Problem Set 6

Research method Problem Set 6 due Sun 16th Dec, 23:59

LIU BOWEN (1004028)

For Problem Set 6, I use the following packages:

```
import matplotlib.pyplot as plt

import numpy as np

import scipy as sp

import scipy.stats as ss
```

Problem 1

Answer:

Based on CLT, the mean of sample distribution is equal to the population mean and the standard deviation of this sample distribution is usually known as the standard error.

The mean is 30/1000=0.03 and standard deviation is $\sigma/\sqrt{4} = 3.5$ and $(\sigma/1000)/\sqrt{4} = 0.0035$ for per person

Problem 2

Answer:

1) Consumer organization wants to test if the mean mileage can achieve 60000 miles or not.

2) H_0: the mean mileage for their premium brand tyre is 60,000 miles

3) H_A: the mean mileage for their premium brand tyre less than 60,000 miles

4) One-tail

5) The standard error is 7000 / $\sqrt{49}$ = 1000

6) The critical value is 60000 − 2000 = 58000 and H0 is rejected if $\bar{x}$ < 58000.

7) The Type-1 error is the true tyre mileage is 60000 miles, but the test result is the tyre mileage is less than 60000 miles. The Type-2 error is that the true tyre mileage is less than 60000 miles, but the sample test result shows the mileage is 60000 miles.

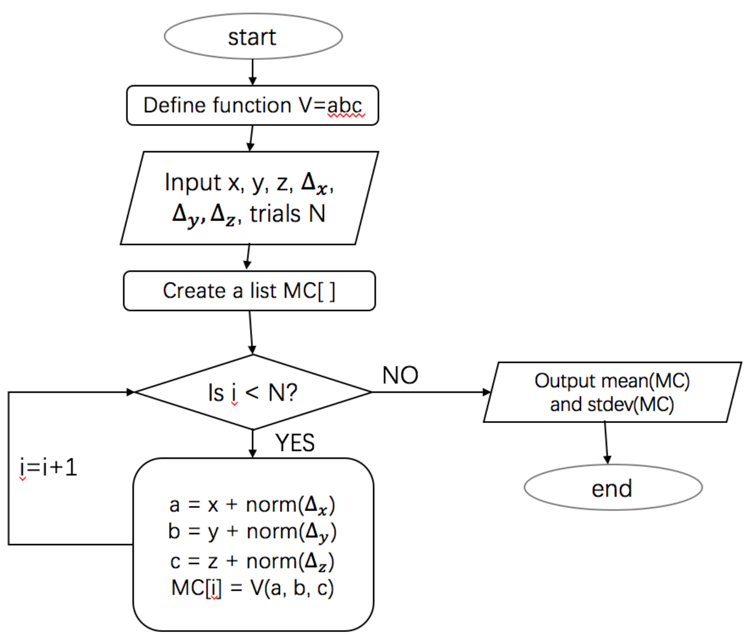
   The probability of Type-1 error is P(Z=-2σ)=0.02275

   For Type-2 error, if we assuming the true population mean $\mu = 57500$, then we have the probability of Type-2 error β=1-P(Z=$\frac{58000-57500}{1000}$)=0.30854 (for assuming 57500 only)

   For the tyre manufacturer, the Type-2 error is more serious problem as bad tyres would cause safety issue.

Problem 3

Answer:



Problem 4

Answer:

From the question, the population variance is unknown, use t-test to do this. $(X-\mu) / (s/\sqrt{N}) = (2750-2500) / (950/\sqrt{85}) = 2.43$.For two-tails t-test, $t_{0.025}(84) = 1.99$ so we reject H_0

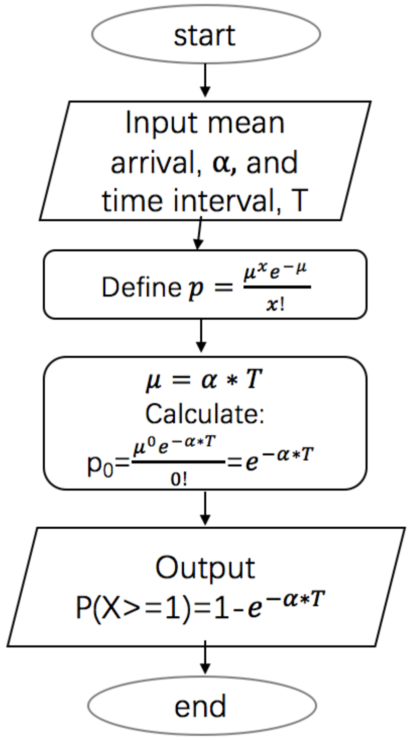
As ths sample size is 85 which is larger than 30, if we use 950 as population variance we also use Z to test. We need reject H_0 if $Z > 1.96$ and $Z<-1.96$, the critical value should be $2500-1.96*950/\sqrt{85}=2298$ and $2500+1.96*950/\sqrt{85}=2702$. Now the $\bar{x}$ is 2750 so we also reject $H_{0.}$

In conclusion, if population variance is known and sample size > 30 we use Z. Also, if population variance is unknown or sample size < 30 t-test should be better.

<u>Problem 5</u>

Answer:

1) Poisson distribution should be used.
2) $P(X<=5)=P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5)=0.369$
3) The patient arrives independently.
4) The flow diagram:



Problem 6

Answer:

```
def calPoE(N1, mean1, var1, mean2, var2):

    deriva1 = N1 / np.sin(mean2/180.0*np.pi) *
np.cos(mean1/180.0*np.pi)

    deriva2 = (N1*np.sin(mean1/180.0*np.pi)) * (-
np.cos(mean2/180.0*np.pi)) / np.sin(mean2/180.0*np.pi)**2

    poe = var1**2 * deriva1**2 + var2**2 * deriva2**2

    N2 = N1 * np.sin(mean1/180.0*np.pi) / np.sin(mean2/180.0*np.pi)

    return np.sqrt(poe)/180.0*np.pi, N2

N1 = 1.0
```

```
mean1 = 22.02

var1 = 0.02

mean2 = 14.45

var2 = 0.02

print calPoE(N1, mean1, var1, mean2, var2)
```

The answer is:

(0.0024133592559739776, 1.502515309544441)

So $n_2 = 1.503 \pm 0.002$