# Erasmus University Rotterdam

## Erasmus School of Law

## Rotterdam Institute for Law and Economics

## EMLE 2020/2021

# Empirical Legal Studies

# Exam

**Prof. Jonathan Klick & Prof. Pieter Desmet**

**March 22, 2021, 9h30 - 11h30**

**PLEASE READ THIS FIRST**

1. The maximum score for the entire exam is **100** points (this will be later converted to a scale of 0-10 points).
2. The exam consists of **2 Parts**.
3. All 6 questions of Part I need to be answered. For Part II, 4 out of 5 questions need to be answered. Indicate clearly which questions of Part II you are answering.
4. Please hand in your answers as well as **all** the exam pages to the supervisor upon leaving.
5. This is an **open book** exam.
6. In case you should have any serious problems with the interpretation of one or more questions, please consult with the supervisor.
7. Please write down your EMLE student ID number on each answer sheet.

**Please check all of the above again before handing in your exam!**
**Good luck!**

PART I

1. In what way are type 1 and type 2 errors related?

**At the most basic level, for given/fixed null and alternate hypotheses, lowering your type I error increases the likelihood you make a type II error and vice versa.**

2. Provide a legal example where a one-sided hypothesis test would make more sense than a two-sided hypothesis test.

**Any legal example where symmetry does not apply will work. For example, in a discrimination case, it is not legally relevant if an employer treated the protected class better than average in terms of pay, hiring, etc. Similarly, in a tort/products liability case, it does not matter legally if the product lowered a risk of harm.**

3. Why do we usually use a normal distribution in hypothesis testing?

**If we have a random sample and the sample size is large enough (and variance is finite), a central limit theorem will apply. If a CLT applies, we know that the sample means will be distributed normally, regardless of the distribution in the population. If a CLT does not apply, using a normal distribution in hypothesis testing is only justified if we're willing to assume that the test statistic is distributed normally in the underlying population.**

4. Comment on the following statement: If you have a non-random sample, in terms of statistical validity, it is better to have a large sample.

**A large sample does not "undo" the bias that arises from a lack of randomness in the sampling procedure. That said, while the sample might be non-random for a given population, it might be random for some underling sub-population, in which case, a larger sample will improve one's confidence in estimates having to do with that specific sub-population.**

5. Intuitively, how is the correlation coefficient between y and x related to the b coefficient in a regression of y = a + b*x .

**For starters, the correlation coefficient and the regression coefficient (in the single x case) must be of the same sign. Formally, the regression coefficient in this case is equal to the correlation coefficient scaled by the ratio of the standard deviations of y and x, so there is nothing one can say about the relationship between the magnitudes of the correlation coefficient and the regression coefficient (without knowing the standard deviations of y and x).**

6. Choose one of the questions 1-5 to count double.

Part II: Choose 4 questions to answer.

7. If you have a random sample of 500 men and 500 women, and you wish to examine the relationship between education and income, discuss the conditions under which it is preferable

to analyze this relationship using a single regression on all 1,000 data points vs. running separate regressions on men and women.

**The intercepts and the slope coefficients might differ between men and women (i.e., one sex may make more on average independently of education AND the effect of education on income may differ by sex). If neither differ by sex, it would be preferable to run a single regression since the increased sample size improves efficiency (i.e., lower standard errors). If either differs, running a single regression that does not account for these differences will give biased estimates, whereas two separate regressions would not generate this problem (assuming there are no other determinants of income that are correlated with education). One could run a single regression with two separate intercepts (a general intercept and a sex effect; or two sex effects and no general intercept) and two slope coefficients (a general education effect and an additional education interacted with sex effect) which would generate comparable results to running two separate regressions. If the slope is the same for both sexes, but the intercept is different, one could run a single regression with two intercepts. If the intercepts are the same, but the slope differs by sex, one could run a single regression with a single intercept and an education effect and an interaction between education and sex.**

8. If you have a random sample of 1,000 people and you have individual data points from each of 20 years for each of those people, if you run a fixed effects regression (i.e., include a separate dummy variable for each person), intuitively explain why it would not be possible to estimate a general regression coefficient for the effect of sex/gender in that fixed effects regression.

**The person fixed effect jointly accounts for all characteristics that do not change for a person across the sample periods (such as sex). To estimate a separate sex effect, it would be necessary for at least some of the individuals to change their sex during the course of the sample. Formally, the sex variable (and all other unchanging variables) would be collinear with the individual fixed effect.**

9. Explain the difference between statistical significance and substantive significance.

**Statistical significance is a concept that captures how (un)likely an estimated effect would arise by mere chance (where the effect is measured relative to some null hypothesis) if the null hypothesis were true. Statistical significance says nothing about importance from a policy or theoretical standpoint. In fact, for a given estimate, whether it is statistically significant will be affected by how precisely it is estimated (significant if large sample size; insignificant if small sample size) independently of how important the effect would be if it were "true." Substantive significance depends on factors outside of the statistical analysis. For example, an estimate can be statistically significant but unimportant (e.g., a huge clinical trial shows that a medicine increases death risk by 0.0000000001%) or statistically insignificant but potentially very important (e.g., a tiny clinical trial estimates that a medicine doubles a death risk but it is not possible to differentiate that doubling from a zero effect because the sample size is 10).**

Questions 10 & 11:

You have data for a quasi-experimental study. The goal of the study is to see to what extent jurors and judges are able to disregard inadmissible evidence in their verdicts (assigning liability). Two independent groups of participants were recruited: jurors (Group = 0) and judges (Group = 1). They were randomly given a case that either contained no inadmissible evidence (Treatment = 0) or inadmissible evidence with the instruction to ignore it (Treatment = 1). The assignment of liability was measured on a scale from 1-7 (higher scores indicating more liability assigned).

10. You first ran a regression model with Group and Treatment predicting Liability. You then ran a model that includes the interaction between group and treatment as well. Stata gives you the output below. Interpret the coefficients of Group and Treatment for both models. Why do they differ between both models?

```
. regress Liability i.Group i.Treatment
```

| Source | SS | df | MS |
|---|---|---|---|
| Model | 88.240569 | 2 | 44.1202845 |
| Residual | 386.320301 | 227 | 1.70185154 |
| Total | 474.56087 | 229 | 2.07231821 |

| | | |
|---|---|---|
| Number of obs | = | 230 |
| F(2, 227) | = | 25.92 |
| Prob > F | = | 0.0000 |
| R-squared | = | 0.1859 |
| Adj R-squared | = | 0.1788 |
| Root MSE | = | 1.3046 |

| Liability | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Group | | | | | | |
| Judge | -.6305017 | .1735481 | -3.63 | 0.000 | -.972473 | -.2885304 |
| Treatment | | | | | | |
| Inadmissable Evidence | 1.0586 | .1720655 | 6.15 | 0.000 | .7195502 | 1.39765 |
| _cons | 2.531788 | .1439524 | 17.59 | 0.000 | 2.248134 | 2.815442 |

```
. regress Liability i.Group i.Treatment Group#Treatment
```

| Source | SS | df | MS |
|---|---|---|---|
| Model | 113.889329 | 3 | 37.9631096 |
| Residual | 360.671541 | 226 | 1.59589177 |
| Total | 474.56087 | 229 | 2.07231821 |

| | | |
|---|---|---|
| Number of obs | = | 230 |
| F(3, 226) | = | 23.79 |
| Prob > F | = | 0.0000 |
| R-squared | = | 0.2400 |
| Adj R-squared | = | 0.2299 |
| Root MSE | = | 1.2633 |

| Liability | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Group | | | | | | |
| Judge | .0401348 | .237124 | 0.17 | 0.866 | -.4271219 | .5073916 |
| Treatment | | | | | | |
| Inadmissable Evidence | 1.644413 | .2216213 | 7.42 | 0.000 | 1.207705 | 2.081121 |
| Group#Treatment | | | | | | |
| Judge#Inadmissable Evidence | -1.347494 | .3361208 | -4.01 | 0.000 | -2.009826 | -.6851626 |
| _cons | 2.234375 | .1579108 | 14.15 | 0.000 | 1.923209 | 2.545541 |

**The effect of group and treatment is significant in model 1, in model 2 only treatment is significant**

**Group: In Model 1, judges are found to assign significantly less liability than jurors (b= -.63 on 7-point likert scale), in model 2 this difference is not significant (b =.04)**
**Treatment: both Model 1 and Model 2 show that those receiving inadmissible evidence assign more liability (Model 1: b = 1.06; Model 2: b = 1.64).**
**In Model 1 the coefficients/slopes/effects of Group (Treatment) are the overall effect of Group (Treatment). In Model 2, the coefficients are calculated for the other predictor equaling zero: The coefficient of Group is that for those who did not receive inadmissible evidence (Treatment = 0); the coefficient of Treatment is that for jurors (group = 0).**
**The coefficient of Group in Model 2 shows that when there is no admissible evidence, judges do not differ from jurors in assigning liability (b =.04); The coefficient of Treatment in Model 2 shows that for Jurors, receiving inadmissible evidence results in more liability assigned (b = 1.64)**

11. You ask Stata for the marginal effects and marginal means. You get the output below. Interpret the interaction. How are the different groups affected by the treatment?

```
. margins Group, dydx(Treatment)

Conditional marginal effects                    Number of obs     =        230
Model VCE      : OLS

Expression     : Linear prediction, predict()
dy/dx w.r.t. : 1.Treatment
```

| | dy/dx | Delta-method Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| **0.Treatment** | (base outcome) | | | | | |
| **1.Treatment** | | | | | | |
| Group | | | | | | |
| Juror | 1.644413 | .2216213 | 7.42 | 0.000 | 1.207705 | 2.081121 |
| Judge | .2969188 | .2527078 | 1.17 | 0.241 | -.201046 | .7948835 |

Note: dy/dx for factor levels is the discrete change from the base level.

```
. margins Treatment#Group

Adjusted predictions                              Number of obs     =        230
Model VCE    : OLS

Expression   : Linear prediction, predict()
```

| | Margin | Delta-method Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Treatment#Group | | | | | | |
| No Inadmissable evidence#Juror | 2.234375 | .1579108 | 14.15 | 0.000 | 1.923209 | 2.545541 |
| No Inadmissable evidence#Judge | 2.27451 | .1768954 | 12.86 | 0.000 | 1.925934 | 2.623085 |
| Inadmissable Evidence#Juror | 3.878788 | .1554998 | 24.94 | 0.000 | 3.572373 | 4.185203 |
| Inadmissable Evidence#Judge | 2.571429 | .1804694 | 14.25 | 0.000 | 2.215811 | 2.927047 |

**The effect of treatment is significant for jurors and not significant for judges**

**Jurors who received inadmissible evidence assign significantly more liability than jurors who did not receive inadmissible evidence (b = 1.64).**

**Judges who received inadmissible evidence do not assign significantly more liability than judges who did not receive inadmissible evidence (b = .29).**