IT3030 Biostatistics midterm exam#2 2019.05.06 ID:________________ Name:_________________

1. (30%) Consider a population of all 20- to 74-year-old males living in the United States. The serum cholesterol levels have a mean value μ=211 and a standard deviation σ=46. (a) For a sample size of n=25 people having an average cholesterol level of 235, determine the 95% conference interval (CI) of it and show that the CI does not contain the actual mean value μ=211. (b) Knowing a smaller n would allow you to have bigger CI, what is the largest possible integer n you would need in order to let the 95% CI contain the actual mean value of 211? (c) If I'd like to limit the CI within 10, what is the smallest possible integer n needed for the sample?

Answer:

(a)

```
>> mu=211;n=25;sigma=46;
>> SEM=sigma/sqrt(n) =       9.2000
>> Z=norminv(0.975) =       1.9600
>> Lower_bound=235-Z*SEM =    216.9683
>> Upper_bound=235+Z*SEM =    253.0317
```

The CI is from 216.9683 to 253.0317. It does not contain the actual mean of 211.

(b) In order to have the 95% CI contain 211, we must have half of the CI as Z*SEM=235-211=24. It is also known that SEM=sigma/sqrt(n). Therefore we have sqrt(n)=sigma/(24/Z)

```
>> sqrt_N=sigma/(24/Z) =      3.7566
>> sqrt_N^2 =      14.1120
```

The largest integer n = 14 to ensure SEM big enough to extend the lower bound of the CI to cover 211.

(c) CI = 10 means Z*SEM=Z*sigma/sqrt(n)=5. Therefore, sqrt(n)= (Z*sigma)/5.

```
>> Z*sigma/5
ans = 18.0317
>> ans^2
ans =    325.1411
```

The smallest integer is n = 326.

2. (30%) Percentages of ideal body weight were determined for 18 randomly selected insulin-dependent diabetics (需要施打胰島素的糖尿病患者) and are shown below. A percentage of 120 means one weighs 20% more than ideal; a percentage of 95 means one weighs 5% less than ideal.
   107, 119, 99, 114, 120, 104, 88, 114, 124, 116, 101, 121, 152, 100, 125, 114, 95, 117
   (a) Determine the mean value and standard deviation of these percentages.
   (b) Compute a two-sided 95% confidence interval for the true mean percentage of ideal body weight for this population.
   (c) Does this CI contain the value 100%? What does the answer tell you?

   Answer:

```
>> X=[107, 119, 99, 114, 120, 104, 88, 114, 124, 116, 101, 121, 152, 100, 125, 114, 95, 117]';
>> PD=fitdist(X,'normal')
PD =
  NormalDistribution

  Normal distribution
        mu = 112.778     [105.605, 119.951]
     sigma = 14.4245     [10.8239, 21.6244]

>> CI=paramci(PD, 0.05)
CI =
  105.6047     10.8239
  119.9509     21.6244
>>
```

   The 95% CI is from 105.6047 to 119.9509. It does not contain 100%, meaning that insulin-dependent diabetics are overweight.

3. (20%) In the class we performed a two-sided test to determine whether a coin clip is fair or not. Suppose we had heads 15 times out of 20 flips. (a) What is the null hypothesis for this test? (b) Determine the p-value for this test, and decide to reject or not to reject the null hypothesis for α=0.05.

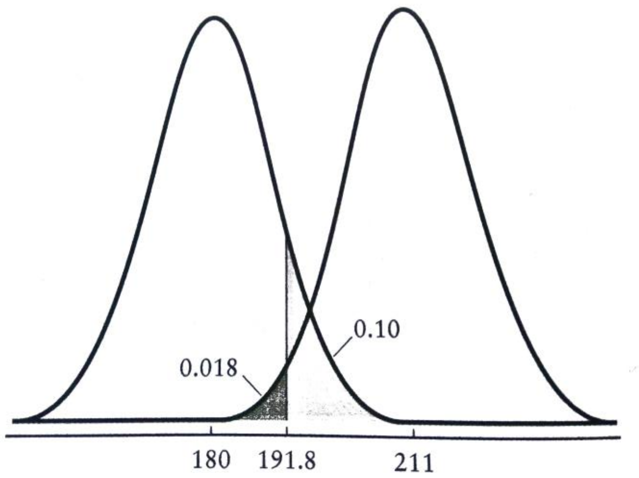
Answer:
(a) H_0: This coin is a fair one (it lands 50%/50% on either head or tail)
(b)

```
>> 1-binocdf(14, 20, 0.5)
ans =
       0.0207
>> ans*2
ans =
       0.0414
```

0.0414

Since this p-value is smaller than 0.05, we'd reject the null hypothesis (that the coin is fair). On other words, this is not a fair coin.

4. (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 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 1-0.018=0.982. (a) Determine the new cut-off value if I reduce $\alpha=0.01$ from testing $H_0=180$. (b) Determine the new β and power of the test.



Answer:

```
>> norminv(0.99, 180, 46/5)
ans =
   201.4024
>> beta=normcdf(ans,211,46/5)
beta =
       0.1484
>> power=1-beta
power =
       0.8516
>>
```