STAT 614 - HW 2_Yunting

Due: Thursday, September 17, 2020, in Blackboard (go to the Homework folder under the Homework/Classwork content area) by 11:59 pm.

Instructions: Please type your solutions to these FOUR problems and upload the document as a pdf file in Blackboard. There is only one file to submit for this assignment.

Notes: This homework continues our discussions of study design and exploratory data analyses and reviews some probability concepts.

Problem 1

Vegetarians have fewer heart attacks than non-vegetarians of the same age and sex. One reason is that vegetarians have (on average) lower blood pressure. But does a vegetarian diet actually cause lower blood pressure? The following studies were used to explain the difference in blood pressure between vegetarians and non-vegetarians.

> **Study 1:** Surveys about diet and health were sent to a random sample of 500 adults in a northwestern city. Of the 233 respondents, those who chose to be vegetarians tended to have lower blood pressure than non-vegetarians.
> **Study 2:** A sample of 200 adults who are not vegetarians were recruited by newspaper advertisement to participate in a study. Blood pressure was measured in all subjects at the start of the study. Half were randomly assigned to eat a vegetarian diet and the other half were assigned to their regular diet. After a year, blood pressure was measured again in each subject and the vegetarian group had, on average, a larger decrease in their blood pressure.
> **Study 3:** Trappist monks are strict vegetarians, while Benedictine monks follow a more standard Western diet. A study of these two groups revealed that Trappist monks have (on average) lower blood pressure than Benedictine monks.

Answer the following questions about the above studies:

(a) Which of these studies, if any, is an example of a randomized experiment?
Ans: Study 1 and Study 2

(b) In which of these studies, if any, were subjects randomly chosen to participate?
Ans: Study1

(c) Which study provides the strongest evidence that a vegetarian diet causes lower blood pressure? Explain your reasoning.
Ans: Study2, because before these 200 adults participate in this study, they are all non-vegetarians, it means they have a similar condition for their dietary habits. After one year, we definitely understand the vegetarian group has a larger decrease in their blood pressure, so only changed their dietary habit that remarkably sees the effect of blood pressure just in one year, which has powerful evidence for this experiment.

(d) For the two remaining studies (that you did not select for part c), pick one and briefly describe why you did not select it in part c.
For me, I would like to explain why study 3 cannot persuade this problem. Firstly, we cannot make sure 100% of Trappist monks are eating vegetables only and 100 % of Benedictine monks eating meals without vegetables every day. Secondly, we cannot know how many people in each group? Does the number of group people is equal? Therefore, Study 3 does not provide the strongest evidence for that.

Problem 2

From *The Statistical Sleuth, Third Edition*, Chapter 1, problems 19 & 20. Please include a description of the method you used for problem 20.
Note: If you don't have a copy of the textbook, go to the Information area in Blackboard where the first two chapters are posted (as a very rough pdf).

19. Write down the names and ages of 10 people. Using coin flips, divide them into two groups, as if for a randomized experiment.

- Did one group tend to get many of the older subjects?
  No, this method cannot filter out an elder one or a young one, as it is randomly selected. I will record the table below question 20.

- Was there any way to predict which group would have a higher average age in advance of the coin flips?
  The traditional way is we need to investigate their age before that, but we don't know which one will be going to which group via coin flips. Hence, we couldn't precisely predict which group has a higher average age.

20. Repeat Exercise 19 using a randomization mechanism that ensures that each group will end up with exactly five people.

- We start to use coin flips, divide the names into two groups. If a coin flips show is head, we can be named H to group 1 and if a coin flips show is tails, we can be named T to group 2 and my end up with HTTHHHHTTH.
- From the below, it can be observed that Group1 and Group 2 have five people each, so we cannot go ahead with coin flips.

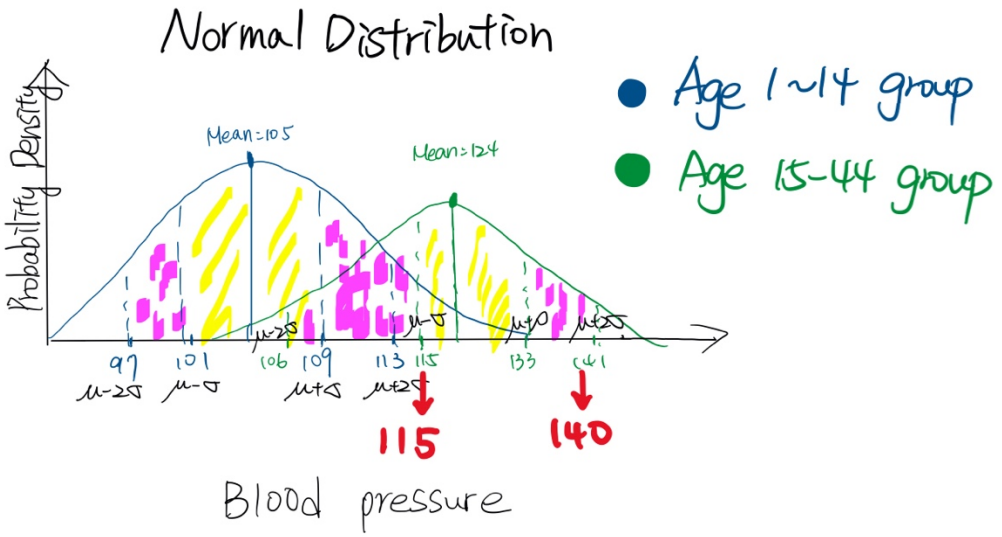
Name	Age	Group	Coin (H = head, T = tail)
Amy	23	1	H
Bob	40	2	T
Coco	69	2	T
David	52	1	H
Eric	11	1	H
Ford	17	1	H
Gary	27	1	H
Helen	34	2	T
Ivy	45	2	T
Jack	73	1	H

Problem 3

People are classified as hypertensive if their systolic blood pressure is higher than a specified level for their age group, according to the last column of the following table. Assume systolic blood pressure is *Normally distributed* with mean and standard deviation given in the table for the age groups 1-14 and 15-44, respectively.

Age group	Mean	St. Dev	Level
1-14	105.0	4.0	115.0
15-44	124.0	9.0	140.0

(a) Sketch a graph of the distribution of systolic blood pressure for both groups (on the same plot). Label the x- and y-axes, the mean, the standard deviation, and indicate the hypertensive cut-off for each group. You may take a picture of your graph and copy-and-paste it into your homework to include it with your solution. DO NOT submit a separate image file! Paste it into your homework document (and I recommend saving your homework document as a pdf to make sure it is readable in Blackboard).



- Z-score: $z = (x - \mu) / \sigma$

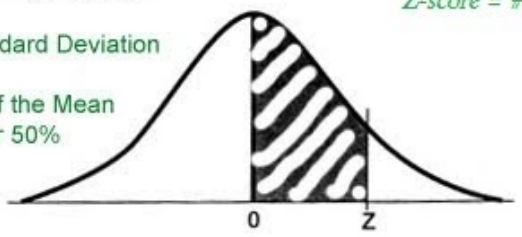
Z-Score formula $z = (X - \mu) / \sigma$

Z = (X-Mean)/Standard Deviation

Note: Each side of the Mean yeilds .50 or 50%

a Z of 3.9 = 50%

Z-score = # of Standard Deviations from the Mean

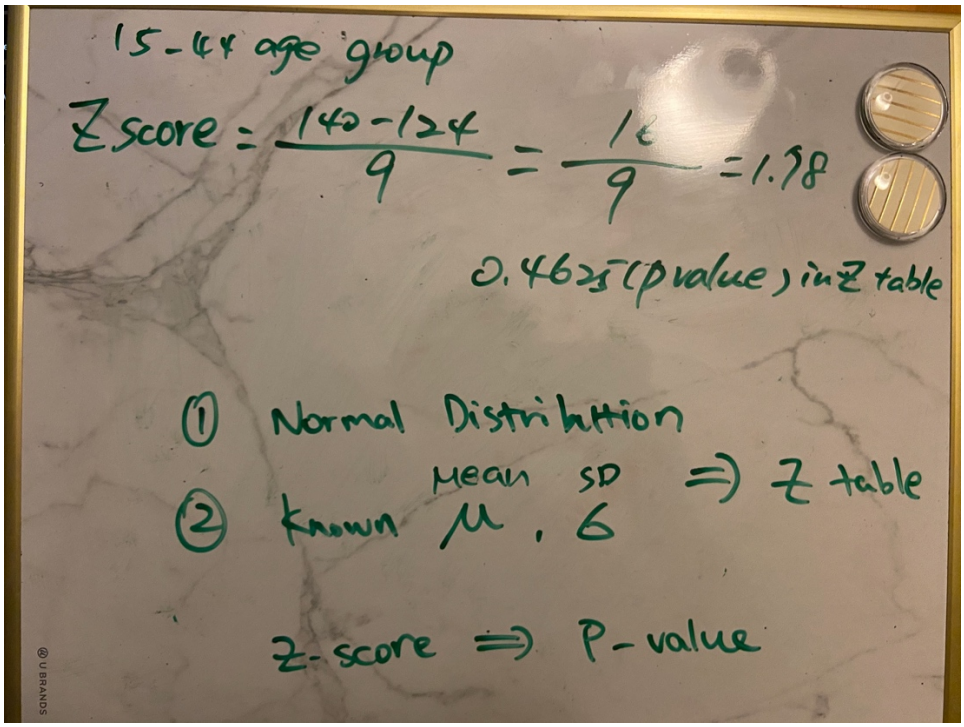


Z score example z=1.96 the shaded area is .475

Areas Under the Standard Normal Curve

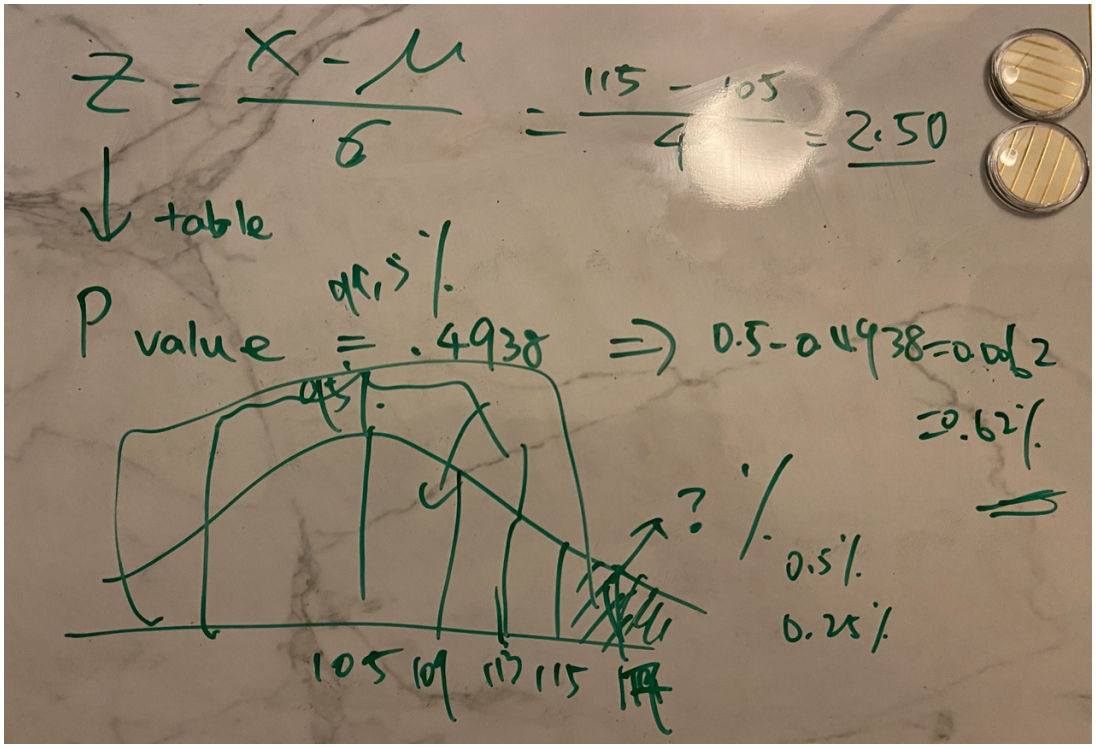
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000									

Why we use the Z table?



(b) What proportion of 1- to 14-year-olds are hypertensive?

Ans: 0.62 %



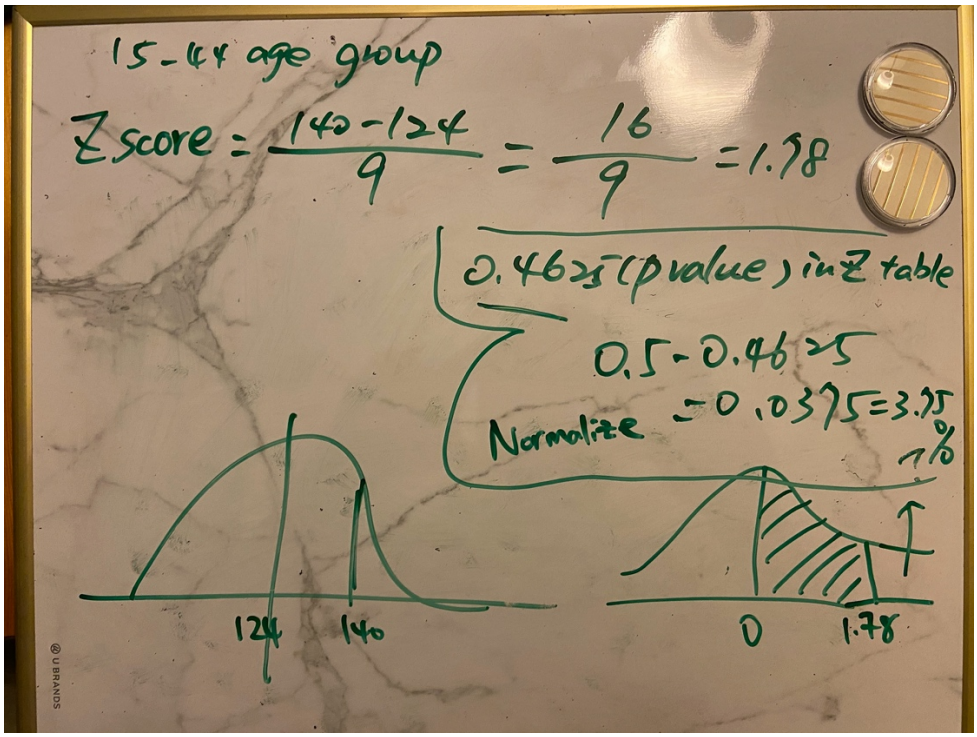
R studio:

```r
x <- 1- pnorm(115, 105, 4)
x
```

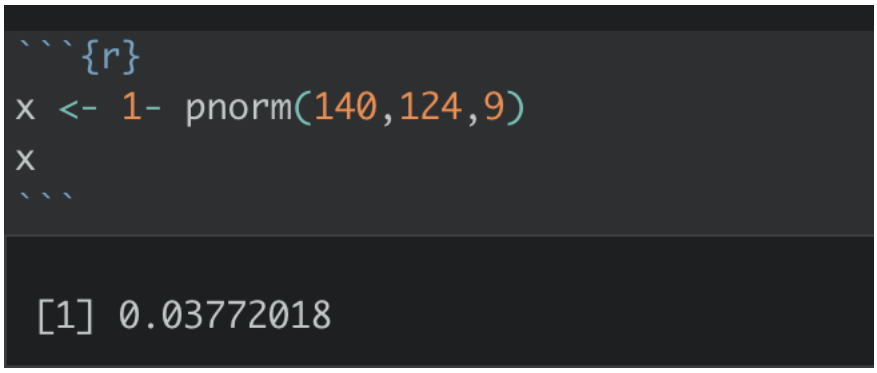
```
[1] 0.006209665
```

(c) What proportion of 15- to 44-year-olds are hypertensive?

Ans: 3.75 %



R studio:

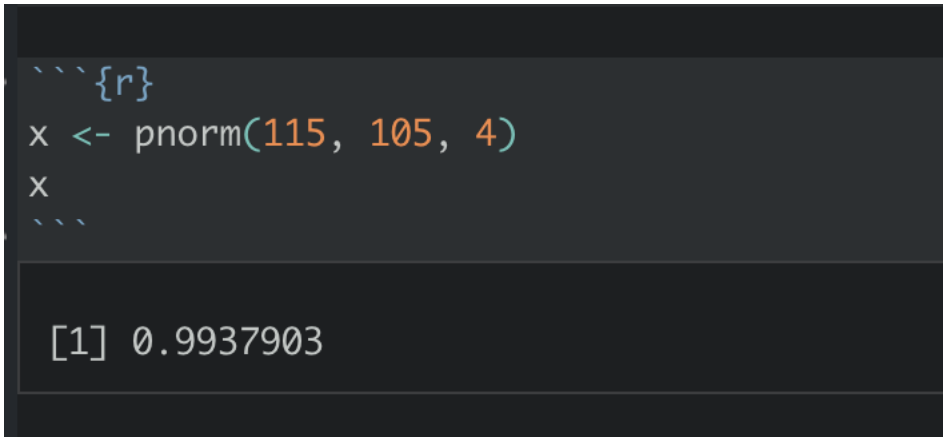


(d) What is the probability that a randomly select 1-14-year-old is not hypertensive?

1- 0.62% = 99.38 %

Ans: 99.38 %

R studio:

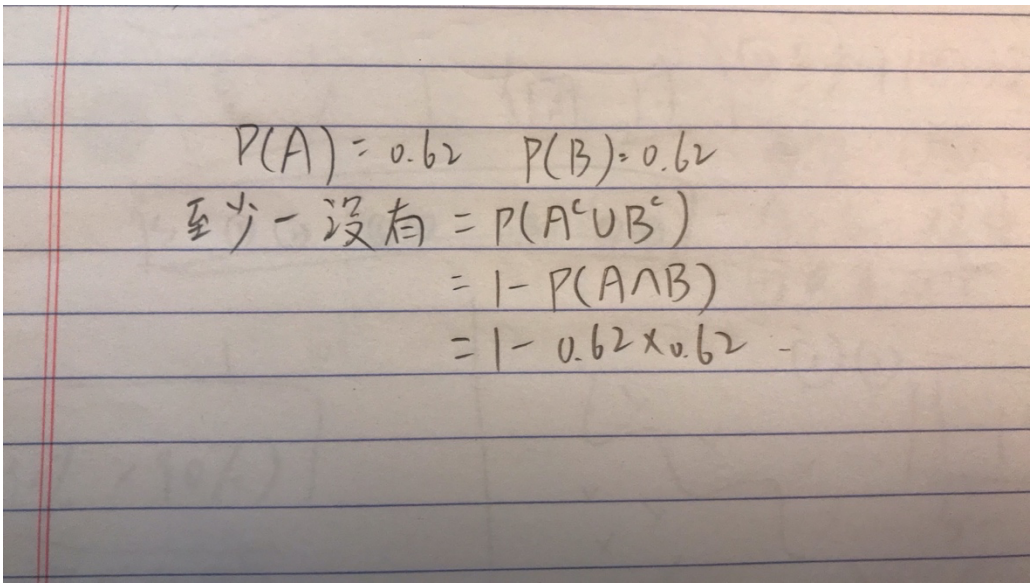


(e) Suppose two people are selected at random from 1-14-year-old age group. What is the probability that at least one of them is not hypertensive?

"At least one" is equivalent to "one or more." So**, P (at least one) = 1 – P(none).**

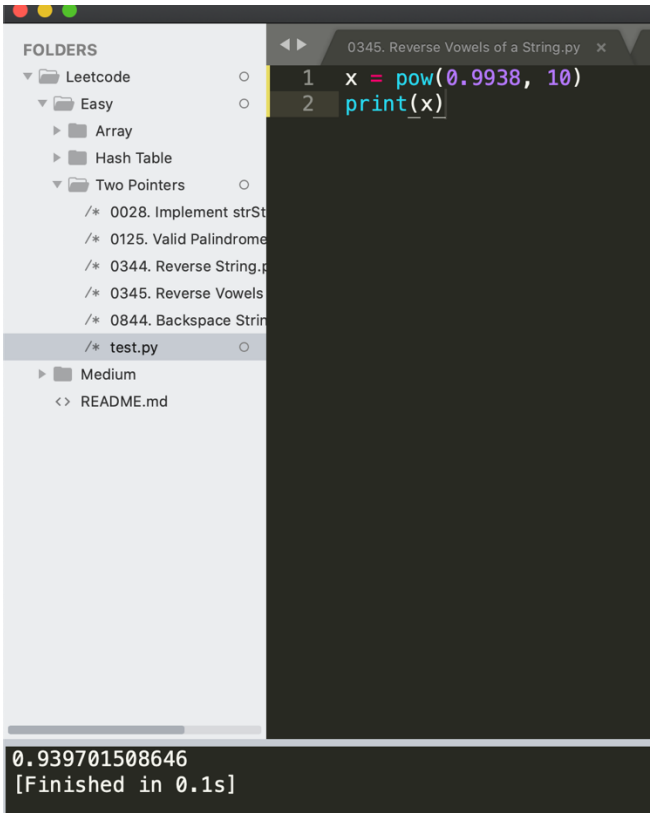
Notice that (1-.62)*(1-.62) = probability that *exactly* two are not hypertensive. At least one not hypertensive includes this but it also includes the cases when one is hypertensive and one is not hypertensive.

- P(at least one) =1- (0.62%*0.62%) = 99.99 %



(f) Suppose ten people are selected at random from 1-14-year-old age group. What is the probability that all of them is not hypertensive?

Ans: 93.97%



Problem 4

Find **one** scholarly article which uses **one** of a normal, Poisson, binomial, or any other distribution of interest to you.

Give the following information from your article:

(a) The full citation of the article (authors, title of article, journal name, year, volume, page #s).

(b) A brief summary of the purpose of the article based on the information in the abstract. You can quote the article but then also put the purpose in your own words.

(c) A brief summary of how the distribution is used in the article. You do not need to go into a lot of detail here. Just try to get a sense of what they are using the distribution for.

(d) How you found the article (for example, through a PubMed or Google Scholar search).

(e) Why you picked that article.

Article of talking about Normal Distribution:

a. Guest, A. M., & McRee, N. (2009). A School-Level Analysis of Adolescent Extracurricular Activity, Delinquency, and Depression: The Importance of Situational Context. *Journal of Youth and Adolescence*, 38(1), 51-62. http://dx.doi.org.proxyau.wrlc.org/10.1007/s10964-008-9279-6

b. I'll quote the article: "In this article they investigate the extent to which the relationship between extracurricular activities and youth development depends on situational contexts. Using a national sample including 13,466 youths in grades 7-12 across 120 schools, they conduct school-level analyses of the association between extracurricular activities, delinquency, and depression." In this case study, they want to explore whether extracurricular activities have a relationship with youth behavior and development.

c. The author indicates that "When looking specifically at the school level, different types of extracurricular activity participation seem to associate with different rates of delinquency and depression in a similar pattern—a near-normal distribution." Extracurricular activities did not significantly affect the behavioral development of adolescents, but the level of the school did affect them in positive emotions or cause them depression.

d. I found this article from the AU database.

e. Because I am a student now, so my motivation is I want to find some topic of campus life with the normal distribution. This kind of topic is closer to my current position. Also, I am enjoying being a student now.

Reference

Guest, A. M., & McRee, N. (2009). A School-Level Analysis of Adolescent Extracurricular Activity, Delinquency, and Depression: The Importance of Situational Context. *Journal of Youth and Adolescence*, 38(1), 51-62. http://dx.doi.org.proxyau.wrlc.org/10.1007/s10964-008-9279-6

Statistics - JilMac-Math. (n.d.). Retrieved September 16, 2020, from https://sites.google.com/site/jilmacmath/statistics