Applied Statistics MATH 661 Assignment #6

October 15, 2019

1 Task 1 IPC Excercise 5.26

1.1 IPC 5.26 What is wrong in each case?

1.1.1 A) The central limit theory states that for large n, the pupulation mean $ $\mu\$isapproximatelynormal.$

This is incorrect because the central limit theory states that for large n, **the sameple mean** $\bar{x}$ is approximately normally **distributed.**

1.1.2 B) For large n, the distribution of observed values will be approximately normal.

There is no theorem or idea to support this statement. The distribution of a random variable depends on the intrinsic nature of the data, not the sample size.

1.1.3 C) For sufficiently large data, the 68-95-99.7 rule says that $\bar{x}$ should be within $\mu \pm 2\sigma$ about 95% of the time.

For sufficiently large data, the 68-95-99.7 rule says that 95% of the **observations** will lie within $\mu \pm 2\sigma$.

1.1.4 D) As long as the sample size n is less than half the population size N, the standard deviation of $\bar{x}$ is $\sigma/\sqrt{n}$

As long as the sample size n is less than half the population size N, the **sample standard deviation** is $\sigma/\sqrt{n}$.

---

2 Task 2 Sleep time of college students. IPC Excercise 5.28 and 5.29

2.1 IPC 5.28

2.1.1 A) What is the standard deviation of the average time?

Since n=120>30, the sameple mean is approximately normally distributed with standard deviation $= \sigma/\sqrt{n}$

[1]:
```
import math
n = 120
sigma = 1.24
ans = sigma/math.sqrt(120)
print('Standard deviation of sample mean is ',ans)
```

```
Standard deviation of sample mean is  0.11319599521773432
```

2.1.2 B) Use the 95 part of the 68-95-99 rule to describe the variability of this sample mean.

The sample mean is centered around 6.78 and 95% of the observations will fall between $6.78 \pm 2\sigma$

[2]:
```
LowerLimit = 6.78 - 2*ans
UpperLimit = 6.78 + 2*ans
print('The sample mean is centered around 6.78 and 95% of the observations \n',
     'Will fall between ',LowerLimit,' and', UpperLimit)
```

```
The sample mean is centered around 6.78 and 95% of the observations
 Will fall between  6.553608009564532  and 7.006391990435469
```

2.1.3 C) What is the probability that your average will be below 6.9 hours?

[3]:
```
z = (6.9-6.78)/ans
print('The z score is',z,'\n','and the probabiliy that the average will be␣
 ↪below 6.9hrs\n',
     'is ', '.8554')
def normpdf(x, mean, sd): # This function defines the pdf of a normal␣
 ↪distribution
    var = float(sd)**2
    denom = (2*math.pi*var)**.5
    num = math.exp(-(float(x)-float(mean))**2/(2*var))
    return num/denom
#  normpdf(6.9,6.78,sigma)
```

```
The z score is 1.0601081758164514
 and the probabiliy that the average will be below 6.9hrs
 is  .8554
```

2.2 IPC 5.29 Determining sample size

Refer to the previous excercise. You want to use a sample size such that about 95% of the averages fall within $\pm.08 hours$ of the true mean $\mu = 6.78$

2.2.1 A) Should the sample size be larger or smaller?

For the averages to fall closer to the true mean, the sample size should be larger.

2.2.2 B) What sample standard deviation do you need?

$2xsigma = .08$
sigma = .08/2 = .04

2.2.3 C) Determine sample size, n.

sigma = $\sigma/\sqrt{n}$
$n = (\sigma/.04)^2$
$n = (1.24/.04)^2$

[4]:
```
n= (1.24/.04)**2
print('The sample size should be >', n)
```

```
The sample size should be > 961.0
```

3 Task 3 IPC 5.31

3.1 IPC 5.31

\$$\sigma_x = \sigma/\sqrt{n}$\$
\$$\sigma_x = .4/\sqrt{5}$\$

[5]:
```
sigma_x = .4/5**.5
print('Sample average is', 250,'and sample standard deviation is ',sigma_x)
```

```
Sample average is 250 and sample standard deviation is  0.17888543819998318
```

4 Task 4 IPC 5.35

4.1 IPC 5.35

4.1.1 A)

$\bar{x}$ is an unbiased estimator for μ. This means that generally, values observed for $\bar{x}$ will be close enough to the true value of μ. That is, values observed in a simple study will approximate the true values of the general topic the study is trying to understand.

4.1.2 B)

A large sample gives more trustworthy results because a large sample is more likely to notice all or most possible values of the population. For example, if you are studying all the colors in the spectrum, a sample size of 4 colours will only observe 4 colors. It would not be representative of the real color line.

[ ]:
```
```