Homework from the course Probability and statistics (NMAI059)
By Jaroslav Jindrak, done in the winter semester 2014/2015

All simulations were done using the R programing language (source code can be found in the file jindrak_jaroslav.R) using the seed 13031992.

Task #1:

According to the simulation of one day in the store warehouse, the total number of orders for the first product was **195** and **113** for the second product.

Task #2:

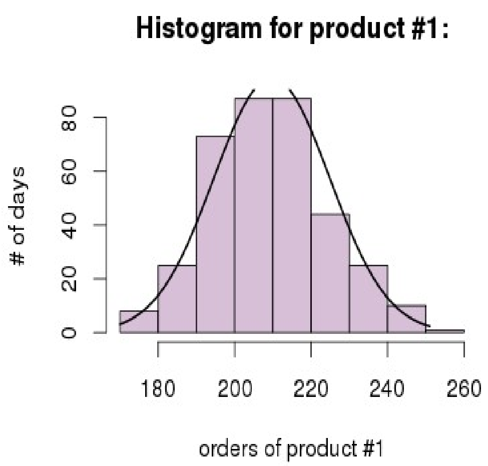
a) I calculated the mean value of the reports I got from 360 simulations for the point estimation of the total number of orders of both products, which is **209.691667** for the first product and **105.391667** for the second product.
To calculate the bounds of the 95% interval estimation, I used the following formula:

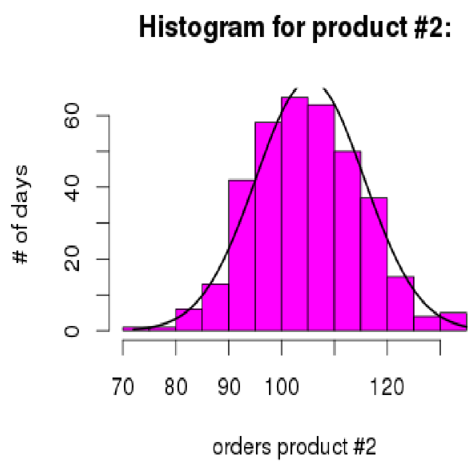
$$ub=\overline{x}-\frac{q*S_n}{\sqrt{(n)}} \qquad lb=\overline{x}+\frac{q*S_n}{\sqrt{(n)}}$$

where $\overline{\mathbf{x}}$ denotes the sample mean, $\mathbf{S_n}$ the standard deviation and **n** is the length of the sample $\mathbf{X_1, X_2, ..., X_n}$ and $P(-q\leq\frac{\overline{x}-\mu}{S_n/\sqrt{(n)}}\leq q)=0.95$, thus q denotes the 97.5th percentile of the distribution. The interval takes the form of $[lb, ub]=[208.377446, 211.005887]$ for the first product and $[lb, ub]=[104.498447, 106.284887]$ for the second product.

b) To test the null hypothesis (μ=196 for the first product and μ=98 for the second product) I decided to use the student's t-test on 5% significance level. The test resulted in a p-value for both products and the hypothesis is rejected if the p-value is lower than the significance level (0.05) in this case. In the case of the first product, the p-value was $\mathbf{1.586167*10^{-48}}$,which is lower than 0.05 and the null hypothesis (μ=196) was rejected. In the case of the second product, the p-value was $\mathbf{2.603116*10^{-34}}$, which is also lower than 0.05 and the null hypothesis (μ=98) was rejected in this case aswell.

c) According to the histograms for the product orders below, I deduced that they both follow the **normal distribution**. The histograms:





d) The Shapiro-Wilk test gave me the p-values of **0.086115** for the first product and **0.358278** for the second product. Since both of these p-values are bigger than the **0.05 significance level**, both passed the normality test, so the distribution of the number of orders of product #1 and product #2 in a day match the normal distribution.

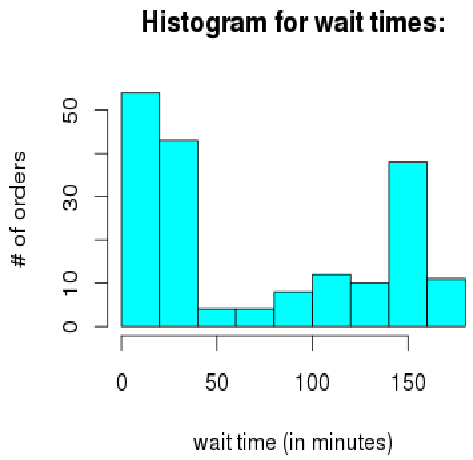
e) The normal distribution curve can be see in the two histograms above, my RStudio sadly cut the top of the curve from the histogram.

Task #3:

a) I must admit that I got a bit confused with the objective of this task, since in the given paper this subtask asks me to calculate the average delivery time of the last order of the day, which, in Czech can be interpreted in several ways (most notably the time when the last order was delivered and the time it took the courier to deliver said last order). I decided to interpret it as the former option and the answer to that is **1:27AM** on the second day or **17** hours and **27** minutes since the start of the shift. (Average delivery time was **31.224267** minutes for all orders.)

   Note: In the task paper there was no mention about the size of the sample we have to use (number of simulations), so I decided to use **100** day simulatons with **7** couriers, which makes the program halt for few seconds as it aproaches the section labeled **Task #3**.

b) The average wait time (time period between the order and expedition) was **69.597902** minutes and the total wait time was **12806.013876** minutes, histogram for these wait times can be seen below:



c) In the case with 7 couriers and no break time between delivery mission, the probability that wait time does not exceed 5 minutes for one order is **15.454449%**. (From the distribution function.)
d) I decided to go with 15 couriers for the optimal number, at first because of the fact that the average number of orders in one hour is 15 and mean value of delivery time being 30 minutes (+ break), then because of the massive cut in order wait times this choice made. For the break length I went with 10 minutes between each delivery mission (which seems quite natural to me, no further mathematical reason).
e) These new values resulted in the total wait time being **220.892746** minutes (down from **12806.013876** minutes - see above) and the average wait time being **0.964597** minutes. In addition to this, the probability that any one wait time exceeds 5 minutes has been extended to **89.947561%**. This is almost *6x* reduction in total wait times with a bit over *2x* increase of wages.