

ECNS 561  
Midterm  
Fall 2015

\_\_\_\_\_Name

**1.)** Proofs of variance and covariance propositions.

**a.) (5 points)** Property Var. 2 states that, for any constants  $a$  and  $b$

$$\text{Var}(aX + b) = a^2\text{Var}(X).$$

Prove this property.

**b.) (5 points)** Property Cov. 2 states that, for any constants  $a_1$ ,  $b_1$ ,  $a_2$ , and  $b_2$

$$\text{Cov}(a_1X + b_1, a_2Y + b_2) = a_1a_2\text{Cov}(X, Y).$$

Using the fact that  $\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$ , prove this property.

2.) (10 points) A Cauchy continuous random variable is characterized by the following pdf:

$$f_X(x) = \left(\frac{1}{\pi}\right) \left(\frac{1}{1+x^2}\right) \quad -\infty < x < \infty.$$

Show that the expected value of a Cauchy random variable  $X$  does not exist (Note that, for any positive number  $M$ ,  $\int_0^M \frac{x}{1+x^2} dx = \frac{1}{2} \log(1+x^2) \Big|_0^M$ ).

3.) Given a positive constant  $k > 0$ , the exponential density function is

$$f_X(x) = \begin{cases} ke^{-kx} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Let  $X$  be a continuous random variable with an exponential density function with parameter  $k$ .

**a.) (7 points)** Solve for  $E[X]$ .

**b.) (8 points)** Solve for  $\text{Var}[X]$ .

4.) Adult males are taller, on average, than adult females. Visiting two recent American Youth Soccer Organization (AYSO) under 12 year old (U12) soccer matches on a Saturday, you do not observe an obvious difference in the height of boys and girls of that age. You suggest to your little sister that she collect data on height and gender of children in 4<sup>th</sup> to 6<sup>th</sup> grade as part of her science project. The accompanying table shows her findings ( $\bar{Y}$ =mean,  $sd$  = standard deviation,  $N$  = sample size).

**Height of Young Boys and Girls, Grades 4-6, in inches**

Boys			Girls		
$\bar{Y}_{boys}$	$sd_{boys}$	$N_{boys}$	$\bar{Y}_{girls}$	$sd_{girls}$	$N_{girls}$
57.8	3.9	55	58.4	4.2	57

a.) (2 points) Let your null hypothesis be that there is no difference in the height of females and males at this age level. Write down the null and alternative hypotheses.

b.) (2 points) Find the difference in height and the standard error of the difference. Note that

$$se(\bar{Y}_1 - \bar{Y}_2) = \sqrt{\frac{sd_1^2}{N_1} + \frac{sd_2^2}{N_2}}.$$

c.) (3 points) Generate a 95% confidence interval for the difference in height.

Critical values	
$p$ -value	two-tailed test
.050	1.96
.025	2.24
.010	2.58
.005	2.81
.001	3.3

d.) (3 points) Calculate the  $t$ -stat for comparing the two means. Using a two-tailed test, is the difference stat. significant? Would the critical value be smaller if you had assumed a one-sided alternative hypothesis? Why? Give no more than a two-sentence answer on the intuition here.

5.) In an effort to restrict youth firearm access, a number of states have passed child access prevention (CAP) laws, which impose criminal liability on gun owners who allow children unsupervised access to firearms. More specifically, these laws target households with children under the age of 18. Proponents of CAP laws argue that they could decrease the incidence of school shootings.

The Department of Justice has asked you to evaluate the relationship between CAP laws and school shooting deaths. To study this relationship, you have data on the census of school-associated shooting deaths in the United States for the period 1990 through 2014.

a.) (5 points) Suppose the mean rates of school-associated shooting deaths (per 100,000 high school student population) are as follows:

	CAP Law States		Non-CAP Law States	
	Pre-Policy	Post-Policy	Pre-Policy	Post-Policy
Shooting death rate	.25	.10	.35	.29

If you were to **only** use within-CAP law state variation in shootings, would you likely be over- or under-estimating the true effect of these laws on school shooting deaths? Why? If you were to use variation from **both** CAP law and non-CAP law states, do the data suggest that CAP laws are effective?

**b.) (10 points)** Now let's assume you are able to obtain more detailed data on each shooting event. Specifically, you now have information on the age of the shooter. This is important because the laws are intended to bind for households with children under a certain age (i.e., 18 years of age). Given these data, you are able to calculate the following mean rates of school-associated shooting deaths:

	CAP Law States		Non-CAP Law States	
	Pre-Policy	Post-Policy	Pre-Policy	Post-Policy
Shooting death rate (where shooter's age < 18)	.32	.15	.40	.32
Shooting death rate (where shooter's age $\geq$ 18)	.18	.05	.30	.26

With these more detailed data, does your inference regarding the effectiveness of these laws change? What does using the data on older shooters buy you? That is, what type of bias does this protect you against? Feel free to give an example of a specific event that, without exploiting the age at which the law binds, would result in a biased estimate.

6.) In matrix notation, suppose we have the following population model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where  $\mathbf{X}$  is an  $n$  by  $k$  matrix of  $k$  independent variables for  $n$  observations,  $\mathbf{y}$  is an  $n$  by 1 vector of observations on the dependent variable,  $\boldsymbol{\beta}$  is a  $k$  by 1 vector of unknown population parameters that we wish to estimate, and  $\boldsymbol{\varepsilon}$  is an  $n$  by 1 vector of errors.

a.) (7 points) Using matrix notation, show that the OLS estimator,  $\hat{\boldsymbol{\beta}}$ , is equal to  $(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y})$ .



**b.) (8 points)** Now, instead of the model above, suppose the true population model is given as

$$(1) \quad \mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}.$$

However, suppose you only have data to estimate the following model

$$(2) \quad \mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{u} \quad \text{where } \mathbf{u} = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}.$$

Using matrix notation, show that the OLS estimator for equation (2) is biased.

**c.) (5 points)** Consider the OLS estimator for equation (2) in part b.) above. Under what two scenarios will the bias in this estimator be eliminated? Be short and brief in your answer.

7.) The following regressions represent alternative specifications for estimating the relationship between local alcohol availability and crime

$$(1) \quad \text{Crime}_c = \beta_0 + \beta_1 \text{Bars}_c + \varepsilon_c,$$

$$(2) \quad \ln(\text{Crime}_c) = \beta_0 + \beta_1 \text{Bars}_c + \varepsilon_c,$$

$$(3) \quad \ln(\text{Crime}_c) = \beta_0 + \beta_1 \ln(\text{Bars}_c) + \varepsilon_c,$$

where *Crime* is equal to the crime rate (per 1,000 population) in county *c* and *Bars* is equal to the number bars per 1,000 population in county *c*. Suppose we estimate these three separate models and obtain the following estimates for  $\beta_1$ :

Local Alcohol Availability and Crime			
	Model (1)	Model (2)	Model (3)
<i>Bars</i>	4.21*** (1.08)	.071*** (.020)	.405*** (.100)
* Statistically significant at 10% level; ** at 5% level; *** at 1% level.			
Notes: Each column represents the results from a separate OLS regression. The dependent variable is equal to the natural log of the crime rate in county <i>c</i> . Standard errors are in parentheses.			

a.) (6 points) Interpret the coefficient estimate in each of the three regressions.

b.) (4 points) Suppose we suspect that the average age of the population within a county is an important predictor of the crime rate and, simultaneously, a determinant of the demand for alcohol. Furthermore, suppose we want to allow for a nonlinear relationship between age and crime. How might we accommodate our regressions above to allow for this possibility? Provide a brief justification for your answer.

**8.) (10 points)** Recall the article by Carrell et al. (2011) that was published in the *Journal of Public Economics* and entitled “Is Poor Fitness Contagious?” Briefly describe the experimental design used in their research and whether a causal interpretation of the results could be threatened by omitted variable bias.