“Babyboom”

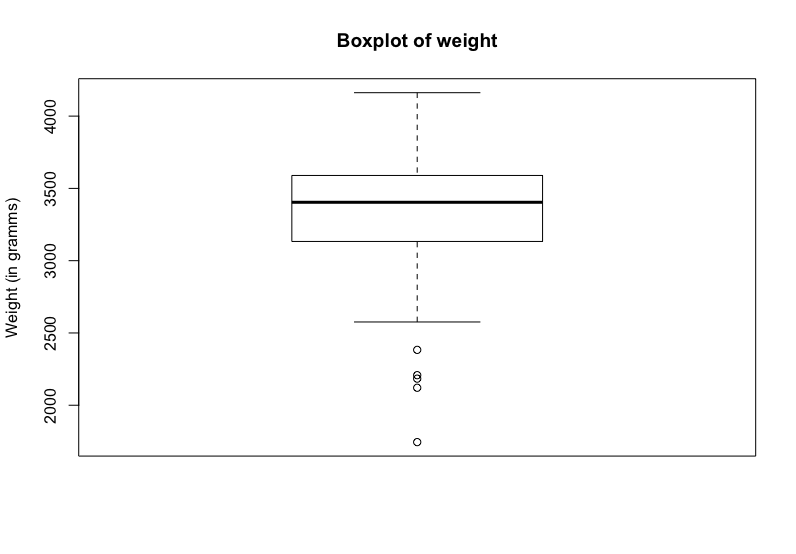
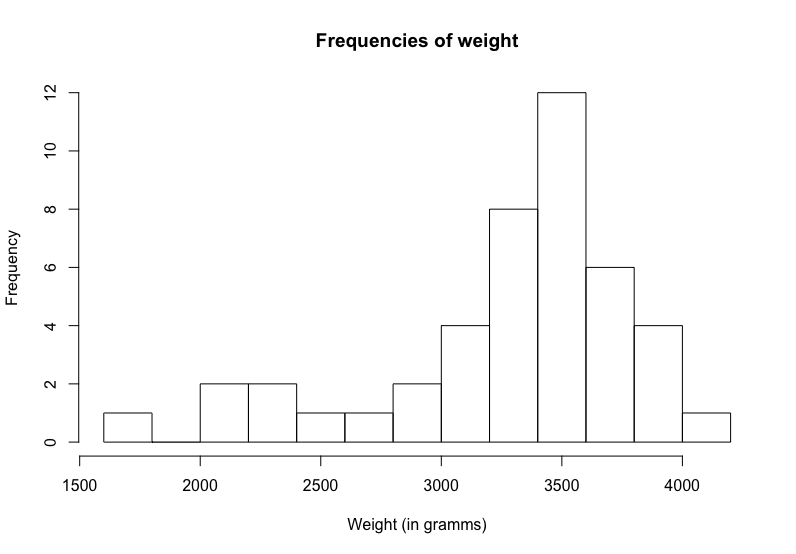
Hypothesis to test:

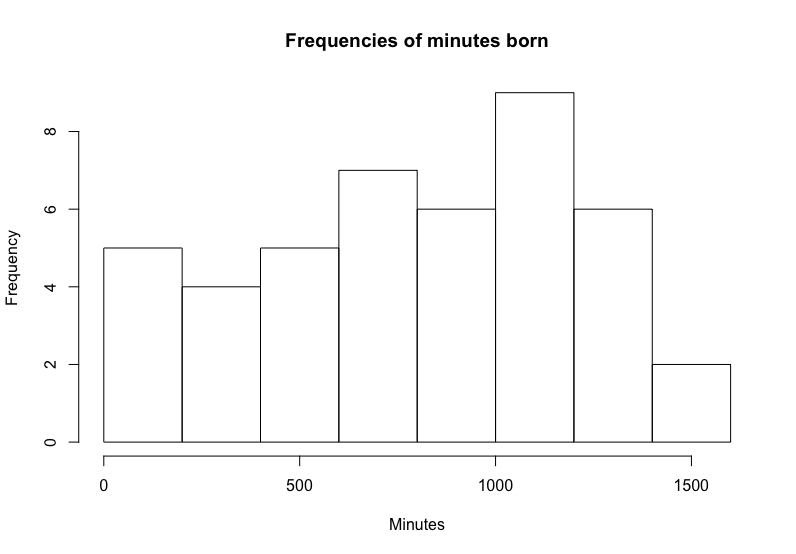
* hypothesis H0: mean weight male = mean weight female
* hypothesis H0: uniform distribution of male and female number of births
* hypothesis H0: uniform distribution number of births during the hours (24 intervals)

Weight whole:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| n | mean | median | sd | mad | min | max | skew | kurtosis | se |
| 44 | 3275,95 | 3404 | 528,03 | 343,96 | 1745 | 4162 | -1,08 | 0,64 | 79,6 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| min | 1 quartile | median | mean | 3 quartile | max |
| 1745 | 3142 | 3404 | 3276 | 3572 | 4162 |





Notes.

There are two peaks in histogram of weights. We should better divide the data by two samples (male and female) and investigate them separately.

Several outliers on boxplot may mean female observations. There are 18 female obs. and 24 male obs. so male observations mean “more” so female observations became outliers.

There is almost nothing interesting in histogram of minutes except the fact that most number of births was after midnight.

“Euro coins’ weight”

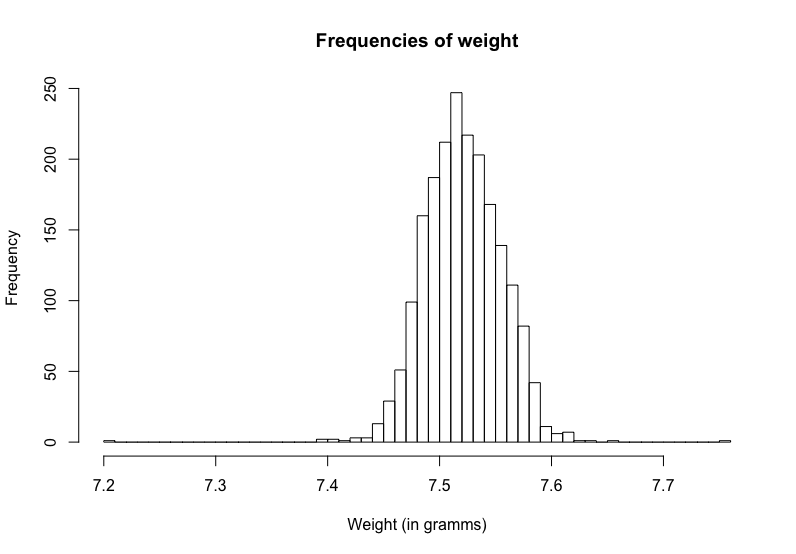
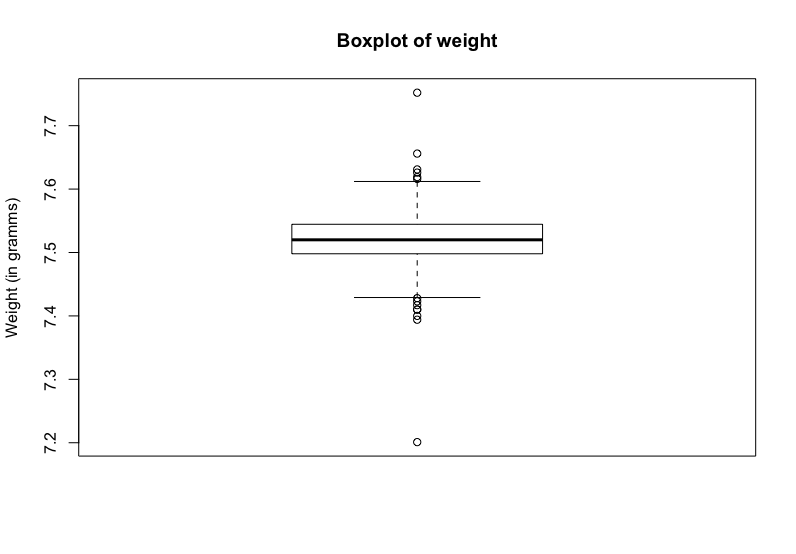
Hypothesis to test:

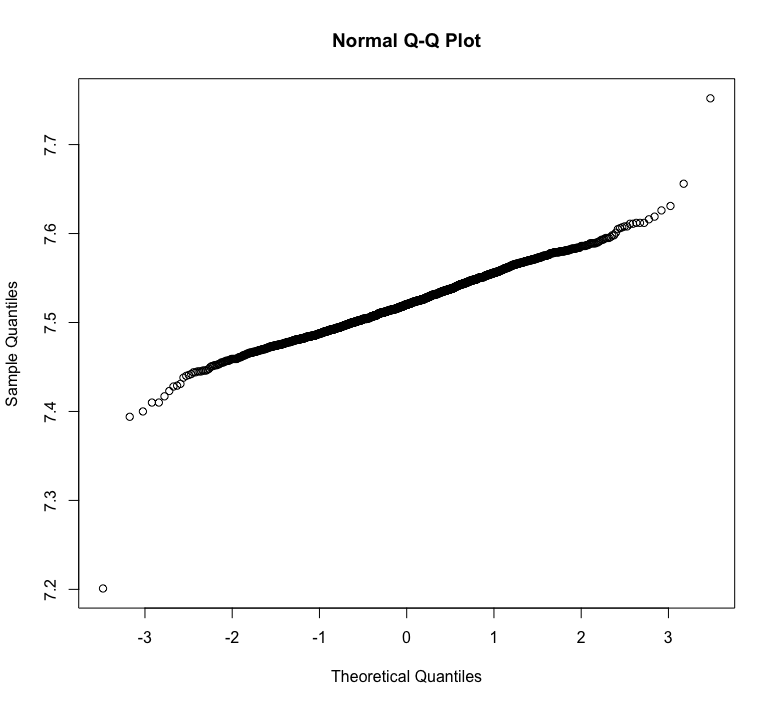
* hypothesis H0: uniform distribution between mean of packages
* hypothesis H0: normal distribution in each package

Weight whole:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| n | mean | median | sd | mad | min | max | skew | kurtosis | se |
| 2000 | 7,52 | 7,52 | 0,03 | 0,03 | 7,2 | 7,75 | -0,19 | 4,42 | 0 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| min | 1 quartile | median | mean | 3 quartile | max |
| 7,201 | 7,498 | 7,52 | 7,521 | 7,544 | 7,752 |





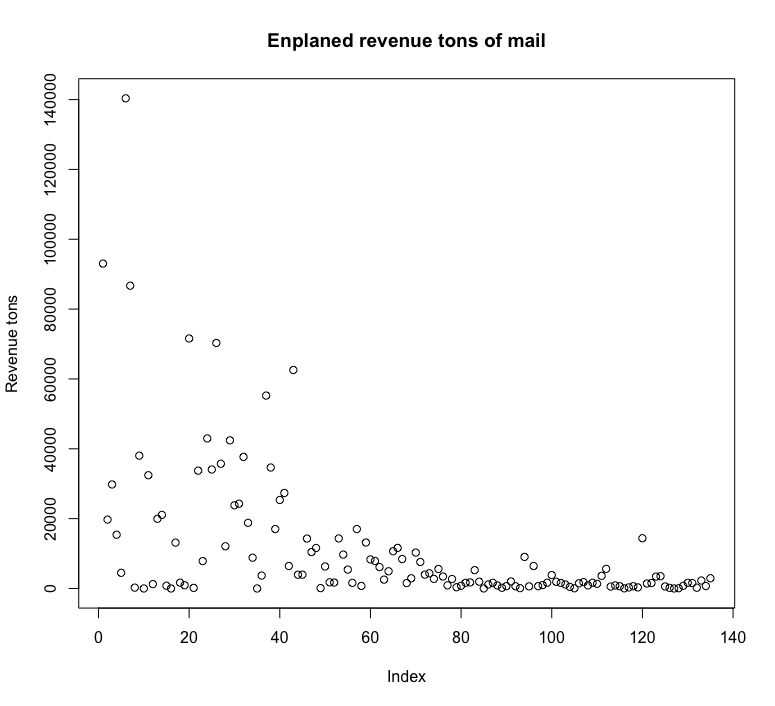
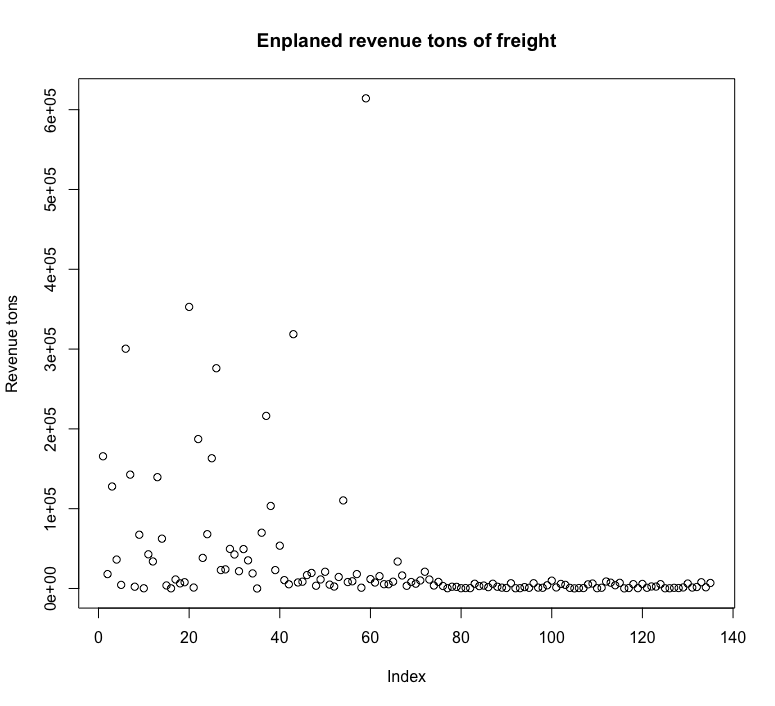
Notes.

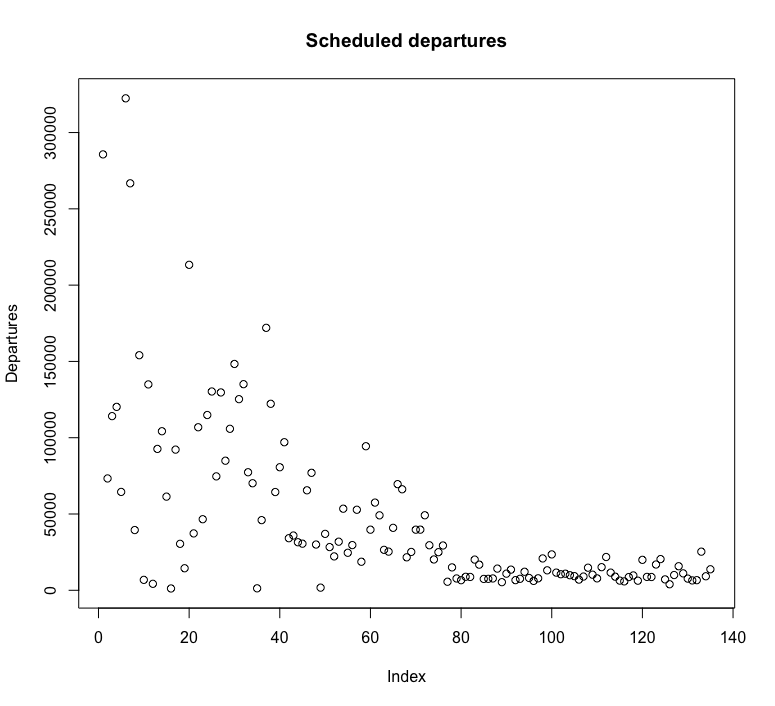
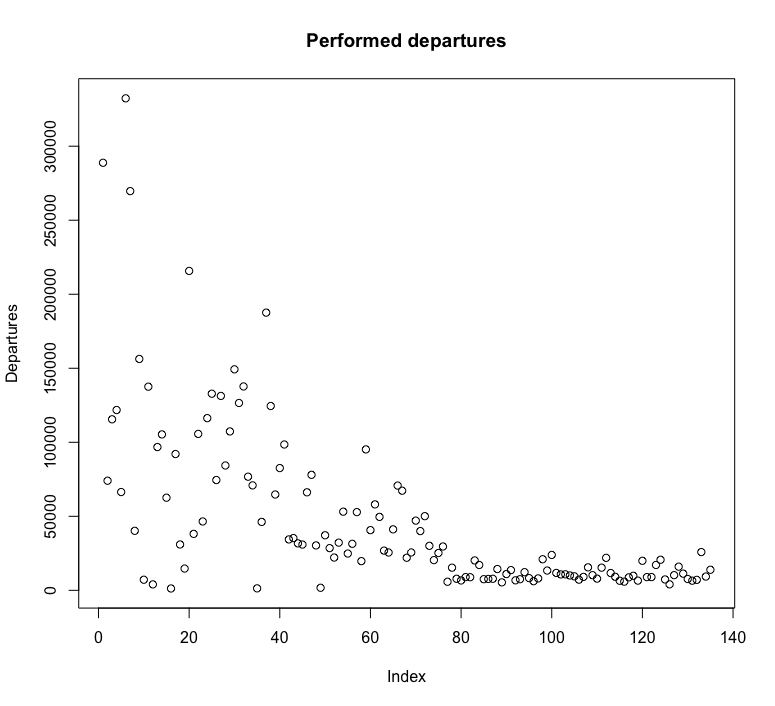
Frequencies histogram has normal distribution shape but with lower tails so there are many outliers. In this case, we can’t use tests with normal distribution assumption.

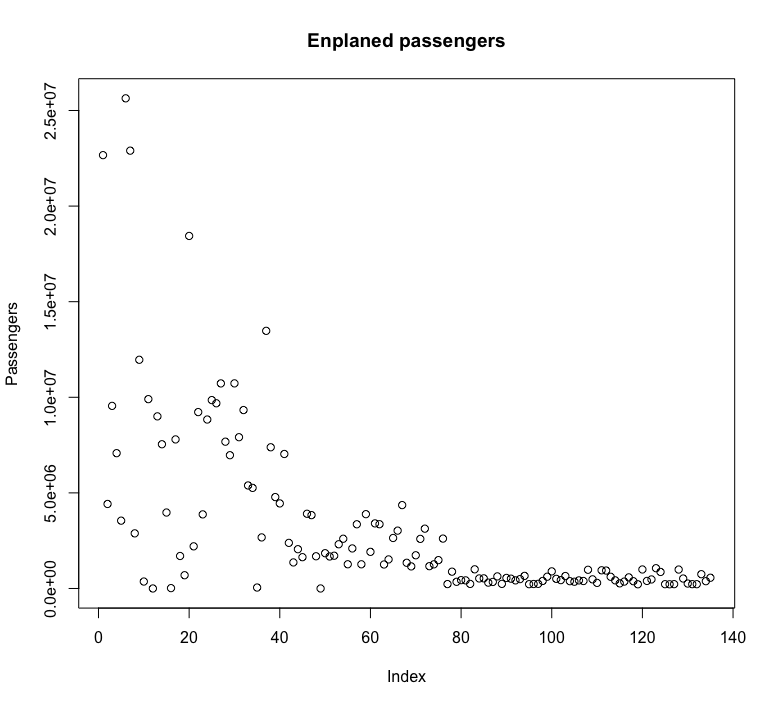
“US airports”

Hypothesis to test:

* hypothesis H0: mean of scheduled = mean of performed
* hypothesis H0: mean of freight = mean of mail
* hypothesis H0: scheduled, performed, pass, freight, mail are exponential distributions
* hypothesis H0: mean scheduled Burbank = mean scheduled County
* hypothesis H0: mean performed Burbank = mean performed Burbank







Variables:

* Enplaned passengers – запланировано пассажиров
* Enplaned revenue tons of freight – запланированный доход тонн грузов
* Enplaned revenue tons of mail – запланированный доход тонн почты
* Scheduled departures – запланированные отправления
* Performed departures – выполненные отправления

More than 1 airport in city:

* Chicago
* Columbia
* Dallas
* Detroit
* Hartford
* Houston
* Lon Angeles
* New York
* San Francisco

Los Angeles has 4 airports. Explore them.

