INFO 7390
Advances in Data Sciences and Architecture
Assignment 2

Student Name: Zihan Zhang
Professor: Nik Bear Brown

Due: Sunday May 27, 2018

Q1 (5 Points) Given a normal distribution with a mean to 33, a standard deviation of 11, and the sample size to 100. What is the probability of finding a value:

a. less than 11 (2 points)

b. greater than 55 (2 points)

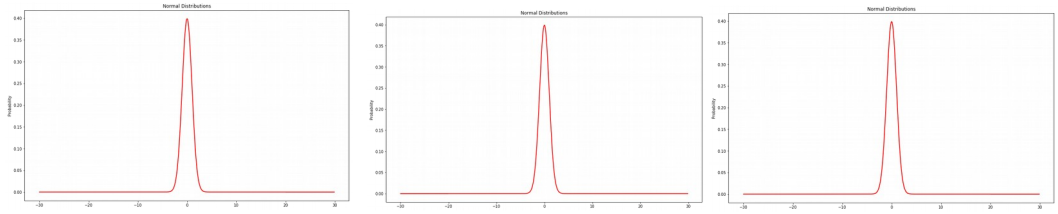
c. less than 11 or greater than 55 (1 point)

Show the calculation as done by hand.

$$z=(x-33)/(11/\sqrt{100})$$

x < 11 => z < -20 => p = 2.7536241186061556e-89
x > 55 => z > 20 => p = 2.7536241186061556e-89
11 > x || x> 55 => -20 > z || z > 20 => p = 5.507248237212311e-89

Q2 (5 Points) Write python code to plot Q1 and calculate Q1.



Calculation is in Assignment2.py

Q3 (5 Points) Given a normal distribution with a mean to 33, a standard deviation of 11, and the sample size to 1000. What is the probability of finding a value:

a. less than 11 (2 points)

b. greater than 55 (2 points)

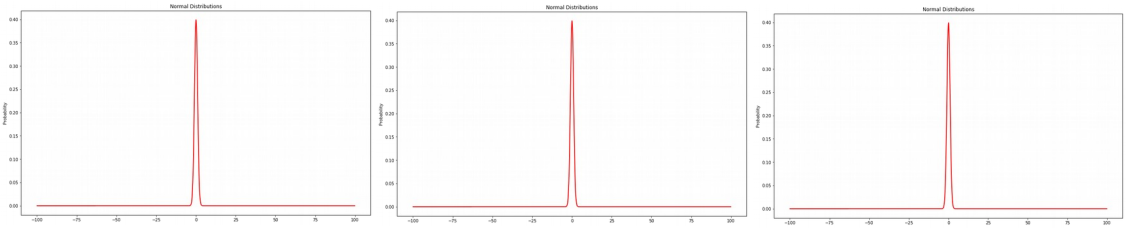
c. less than 11 or greater than 55 (1 point)

Show the calculation as done by hand.

$$z=(x-33)/(11/\sqrt{1000})$$

x < 11 => z < -63.24 => p = ?
x > 55 => z > 63.24 => p = ?
11 > x || x > 55 => -63.24 > z || z > 63.24 => p = ? * 2
p is too small that python cannot show it

Q4 (5 Points) Write python code to plot Q3 and calculate Q3.



Calculation is in Assignment2.py

Q5 (5 Points)
The one-year rate of return to shareholders was calculated in a sample of 55 tech stocks. The data, is below and in the file tech_stocks.csv.

```
[23.72502353842273, 21.62401646603374, -0.7463274288122257,
1.7178830450828002, -2.634776050958738, -2.792138753758266,
-10.514395560878746, 8.720529920419578, 18.782813772780308,
5.825456165455785, 11.172228117978728, 11.97032962928146,
-30.981624884074883, 8.428109006257554, 13.715597227579686,
-7.14438096845215, 35.38150590002323, 5.951675701660346,
-2.128337264991565, 12.952160066221724, -9.52841782146271,
9.27768703224383, -10.489029625059331, 1.7170477394203232,
```

```
11.717280979491225, 18.84977052950971, 12.645227894971965,
-2.444524930791145, -4.870684454119193, 9.384408019477661,
13.450953108385315, 23.714466213916317, 5.7140681189301255,
-14.73667486810843, 6.455693762385872, 9.715370033540502,
11.133859293104898, 5.12584305942378, -3.6547977197096486,
15.65791149754521, 17.045514919166266, 20.86418259486488,
28.498593533062984, 15.689734619702122, 7.954721816163218,
-3.113512775937407, 12.86046371264133, 2.467429173851536,
-2.682786932363779, -1.9362359856511269, 5.912048015521583,
24.003261208189425, 9.7084789611135, -6.91532401310932, 21.426117689357]
```

A. (1 point) Specify the null and alternative hypotheses tested for determining whether the true mean one-year rate of return for tech stocks exceeded 10%.

null hypotheses: mu >= 10
alternative hypotheses: mu < 10

B (3 points) Calculate the observed significance level of the test.

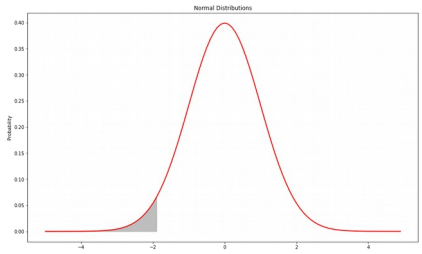
mu = 6.975336218998899
sigma = 11.903902235771346
z = (miu - 10) / (sigma / 7.416198487095663) = -1.884382659765678
According to z value table, probability is 0.030054
So p-value is 0.030054

C. (1 point) Interpret the result.

If alpha > p(0.05 for example) then the observed result falls in rejection region, so we reject null hypothesis
else if alpha < p(0.01 for example) then the observed result falls in accept region, so we accept the null hypothesis

Show the calculation as done by hand.

Q6 (5 Points) Write python code to plot Q5 and calculate Q5.



Calculation is in Assignment2.py

Q7 (5 Points) A company has placed an order for 5,000 laptops with a supplier on the condition that no more than 1% of the devices will be defective. To check the shipment, the company tests a random sample of 100 laptops and finds that 2 are defective.

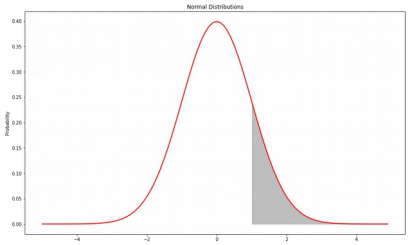
You can choose a population variance.

Does this provide sufficient evidence to indicate that the proportion of defective can laptops in the shipment exceeds 1%? Explicitly state your null and alternative hypothesis.

Null hypothesis: p0 >= 1%
alternative hypothesis: p0< 1%

Q8 (5 Points) Write python code to plot Q7 and conduct a hypothesis test on Q7.

$$z=(0.02-0.01)/\sqrt{0.01\times(1-0.01)/100} = 1.01$$



Suppose alpha is 0.05, zalpha is 1.6448536269514729 > 1.01 which means H0 is not being rejected

Calculation is in Assignment2.py

Q9 (5 Points) An ultra-marathon runner ran 103 miles per week as reported by runners world. A random sample of 500 ultra-marathon runners had a mean of 101 miles per week ran when asked.

Let m denote mean distance for all ultra-marathon runners.

A (3 Points). Perform the hypothesis test

Ho: m=103 miles per week ran

Ha: m ≠103 miles per week ran

at the 5% significance level. Assume the standard deviation is 60 miles.

This is a two-sides test, so alpha is 0.025 each side

$z=(101-103)/(60/\sqrt{500})$ = -0.7453559924999299

-0.7453559924999299 > -1.9599639845400545 which means H0 is not rejected

[ ]:

B (2 Points). Find a 95% confidence interval for m.

$x \pm 1.96 \times 60/\sqrt{500}$

$x \pm 5.259231883079505$

$101 \pm 5.259231883079505$

Show the calculation as done by hand.

Q10 (5 Points) Write python code to plot Q9 and conduct a hypothesis test on Q9.