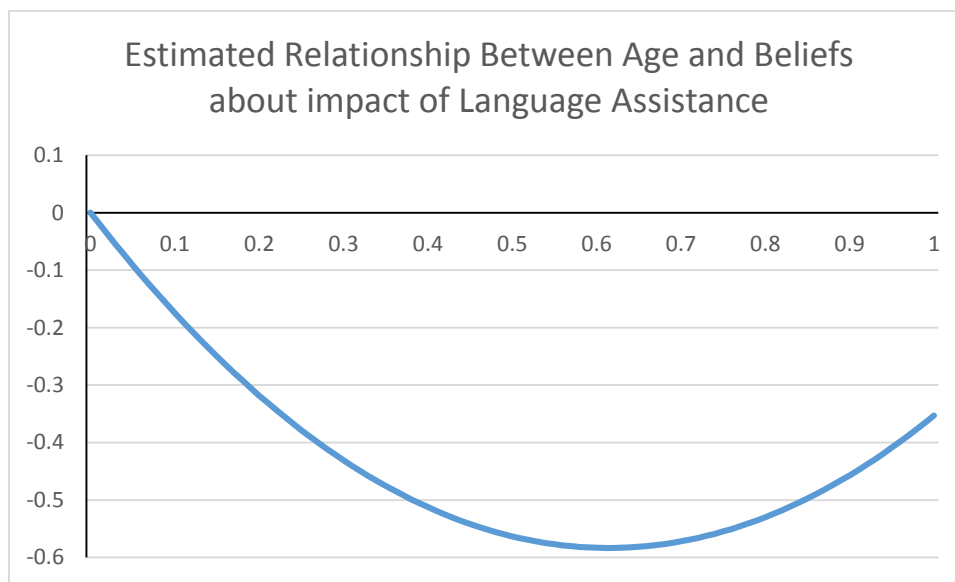# 17.871: Solutions for Problem Set 5

Part I

(1)

- <u>age01:</u> age is strongly negatively related to the belief that voting assistance improves elections. A one percentage point increase in age is associated with being 0.019 points lower on the four-point scale for supporting voter assistance
- <u>age01sq:</u> overall. the negative impact of age on the belief that voting assistance improves elections increases with age, up to a point, and then decreases. The squared term tends to dampen the negative impact of age, but has very little impact at low values of age. However, as age increases, it has a larger impact in dampening the effect, eventually causing the total effect of age to begin falling with age.
- [Note: it's easiest to understand squared terms by plotting the effect. Here it is in excel, where we simply calculated -1.9*age + 1.56*age-squared over the whole range of age:]

Estimated Relationship Between Age and Beliefs about impact of Language Assistance

- <u>liberal01:</u> liberal ideology is very weakly positively related to the belief that voting assistance improves elections. A one percentage point increase in liberalism is associated with being 0.0019 points higher on the four-point scale for supporting voter assistance
- <u>dem01:</u> identifying more with the Democrats is quite strongly positively related to the belief that voting assistance improves elections. Moving from being a Republican to a Democrat is associated with being 0.5 points higher on the four-point scale for supporting voter assistance
- <u>fraudscale01:</u> believing that fraud occurs frequently is strongly negatively related to the belief that voting assistance improves elections. A one percentage point increase in the belief that fraud occurs regularly is associated with being 0.0069 points lower on the four-point scale for supporting voter assistance

- <u>votedin12:</u> having voted in 2012 is weakly negatively related to the belief that voting assistance improves elections. Moving from having not voted to having voted is associated with being -0.19 points lower on the four-point scale for supporting voter assistance
- <u>black:</u> being black is positively related to the belief that voting assistance improves elections. Moving from being white to black is associated with being 0.28 points higher on the four-point scale for supporting voter assistance
- <u>hispanic:</u> being hispanic is quite strongly positively related to the belief that voting assistance improves elections. Moving from being white to hispanic is associated with being 0.4 points higher on the four-point scale for supporting voter assistance
- <u>asian:</u> being asian is strongly positively related to the belief that voting assistance improves elections. Moving from being white to asian is associated with being 0.71 points higher on the four-point scale for supporting voter assistance
- <u>other_race:</u> being non-white but not black, asian or hispanic is basically unrelated to the belief that voting assistance improves elections. Moving from being white to being in the 'other' racial category is associated with being just 0.092 points lower on the four-point scale for supporting voter assistance
- <u>constant:</u> a voter who scored zero for every variable in the regression is predicted to be at 2.16 on the four-point scale for supporting voter assistance

(2)

The confidence interval is equal to: coefficient $\pm 1.96 *$ standard error

- for age01 it is $-1.9 \pm 1.96 * 0.519 = [-2.92, -0.88]$
- for age01sq it is $1.55 \pm 1.96 * 0.545 = [0.48, 2.62]$
- for liberal01 it is $0.194 \pm 1.96 * 0.0429 = [0.11, 0.28]$

Part II

(1)

You would expect the average to be the same in Sample 2. The sample mean of a random sample from the population provides an unbiased estimate of the population mean, meaning that if you took many sample means from repeated samples, the distribution of these sample means would be centered around the true mean. This property of *unbiasedness* holds regardless of the sample size.

Therefore, there is no reason to expect that the average in Sample 2 would be systematically different to Sample 1. It's very unlikely to be exactly the same; there's just no reason to think it will definitely be higher or lower.

(2)

You would expect the standard error of Sample 2 to be lower. That is because the standard error is calculated as the standard deviation of the sample divided by the square root of the sample size (n). The sample standard deviation does not depend on n, and therefore as n increases, the standard error decreases. This property is known as *consistency* and is one reason why larger samples are preferable.

We can also calculate an estimate: since the standard error is equal to the standard deviation divided by the square root of the sample size, we know that the standard deviation of Sample 2 was equal to $3950 * \sqrt{100} = 39500$. Therefore, assuming the standard deviation of Sample 2 is the same, its standard error will be $\frac{3950}{\sqrt{10000}} = 395$

Part III

The sample size would need to be 1,000 as well. This follows immediately from the formula for confidence intervals.

Part IV

Here we need to calculate a two sample t-test to compare the two proportions. The formula for a two-sample t test is:

$$t = \frac{\mu_1 - \mu_2}{\sqrt{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]}}$$

where μ_1 and μ_2 are the means from samples 1 and 2, s_1^2 and s_2^2 and s_2 are their variances, and n_1 and n_2 are the sample sizes.

In this case, however, the variance of the figure from the official election statistics is zero, since they cover the entire population. Therefore we only need to calculate the standard error from the CCES, which is $\sqrt{\frac{(0.359)(1-0.359)}{56200}} = 0.00202$, where we have used the fact that the variance of a proportion *p* is equal to *p*(1-p)*

Therefore, the t statistic is equal to (0.632-0.359)/0.00202 = 134.9. This far exceeds the critical values for tests of any standard significance level, so there is very strong evidence that the two estimates are different.