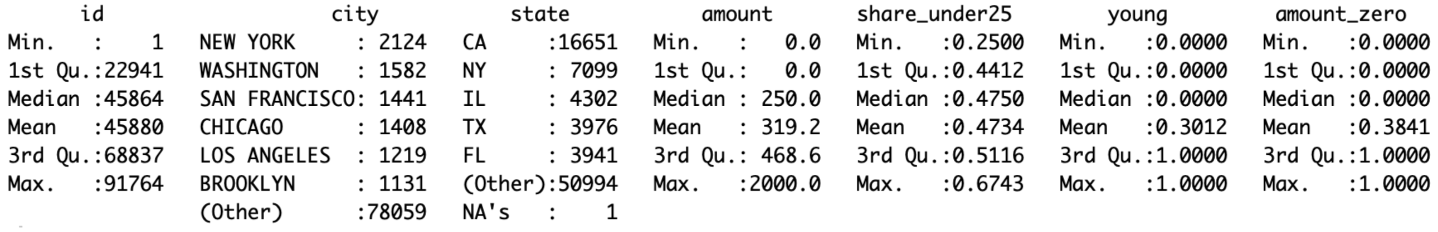
ECOM20001: Econometrics 1

Assignment 1: Suggested Solutions

---

1. Summary statistics reported below along with standard deviations for amount, share_under25, and young. Interpreting the sample means, a typical donor donates $319.20 to the Democratic Party, lives in a ZIP code where 47.34% of people are under 25 years old, and 30.12% of donors live in ZIP codes classified as 'young,' that is where more than 50% of people are less than 25 years old. Based not the 30.12% sample mean for young alone, we can conclude that 0.5 is not the sample median for young; if it was, the mean for young would be 50% (with half the sample classified as young=1 and the other half the sample classified as young=0).

   Summary statistics with means, min, max values for amount, share_under25, and young:



```
       id                   city           state           amount        share_under25        young          amount_zero
 Min.   :    1   NEW YORK     : 2124   CA     :16651   Min.   :   0.0   Min.   :0.2500   Min.   :0.0000   Min.   :0.0000
 1st Qu.:22941   WASHINGTON   : 1582   NY     : 7099   1st Qu.:   0.0   1st Qu.:0.4412   1st Qu.:0.0000   1st Qu.:0.0000
 Median :45864   SAN FRANCISCO: 1441   IL     : 4302   Median : 250.0   Median :0.4750   Median :0.0000   Median :0.0000
 Mean   :45880   CHICAGO      : 1408   TX     : 3976   Mean   : 319.2   Mean   :0.4734   Mean   :0.3012   Mean   :0.3841
 3rd Qu.:68837   LOS ANGELES  : 1219   FL     : 3941   3rd Qu.: 468.6   3rd Qu.:0.5116   3rd Qu.:1.0000   3rd Qu.:1.0000
 Max.   :91764   BROOKLYN     : 1131   (Other):50994   Max.   :2000.0   Max.   :0.6743   Max.   :1.0000   Max.   :1.0000
                 (Other)      :78059   NA's   :    1
```

Standard deviations from the R code are:

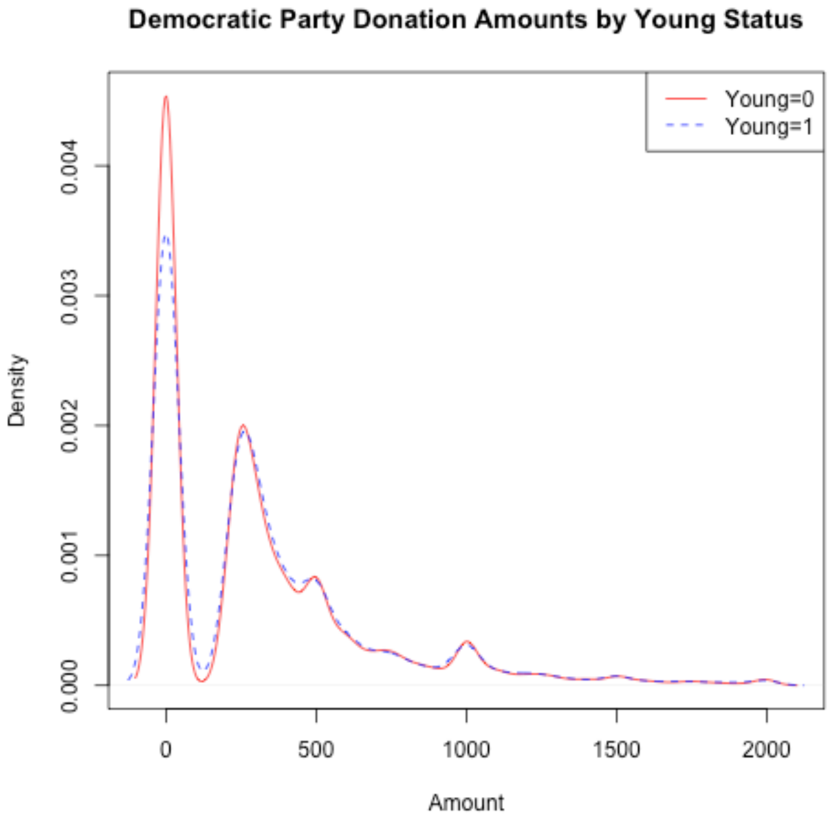
- $374.97 for amount
- 0.045 for share_under25
- 0.459 for young

2. 95% confidence intervals for the mean of each respective variable

   - [$316.73, $321.71] for amount
   - [0.4730,0.4736] for share_under25
   - [0.2981,0.3042] for young

3. Densities displayed in the figure on the next page. Both distributions of amount when young=1 and young=0 are right-skewed and bi-modal with modes at amount=$0 and amount=$250. That is, a large share of donors donate nothing,

but if they donate at all, they tend to donate $250. The largest difference between the two densities exists at amount=$0, with the young=0 group of donors (e.g,. people living in areas with 'older' populations) being relatively more likely to donate $0; the distributions are otherwise quite similar.

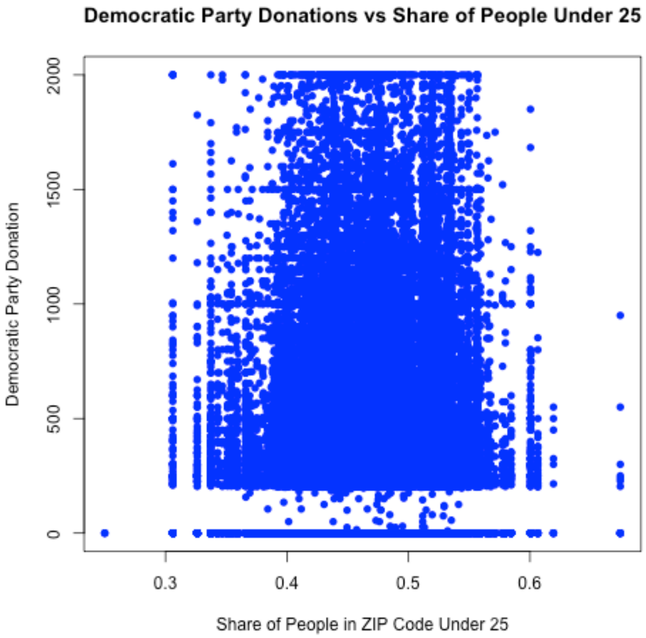


4. The difference in means is $11.71, 95% CI is [$6.27, $17.15], p-value for the test is < 0.01, implying a statistically significant difference in means at the 5% level. Interpretation is that donors living in 'young' (young=1) areas tend to donate $11.71 more to the Democratic Party than donors living on 'old' areas (young=0).

5. The difference in means is -0.026, 95% CI is [-0.033, -0.019], p-value for the test is < 0.01, implying a statistically significant difference in means at the 5% level. Interpretation is that donors living in 'young' (young=1) are 2.6 percent less likely to just donate $0 to the Democratic Party than donors living on 'old' areas (young=0), as foreshadowed by the conditional density plots from question 3.

6. Scatter plot presented on the next page, which visually does not immediately reveal a clear positive or negative relationship between amount and

share_under25. From the R code, the correlation coefficient is computed as 0.015, suggesting a weak positive relationship at best.



7. Summarising the results from the single linear regression of amount on share_under25:

- Intercept:
  - Estimate is $259.05, which in words means the predicted mean donation for the Democratic Party is $259.05 when share_under25=0.
  - It has a standard error of 13.37, t-statistic of 19.4, and the p-value for a 2-sided test of the null that the intercept equals 0 is less than 0.01 meaning we reject the null at the 5% level.
  - The 99% confidence interval for the intercept estimate is [$259.05-2.58 x 13.37,$259.05+2.58. x 13.37] = [$224.56,$293.54]
- Predicted change in amount for a one-unit change in share_under25, which we can read off directly from the single regression output in R, is:
  - Estimate is $127.10, which in words means the predicted change in amount if share_under25 increases from 0 to 1 (e.g., changes from the minimum to maximum theoretical value for share_under25) is $127.10.

    - It has a standard error of 28.12, t-statistic of 4.52, and the p-value for a 2-sided test of the null that this predicted change equals 0 is less than 0.01 (all from the R output) meaning we reject the null at the 5% level.
    - The 99% confidence interval for the predicted change in amount is [$127.10-2.58 x 28.12,$127.10+2.58 x 28.12] = [$54.55,199.65]
  - Predicted change in amount for a one-standard deviation change in share_under25, which we need to do auxiliary calculations for (e.g., it cannot be read directly from the R output):
    - First thing to recall from question 1 that the standard deviation of share_under25 is 0.045.
    - Given this, the estimate of the predicted change in amount from a 0.045 change in share_under25 is: 0.045 x $127.10=$5.72. In words, a 0.045 one-standard deviation in share_under25 corresponds to a predicted $5.72 increase in a donor's donation to the Democratic Party.
    - Computing the 99% confidence interval around the $5.72 predicted change, (see slide 32 of lecture note 5) we obtain a confidence interval of [($127.10-2.58 x 28.12) x 0.045,($127.10+2.58 x 28.12) x 0.045] = [$2.45, $8.98]. We can also use this to test for statistical significance. Given that 0 does not lie within the 99% CI, we can conclude that the predicted $5.72 is statistically significantly different from 0 at the 1% level.
    - The final part regarding Obama's data analytics team: it would be more useful to present the predicted changes in donations for a one-standard deviation change in share_under25 than a one-unit change in share_under25 because the former is a "standard" change in share_under25 in the data, whereas the latter is an extreme 0 to 1 change from the theoretical min to the max in the data, which is virtually impossible in reality.

8. Submitted R code should be similarly organised and commented as the solution R code for full marks; see as1.R from Canvas.