**生统第三次作业**

一、实验室欲购进一批灯泡，打算在两个供货商之间选择一家购买。选购考虑的主要因素就是灯泡使用寿命的方差大小，为此需要对供货商提供的20个样品进行检测，得到的数据如下表所示。(20’)

|  |  |
| --- | --- |
| 供货商1 | 供货商2 |
| 6802 | 5884 |
| 5730 | 5871 |
| 5823 | 5797 |
| 5915 | 5957 |
| 5774 | 5803 |
| 5880 | 5862 |
| 5870 | 5814 |
| 5773 | 5885 |
| 5830 | 5856 |
| 5841 | 5940 |
| 5763 | 5945 |
| 5851 | 5803 |
| 5789 | 5864 |
| 5796 | 5851 |
| 5818 | 5714 |
| 5685 | 5943 |
| 5602 | 5830 |
| 5841 | 5858 |
| 5723 | 5922 |
| 5757 | 5866 |

1、检验两家供货商的灯泡使用寿命的方差有无显著差异（ɑ=0.05）（10’）

2、择最合适的检验方法检验两家供应商的灯泡使用寿命有无差别。（10’）

二、R language application (25’)

Please use R to resolve the following issues and display your R code and results.

1. For a normal random variable X with mean 4.0, and standard deviation 1.0,

1. find the probability that X is less than 2.0. (4’)
2. find the value K so that P(X>K) = 0.05. (4’)

2. When tossing a fair coin 8 times,

1. find the probability of seeing no heads (Hint: this is a binomial distribution.) (3’)
2. find the probability of seeing exactly 4 heads. (3’)
3. find the probability of seeing more than 5 heads. (3’)

3. Simulate a sample of 1000 random data points from a normal distribution with mean 100 and standard deviation 8, and store the result in a vector.

1. plot a histogram and a boxplot of the vector you just created. (4’)
2. using the data above, test the hypothesis that the mean equals 100 (using t.test). (4’)

三、R language application. (25’)

（Please use R to read-in and manipulate data, code and results should be displayed.）

In order to detect air quality, a city's environmental protection department conducted a random test of PM2.5 in the air every few weeks. It is known that the average value of PM2.5 per cubic meter of air in the city is 82ug/m3. In the most recent test, the value of PM2.5 per cubic meter of air(ug/m3) is shown in the homework3\_data.

1)Show your work directory (2’)

2)use R to read in the data（3’）

3)use boxplot to show the PM2.5 distribution in every month (10’)

4)get the data of month equal to March and store in data\_march (10’)

四、Suppose we draw a sample of size 20 of birthweights from a hospital, the details can be found in the homework data. The mean of national-wide birthweights is 118.(30’)

1)What is the probability that the mean birthweight of the sample falls between 100.0 and 126.0? Please list the formulas to calculate this and also the R code for it.(5’)

2)What is the 95% confidence interval of the sample mean?(5’)

3)Can we say the underlying mean birthweight from this hospital is higher than the national average?

Please list the formulas for this and also the R code for it.(5’)

4)Test the hypothesis that the mean birthweight of sample size 20 is different from the national average (Significance level 0.05). Please list the formulas for this and also the R code for it.(5’)

5)Compute the power of the test performed in (4) with significance level 0.05.(5’)

6)To see the significance difference between the sample mean and the national mean and ensure the

type II error to be β=0.05, what is the appropriate sample size with significance level is 0.01?(5’)