IT3030 Biostatistics midterm exam#2 2020.05.19 ID:________________ Name:________________

Write down important intermediate steps and results, including MATLAB or Excel commands used to illustrate your solution strategies. **DO NOT** just show the final result.

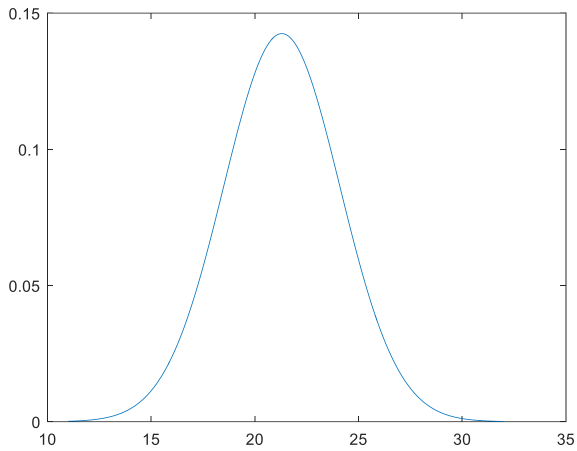
1. (20%) Given the BMI for university students in a normal distribution with the mean value 21.3 and standard deviation 2.8 (shown below). If we were graphing a box plot of this BMI distribution, determine:

(a) the median BMI value.

(b) the BMI that marks the lower 25% of the distribution.

(c) the BMI that marks the upper 25% of the distribution.

(d) Consider BMI from 18 to 24 as normal, determine the percentage of students having BMI falling within this range.



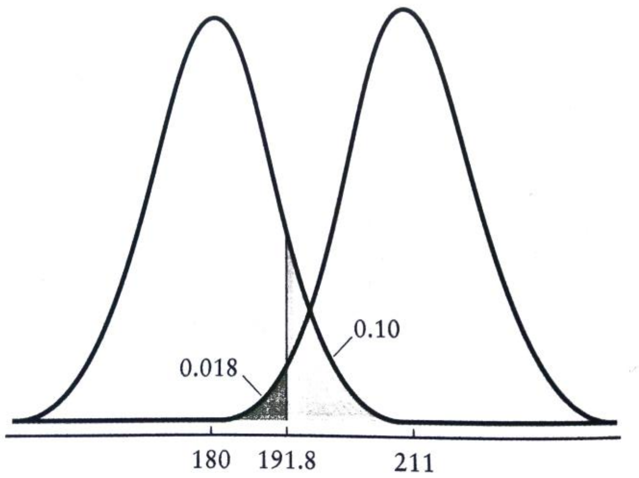
Ans:

```
>> norminv(0.5, 21.3, 2.8)
ans =      21.3000
>> norminv(0.25, 21.3, 2.8)
ans =      19.4114
>> norminv(0.75, 21.3, 2.8)
ans =      23.1886
>> normcdf(24, 21.3, 2.8) - normcdf(18, 21.3, 2.8)
ans =       0.7133
```

2. (20%) We talked about the two distributions of mean values of $n=25$ samples in class – the left bell centered at 180, the right bell centered at 211. Both have population standard deviation $\sigma=46$. We showed in class (also the figure below) that for choosing $\alpha=0.1$ from testing $H_0=180$, the cut-off value would be 191.8, the type II error β would be 0.018 (the black area) and the power would be 0.982.

(a) Determine the new cut-off value if I reduce α to 0.01 from testing $H_0=180$.

(b) Determine the new β and the power of the test.




```
>> norminv(0.99, 180, 46/5)
ans =    201.4024

>> normcdf(ans, 211, 46/5)
ans =        0.1484
```

Power = 1 – 0.1484 = 0.8516

3. (30%) A research study was conducted to examine the differences between older and younger adults towards their satisfaction in life. Ten older adults (over the age of 70) and ten younger adults (between 20 and 30) were give a life satisfaction test. Scores on the measure range from 0 to 60 with high scores indicative of high life satisfaction; low scores indicative of low life satisfaction. For older adults the scores are (45, 38, 52, 48, 25, 39, 51, 46, 55, 46). For younger adults the scores are (34, 22, 15, 27, 37, 41, 24, 19, 26, 36). We would like to perform a t-test to see if there exists statistically significant score difference for the older people from the younger people.

(a) Compute the mean value and standard deviation for each of the two groups.

(b) What is your null hypothesis H_0 and alternative hypothesis H_A?

(c) What is the computed t-value?

(d) Is this a two-sided test or one-sided? Why?

(e) For $\alpha = 0.05$, determine the critical t value for having statistically difference between the two groups.

(f) Compute the p-value for this test. Would you reject your null hypothesis? Why?

Answer:

```
>> OLD=[45, 38, 52, 48, 25, 39, 51, 46, 55, 46];
>> YOUNG=[34, 22, 15, 27, 37, 41, 24, 19, 26, 36];
>> OLD_mean = mean(OLD) =      44.5000
>> YOUNG_mean = mean(YOUNG) = 28.1000
>> OLD_std = std(OLD) =       8.6827
>> YOUNG_std = std(YOUNG) =       8.5434
```

Ho : $\mu_0 = 28.1$, HA : μ_0 != 28.1,

```
>> t=(44.5-28.1)/(8.5434/sqrt(10)) =      6.0703
```

This is a two-sided test since we concern about the score difference, not whether one group is having bigger score that the other.

The critical t-value = tinv(0.975,9) =   2.2622

p-value =   2*(1-tcdf(6.0703, 9)) =   2*9.2916e-05 = 1.8583e-04

4. (30%) An outbreak of Salmonella(沙門氏菌)-related illness was attributed to ice cream produced at a certain factory. Scientists measured the level of Salmonella in 9 randomly sampled batches of ice cream. The levels (in MPN/g) were:

0.593 0.142 0.329 0.691 0.231 0.793 0.519 0.392 0.418

(a) (5%) Determine the mean value and standard deviation from these 9 measured levels.

(b) (10%) We are only interested in knowing a lower bound of Salmonella level in the ice cream. Determine this lower bound based on a 95% CI estimation. (Hint: This is a one-sided CI. Recall that we do not have population mean and standard deviation. So a t-distribution will be used.).

(c) (10%) A one-sided null hypothesis is established that the mean Salmonella level in ice cream is 0.3. Compute the p-value for this test, and decide whether you should reject this null hypothesis or not.

(d) (5%) Compare your results from (b) and (c). Are they consistent or not. Explain why?

Ans:

(a)

```
>> X=[0.593 0.142 0.329 0.691 0.231 0.793 0.519 0.392 0.418 ];
>> mu=mean(X) =        0.4564
>> std=std(X) =        0.2128
```

(b)

```
>> N=9;
>> t=tinv(0.05,N-1) =       -1.8595
>> MIN=mu+t*std/sqrt(N) =       0.3245
```

(c)

```
>> t=(mu-0.3)/(std/sqrt(N)) =        2.2051
>> 1-tcdf(t,N-1) =        0.0293
```

Here the p-value = 0.0293, which is smaller than 0.05. This indicates the sample mean value 0.4564 is distinctly different than the hypothesized 0.3.

(d) This coincides the result in (b) that the 95% CI does not cover 0.3 (lower bound is at 0.3245).