New York University Tandon School of Engineering

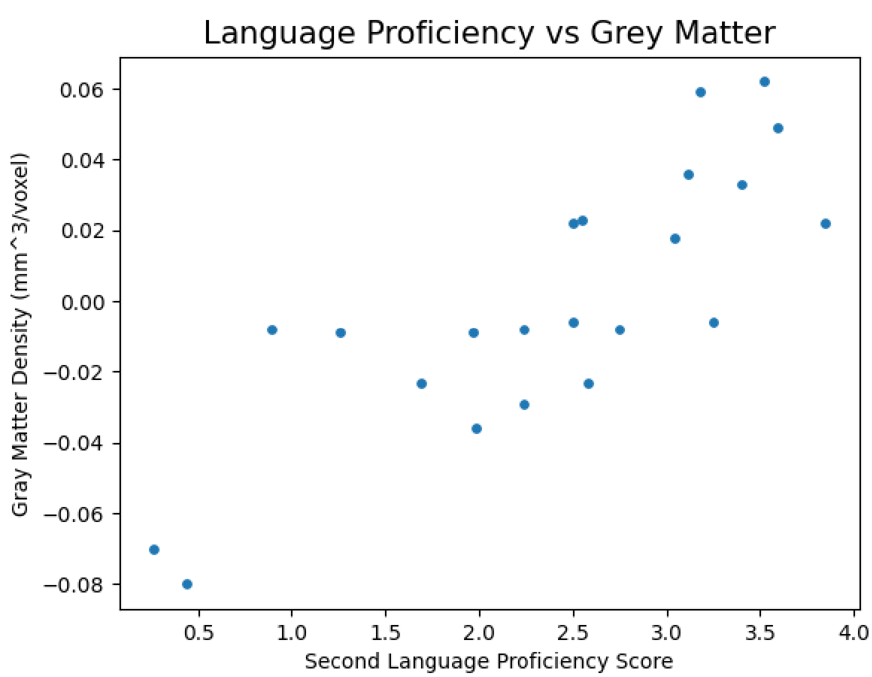
Biomedical Engineering
Applied Mathematics and Statistics for Biomedical Engineering

Fall 2021
Professor Mirella Altoe
Tuesday 5:00-7:30PM Rogers Hall 325

Computer Lab Assignment #10

QUESTION 1: Does learning a second language change brain structure? Mechelli et al. (2004) tested 22 native Italian speakers who had learned English as a second language. Proficiencies in reading, writing, and speech were assessed using a number of tests whose results were summarized by a proficiency score. Gray-matter density was measured in the left inferior parietal region of the brain using a neuroimaging technique, as mm 3 of gray matter per voxel. (A voxel is a picture element, or "pixel," in three dimensions.) The data are listed in the accompanying table. **Data available online.**

A. Display the association between the two variables in a scatter plot.



B. Calculate the correlation between second language proficiency and gray-matter density.

	proficiency	greymatter
proficiency	1.000000	0.818313
greymatter	0.818313	1.000000

The correlation between second language proficiency and gray-matter density is 0.818.

C. Test the null hypothesis of zero correlation.

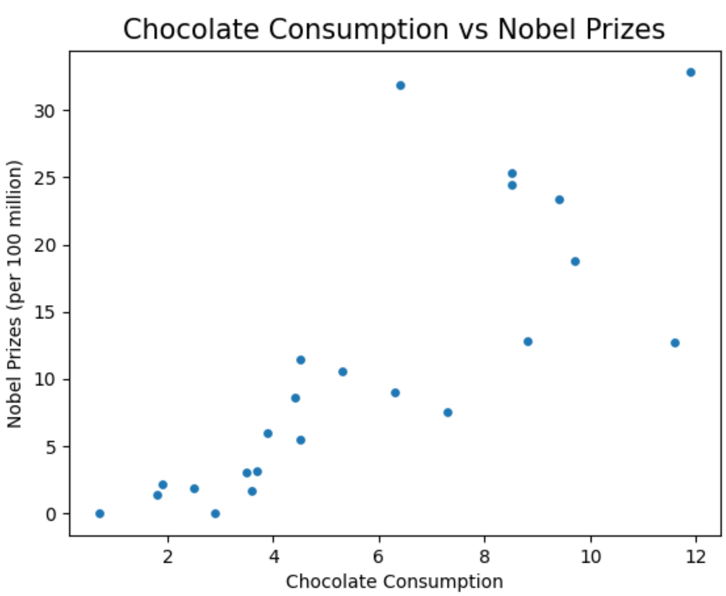
$$t = \frac{r-0}{SE_r} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.818}{\sqrt{\frac{1-0.818^2}{22-2}}} = \boxed{6.35} \quad \cancel{6.67}$$

$$t_{(0.05)(2)(20)} = \underline{2.086}$$

Our test statistic of ~~6.67~~ 6.35 is greater than the critical value of 2.086. Thus, we reject H_0 and conclude that Grey matter increases w/ second language proficiency

QUESTION 2: There is evidence that higher consumption of foods containing chemicals called flavonols—including cocoa, red wine, green tea, and some fruits—increases brain function in several ways. Messerli (2012) asked whether chocolate consumption in a country is correlated with the number of Nobel Prizes for the country over all time. The data are below. Both chocolate consumption and number of Nobel Prizes are scaled to the number of people in each country. **Data available online.**

A. Plot and examine these data. What challenges do you anticipate if your goal is to test whether chocolate consumption and number of Nobel Prizes are correlated?



B. Without transforming the data, test for an association between the two variables using an appropriate method.

```
Spearman's P: 0.9003708556969835
P value: 4.8729050551004556e-09
```

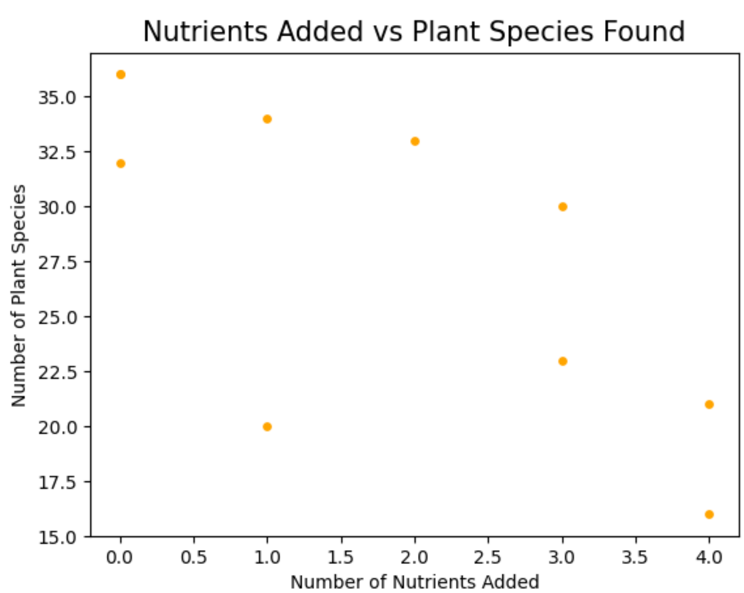
P value is significantly lower than 0.05

C. Interpret the findings of the study appropriately. Does chocolate consumption increase the probability of winning a Nobel Prize? Should it be recommended as a national priority, based on these data?

Since the p value is less than 0.05, we can reject H_0 and accept the alternate hypothesis of chocolate consumption increases Nobel Prize probability. However, this is probably not the case should not be recommended as a national priority. We should not test for arbitrary associations such as this.

QUESTION 3: You might think that increasing the nutrient resources available would elevate the number of plant species that an area could support, but the evidence suggests otherwise. The data in the accompanying table are from the Park Grass Experiment at Rothamsted Experimental Station in the U.K., where grassland field plots have been fertilized annually for the past 150 years (collated by Harpole and Tilman 2007). The number of plant species recorded in 10 plots is given in response to the number of different nutrient types added in the fertilizer treatment (nutrient types include nitrogen, phosphorus, potassium, and so on). Data available online.

A. Draw a scatter plot of these data. Which variable should be the explanatory variable (X) , and which should be the response variable (Y)?



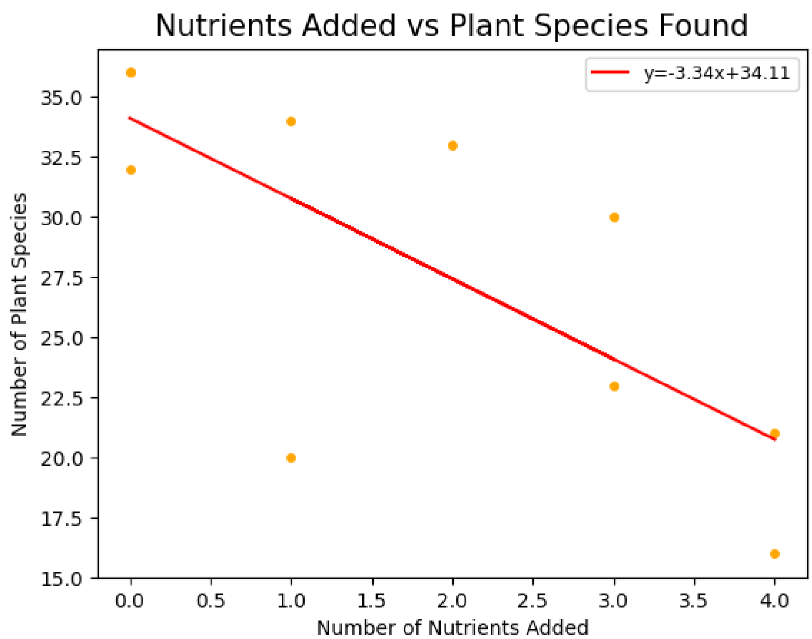
The explanatory variable would be nutrients, and the response variable would be number of plant species found.

B. What is the rate of change in the number of plant species supported per nutrient type added? Provide a standard error for your estimate.

The expected change per nutrient type added is about -3.34.

```
Standard Error:  1.0984101945414404
```

C. Add the least squares regression line to your scatter plot. What fraction of the variation in the number of plant species is "explained" by the number of nutrients added?



The number of nutrients added accounts for ~53.60% of the variation in number of plant species.

D. Test the null hypothesis of no treatment effect on the number of plant species.

	nutrients	species
nutrients	1.000000	-0.732106
species	-0.732106	1.000000

$$t = \frac{r - 0}{SE_r} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

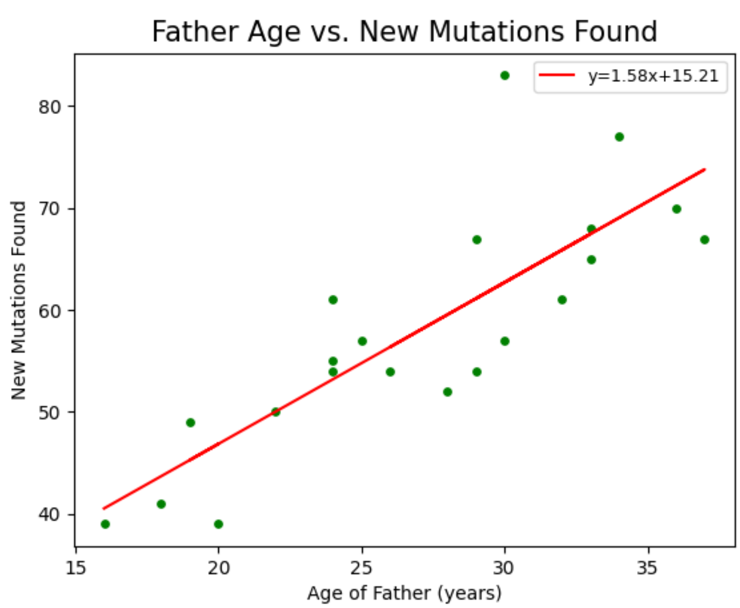
$$= \frac{-0.732}{\sqrt{\frac{1+0.732}{10-2}}}$$

$$= \boxed{|-1.57|} \qquad t_{(0.05)(2)(8)} = 2.306$$

Our test statistic of 1.57 is less than the critical value of 2.306. Thus, we fail to reject H_0, number of nutrients added has no effect on number of plants found.

QUESTION 4: Dads transmit many more new mutations than do mothers to their babies at conception. These mutations occur from copying errors during sperm production. There is increasing interest in the effect of father age on this process. As part of a larger study into the genetics of mental illness, Kong et al. (2012) used complete-genome sequencing of 21 father-child pairs to tally the total number of new mutations inherited from each father (in this particular sample, all the offspring were afflicted with schizophrenia). These counts are listed in the following table along with fathers' ages at offspring conception. <u>Data available</u> <u>online.</u>

A. Graph the relationship between number of new mutations (Y) and father's age (X) . Add the regression line to your plot.



B. Based on these data, how rapidly does the number of new mutations increase with father's age? Provide a standard error for your estimate.

New mutations increase with father's age by a factor of ~1.58.

```
Standard Error:  0.2542534489600477
```

C. What is the predicted mean number of new mutations from fathers 36 years of age? How does this compare with the predicted number for fathers only 18 years old?

$y = 1.58x + 15.25$

Case 1: 36 yrs	Case 2: 18 yrs
$1.58(36) + 15.25$	$1.58(18) + 15.21$
$= \boxed{72.09 \text{ mutations}}$	$= \boxed{43.65 \text{ mutations}}$

D. What fraction of the variation among fathers in the number of new mutations is explained by father's age?

The age of the father accounts for ~67.1% of the variation in new mutations found.