

**FORK 1002 Statistics**  
**Exercise Set 4: Qualitative Explanatory Variables**

1. In a survey ten students were asked about the time (in hours) they spend on average during a week on Facebook, and about the mobilephone/smartphone they use most frequently:

Student no.	Time spent on Facebook	Brand
1	1	iPhone
2	40	iPhone
3	2	iPhone
4	10	Nokia
5	15	iPhone
6	2	Sony Ericsson
7	35	Nokia
8	35	Sony Ericsson
9	2	Sony Ericsson
10	1	iPhone

Make three dummy variables  $D1$ ,  $D2$  and  $D3$ , one variable for each phone brand.

2. Let  $Y_i$  be the hourly wage of person  $i$ , let  $X_i$  be the number of education of person  $i$ , and let  $D1_i$  and  $D2_i$  be dummy variables associated with the values of the qualitative variable gender.  $D1_i$  is equal to 1 if person  $i$  is woman and 0 otherwise, whereas  $D2_i$  is equal to 1 if person  $i$  is a man and 0 otherwise. Which of the following models (below) cannot be estimated due to the “dummy-trap” problem? Propose a solution in the cases where there is a dummy-trap problem.

- (a)  $Y_i = B_1 + B_2D1_i + u_i$
- (b)  $Y_i = B_1D1_i + B_2D2_i + u_i$
- (c)  $Y_i = B_1 + B_2D1_i + B_3D2_i + u_i$
- (d)  $Y_i = B_1 + B_2D1_i + B_3X_i + u_i$
- (e)  $Y_i = B_1D1_i + B_2D2_i + B_3X_i + u_i$
- (f)  $Y_i = B_1 + B_2D1_i + B_3D2_i + B_3X_i + u_i$

3. Let  $D1$  and  $D2$  be dummy variables associated with two of the values of a qualitative variable that can assume 3 values.

- (a) What are the predicted values of the estimated model

$$\hat{Y} = 0,5 + 0,2D1$$

for  $D1$  equal to 0 and 1, respectively?

(b) What are the predicted values of the estimated model

$$\widehat{Y} = 0,3 + 0,1D1 + 0,6D2$$

for the four value-pairs  $(D1, D2) = (0, 0)$ ,  $(D1, D2) = (1, 0)$ ,  $(D1, D2) = (0, 1)$  and  $(D1, D2) = (1, 1)$ ?

(c) What are the predicted values of the estimated model

$$\widehat{Y} = 0,2 + 0,3D1 + 0,5D2 + 0,1D1 \cdot D2$$

for the four value-pairs  $(D1, D2) = (0, 0)$ ,  $(D1, D2) = (1, 0)$ ,  $(D1, D2) = (0, 1)$  and  $(D1, D2) = (1, 1)$ ?

4. The file *infidelity\_sample.xls* contains data about the extramarital affairs during the year of 150 married persons. The observations are from a dataset that was used in Ray Fair, “A theory of extramarital affairs”, *Journal of Political Economy*, volume 86. The variables in the dataset are:

<i>naffairs<sub>i</sub></i> :	The number of extramarital affairs of person <i>i</i>
<i>male<sub>i</sub></i> :	Dummy-variable equal to 1 if person <i>i</i> is a man, 0 otherwise
<i>age<sub>i</sub></i> :	Age in years of person <i>i</i>
<i>yrsmarr<sub>i</sub></i> :	The number of years that person <i>i</i> has been married
<i>kids<sub>i</sub></i> :	Equal to 1 if person <i>i</i> has children, 0 otherwise
<i>relig<sub>i</sub></i> :	Degree of religiousness of person <i>i</i> : 1 = anti-religious, 2 = non-religious, 3 = a little bit religious, 4 = somewhat religious, 5 = very religious
<i>educ<sub>i</sub></i> :	Years of education
<i>ratemarr<sub>i</sub></i> :	Own perception of how well the marriage is going: 1 = very bad, 2 = below average, 3 = average, 4 = above average, 5 = very good

(a) Consider the following estimated model:

$$\widehat{naffairs}_i = \underset{(0.50)}{6.21} - \underset{(0.70)}{0.72}male_i.$$

The numbers in parentheses are the standard errors of the parameter estimates. What is the predicted number of extramarital affairs for a woman and for a man, respectively? Is the difference significant at a 10% significance level?

- (b) Consider the following estimated model:

$$\widehat{naffairs}_i = \underset{(1.23)}{6.92} - \underset{(1.65)}{2.05}male_i - \underset{(1.35)}{0.85}kids_i + \underset{(1.83)}{1.62}male_i \cdot kids_i.$$

What is the predicted number of extramarital affairs for a woman with children? What is the predicted number of extramarital affairs for a man without children? And for a man *with* children?

- (c) Consider the following estimated model:

$$\widehat{naffairs}_i = \underset{(0.71)}{3.76} + \underset{(0.07)}{0.23}yrsmarr_i - \underset{(0.06)}{0.02}yrsmarr_i \cdot male_i.$$

How many more extramarital affairs does a man and a woman, respectively, have on average for each year more of marriage? Is the difference significant at a 5% significance level?

## 5. Computer exercises:

- (a) Make a new data file [Hint for SPSS: File → New → Data]
- (b) Make a new variable called *brand*: Let 1 stand for Nokia, let 2 stand for Sony Ericsson and let 3 stand for iPhone [Hint for SPSS: Click on “Data View”, enter the values into an empty column, click on “Variable View”, write the name (*brand*) in the the “Name” column]
- (c) Make three dummy variables, one for each phone brand [Hint for SPSS: Transform → Compute Variable..., write *D1* in the “Target Variable” box, write “brand = 1” in the “Numeric Expression” box, press “OK”. Same procedure for the two other values]
- (d) Estimate the model in 4(b) [Hint for SPSS: Before estimation the variable *male · kids* needs to be constructed: Transform → Compute Variable..., write (for example) “MaleKids” in the “Target Variable” box, write “male\*kids” in the “Numeric Expression” box, press “OK”, estimate model via Analyze → Regression → Linear...]