Statistical techniques

An explanatory guide to some of statistical techniques, with examples

This document was prepared and issued
by Ciprian Raileanu on 5th April 2005

Pareto Diagram

This technique is used on a large scale because of its simplicity and the power of provided results. It may be used when you want to prioritize resources for a given problem or when you want to segregate major problems from minor problems. This technique comes from Vilfredo Pareto, an economist who in 19th century found that approximately 80% of the wealth and prosperity of Italian population was in the hand of 20% of the population. Further studies in a wide range of industry and problems shown that this rule, 80/20 is perfectly applicable.

A modern translation of this rule may sound like this: 80 percent from the total problems have a root in 20 percent of the total number of causes. Therefore, if you are focusing on those 20% causes and you will somehow prevent them, you will end up in solving 80% from total number of problems.

Applying this technique is simple. Guidance is provided below:

- collection of data related to identified problems (e.g. for 50 batches what are the nonconforming products);
- computing the total number of appearances of problems;
- rearranging the sums in descending order;
- computing the frequency for each type of cause;
- computing cumulated frequency;
- representing the data in the chart;
- interpretation of the chart.

Basic rules for interpreting the chart are:

- the important aspects on which you have to focus are on the left side of the chart;
- a steeper line of cumulating frequencies indicates a break in the range of plotted items. In other word the item in the left has a more crucial importance than the consequent item.

An example for application Pareto diagram is shown below.

Theme definition: Upon an internal audit, a number of non-conformances were found. The Quality Manager is interested in identifying which are the most relevant clauses to which the company is not complying, so that to focus especially on them.

Data collection is shown on the right.

No.	Area	ISO 9001:2000 Clause	Description
1	Sales	7.2.1	Inadequate determination of legal requirements
2		7.2.2	Insufficient determination of customer requirements for orders 5155, 5167, 5168
3		4.2.4	Incomplete records for customer orders no. 5155, 5167, 5168
4		6.2.2	Insufficient knowledge regarding applicable procedures for sales employees
5	Production	7.5.1	Three production employees are not wearing the protection equipment
6		8.2.4	Inspection of the product is not made according to documented specification
7		4.2.4	Record no F0244 is not visible enough
8		4.2.3	Drawing DS234221 not approved and found to be in use
9		4.2.4	Found a set of control sheets stored in inadequate conditions
10		4.2.4	Found control sheet records which are more than two years old, not destroyed yet
11		6.2.2	Training Plan not followed. No training was carried out within last two months

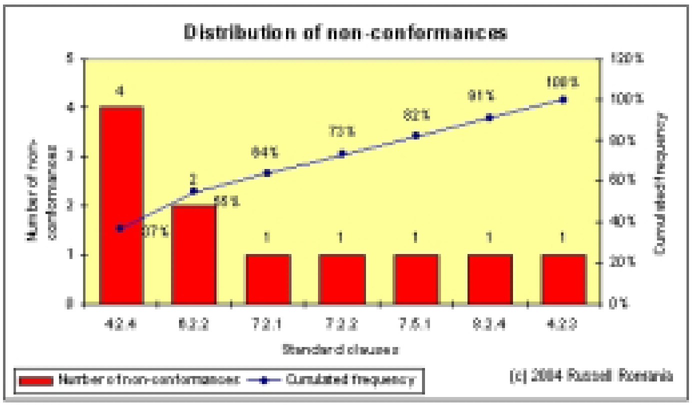
Grouping of collected data is made on standard clauses. The result is shown in the table on the right.

ISO 9001:2000 Clause	Number of nonconformances
7.2.1	1
7.2.2	1
4.2.4	4
6.2.2	2
7.5.1	1
8.2.4	1
4.2.3	1
Total	11

Sorting data in descending order and computing the individual and cumulated frequencies lead to:

ISO 9001:2000 Clause	Number of nonconformances	Frequency	Cumulated frequency
4.2.4	4	37%	37%
6.2.2	2	18%	55%
7.2.1	1	9%	64%
7.2.2	1	9%	73%
7.5.1	1	9%	82%
8.2.4	1	9%	91%
4.2.3	1	9%	100%

Representation of data is shown in the following diagram. As was stated initially, this type of chart helps you to prioritize things. Analyzing the chart induces the conclusion that the organization is facing two major problems. The biggest one is related to control of records and the second one is related to human resource management, in terms of training. All the other problems can be treated as incidental, however, without diminishing their importance. What the company has to do as a priority is to find solutions for proper recording of data and secondly, to do something for better knowledge of tasks and responsibilities. What's interesting is that improving these areas, other problems will be corrected almost automatically, like use only of approved documents and better documentation of customer requirements (these are as a result of training process, where more awareness will appear). This is a good example of better understanding of 80/20 rule (solving a main problem – left side of the chart – also solves other smaller problems).



Pay special attention to the following: the diagram is showing what is the "spread" of an issue, not strictly its gravity!!! To understand better this idea, lets' go back to our example. Is it a bad thing that several records are not as they should be? Obviously, the answer is yes. Is it a bad problem that employees does not know as it should be normal what they have to do? Obviously, once again the answer is yes. But which one of these two problem is worst than the other? Tough question... As was stated in the above paragraph, lack of training may lead to bad records. Therefore, training may become a priority of the company. Another aspect... How bad is that workers didn't wear their working equipment? Usually, this is a big problem, because it may lead to accidents, also to severe penalties for that. Even if in the diagram that was mentioned on the right side, the gravity of the problem is huge, therefore, it has to be corrected somehow...

Control Charts

Control charts are used widely in order to determine whether a process is or is not under statistical control. In other words, control charts are used to understand how a process is evolving. Basic fundament for control chart is the representation of collected observation on a chart, along with drawing the statistical limits. There is a fundamental difference between statistical limits of control and specification limits (these are usually imposed by customer requirements).

In order to use control charts, some terms must be understood:

- **Population** – represents the overall number of "items". (e.g. all products made during a working shift);
- **Sample** – a set of "items" extracted from the population. In theory, in order to get better results for statistical analyze, a larger number of samples should be taken. Most often, this is not practical, either due to high costs, destructive methods for testing, or simply small batches. Recent studies (Central limit Theorem) indicate that a sample size of 30 or more is sufficient for analyzing the population.
- **Mean** – the average of observation;
- **Standard deviation** – represents a measure of the variation of data.

Please remember that a batch represent a group of the same kind of goods, made under the same conditions. Conditions include workers, machines, environment, measurements, materials used, and working methods. As an example, if a factory is running with the same machines, at the same parameters, under the same environmental conditions, with materials of the same kind, and verifies the products in the same way, but a change in working shifts occurs, then, the batch is ending upon working shift change.

The result of a sampling process may lead to individual observations or to subgroups. Conditions where samples result in individual data are (no particular order):

- long time between sampling;
- cycle of production is very long;
- no correlation to time of observed data;
- expensive process of sampling;
- destructive method for inspection;
- each sample comes from a distinct batch;
- each sample is uniquely identified with a certain time period.

Where practical, samples may by grouped, resulting what is called subgroups. Analyze can be made working with subgroups. Example: from 20 batches of products, are randomly extracted 5 items. Each group of 5 items, extracted from a batch represents a subgroup.

If the number of subgroups is less than 5, Xbar - R charts must be used (Xbar stand for mean of each subgroup, R stands for Range of Subgroups). For subgroups larger than 5, Xbar - S charts must be used (S stands for Standard Deviation of Subgroups). Both types of charts measure the variability of the process. As a particular observation, S charts are more sensitive for small process mean shifts. S chart is preferred tool when the production output is high and collection of data is simple and/or inexpensive. Based on the high sensitivity, S chart is used when a more rigorous control is needed.

For individual observations, where no grouping is possible, I - MR charts are used (I stands for Individual and MR stands for Moving Range).

Statistical control limits are determined as a principle using the +/-3 sigma from the mean, where sigma is the standard deviation of data. After plotting the data and the control limits in the chart, if there is any point outside control limits, the process is not under statistical control, and there are special causes which lead to it. However, if all the points are within control limits, there may still exist a lack of control when:

- 9 consecutive observations on the same side of center line;
- 6 consecutive observations, all decreasing or increasing;
- 14 consecutive observations alternating up and down;
- 2 out of 3 consecutive observations, more than 2 sigma on the same side of central line;
- 4 our of 5 consecutive observations more than 1 sigma from the center line, on the same side;
- 15 consecutive observations within 1 sigma from center line, no matter which side;
- 8 consecutive observations more than 1 sigma from the center line, mo matter which side.

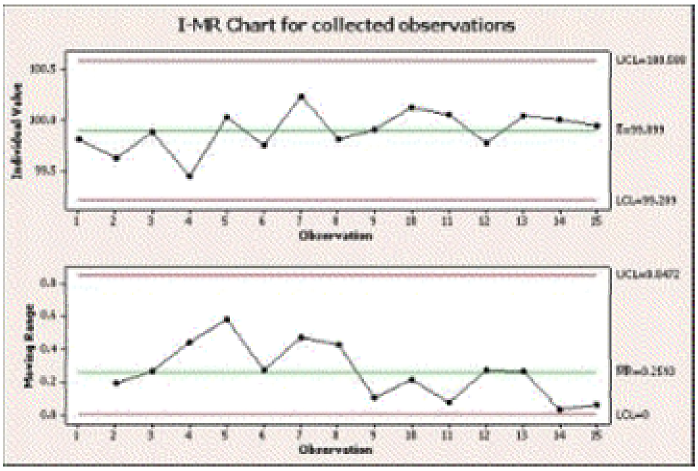
Mentioned values can be customized depending on the purpose of the study. As an example, for the first mentioned test, increasing the number of observation (let's say 15 instead of 9) will lead to a lower precision of determining the lack of control over the process.

An example of I-MR chart is presented below. A company is manufacturing high precision tubes. Quality Control department is interested to determine if the production process in under control. For simplicity, we assume that from each batch is measured the outer diameter of a random selected tube. Measured data are shown in the following table.

No.	Observation
1	99.82
2	99.63
3	99.89
4	99.45
5	100.03
6	99.76
7	100.23
8	99.81
9	99.91
10	100.12
11	100.05
12	99.78
13	100.01
14	100.04
15	99.95
Mean = 99.90	
Sigma I = 0.176507	
Sigma MR = 0.169733	

Results of needed computations are included on the last row of the table.

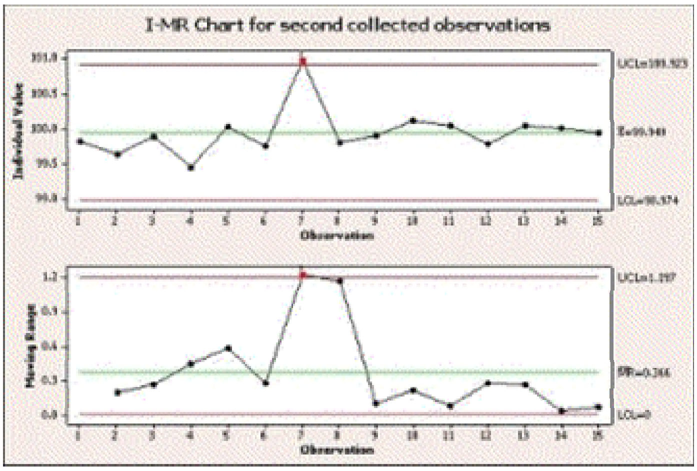
Analyzing I chart and MR chart (below) is observed that all points are within control limits determined as +/-3 sigmas from the center line. This indicates the existence of a control in the process that produces those tubes.



Let's assume that one of the measured values at position 7 is 100.98 instead of 100.23. Graph has changed as shown below:

No.	Observation
1	99.82
2	99.63
3	99.89
4	99.45
5	100.03
6	99.76
7	100.98
8	99.81
9	99.91
10	100.12
11	100.05
12	99.78
13	100.01
14	100.04
15	99.95
Mean = 99.95	
Sigma I = 0.297536	
Sigma MR = 0.286116	

In this case, I chart shows that at observation no 7, the measured value falls outside the upper control limit. The MR chart shows a very close value to the upper limit. As a conclusion, measured data indicate a process outside of control, meaning that a special cause appeared. Supplemental investigation is needed in order to determine what caused that value, outside control limits.



Pay special attention that using control charts is investigating the variation on the process, in other words if it's in control or not. Control charts are not used to assess if measured component falls within the specification imposed by customer or not. You noted that within this paper, no values were mentioned regarding the specification limits...

Process Capability

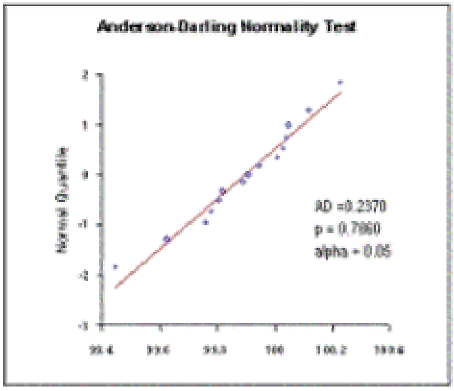
In previous section we have investigated how wide a process is performing, what is its variation. In other words, we investigated whether there exist a control or not for that process.

Establishing that a process is under control, a normal question arises: on a long term, how measured data fit customer requirements? Here we face the specification limits, those values which are acceptable for the client. How we will do that? By using process capability studies. Please note that it makes no sense to investigate process capability as long as the process in not under statistical control. The existence of the control over the process is a fundamental assumption for process capability study.

Before performing process capability analyze, it's important to determine the normality of data. All studies of process capability rely on the assumption that data are normally distributed, because the estimation of nonconforming fraction from population is determined using standard normal tables. If analyzed data does not follow a normal distribution, there is the risk of overestimating or underestimating the proportion of nonconformities. After all, it's nothing wrong in overestimating the proportion of nonconformities. The worst scenario is the underestimation of it, resulting in more nonconformities that predicted. In other words, capability study indicates a capable process, when in fact is not capable.

Normality of data can be determined in several ways, described in many statistical books. Usual methods for determining the normality are Darling – Anderson test, chi-squared goodness-of-fit, Kolmogorov-Smirnov (pretty similar to Darling-Anderson test), normal probability plot, and with a little more precaution, histograms. The basic of normality is the following: plotting the data on a normal probability graph, those data should fit a straight line. If plotted points do not follow the line, then you cannot assume that data is normally distributed. More investigation should be carried, in order to determine if data comes from a non-parametric distribution (binomial, Weibull, etc.). Where applicable, data may be transformed using a Box-Cox transformation, but this has to be done so that to make sense (square root of data, inverse of data). A coefficient of let's say 2.35 has no practical utility, being intended only for theoretical use.

Making as input data those from Control Charts, example one, the Darling-Anderson test value is 0.237 and the p value is 0.786. Assuming the alpha level of 0.05 (which is common practice), and observing that p value is greater than selected alpha level, we can conclude that data is normally distributed. A representation of data distribution is shown on the right.

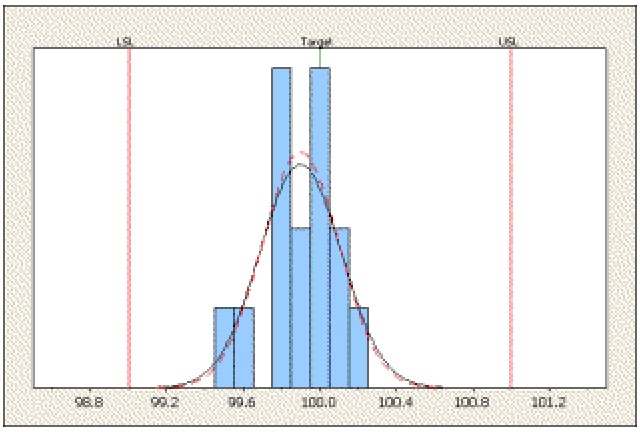


Passing these two mandatory conditions for given example (process under control and normal distribution of data), we can determine the process capability. First of all, we must be aware of specification limits, imposed by the customer. For those precision tubes, lower specification limit (LSL) set by the client is 99.00 and upper specification limit (USL) is 101.00. What customer ideally wants is a tube with outer diameter of 100.

In a simplified method of computing capability indices, we will have to correct the standard deviation, because the number of observation is less than 60. If you are using Minitab or other equivalent software products, all computation is performed automatically, in a more complex manner. For given set of data, standard deviation (sigma - σ) is 0.1991. Needed correction is $\sigma/c4$, where c4 is 0.9823. Corrected sigma is 0.2027.

First capability index is Cp, which represent the potential performance of the process, by dividing the tolerance interval to natural process spread. The relation is $Cp = (USL - LSL)/(6\sigma)$. Remember that sigma is that which was corrected with c4 coefficient. The result is Cp = 1.64. As a conclusion, the process is potentially capable of performing at a level of 1.64.

The actual performance of the process is computed by selecting the minimum between C_{pU} and C_{pL}, which are capability indices which take in consideration the physical limits imposed. C_{pU} and C_{pL} are determined as follow: C_{pU} = (USL – mean)/(3σ) and C_{pL} = (mean – LSL)/(3σ). Doing the computation is resulting C_{pU} = 1.81 and C_{pL} = 1.48. Minimum between these two values is 1.48, which means that C_{pk} = 1.48. This is the real (actual performance of the process).



From where comes the difference between potential and actual performance of the process? The answer is simple. C_p does not take in consideration the mean of the process, which is a real thing. In other words, C_p "believes" that the process is perfectly centered, while C_{pk} makes a difference when sensing the mean of the process in correlation with specification limits.

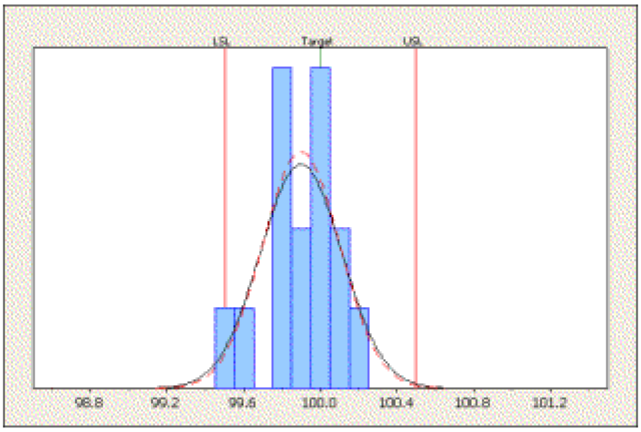
One more thing... The customer wants ideally to get those tubes with the outer diameter of 100. This is the target T. We can compute the capability indices for the target, using the formula $C_{pm} = C_p/(1+((mean-T)^2)/(\sigma^2))$. Doing the computation is resulting C_{pm} = 1.48. s

What is the significance of these indices? Indices are correlated with number of defects! The following table gives these correlations:

C_p	Capability
0.50	86.64%
0.62	93.50%
0.68	96.00%
0.75	97.50%
0.81	98.50%
0.86	99.00%
0.91	99.35%
1.00	99.73%
1.33	99.964%

If a capability index is 1.33, the capability of the process is 99.964%; is expected that 0.036% of the process to produce nonconforming products. This is equivalent of producing 36 nonconforming products within one million products produced. Many industries works with a value of capability index of 1.33 as a minimum requirement. Top companies are using minimum level set to 1.66 or even 2.

What would happen if customer requirements are tighter than those initially imposed? Let's assume that the client wants a specification range of 99.5 … 100.5. Redoing computations, results are: C_p = 0.82, C_{pk} = 0.66 and C_{pm} = 0.74. A drastic reduce-tion in process performance is observed, resulting approximately 26,100 non-conforming products from a million produced. Analyzing the chart is observed that each tail of the distribution curve fall outside specification limits (more on the left tail, corresponding to lower specification limit). What would be the solution for complying with customer requirements? They will have to improve the process, in terms of narrowing the variation of data near the target specification. Yes, this will imply some costs, possible in terms of equipment, training, process technology, etc.



What initially appears to be a loss in business is turning to be a gold mine. As an example, if a tube is sold with a price of 15$, at a monthly production of 700,000 pieces and investment in improvements costs 200,000$, here is the saving:

- number of defective products for given production is 700,000 x 26,100 / 1,000,000 = 18,270, with a total cost of 274,050$;
- cost of investment: 200,000$;
- after the investment, in the first month, the profit is 74,050$;

- in next months, savings are turned to the equivalent in bad defects, prior the implementation of improvement.

Note: assumption is that investment in improvement will lead to zero defects. However, this is not possible all the time. Just think that investment in improvement will make a saving of 25%. On a month, this means 274,050 x 25% = 68,512$. On a year, savings are 822,144$. Intelligent stuff, right?

Entire previous discussion was for process capability, where, common practice imply the use of 6 sigmas. For machinery capability, common practice impose an 8 sigmas. The way of performing the studies is pretty similar to process capability.

Statistical Control for Attribute Data

Life is not made only by numbers. Many industries are performing testing and inspection focused on aspects like Good – Bad, Go – No Go, Pass – No Pass. These industries are using attributes to define the quality of their products. Several statistical methods were developed for this kind of data. As fundamental idea, these methods are relying on method for numeric data.

Most used statistical methods for attribute data are the control charts, like P chart, NP chart, C, chart, and U chart. P and NP charts are used for defectives, while C and U charts are used for defects. Please note the difference in used terms: defective and defect. A defective product is one which is not acceptable in terms of quality, while for complex products, a defect of a product does not necessarily lead to a defective one.

Charts for defectives	Charts for defects
P chart shows the proportion of defectives in each analyzed subgroup	C chart used when the subgroup size is constant, for the number of defects in each analyzed subgroup
NP chart shows the number of defectives in each analyzed subgroup	U chart used when the subgroup size is not constant, for the number of defects per sampled unit in each analyzed subgroup

Each type of chart is detecting the presence of special causes of variation, indicating either the control or lack of control in process. As is for numerical data, the strongest indication of lack of control is finding an observation which falls outside of +/- 3 sigmas from the mean of the process. Please note that we didn't mention the specification limits; the main use of control charts are to determine variation of the process related to the mean.

P Chart

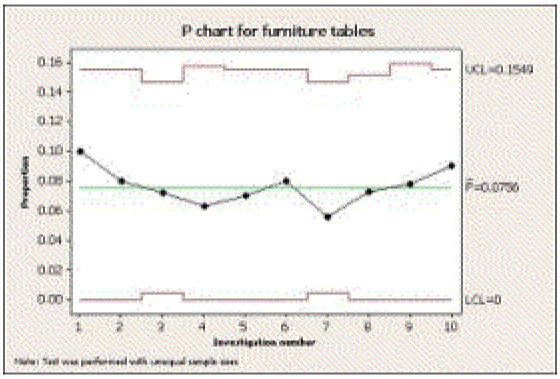
The proportion of defectives in each subgroup is shown in this type of chart and indicates the lack of control in the process. Suppose that a furniture factory produces tables. The surface of it should be free of defects, meaning no scratches on it. Sampling the batches, quality inspector collected the following data:

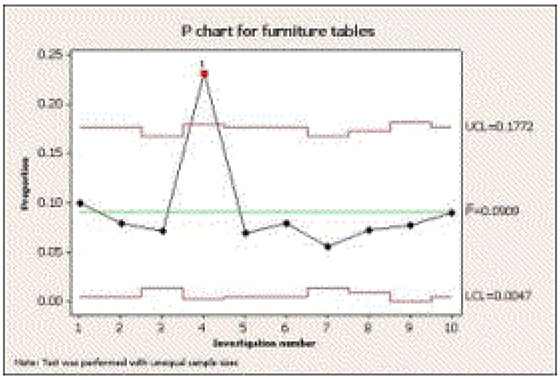
Batch no.	Batch size	Number of defectives
1	100	10
2	100	8
3	125	9
4	95	6
5	100	7
6	100	7
7	125	7
8	110	8
9	90	7
10	100	7

Needed computations are:

- divide the number of defectives to batch size, thus resulting the proportion of detectives in each batch (10/100 = 0.1, 8/100 = 0.08, 9/125 = 0.072, etc.);
- sum separately the number of all batches and number of defectives (e.g. sum of batches is 1045 and sum of defectives is 79);
- divide the sum of defectives to the sum of batches and the result is mean of the proportion of defectives (in the example, the average is 0.07559);
- compute the lower and upper control limits (LCL and UCL) using the formulas: LCL = p -3 sqrt(p(1-p)/n_i), where n_i is the batch size and p is the mean computed previously. UCL = p +3 sqrt(p(1-p)/n_i);
- if LCL is lower than 0 set it to 0. If UCL is greater than 1, set it to 1;
- plot the computed data on the chart and observe if there is any observation that fall outside the control limits.

P chart is shown above, near the table. Because all results fall within control limits, is concluded the existence of a control in production process. Assuming that run number 4 indicates a number of defectives of 22, for the same batch size of 95, the resulted P chart is shown below. Is observed that at inspection number 4, the proportion of defectives go beyond the upper control limit. This indicates a lack of control, and is a strong suggestion that a 100% inspection of batch no. 4 should be undertaken.





NP Chart

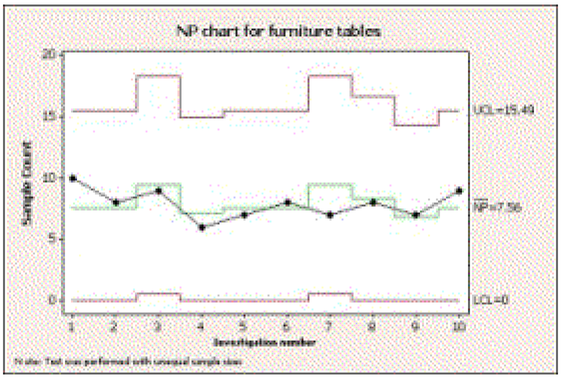
NP chart plots the number of defectives and indicate the lack of control in the process. For a better understanding of this type of chart, let's work on the example for P chart as shown below:

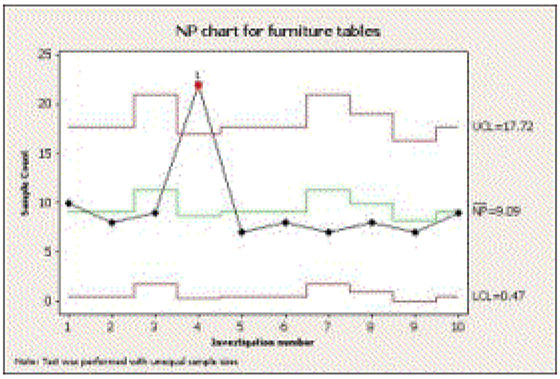
Batch no.	Batch size	Number of defectives
1	100	10
2	100	8
3	125	9
4	95	6
5	100	7
6	100	7
7	125	7
8	110	8
9	90	7
10	100	7

Needed computations are:

- compute the proportion resulted by dividing the number of defectives to batch size (pi = xi/ni, where xi is the number of defectives, ni is the size of batch);
- compute the center line value, by multiplying each resulted proportion with batch size (ci = pi * ni);
- compute the lower and upper control limits as follow: LCL = ci – 3 * sqrt(ci * (1- pi)), UCL = ci + 3 * sqrt(ci * (1- pi));
- if LCL is less than 0, set it to 0. If UCL is greater than ni, set it to ni;
- plot the data on the chart and observe if there are any points that fall outside the control limits.

Resulted chart is shown above, near the table. Analyzing the chart is observed that no run fall outside the control limits, which is an indication of good control in the process. If we do the same alteration of data as in the example of P chart, by setting the defective number to 22 for run no. 4, the resulting chart shown above on the right. There is evidence that a lack of control does exist. Further investigation of causes of lack of control is needed.





C Chart

This type of chart is used for plotting the number of defects (not defectives) in each subgroup, when the subgroup size is the same for each run. It also detects the presence of special causes of variation. Just to remind you: a complex product is not necessarily a defective one, even if some defects still exists.

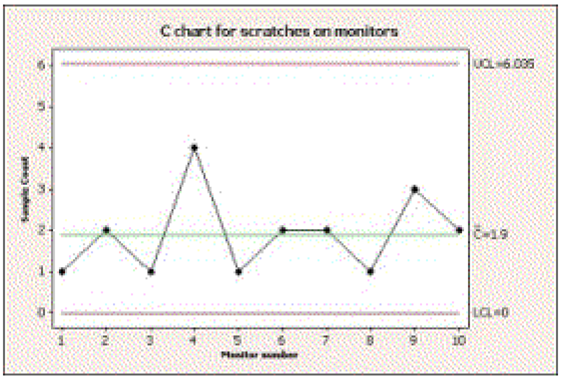
An example where C chart is used is the following: a refurbishing computer monitor company controls the polishing of monitor screens, in order to detect the surface of scratches in the entire visible surface of the monitor. An important assumption is that all monitors are the same size, let's say 17". What the quality control is doing is to determine surface scratched after refurbishment. The result is reported to entire surface of monitor. There is a certain limit of permitted scratches; if that limit is exceeded, the monitor is declared low grade and sold out to a lower price. An example of collected data is the following:

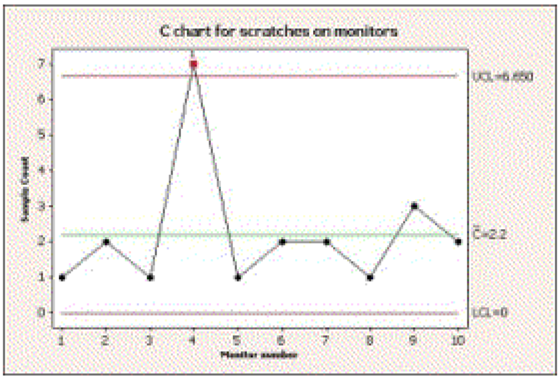
Monitor number	Scratches [%]
1	1%
2	2%
3	1%
4	4%
5	1%
6	2%
7	2%
8	1%
9	3%
10	2%

Needed computations are:

- mean of observations is the average of collected observation (sum of all defects, divided to number of investigations (e.g. [1+2+1+4+1+2+2+1+3+2]/10 = 1.9);
- lower control limit is determined using the formula LCL = mean – 3 * sqrt(mean), while upper control limit is LCL = mean + 3 * sqrt(mean);
- if lower control limit LCL is less than 0, set it to 0.

Resulting shows that collected data fall within control limits, meaning that the process is under control (the above graph, near the table). Supposing that monitor number 4 shows a scratched surface of 7 percent, the resulting chart shows the observation number 4 outside the control limits, meaning a lack of control in polishing process (see the diagram on the right). Further investigations have to be carried out, in order to determine the cause of that problem.





U Chart

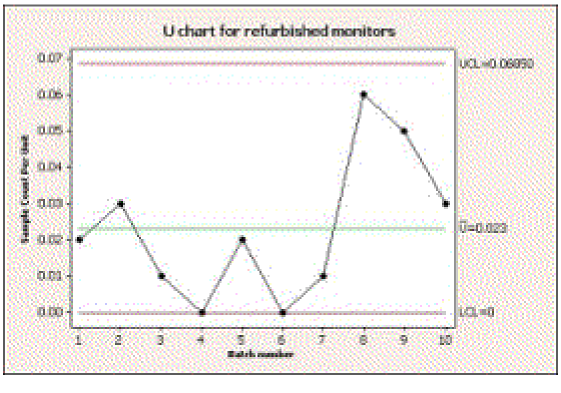
This type of chart plots the number of defects per sampled unit and indicate the lack of control in the process. Let's go to the same example of refurbishing computer monitors company. The monitor is assembled, containing main parts like screen, case, wires, desk support and electronic. Each of these components is checked by quality control personnel. For each monitor, the number of defects found is recorder. One more assumption: the size of each subgroup is 100, for each of them 2 monitors is checked.

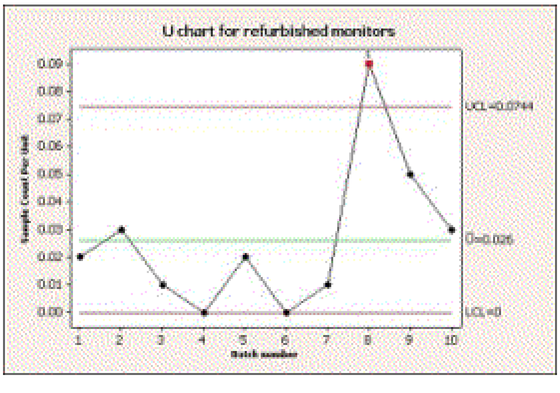
Test no.	Scratched tubes	Monitor case	Connecting wires	Electronics	Desk support	Total of defects
1	0	0	1	1	0	**2**
2	1	0	0	1	1	**3**
3	0	0	0	0	1	**1**
4	0	0	0	0	0	**0**
5	0	1	0	0	1	**2**
6	0	0	0	0	0	**0**
7	1	0	0	0	0	**0**
8	1	2	0	3	0	**6**
9	2	0	1	2	0	**5**
10	1	0	0	2	0	**3**

Needed computations are:

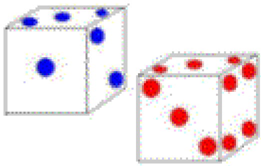
- percent of defects of each subgroup (e.g. u1 = 2/100 = 0.02, u2 = 3/100 = 0.03, etc.);
- mean of the process, as sum of all defects, divided to sum of all sample sizes (e.g. mean = [2+3+1+0+2+0+1+6+5+3]/[100+100+100+100+100+ 100+100+100+100+100] = 0.023;
- lower control limit, using the formula: LCL = mean – 3 * sqrt(mean/n_i), where n_i is sample size (e.g. LCL = 0.023 – 3 * sqrt(0.023/100) = - 0.0225. If LCL is less than 0, set it to 0;
- upper control limit, using the formula UCL = mean + 3 * sqrt(mean/n_i), where n_i is sample size (e.g. UCL = 0.023 + 3 * sqrt(0.023/100) = 0.0685.
- plot the data on the chart and observe if there is any data falling outside control limits.

Resulting shows that collected data fall within control limits, meaning that the process is under control (the above graph, near the table). In order to point the lack of control, let's assume that for test number 8, total number of found defects is 9 (2 scratched tubes, 3 cases, 1 wire, 3 electronics, 1 desk support). Redoing the computations and plotting the chart (on the right) we observe that for batch number 8, there is evidence for a lack of control (too many defects compared with the average rate of them). Investigations should be undertaken, in order to find the cause of the problems.





Dice Experiment



Variation of a process is not a positive aspect in financial terms. To prove this theory, let's take the example of dices. All of us know that a dice has 6 faces, each of them having a number from 1 to 6, arranged so that on opposite faces, the sum of those numbers is 7 (or the average of these two numbers is 3.5). OK, let's take the dice and roll it 20 times. Record each run of the dice (in our experiment, it's the first column in the table, namely Regular Dice).

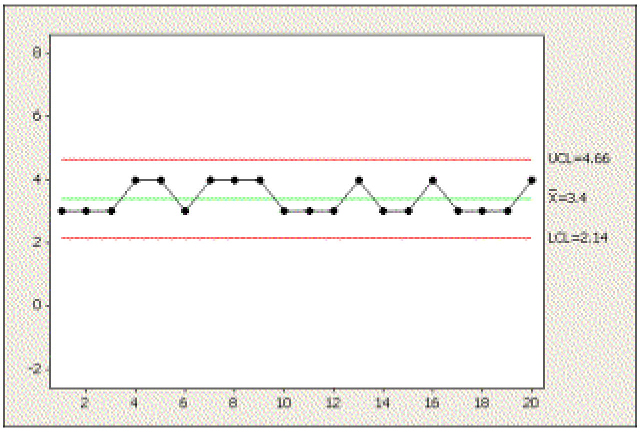
Regular dice	Modified dice
1	3
4	3
1	3
2	4
2	4
3	3
5	4
4	4
1	4
4	3
5	3
5	3
5	4
1	3
5	3
1	4
2	3
5	3
5	3
3	4

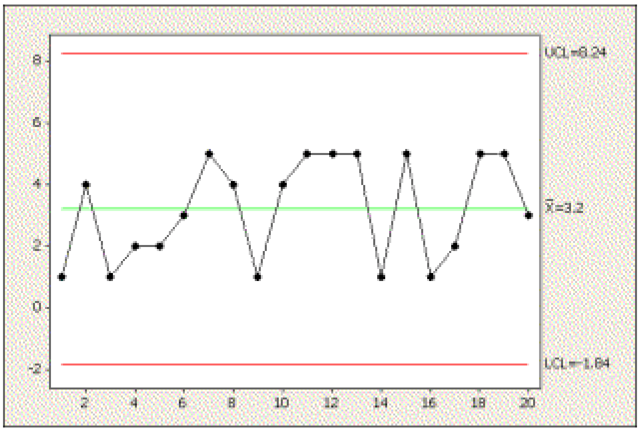
Now, let's alter the dice, so that to keep the same rule, but using the closest numbers to average value of two opposite faces of regular dice. These numbers will be 3 and 4. So, on each pair of two opposite faces of the dice, we will have the values 3 and 4. Now let's make the same number of runs as in the first case and record the values (second column in our experiment, namely Modified Dice).

Let's analyze the performance of each trial. Computed statistics (the average of runs and the variation of them) are presented in the following table:

	Regular dice	**Modified dice**
Mean (average)	3.2	3.4
Variation (stDev)	1.673	0.503

Plots of I-charts for each series are presented below (left chart for Regular Dice, right chart for Modified Dice):





Analyzing the charts and computed statistics lead to following conclusions:

- the first impression when comparing the average of each dice is that they are close to each other, in term of means (averages). There is a 6.25% difference in favor of Modified Dice. In terms of variation, a highly one is observed for Regular Dice, while for the Modified Dive the variation is small.
- both of I-charts show a controlled process, because no observation fall outside the control limits;
- both charts were plotted on same scales, in order to present visually the differences for variation. Is clearly observed that modified dice have a much lower variation than regular dice. These charts underline the need of lowering the variation in the process, with gains in terms of efficiency (means comparison).

Now, let's try to apply the data to a real situation. A truck transporting company has two loaders, first of them is loading the truck with quantities described in column 1, and the second one with quantities described in column 2 (all of these because skills and qualification of loader operators. Transportation of product is made on the same length of road, with the same driver, under the same conditions. What are the benefits? Because of lower variation in loading provided by the second loader, the company makes benefits in terms of either duration, or costs, or profits. That difference of means of 6.25% is in the favor of the second alternative, where variation is smaller. Assuming that in one year of operation, the profit made by the company using the first loader is 100.000$, in the second alternative the profit is 106.250$. Is that good? Of course...

The basic rule learned from this example is: narrow the variation of the process and expect an increase of profits.

Some of you may say: it's possible that in twenty runs, with the first dive, fifteen times of get the number 6. Of course, the mean of the process will be higher than with Modified Dice, but this is pure luck. Every business should be run so that to make that product or service with as smaller variation of characteristics as possible. This way, high confidence will arise meaning that established objectives will be achieved. With a higher variation, reaching planned objectives is more risky.