American University
The Department of Computer Science
Spring 2021

Introduction to Simulation and Modeling, CSC 432/632

Exam 2
(Total of 20 points)

Instructions: Copy and paste the following page to the top of a word file and after that use as many pages as required to write your answers. Save the file as "lastname_Exam 2" and Use WORD or PDF documents to submit your answers in the Blackboard. There is no minimum and maximum number of words for each question, but your answers need to be comprehensive and address the question. Please submit additional .py file and provide adequate comments within the code you write.

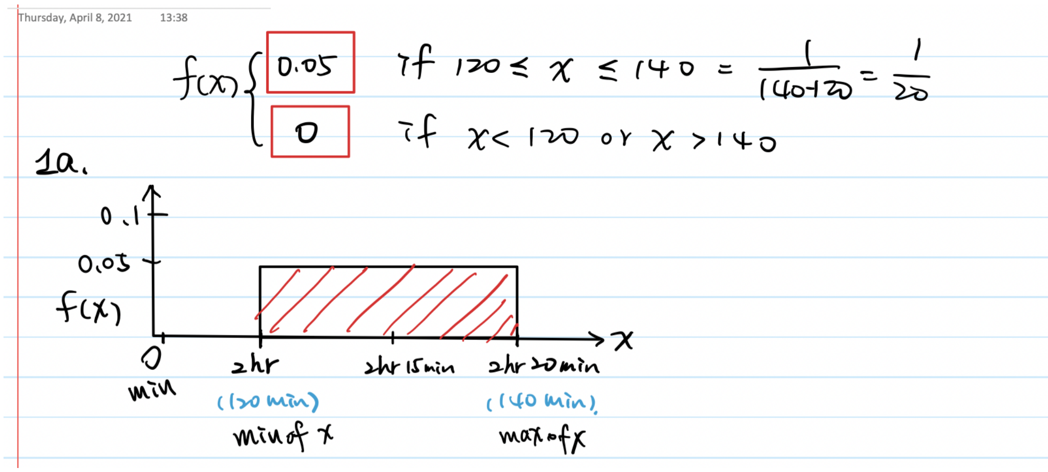
*"**Pledge:** I understand that this exam is an individual task and I have done it by myself and have not received any extra help including coding, writing the answers, or any other general advises. I have also not used the Internet resources, and a in rare situation, if I needed to do so, I have provided the Internet resources at the end of the question. I understand that this exam is closed-book and closed-note and I have not used these resources in any format. Signature: (Write your complete name as your signature) Yu Tang Chi" (lack of pledge signature results in a 0.5 grade deduction and a follow up email to confirm that the pledge is valid).*

Question 1 (3 points)
RUBRIC: (-Write down all steps of the answer. Write down your final answers and explain as needed. Lack of each item results in 0.5-to-1-point deduction)

Delta Airlines quotes a flight time of 2 hours, 5 minutes for its flights from Cincinnati to Tampa. Suppose we believe that actual flight times are uniformly distributed between 2 hours and 2 hours, 20 minutes.

a) Show the graph of the probability density function for flight time.

The actual fight times x follow a uniform distribution, which is between 120 minutes to 140 minutes.



b) What is the probability that the flight will be no more than 5 minutes late?

P(X <= x) = x-a/ b-a
P(X < 2hr 5mins + 5 mins) = P(X < 130) = 130 - 120 / 140 - 120 = 10 / 20 = .5
Ans: The probability that the flight will be no more than 5 minutes late is .5

c) What is the probability that the flight will be more than 10 minutes late?

P(X > x) = 1 - (x-a/ b-a)
P(X > 2hr 5mins + 10 mins) = P(X > 135) = 1-(135 - 120) / (140 -120)
= 1 - 15/20 = .25
Ans: the probability that the flight will be more than 10 minutes late is .25

d) What is the expected flight time?

Expected fight time = the mean of flight time

mu = (120+140) / 2 = 130
Ans: the expected flight time is 130 mins.

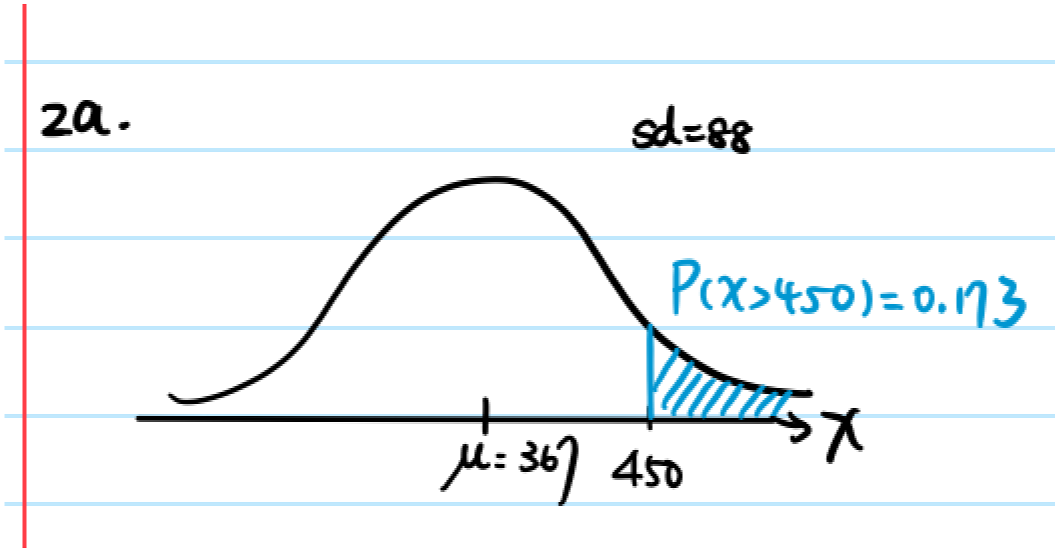
Question 2 (4 points)
RUBRIC: (-Write down all steps and formula. -Write down (or copy) the normal distribution formula in Excel or Python. Write down your final answers. For each item Sketch or draw the normal distribution and highlight the desired area. Lack of providing the graph for each question results in 0.5-1-points grade deduction).

Automobile repair costs continue to rise with an average 2015 cost of $367 per repair (U.S. News & World Report website). Assume that the cost for an automobile repair is normally distributed with a standard deviation of $88. Answer the following questions about the cost of automobile repairs.

I use Excel to do question 2, please see the **excel file** for detailed formulas. We know that the **mu = 367, sd = 88**, and the automobile repair costs follow a normal distribution.

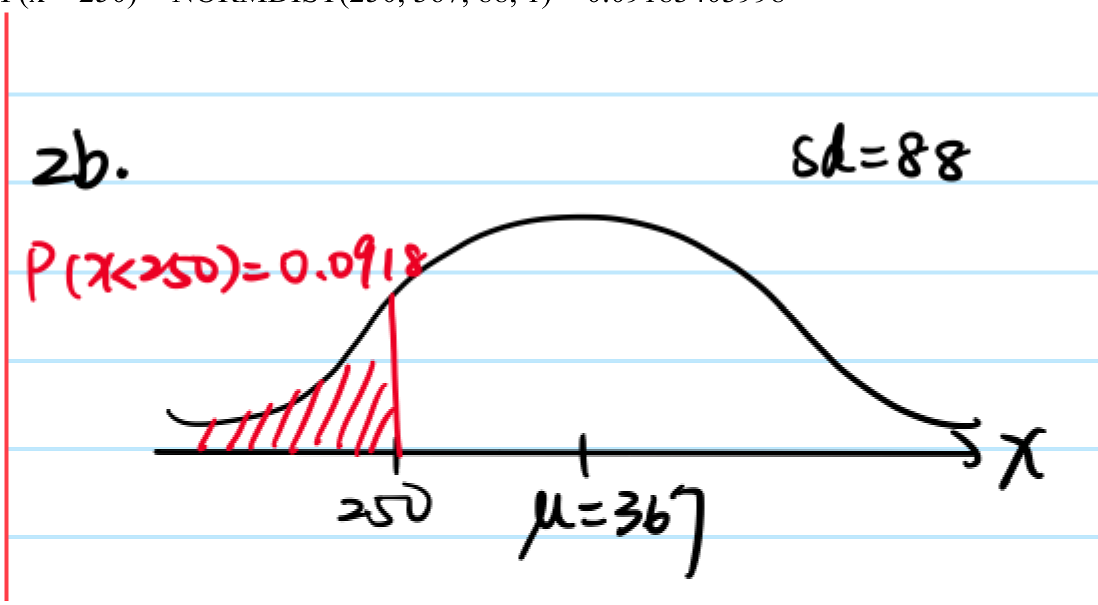
a) What is the probability that the cost will be more than \$450?

P(x > 450) = 1- P(x<=450) = 1 - NORMDIST(450, 367, 88, 1) = 0.1727939558



b) What is the probability that the cost will be less than \$250?

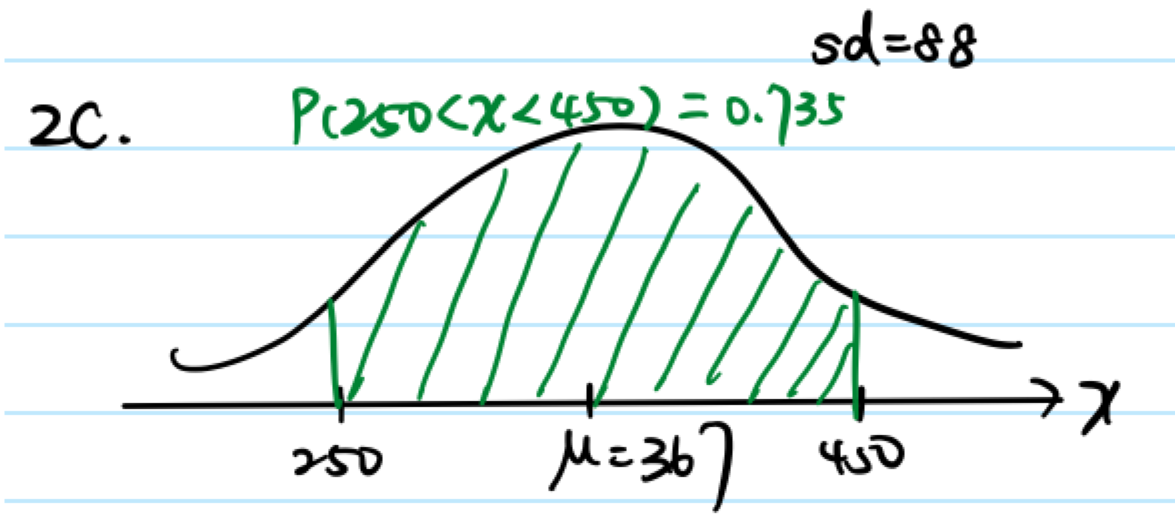
P(x < 250) = NORMDIST(250, 367, 88, 1) = 0.09183403998



c) What is the probability that the cost will be between \$250 and \$450?

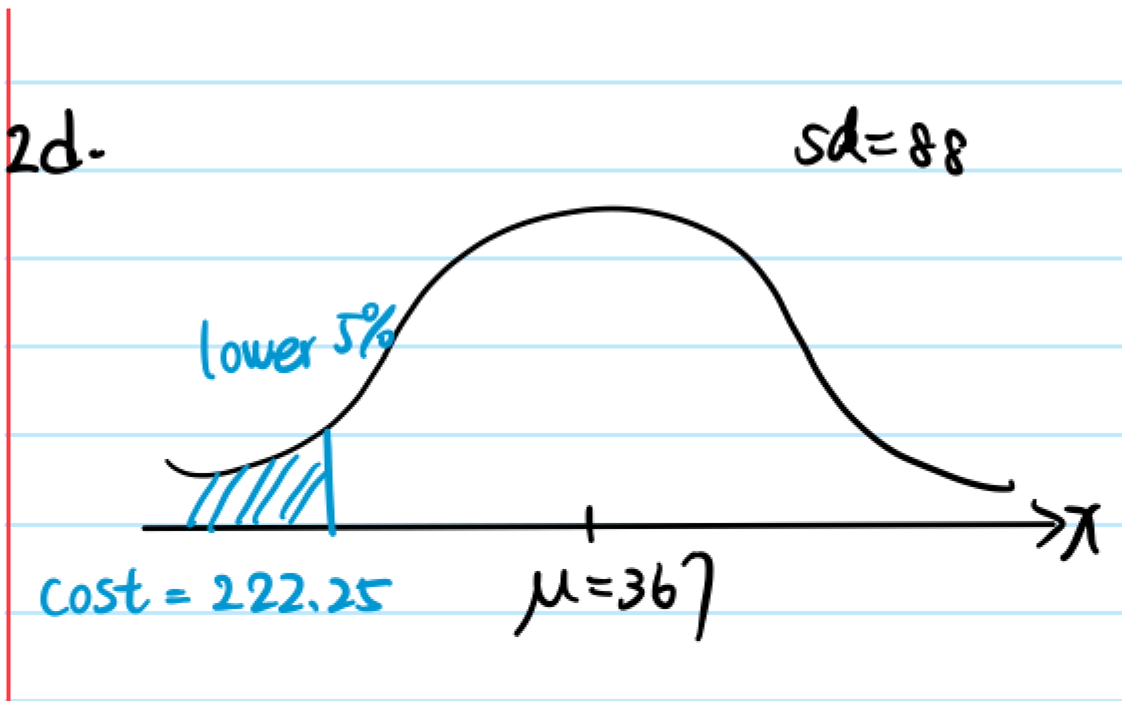
P (250 < x < 450) = P(x < 450) - P(x < 250)
= NORMDIST(450, 367, 88, 1) - NORMDIST(250, 367, 88, 1)

= 0.7353720042



d) If the cost for your car repair is in the lower 5% of automobile repair charges, what is your cost?

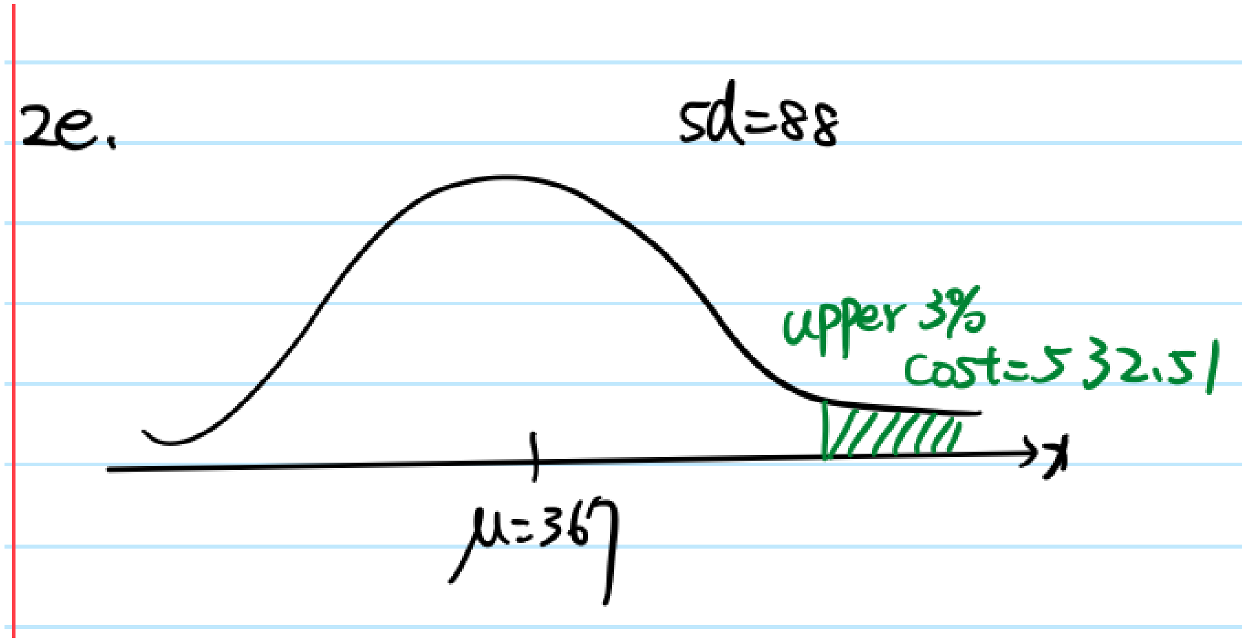
P(x < cost) = 0.05, cost = ?
cost =NORMINV(0.05, 367, 88) = 222.252881



e) If the cost for your car repair is in the upper 3% of automobile repair charges, what is your cost?

1 - upper 3% = lower 97 % (probability)

P(x < cost) = 0.97, cost = ?
cost =NORMINV(0.97, 367, 88) = 532.5098373



Question 3 (3 points)

RUBRIC: (-Write down all steps and formula. -Write down (or copy) the normal distribution formula in Excel or Python. -Write down your final answers. -For each item Sketch or draw the normal distribution and highlight the desired area. Lack of providing the graph for each question results in 0.5-1-points grade deduction).

Sales personnel for Skillings Distributors submit weekly reports listing the customer contacts made during the week. A sample of 65 weekly reports showed a sample mean of 19.5 customer contacts per week. The standard deviation was 5.2.

n = 65, $\bar{x}$ = 19.5, sd = 5.2
According to the t-table, we know that if the degrees of freedom is 64, the t-score is 1.671 when alpha = .1, and the t-score is 2 when alpha = .05. Also, the confidence intervals are always two tailed. I use t-table to calculate part a.
T-table: https://www.statisticshowto.com/tables/t-distribution-table/

a) Provide 90% and 95% confidence intervals for the population mean number of weekly customer contacts for the sales personnel.

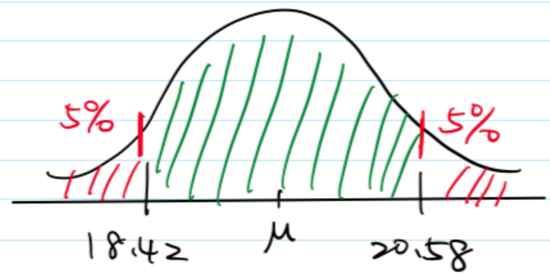
Formula: $x \pm t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}}$

90%CI $= 19.5 \pm t_{0.1/2,64} \cdot \frac{5.2}{\sqrt{65}}$

$= 19.5 \pm 1.671 \times \frac{5.2}{\sqrt{65}}$

$= 19.5 \pm 1.077763$

$= [18.42224, 20.57776]$

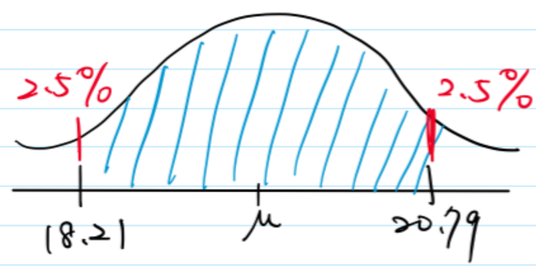


95%CI: $= 19.5 \pm t_{0.05/2,64} \cdot \frac{5.2}{\sqrt{65}}$

$= 19.5 \pm 2 \cdot \frac{5.2}{\sqrt{65}}$

$= 19.5 \pm 1.289961$

$= [18.21004, 20.78996]$.



b) Provide the conclusion that you can make based on this interval estimation.
Based on the result, we have 90 % confidence that the population mean of weekly customer contacts is between 18.42 to 20.58. Also, we have 95 % confidence to say that the population mean of weekly customer contacts is between 18.21 to 20.79.

c) Repeat Part a) using a Python code. Provide the code file in a .py file and provide comments within the code. For this problem in python use a t-test (as described in the class).

```python
# for small samples (<50) we use t-statistics
xbar = 19.5 # sample mean
sd = 5.2
n = 65 # sample size
df = n-1 # degrees of freedom

# define T-score function
def Tscore(ci, n):
  t = stats.t.ppf(1- ((100-ci)/2/100), n-1) # CI, df
  return t

# print(Tscore(90, 65))

# 90% CI
ci = 90
tscore = Tscore(ci, n)
lower = xbar - tscore * sd/math.sqrt(n) #lower level
upper = xbar + tscore * sd/math.sqrt(n) # upper level
print("The population mean is {} to {} with {}% confidence intervals".format(lower, upper, ci))

# 95% CI
ci = 95
tscore = Tscore(ci, n)
lower = xbar - tscore * sd/math.sqrt(n) #lower level
upper = xbar + tscore * sd/math.sqrt(n) # upper level
print("The population mean is {} to {} with {}% confidence intervals".format(lower, upper, ci))
```

```
The population mean is 18.423518944677575 to 20.576481055322425 with 90% confidence intervals
The population mean is 18.21150308968386 to 20.78849691031614 with 95% confidence intervals
```

Question 4 (4 points)
RUBRIC: (-Write down all steps and formula. -Write down (or copy) the normal distribution formula in Excel or Python. -Write down your final answers).

Many medical professionals believe that eating too much red meat increases the risk of heart disease and cancer. Suppose you would like to conduct a survey to determine the yearly consumption of beef by a typical American and want to use **3 pounds** as the desired margin of error for a confidence interval estimate of the **population mean** amount of beef consumed annually. Use **25 pounds** as a planning value for the population standard deviation and **recommend a sample size** for each of the following situations.

Note: margin of error = 3, sd = 25
Z-score: https://www.medcalc.org/manual/values_of_the_normal_distribution.php

- Margin of error: $E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$
- Necessary Sample Size : $n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2}$

a) A 90% confidence interval is desired for the mean amount of beef consumed.

4a. 90% CI : $\alpha = 1 - 0.9 = 0.1$

$\alpha/2 = 0.1/2 = 0.05$

$Z_{(0.05)} = 1.645$

$n = (1.645)^2 \cdot 25^2 / (3)^2 = 1691.2661/9 = 187.9184$

$\doteqdot 188$ #

b) A 95% confidence interval is desired for the mean amount of beef consumed.

4b. 95% CI : $\alpha = 1 - 0.95 = 0.05$

$\alpha/2 = 0.05/2 = 0.025$

$Z_{(0.025)} = 1.96$

$n = (1.96)^2 \cdot 25^2 / 3^2 = 2401/9 = 266.7778$

$\doteqdot 267$ #

c) A 99% confidence interval is desired for the mean amount of beef consumed.

4c. 99% CI: $\alpha = 1 - 0.99 = 0.01$

$$\alpha/2 = 0.01/2 = 0.005$$

$$Z(0.005) = 2.575$$

$$n = (2.575)^2 \cdot 25^2 / 3^2 = 4144.141 / 9 = 460.4601$$

$$\doteqdot 460 \#$$

d) For the above three parts Part a-c, write a Python code to calculate the sample size.

4d

For the above three parts Part 4a-4c, write a Python code to calculate the sample size.

```
# Z-score for each confidence interval
z90 = 1.645
z95 = 1.96
z99 = 2.575

margin = 3
sd = 25

# 4a
a = z90**2*sd**2/margin**2
print(round(a))

# 4b
b = z95**2*sd**2/margin**2
print(round(b))

# 4c
c = z99**2*sd**2/margin**2
print(round(c))


```

```
188
267
460
```

e) When the desired margin of error is set, what happens to the sample size as the confidence level is increased? Would you recommend using a 99% confidence interval in this case? Discuss.

As we can see from the result above, as the confidence level increases, the appropriate sample size increases in order to keep the margin of error constant. The 99 percent confidence interval is the most precise when compared to the 90 percent, 95 percent, and 99 percent confidence intervals. The 99% CI also recommends 460 sample sizes, which is not a large number, implying that we can consider using a 99 percent confidence interval to determine a typical American's yearly consumption of beef.

Question 5 (3 points)

RUBRIC: (-Write down all steps and formula. -Write down (or copy) the normal distribution formula in Excel or Python. -Write down your final answers. -For the hypothesis test provide a graph of critical values. Lack of providing the graph for each question results in 0.5-1-points grade deduction).

A shareholders' group, in lodging a protest, claimed that the mean tenure for a chief executive office (CEO) was **at least nine years**. A survey of companies reported in The Wall Street Journal found a **sample mean tenure of x = 7.27** years for CEOs with a standard deviation of **s = 6.38** years.

Based on the information above, we know that the sample mean xbar is 7.27, sd = 6.38, mu = 9

a) Formulate hypotheses that can be used to challenge the validity of the claim made by the shareholders' group.

Hypothesis testing: Ho: mu >= 9 years vs Ha: mu < 9 years. Also, this is a one-sided t-test.

b) Assume 85 companies were included in the sample. What is the p-value for your hypothesis test?

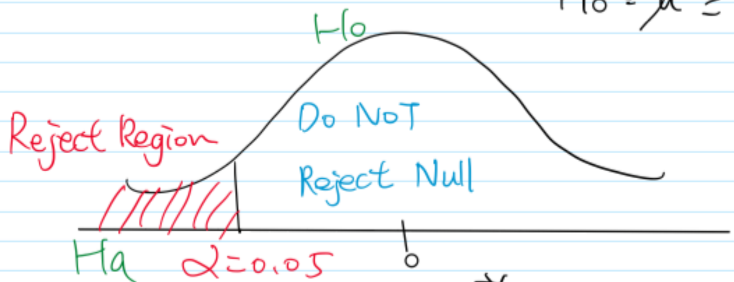
5b. sample size = n = 85 , p-value = ?

$$t^* = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{7.27-9}{6.38/\sqrt{85}} = \frac{-1.73}{0.692} = -2.5$$

critical value

Lower-tailed (95% CI)

$H_0 : \mu \geq 9$ vs H_a $\mu < 9$



$t = -1.96$ $\Rightarrow t^* = -2.5 < t$ ∴ reject Ho

Now we know, the t is -2.5. And the degrees of freedom is n -1 = 84. The T-score calculator is then used to calculate a p-value with 84 degrees of freedom and a -2.5 t-value on a one-sided t-test. Finally, the p-value is 0.007182 with 95 % confidence interval. With the small p-value (< 0.05), we have evidence to reject the null in favor of the alternative hypothesis, meaning that the mean tenure for a CEO was NOT at least nine years or we can say the mean tenure for a CEO was less than nine years.

T-score calculator: https://www.socscistatistics.com/pvalues/tdistribution.aspx

c) At alpha = .01, what is your conclusion? Draw the graph and show critical values.

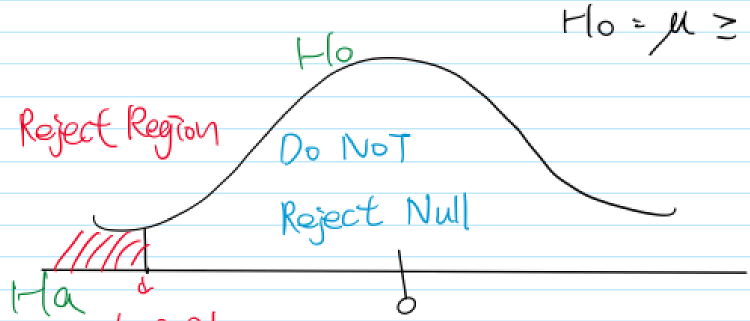
If the p-value is less than the significance level, we reject the null hypothesis. In the question 5b, the significance level is .05. Because .007182 < .05 so we reject the null in question 5b.

Now, in question 5c, the significance level is .01. Because .007182 is also less than .01, so we still can reject the null hypothesis, meaning that there is no sufficient evidence to conclude that the mean tenure for a CEO was at least nine years with 99 % confidence interval.

5C
i.

Lower-tailed (99% CI)

$H_0: \mu \geq 9$ vs H_a $\mu < 9$



$\alpha = 0.01$

$t = -2.326$ $\Rightarrow t^* = -2.5 < t$ $\therefore$ reject Ho.

Question 6 (3 points)

RUBRIC: (-Write down all steps and formula. -Write down (or copy) the normal distribution formula in Excel or Python. -Write down your final answers. -For the hypothesis test provide a graph of critical values. Lack of providing the graph for each question results in 0.5-1-points grade deduction).

The U.S. Bureau of Labor Statistics reports that 11.3% of U.S. workers belonged to unions in 2013. Suppose a sample of 400 U.S. workers is collected in 2018 to determine whether union efforts to organize have **increased** union membership.

a) Formulate the hypothesis that can be used to determine whether union membership increased in 2018. Clearly write down the Hypothesis test including H0 and Ha.

Because the question indicates the 11.3% of U.S. workers belonged to unions in 2013. In other words, the probability is 0.113. The question assumes workers belonged to unions is greater than 0.113, so we set workers > 0.113 as our alternative hypothesis and workers = 0.113 as our null hypothesis.

Hypothesis testing:
Ho: percentage of workers belonged to unions in 2013 = 0.113
Ha: percentage of workers belonged to unions in 2013 > 0.113.

b) If the sample results show that 52 of the workers belonged to unions, what is the p-value for your hypothesis test?

n = 400, sample results = x = 52, workers = p = 0.113, significant level = alpha = 0.05

The right-tail p-value is 0.1401.

6b.

$$\hat{p} = \frac{x}{n} = \frac{52}{400} = 0.13 \text{ (sample proportion)}$$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.13 - 0.113}{\underbrace{\sqrt{\frac{0.113(1-0.113)}{400}}}_{0.015829964}} = \frac{0.017}{0.0158} = 1.075949 = 1.08$$

$$P(Z > 1.074) = 1 - P(Z \le 1.074)$$

$$= 1 - \boxed{0.8599} \Rightarrow \text{from ztable : } Z = 1.08$$

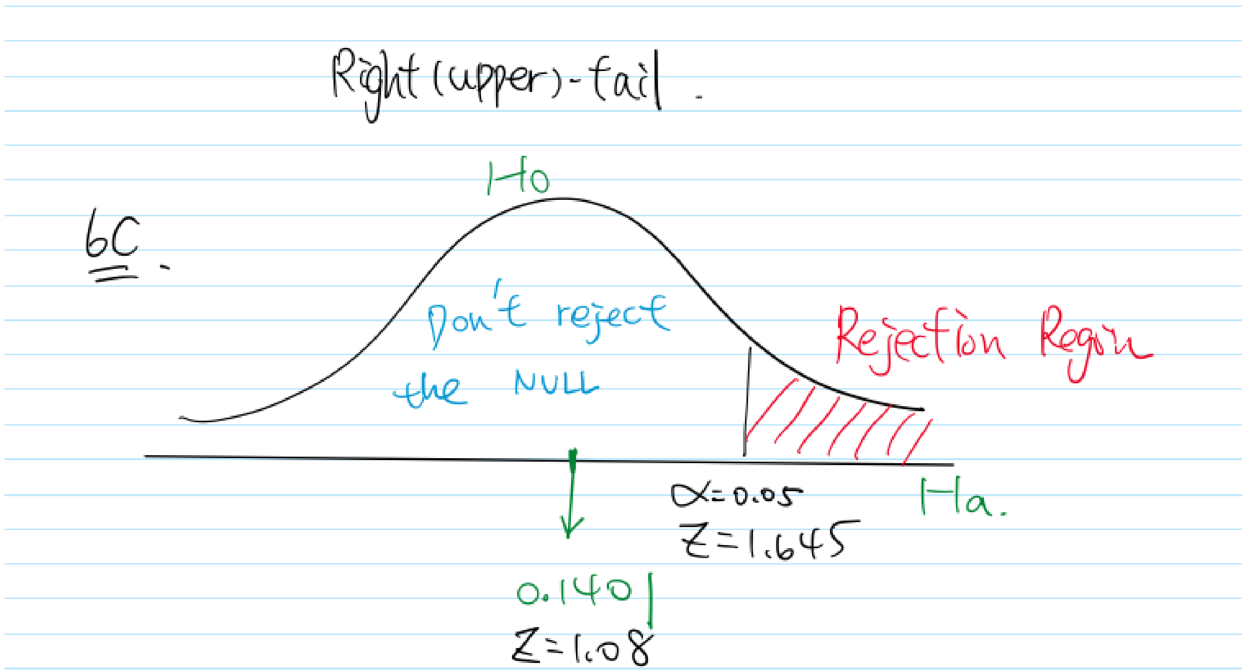
$$= 0.1401$$

\#

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830

c) At alpha = .05, what is your conclusion?

As previously stated, the p-value is 0.1401, which is greater than alpha = 0.05. Thus, we fail to reject the null hypothesis, meaning that there is no evidence to conclude that the union efforts to organize have increased union membership.



d) Write a Python code to determine and address the hypothesis tests that you developed in Part a) and manually answered in Part b).

6d.

Write a Python code to determine and address the hypothesis tests that you developed in Part 6a) and manually answered in Part b).

```
# Ho: workers = 0.113 vs Ha: workers > 0.113.
n = 400
x = 52
p = 0.113
alpha = 0.05

# compute the sample proportion
phat = x/n
# print(phat)

# compute the z-score
z = (phat-p) / np.sqrt(p*(1-p)/n)
# print(z)

# find the p-value
# if two-sided, needs to *2, otherwise is not
pvalue = stats.norm.sf(abs(z))
print(" The p-value is",pvalue)

# state the hypothesis test
if pvalue < alpha:
  print("We have 95 % confident to reject the null hypothesis")
else:
  print("We are 95 % confident fail to reject the null hypothesis")
```

```
 The p-value is 0.14142596913732158
We are 95 % confident fail to reject the null hypthesis
```