**Exercise 1**

**Exercise 2**

1. Given the boxplot and histogram of light1879 dataset, we can observe some outliers and the QQ line is not straight; So, we can presume that this dataset doesn't follow a normal distribution. Similarly, we can observe some outliers in the boxplot of 1882 data set diagram and the QQ line is not straight in the histogram of 1882 dataset. We can consider that this dataset is not in normal distribution. As we can see from the histogram of light.txt dataset, the points are approximately on a straight line then this dataset can be assumed to be sampled from a normal distribution.
2. The dataset light.txt follows the normal distribution so mean and median are equal. The confidence interval for μ with 95% confidence is measured as below.
3. –
4. The p-value of t-test is greater than 0.05 and therefore we cannot reject the null hypothesis that the value of speed measured by Michelson and Newcomb is equal to the currently most accurate value, 299792.458 km/s.

**Exercise 3**

1. According to the QQ-plot of KLM data, the distribution is not taken from normal population. For skewed distribution, the mean is highly influenced by the high/low values. In such cases, it is better to test location in terms of the median, instead of the mean. Therefore, the sign test is the best choice which prerequisites are all satisfied. The data are a random sample from a population with a certain median m. we test null hypothesis . The test statistics is which has the binary (N,0.5)- distribution under . Conclusion: As the p-value is 0.01, we can reject the null hypothesis. It means that the median of this population is bigger than 31.
2. This test is similar to sign test except the probability of success in binary distribution. we test null hypothesis . The test statistics is which has the binary (N,0.1)- distribution under . Conclusion: As the p-value is 0.007, we can reject the null hypothesis. It means that more than 10% of the parts arrives after the maximum delivery period of 72 days.

**Exercise 4**

1. Given the histograms and QQ-plots of seeded and unseeded cloud data, we can't assume that they are in a normal distribution. As a result, we can't apply the two samples t-test in this case because the assumption of the normal distribution in t-test was violated. Mann- Whitney test and the Kolmogorov-Smirnov test can be adopted in this case for the reason that both don't assume observations are from normal distribution. Wilcoxon signed rank test: p-value = 0.01383 < 0.05, we can conclude that of equal means is rejected. Kolmogorov-Smirnov test: p-value = 0.01905 < 0.05, we can conclude that of equal means is rejected.
2. Similarly, we can't assume that the square root of seeded and unseeded cloud data is normal distribution. The assumption of the normal distribution in the two samples t-test was violated so we shouldn't apply t-test to the data. Mann- Whitney test and the Kolmogorov-Smirnov test can be adopted in this case for the reason that both don't assume observations are from normal distribution. Wilcoxon signed rank test: p-value = 0.01383 < 0.05, we can conclude that of equal means is rejected. Kolmogorov-Smirnov test: p-value = 0.01905 < 0.05, we can conclude that of equal means is rejected.
3. After transforming by square root of the square root of the values, seeded clouds data don't still follow the normal probability distribution, which can be observed from the histogram and QQ-plot. After transforming by square root of the square root of the values, unseeded clouds data can be considered to follow the normal probability distribution, which can be observed from the histogram and QQ-plot. Square root of the square root of the values in seeded clouds doesn't follow the normal distribution and we can't use the two samples t-test in this situation. Mann- Whitney test and the Kolmogorov-Smirnov test can be applied again in this case due to the fact that both don't assume observations are from normal distribution. Wilcoxon signed rank test: p-value = 0.01383 < 0.05, we can conclude that of equal means is rejected. Kolmogorov-Smirnov test: p-value = 0.01905 < 0.05, we can conclude that of equal means is rejected.

**Exercise 5**

1. Based on the pairs is possible to see a potential correlation in the following pairs: migration x age, migration x weight, migration x wrist. This conclusion is based on the fact that the plot from these pairs resembles a straight line passing through the origin.
2. Tests will be conducted using Spearman rank correlation test which doesn't assume normality between the two variables.

* Test 5.2.1 (migration x age): As can be seen from the test, p-value = 0.0021, which led us to reject the null hypothesis that rho is equal to 0. In fact, the calculated rho based on the samples is 0.4760.
* Test 5.2.2 (migration x weight): As can be seen from the test, p-value = 0.02861, which led us to reject the null hypothesis that rho is equal to 0. In fact, the calculated rho based on the samples is 0.3506.
* Test 5.2.3 (migration x length): As can be seen from the test, p-value = 0.6087, which led us to fail to reject the null hypothesis that rho is equal to 0 - considering a o.05 confidence level (despite the fact that R's output states that H0 can be rejected). The calculated rho based on the samples is 0.0845 which is very close to 0 -> So it is indeed possible to conclude that these variables are not correlated to each other.
* Test 5.2.4 (migration x wrist): As can be seen from the test, p-value = 0.1797, which led us to fail to reject the null hypothesis that rho is equal to 0 - considering a o.05 confidence level (despite the fact that R's output states that H0 can be rejected). However, the calculated rho for the sample is different from 0 (in fact rho = 0.2193).
* Test 5.2.5 (migration x diastolic): As can be seen from the test, p-value = 0.6494, which led us to fail to reject the null hypothesis that rho is equal to 0 - considering a o.05 confidence level (despite the fact that R's output states that H0 can be rejected). However, the calculated rho for the sample is different from 0 (in fact rho = 0.0751).

**Exercise 6**

1. To study the data, we plot the running time before and after both soft and energy drink separately. It includes QQ-plot, histogram and box plot of all of them. The median of running time increased after soft drink. Moreover, the running time after drinking spread in larger range rather than before drinking. As the sample size is really small, it is hard to conclude the normality of population. QQ-plot of before and after for soft drink demonstrates the normality of population while the histograms contradicts this idea. In case of energy drink, the histograms and QQ-plots prove that the sample is not taken from normal population for both before and after energy drink. The median of running time decreased after energy drink. Additionally, the running time after drinking spread in larger range than before.
2. Since we can’t assume normality for both soft drink and energy, we should apply permutation test for two cases separately. We generate 1000 randomly chosen permutations to estimate the distribution of our test statistic under . Conclusion: In case of soft drink, since the p-value is about 0.16 we can’t reject null hypothesis. Therefore, soft drink doesn’t affect the running time. In terms of energy, the p-value is 0.29 and still greater than critical value. So, it means that energy drink can’t influence running time either.
3. In such case, we have two different groups without normal population. Therefore, we apply Man-Whitney test. The sample stems from soft drink population (S) and similarly originates from energy drink (E). We test null hypothesis that the populations are the same. Conclusion: of equal means is not rejected. The underlying distribution of time difference in soft drink is similar to energy drink and drink type doesn’t affect running time.
4. Time differences (after - before) are negative numbers for some cases. It may because of dependency of two measurement (before and after) in a short period of time. Additionally, the sample size for both cases seems small. The distribution of time difference is not normal in soft and energy drink and we can’t use t-samples t-test. **Are these correct?**
5. Yes, we have similar objections here.
6. If we apply t-test in part 3, we should assume normality of time difference in both drink types. Instead, we can apply Man-Whitney test. **(I don’t understand the question)**

**Exercise 7**

1. According to the QQ-plot of these three drugs, we just can assume samples in drug 3(cyclopropane) are taken from normal population. Drug 1 and 2 are certainly not normal.
2. The estimated concentration for drug 1 is equal 0.4340. drug 2 and drug 3 are 0.399 and 0.015 respectively. Since the p-value is 0.11, we reject null hypothesis which means for some (i, j).
3. In this part, the p-value is 0.059. Therefore, the null hypothesis in not rejected (but not with a great difference). . As we concluded in first part, the samples are not taken from normal population. However, ANOVA assumes normality to test the data. Therefore, the conclusion of ANOVA is against Kruskal-Wallis result.