Homework 4

Here we present solutions to the problems posted in the fourth homework assignment. The solutions and related commentary are put in italics. Remember that problems can have several different but correct ways of solving them.

Multiple choice questions

1. The central horizontal line on control charts

   A is computed as the overall average of all available data;

   B is computed by taking the midpoint between the maximal and minimal values in the data;

   C is not affected by values in the data that are close to it;

   ***D** is computed by eliminating the effect of unusually large or small values;

   E is evaluated by properly scaling the standard deviation.

   Solution: *The control charts are designed to detect unusual (from the point of view of chance variation) behavior of the process. Thus the iterative procedure of determining the mean, range and standard deviation of the data is attempting to eliminate the influence of unusually large or small data.*

2. The upper and lower control limits

   ***A** are approximately set up so that more than 99% data falls in between them if process is in control;

   B are approximately set up so that more than 99% data falls in between them irrespectively if process is or is not in control;

   C are based only on the value of the overall mean;

   D do not vary for different subsample sizes in $\bar{X}$ charts;

   E are obtained by using the largest and the smallest data points.

   Solution: *The upper and lower control limits are setup based on the three sigma rule that under the assumption of normal chance variation should give more than 99% data within the control "belt".*

3. The occurrence of an 'out-of-control' point on a control chart

   A is said to be a false alarm if it not as extreme as other cases of 'out-of-control' occurences;

   ***B** is not likely to happend so is considered statistically significant for detecting an assignable cause of variation;

   C is said to be statistically significant only if the process was not properly centred in the first place;

   D is less likely when using "2-σ limits" than when using "3-σ limits";

E is less likely when the process is off centre than when it is properly centred.

Solution: *As discussed in the previous solution, the chances of a process that is under control to go outside the control "belt" are less than one percent. Thus finding a point outside this point is a statistically significant deviation from the control state.*

4. When using numbers of defective items in subgroups sampled from a process as the basis for a control chart, which one of the following is correct?

A *np* charts are so called because the false alarm rates are based on normal probabilities, even though the Normal model is not strictly correct;

***B** *np* charts are so called because the expected number of defects is n times the proportion of defects characteristic of the process;

C the lower control limit in an *np* chart calculated from the standard formula can never be less than 0;

D it is not possible to get a value below the lower control limit when using an np chart because the lower control limit cannot be less than 0;

E it is not possible to calculate an accurate standard error formula for use in an np chart because there is no σ involved.

Solution: *If there is n observations and in each of the chances for a defect are p, then on average we should observe np defects, so that np stands for the average number of defects also know as the expected number of defects.*

Problems

1. A certain process is observed and recorded daily.
   - How improbable is a point outside the 3σ limits when the process is in control?
   - How often will such a point occur when observed daily (find expected frequency in terms of days)?
   - How improbable is a point outside the 2σ limits when the process is in control?
   - How often will such a point occur?

Solution: *We assume that when the process is under control, then it follows the normal models for which chances to go further than 3σ from the mean are 0.2699796% (this precise number was found by R-software but you could find approximately this value from the tables using value $z = 3$), it is less than 3 out of 1000 days when we should observe such a point on average. Deviating 2σ is more probable, namely, 4.550026% (again you can use tables to find this value). This mean that there should be on average less than 46 days out of 1000 that you would see values outside two sigma "belt".*

2. Suppose that for the above process $\bar{X}$ and R control charts have been created based on the subsample means of weekly observation (thus the subsample size is $n = 7$).
   - Using the AIAG chart that is presented in Figure 4.6 of the textbook, identify the values of A_2, D_3 and D_4 for this process (see also our lecture slides where these values have been presented).

- After analysing twelve month data it has been found out that the process has been in control for the entire year and the average value of the range was 6.3 while the mean value was 78. Find the control limits for the $\bar{X}$ and R charts and present them on a graph for a weekly control chart.

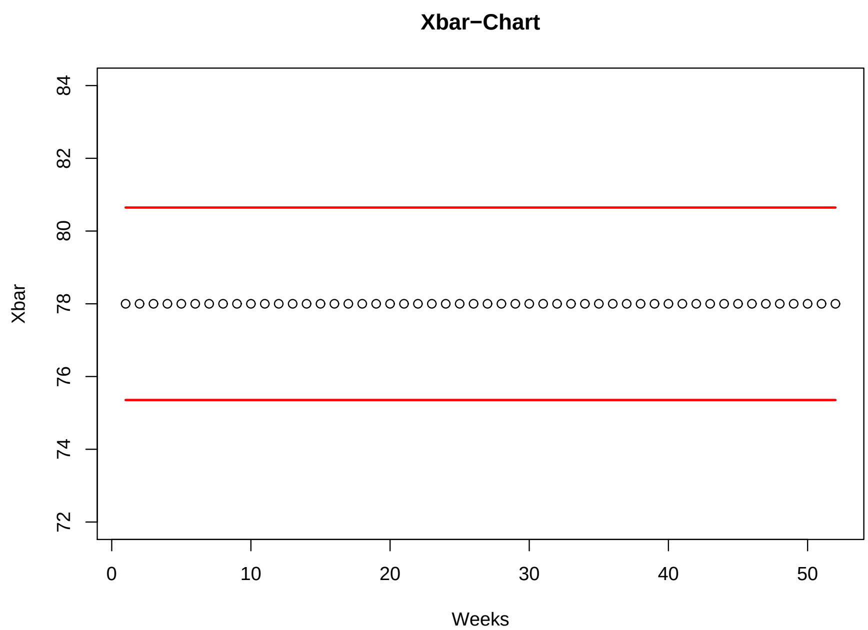
- From the obtained data calculate the standard deviation for this process.

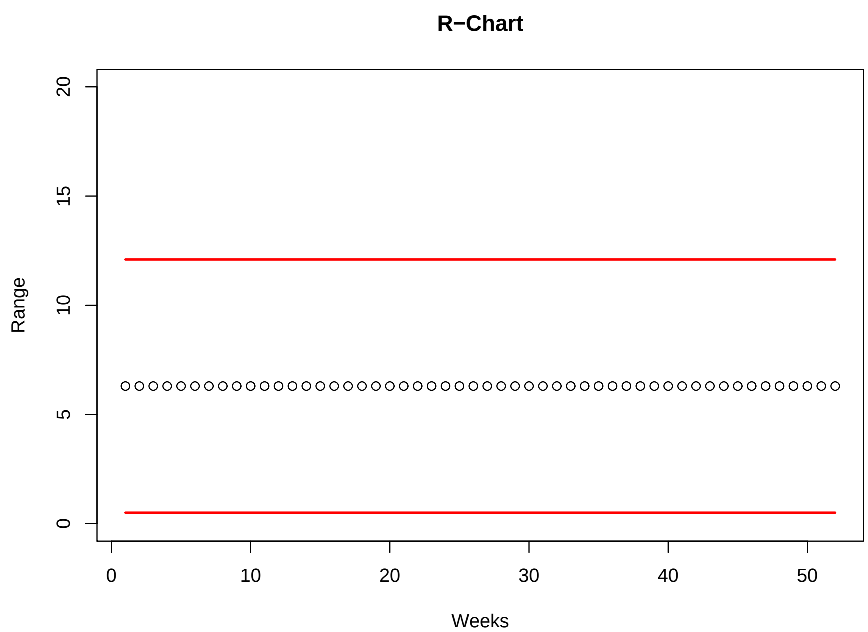
- Using the obtained values compute the percentage of the items that will not fall within the $\bar{X}$ control belt if the process is in control.

Solution:

- *It is easy to find out from the information at the bottom-right corner of the chart that for the sample size* $n = 7$ *the values are* $A_2 = .42$, $D_3 = .08$, $D_4 = 1.92$.

- *The central line for* $\bar{X}$ *chart should be at the overall mean, i.e. at* 78 *and the upper and lower control limits are calculated as follows:* $UCL = 78 + 0.42 * 6.3 = 80.646$, $LCL = 78 - 0.42 * 6.3 = 75.354$, *respectively. Similarly, for the* R *chart, we get the center at* 6.3 *and the upper and lower control limits are* $UCL = 0.08 * 6.3 = 0.504$, $LCL = 1.92 * 6.3 = 12.096$, *respectively. The graphs are presented on the following figures*





- *We recall from the lecture that by the three-sigma rule we have the following relation*

$$A_2 * \bar{R} = 3 * \sigma/\sqrt{n}$$

which in our case translate to

$$0.42 * 6.3 = 3 * \sigma/\sqrt{7}$$

yielding $\sigma = 0.42 * 6.3 * \sqrt(7)/3 = 2.333553.$

- *Finally, if a single item is represented by the value* X *following the normal model with the mean 78 and standard deviation* 2.33*, than chance for* X *to fall outside the control interval* $[75.354, 80.646]$ *are given as* $2(1-0.8719) = 0.2562$*, where by standardization* 0.8719 *is value from the table for* $z = (80.646 - 78)/2.33 = 1.14$*. We conclude that there will be more than 25% individual observations falling outside the control "belt".*

3. Assuming a value of 7.3 [mm] for σ, use the Normal table to predict the proportion of clips whose gaps fail to meet the specification limits of 50[mm] to 90[mm]

   - when the process mean is 74[mm],
   - when the process mean is 67[mm].

   Solution:

   - *The process to meet the specification means that* $(50-74)/7.3 < (X-74)/7.3 < (90-74)/7.3$*, so that we need to find frequencies at which standardized date* Z *are between* -3.287671 *and* 2.191781 *which from the tables can be found as* $0.9857 - 0.0005 = 0.9852$*. So the percentage of items that will fail is* 1.48%*.*
   - *We process similarily and take* $(50-67)/7.3 < (X-67)/7.3 < (90-67)/7.3$*, so that we need to find frequencies at which standardized date* Z *are between* -2.3287 *and* 3.1506 *which from the tables can be found as* $0.9992-1+0.9901 = 0.9893$*. So the percentage of items that will fail is* 1.07%*.*

4. The data on volume measurements for a sample of 50 glasses are given below

```
Sample Run 1
1 503.5
2 507.7
3 506.1
4 505.6
5 504.1
6 504.2
7 504.3
8 503.0
9 504.2
10 507.1
11 500.2
12 509.2
13 506.2
14 501.3
15 506.4
16 506.0
17 506.1
18 505.0
19 504.4
20 504.0
21 505.0
22 502.8
23 502.4
24 506.5
25 504.0
26 503.5
27 502.2
28 502.5
29 505.3
30 507.8
31 507.7
32 506.2
33 506.9
34 502.5
35 505.0
36 505.4
37 507.3
38 506.2
39 507.6
40 506.5
41 504.1
42 506.5
```

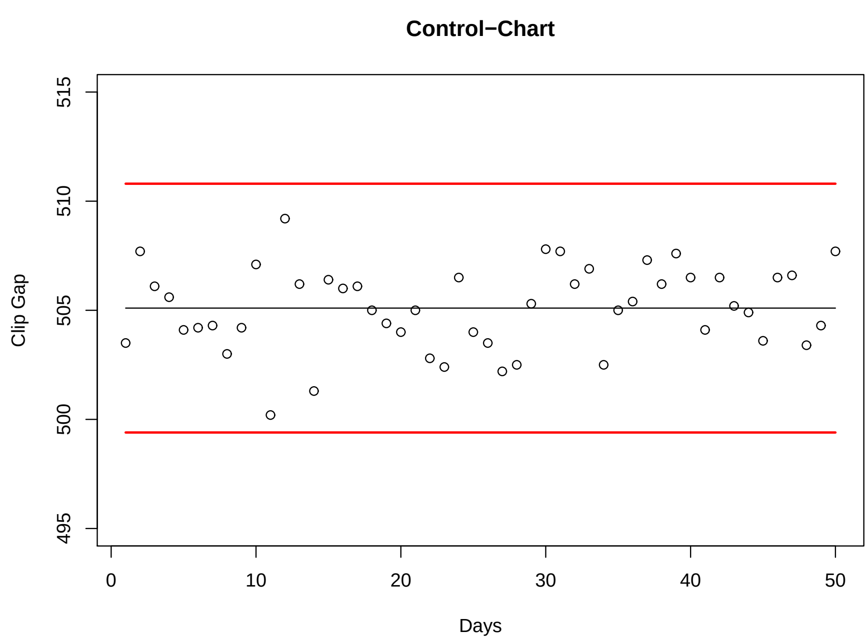
```
43 505.2
44 504.9
45 503.6
46 506.5
47 506.6
48 503.4
49 504.3
50 507.7
```

The mean and standard deviation were calculated as 505.1 and 1.9, respectively.

- Calculate control limits,
- Make a control chart for the individual measurements and plot the above values on it.
- Report on the state of statistical control of the process.

The following values can be found useful when working out this problem: the sum of the above values is 25254.7 and the sum of their squares is 12756174.
Solution:

- *We first look at the data themselves and observe that the values at days 11 and 12 gives the smallest (500.2) and largest (509.2) values, respectively. The value at day 11 is most deviating from the mean. We compute the mean and standard deviation without this extreme value:* $(25254.7-500.2)/49 = 505.1939$ *and* $\sqrt{(12756174 - 500.2^2)/49 - 505.1939^2} = 1.76$. *The corresponding control limits are* $505.1939 - 3 * 1.76 = 499.9139$ *and* $505.1939 + 3 * 1.76 = 510.4739$. *We observe that all the data are within the limits including the removed one so we should not exclude it in the calculation and the three sigma rule for calculating the control limits leads to the limits:* $505.1 - 3 * 1.9 = 499.4$ *and* $505.1 + 3 * 1.9 = 510.8$.
- *The following figure represents the individual data on three-sigma control chart*



.

- *By inspecting the chart for any anusual behavior we conclude that the process is in control.*