ECO220Y1Y, APRIL 2022, FINAL EXAM: SOLUTIONS

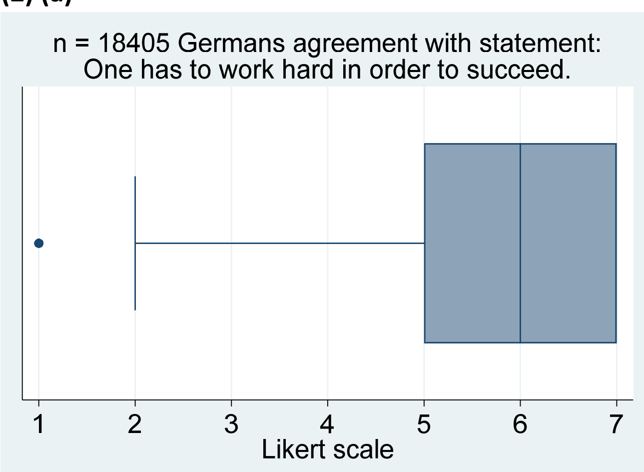
(1) (a) About 33% of managers predicted that they would be in the top (fifth) quintile, which is the best 20% of managers, which would be consistent with some managers being overconfident.

(b) Among those managers whose actual performance is in the bottom (first) quintile (worst), 14% of them predicted that they would be in the top (fifth) quintile (best), which is consistent with a huge overestimation of themselves.

(c) Panel A would have all five bars at a height of 0.2 if they could perfectly predict: exactly 20% would be in each quintile. Panel B would have 1's on the diagonal from the bottom left to the top right and 0's in the off diagonal: for each fourth quarter quintile 100% would of managers would have correctly predicted themselves being in that quintile.

(d) The values on the diagonal from the bottom left to the top right would be smaller than they are now – fewer managers correctly predicting – and the off-diagonal elements would be bigger than they are now – more managers either over or under predicting their future performance. This would weaken the positive correlation between actual performance and predicted performance.

(2) (a)



(b) The bar height is about 0.66 and the bar width is about 1/6 so the bar area is about 0.11, which for a density histogram is the fraction in the bin. 11% of 18,405 is about <u>2,025</u> respondents.

(c) It measures the variation of opinions across Germans in the sample about how much they agree that they have little control over things in their life: it is an *estimate* of the variation in the population. The sample standard deviation of 1.5 is not expected to either increase or decrease: it should be about the same aside from any random sampling error.

(d) Neither is correct because both Questions 1) and 5) measure a person's sense of an locus of control and they are surely positively correlated. We must use $\sqrt{V[X_1] + V[X_5] + 2 * SD[X_1] * SD[X_5] * CORR[X_1, X_5]}$ and both answers forget about the correlation. Answer #1 makes an extra mistake because we can NEVER add standard deviations even when there is zero correlation between two variables.

(e) In 2010 in Germany, on average adults with an average level for the locus of control (the standardized value is zero), donate about 68.26 euros to charity in 2009. People with a one standard deviation higher locus of control (more sense that they control their own destiny), donate about 16.42 euros more on average, which is a sizeable (24%) increase.

(f) After controlling for income, sex, age, and level of education, people with a one standard deviation higher locus of control, donate about 6.72 euros more on average. The excerpt says that the slope estimate in Column (1) is likely suffering an endogeneity bias because of lurking/unobserved/confounding/omitted variables. Column (2) controls for these variables. The excerpt means that the authors think the estimate in Column (2) is superior to the estimate in Column (1) in figuring out the causal relationship between charitable giving and a person's sense of the locus of control.

(g) After controlling for a measure of locus of control, sex, age, and years of education, Germans with income that is 10% higher on average give away 9.66 euros less in the hypothetical dictator game.

(h) After controlling for a measure of locus of control, income, age, and years of education, females on average give away an amount of money that is 0.22 standard deviations higher than males in the hypothetical dictator game.

(i) First, create a dummy variable for each answer: e.g. a variable named muslim that is 1 if the person picked that answer and zero otherwise. Next, include THREE of the religion dummy variables as explanatory (x) variables in the multiple regression remembering that you must leave one out to serve as the reference (aka omitted) category.

(3) (a)

$$H_0: p_F - p_C = 0$$

$$H_1: p_F - p_C \neq 0$$

$$\bar{P} = \frac{X_F + X_C}{n_F + n_C} = \frac{0.56 * 2{,}006 + 0.53 * 2{,}022}{2{,}006 + 2{,}022} = \frac{1123.36 + 1071.66}{4{,}028} = \frac{2{,}195.02}{4{,}028} = 0.54495$$

$$z = \frac{\hat{P}_F - \hat{P}_C}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n_F} + \frac{\bar{P}(1-\bar{P})}{n_C}}} = \frac{0.56 - 0.53}{\sqrt{\frac{0.54495(1-0.54495)}{2{,}006} + \frac{0.54495(1-0.54495)}{2{,}022}}} = \frac{0.03}{0.01569} = 1.91$$

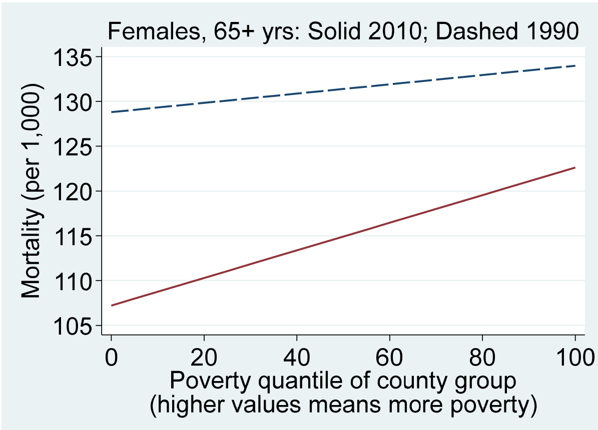
$$P - value = P(Z < -1.91) + P(Z > 1.91) = 2 * (0.5 - 0.4719) = 0.056$$

There is a statistically significant difference at a 10% significance level (and *almost* a 5% level) in the fraction of people in Canada and France willing to either limit their flying "a lot" or "a great deal." However, the difference is only 3 percentage points – in France 56% versus 53% in Canada – making these two countries very similar to each other relative to high-income countries in general, which means that this difference is *not* particularly economically significant.

(b) As of 2021 – 2022, we are 95% confident that people in Italy are between 35.2 and 40.8 percentage points more likely to say they are willing to limit their beef/meat consumption either "a lot" or "a great deal" in comparison with people in Japan. This is a huge difference: among the high-income countries Italians are by far the most willing to adopt the climate friendly behavior of limiting meat consumption (62%) whereas the Japanese are the least willing to do so (24%).

(c) $P\left(\hat{P} < 0.27 \,\middle|\, p = 0.42, n = 1{,}717\right) = P\left(Z < \frac{0.27-0.42}{\sqrt{\frac{0.42(0.58)}{1{,}717}}}\right) = P\left(Z < \frac{-0.15}{0.0119}\right) = P(Z < -12.6) \approx 0$

(4) (a)



Left endpoint of 2010 line: (0, 107.2167507)

Right endpoint of 2010 line: (100, 122.61887) where later is 107.2167507 + 0.154021206*100

Left endpoint of 1990 line: (0, 128.79743) where later is 107.2167507 + 21.58067632

Right endpoint of 1990 line: (100, 133.98009) where later is 107.2167507 + 21.58067632 + (0.154021206 - 0.102194588)*100

(Note: Following Currie and Schwandt (2016), the vertical axis is three-year mortality.)

(b) Yes, the coefficient on the interaction term is highly statistically significant at the 1% level, because Excel reports that the P-value for the test of statistical significance – $H_0: \beta_3 = 0$ versus $H_1: \beta_3 \neq 0$ – is only 0.007. This means that we can conclude that there IS a difference in slopes for the relationship between poverty and mortality rates comparing 1990 with 2010 for women aged 65 and over in the U.S.: in other words, mortality inequality is different. If it were not statistically significant, we would not be sure there is a difference in mortality inequality comparing 1990 with 2010. [In fact, the point estimate means mortality inequality got worse for this sex-age-group from 1990 to 2010.]

(5) (a)

$H_0: \beta_1 = \beta_2 = \beta_3$

H_1: Not all slope coefficients are zero

Use the $F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$ test statistic. The F table gives the critical value of 4.94. Hence, we need an F test statistic at least that big for the regression to be statistically significant overall at a 1% significance level. Solving $4.94 = \frac{R^2/3}{(1-R^2)/(24-3-1)}$ for the R^2 yields 0.426.

(b) With a small sample (like in Part (a)) we need a large R-squared of at least 0.426 – a good fit – even to rule out no relationship at all between any of the x variables and y. With small sample sizes, which are subject to lots of sampling error, we need quite dramatic evidence to rule out the null. In contrast, with a large sample size, we have less sampling error and hence we can prove that there is a relationship between y and the combination of the five x variables even if we have an R-squared as tiny as 0.005. With very large sample sizes even results that are tiny and not at all economically significant will become statistically significant.