C D E F
Question 9: A B C D E F
Question 10: A B C D
Question 11: A B C D
Question 12: A B C D
Question 13: A B C D
Question 14: A B C D
Question 15: A B C D
Question 16: A B C D
Question 17: A B C D
Question 18: A B C D
Question 19: A B C D
Question 20: A B C D
Question 21: A B C D