OPRE 6302. OM : Quality

1 Computing the Standard Deviation of Sample Means

Quality control charts are based on sample means not on individual values within a sample. A sample is a group of items, which are considered all together for our analysis. Items within a sample lose their individual characteristics in the analysis. Rather a summary statistic, e.g. sample mean, is used to represent the information in the sample. See the examples of samples below:

1. A section of Operations Management students in the current semester is a sample of students. Then the sample size is the number of students in the section. Different sections constitute different samples. The number of sections offered in the current semester would be the number of samples.

2. Voters surveyed by a given polling agency on a single day is a sample. The sample size is the number of voters surveyed on that particular day. Polls made on different days constitute different samples. The number of the polls is the number of samples.

3. Customers buying a particular brand of perfume over a specified month can be considered as a sample. The sample size is the number of customers buying the perfume over the specified month. Another sample can be generated by considering customers buying another brand of perfume. If we consider four brands of perfumes, we end up with four samples.

The number of samples and the sample size can potentially be confusing. Sample size is the number of items within a group. Number of samples is the number of groups.

Example 1: After a midterm exam for a course that is given to five sections of a course, the average exam grade $\bar{x}_j$ in section j is computed and reported below.

	Sec 1	Sec 2	Sec 3	Sec 4	Sec 5
Average grade	68	72	74	82	71

Suppose that there are 50 students in each section and use $x_{i,j}$ to denote the ith student's grade in Sec j. Then the average grades are computed by

$$\bar{x}_j = \frac{\sum_{i=1}^{50} x_{i,j}}{50} \qquad \text{for } j \in \{1, 2, 3, 4, 5\}.$$

Since all 50 grades within a section are reduced to a single summary statistic (the sample mean), all the students within a section are represented merely by the section's summary statistic (the sample mean); Individual student grades are immaterial for an analysis that checks if a certain sectin is performing better than the others. Clearly, the sample size is 50 and the number of samples is 5. □

There are two ways to compute the standard deviation $\sigma_{\bar{x}}$ of sample means. The first way requires the knowledge of the standard deviation σ_x of the individual values within a sample, the second way does not require σ_x.

1.1 Computing $\sigma_{\bar{x}}$ with known σ_x

In order to understand what we have and what we want, first recall that

$$Var(X) = \sigma_x^2 \text{ and } Var(\bar{X}) = \sigma_{\bar{x}}^2.$$

Note that $Var(X)$ is known and we want to compute $Var(\bar{X})$.

In order to perform this computation, we need to recall the following lemma from statistics:

Lemma 1. *i) If X is a random variable and c is a constant, then $Var(c \cdot X) = c^2 \cdot Var(X)$.*
ii) If X_1 and X_2 are two independent random variables, then $Var(X_1 + X_2) = Var(X_1) + Var(X_2)$.

Proof: i) First convince yourself that the mean of cX would be $c\bar{x}$ where $\bar{x}$ is the mean of X. We start with $Var(c \cdot X)$ and use the definition of variance

$$Var(cX) = \frac{1}{n}\sum_{i=1}^{n}(cx_i - c\bar{x})^2 = c^2\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = c^2 Var(X).$$

ii) Again by using the definition

$$\begin{aligned}
Var(X_1 + X_2) &= \frac{1}{n}\sum_{i=1}^{n}(x_{1,i} + x_{2,i} - \bar{x}_1 - \bar{x}_2)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\{(x_{1,i} - \bar{x}_1)^2 + (x_{2,i} - \bar{x}_2)^2 + 2(x_{1,i} - \bar{x}_1)(x_{2,i} - \bar{x}_2)\} \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_{1,i} - \bar{x}_1)^2 + \frac{1}{n}\sum_{i=1}^{n}(x_{2,i} - \bar{x}_2)^2 + 2\frac{1}{n}\sum_{i=1}^{n}(x_{1,i} - \bar{x}_1)(x_{2,i} - \bar{x}_2) \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_{1,i} - \bar{x}_1)^2 + \frac{1}{n}\sum_{i=1}^{n}(x_{2,i} - \bar{x}_2)^2 + 0 \\
&= Var(X_1) + Var(X_2)
\end{aligned}$$

The fourth equality is due to the fact that X_1 and X_2 are independent so the sum of the cross products is zero. This sum would be the covariance of X_1 and X_2, if X_1 and X_2 were not independent. □

Now Lemma 1 can be used to relate the variance of the sample mean to the variance of the observation within the samples. We start with the definition ofthe sample mean, proceed as follows

$$\begin{aligned}
Var(\bar{X}) &= Var\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) \\
&\overset{Lem.1.i}{=} \left(\frac{1}{n}\right)^2 Var\left(\sum_{i=1}^{n}X_i\right) \\
&\overset{Lem.1.ii}{=} \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n}Var(X_i) \\
&= \frac{n}{n^2}Var(X) \\
&= \frac{1}{n}Var(X) \qquad (1)
\end{aligned}$$

where we use the fact that each individual observation has the same variance as the other individuals: $Var(X_1) = Var(X_2) = Var(X_i) = Var(X)$ where X stands for a generic observation and represents one of $X_1, X_2, \ldots X_n$. This fact is assumed when constructing samples; otherwise, we would be grouping "apples" with "oranges".

Given (1) which relates varainces, relating the standard deviations is easy. Just take the square root of the both sides in (1) to arrive at

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{n}}\sigma_x. \tag{2}$$

Example 2: Refer to Example 1 and suppose that the indivudual scores has a standard deviation of 20, compute the standard deviation of the sample means.
Solution: We are given $\sigma = 20$, sample size is already known as $n = 50$. Then by using (2),

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{n}}\sigma_x = \frac{1}{\sqrt{50}}20. \ \square$$

1.2 Computing $\sigma_{\bar{x}}$ with unknown σ_x

This method is rather direct; Without σ_x, the only information available is the population of the sample means $\{\bar{x}_1, \bar{x}_2, \ldots \bar{x}_m\}$ where the number of samples is denoted by m. We could use this population to estimate the standard deviation of the sample means. First let us compute the variance:

$$Var(\bar{X}) = \frac{1}{m}\sum_{j=1}^{m}(\bar{x}_j - \bar{\bar{x}})^2$$

where $\bar{\bar{x}}$ is the grand mean which can be computed by

$$\bar{\bar{x}} = \frac{1}{m}\sum_{j=1}^{m}\bar{x}_j.$$

Finally the standard deviation of the sample mean is

$$\sigma_{\bar{x}} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}(\bar{x}_j - \bar{\bar{x}})^2}. \tag{3}$$

Example 3: Refer to Example 1 and compute the standard deviation of the sample means from the population $\{68, 72, 74, 82, 71\}$.
Solution: First we compute the grand mean

$$\bar{\bar{x}} = \frac{1}{m}\sum_{j=1}^{m}\bar{x}_j = 73.4.$$

Then the standard deviation of the sample means by (3) is

$$\sigma_{\bar{x}} = \sqrt{\frac{1}{5}\{(68-73.4)^2 + (72-73.4)^2 + (74-73.4)^2 + (82-73.4)^2 + (71-73.4)^2\}}.$$

1.3 Remark

When σ_x is unknown, you must use (3) to compute $\sigma_{\bar{x}}$. In this case, you do not have any choice. When σ_x is known, you have to choose between equations (2) and (3). Unless otherwise is specified, use (2) to find $\sigma_{\bar{x}}$. Rationale here is that the computation in (2) is exact whereas (3) gives you only an estimate. The general principle applies: use the information available to you as much as possible and refrain from estimation unless absolutely necessary.

2 Solved Exercises

2.1 Statement of Exercises

1. [Process distribution and sampling distributions] An industrial process that makes 3-foot sections of plastic pipe produces pipe with an average inside diameter of 1 inch and a standard deviation of 0.05 inch.
   a) If you randomly select one piece of pipe, what is the probability that its inside diameter will exceed 1.02 inches, assuming the population of the pipe lengths is normal?
   b) If you select a random sample of 25 pieces of pipe, what is the probability that the sample mean will exceed 1.02 inches?

2. [Shifting Normal Distribution] In an operations management class of 50 students, the midterm exam scores turn out to be normally distributed with mean 65 and standard deviation of 10. The instructor announces that he will give a grade of A to the students who scored more than mean plus half of the standard deviation.
   a) How many students receive a grade of A?
   b) Students feel better when their midterm score is higher. Suppose that the instructor adds 15 points to everybody's exam score, effectively making the mean score 80. Then how many students receive a grade of A?
   c) For the same course, the final scores turn out to be normally distributed with mean 80 and standard deviation 5. The instructor computes the overall grade with the formula 0.4 Midterm score + 0.6 Final score and without shifting the grade distribution as described in b). How many students receive a grade of A?
   d) Repeat part c) assuming that the instructor shifts the midterm scores as in b).

3. [Control charts for means and ranges] Processing new accounts at a bank is intended to average 10 minutes each. Five samples of four observations each have been taken.

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
	10.2	10.3	9.7	9.9	9.8
	9.9	9.8	9.9	10.3	10.2
	9.8	9.9	9.9	10.1	10.3
	10.1	10.4	10.1	10.5	9.7
Total	40.0	40.4	39.6	40.8	40.0

   Use the sample data above in conjunction with Table 8.2 captioned "Control Chart Parameters for 99.7 percent Confidence" on p.159 of the textbook to construct upper and lower control limits for both a mean chart and a range chart. Do the results suggest that the process is in control?

4. [p-chart lawnmower] Before shipping lawnmowers to dealers, an inspector attempts to start each mower and notes any that does not start on the first try. For mower samples of size is 100, and an average of 4 does not start. Provide a p-chart with 2-sigma limits.

5. [p-chart billing] BillMatters is a company that provides outsourced electronic payment processing services. An inspector at BillMatters has counted the number of defective monthly billing statements in each of the following 20 samples:

Sample	Number of Defectives	Sample	Number of Defectives	Sample	Number of Defectives	Sample	Number of Defectives
1	4	6	11	11	8	16	10
2	10	7	10	12	12	17	8
3	12	8	22	13	9	18	12
4	3	9	13	14	10	19	10
5	9	10	10	15	21	20	16

where each sample contains 100 statements.[1]
a) Construct a p-control chart that will place 99.74% of the chance (random) variation in the process between its control limits when the process is in control.
b) Is the process in control?
c) The inspector later learns that she was given incorrect numbers for samples 7, 8, 14 and 15. The error happened by mixing sample 7 and 8 and also by mixing sample 14 and 15. Actually the number of defectives in these samples were: Sample 7 had 14 defectives; Sample 8 had 18 defectives; Sample 14 had 18 defectives; Sample 15 had 13 defectives. With these correct number of defectives, repeat a) and b). Comment if you expect the control chart to be different.

6. [X-bar-chart at a Dry Cleaner] Twenty samples of $n = 8$ have been taken from a dry cleaner's ironing operation. Each sample included durations for ironing 8 shirts. The average sample range for the 20 samples was 0.016 minute and the average mean was 3 minutes. Determine three-sigma control limits for this process.

2.2 Solutions

ANSWER for Exercise 1:

From the question statement, the mean and standard deviation are: $\mu = 1$ and $\sigma = 0.05$.
a) By using Excel, the desired probability is $1 - Normdist(1.02, 1, 0.05, 1) = 0.3446$. We can also use the standard Normal distribution table:

$$z = \frac{x - \mu}{\sigma} = \frac{1.02 - 1.00}{0.05} = 0.4$$

In the standard Normal distribution table, look up for $\mathrm{P}(z > 0.4) = 1 - \mathrm{P}(z \leq 0.4) = 1 - 0.6554 = 0.3446$
b) The sample mean is $\bar{x}$, it is distributed by a Normal distribution with $\mu_{\bar{x}} = 1$ and $\sigma_{\bar{x}} = \sigma/\sqrt{25} = 0.01$. By using Excel, the deired probability is $1 - Normdist(1.02, 1, 0.01, 1) = 0.0228$. We can also use the standard Normal distribution table:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{1.02 - 1.00}{0.05/\sqrt{25}} = 2$$

Using the standard Normal distribution table, look up for $\mathrm{P}(z > 2) = 1 - \mathrm{P}(z \leq 2) = 1 - 0.9772 = 0.0228$. □

ANSWER for Exercise 2

a) We find the probability that a randomly seleted student scores more than mean+half of the standard deviation=70. This probability is $1 - Normdist(70, 65, 10, 1) = 1 - Normdist(5, 0, 10, 1) = 1 - Normdist(0.5, 0, 1, 1) = 0.3085$. This corresponds to roughly 15, 16 because 50*0.3085=15.425.

[1] This question is inspired by BillMatrix CEO Scott Walker's presentation at SOM on April 9, 2008.

b) After the shift one has to score above 70+15 to receive A grade and the mean becomes 80. The shift preserves the variance as it is at 10. The probability of receiving A is $1 - Normdist(85, 80, 10, 1) = 1 - Normdist(70, 65, 10, 1) = 0.3085$. This corresponds to roughly 15, 16 because 50*0.3085=15.425. As expected shifting the scores by a constant does not affect the grade distribution.

c) Let X_1 and X_2 be the midterm and final scores. The overall score is $0.4X_1 + 0.6X_2$. The mean of the overall score is

$$\mathrm{E}(0.4X_1 + 0.6X_2) = 0.4\mathrm{E}(X_1) + 0.6\mathrm{E}(X_2) = 0.4 * 65 + 0.6 * 80 = 74.$$

The variance of the overall score is

$$Var(0.4X_1 + 0.6X_2) = 0.4^2 Var(X_1) + 0.6^2 Var(X_2) = 0.16 * 100 + 0.36 * 25 = 25,$$

so the standard deviation of the overall score is 5. One receives a grade of A by having an overall score of 74+5/2. The probability of this event is $1 - Normdist(76.5, 74, 5, 1) = 1 - Normdist(2.5, 0, 5, 1) = 1 - Normdist(0.5, 0, 1, 1) = 0.3085$. Still 15-16 students receive A grade.

d) If X_1 is shifted by 15, $\mathrm{E}(X_1) = 80$ and $Var(X_1) = 10$. Then $\mathrm{E}(0.4X_1 + 0.6X_2) = 80$ and $Var(0.4X_1 + 0.6X_2) = 25$. The probability of receiving an A is $1 - Normdist(82.5, 80, 5, 1) = 1 - Normdist(76.5, 74, 5, 1) = 0.3085$. Once again 15-16 students receive A grade out of 50 students. □

ANSWER for Exercise 3:

First determine the mean $\bar{x}_j$ and range R_j of sample j. Note that the sample size is $n = 4$.

$$\bar{x}_j = \frac{\sum_{i=1}^{4} x_{i,j}}{n} \quad \text{and} \quad R_j = \max\{x_{1,j}, x_{2,j}, x_{3,j}, x_{4,j}\} - \min\{x_{1,j}, x_{2,j}, x_{3,j}, x_{4,j}\}$$

Sample	Mean	Range
1	40.0/4=10.0	10.2-9.8=0.4
2	40.4/4=10.1	10.4-9.8=0.6
3	39.6/4=9.9	10.1-9.7=0.4
4	40.8/4=10.2	10.5-9.9=0.6
5	40.0/4=10.0	10.3-9.7=0.6

Now we can compute the average of the sample means and the sample ranges:

$$\bar{\bar{x}} = \frac{10.0 + 10.1 + 9.9 + 10.2 + 10.0}{5} = 10.04 \quad \text{and} \quad \bar{R} = \frac{0.4 + 0.6 + 0.4 + 0.6 + 0.6}{5} = 0.52$$

We obtain factors A_2, D_3, D_4 from Table 8.2 for $n = 4$: $A_2 = 0.73, D_3 = 0, D_4 = 2.28$. With these, the upper and lower limits are

$$UCL_{\bar{x}} = \bar{\bar{x}} + A_2\bar{R} = 10.04 + 0.73 * 0.52 = 10.42$$

$$LCL_{\bar{x}} = \bar{\bar{x}} - A_2\bar{R} = 10.04 - 0.73 * 0.52 = 9.66$$

$$UCL_R = D_4\bar{R} = 2.28 * 0.52 = 1.19$$

$$LCL_R = D_3\bar{R} = 0 * 0.52 = 0$$

The smallest sample mean is 9.9, ant the largest is 10.2. Both are well within the control limits. Similarly, the largest sample range is 0.6, which is also within the control limits. Hence, the results suggest that the process is in control. □

ANSWER for Exercise 4:

The inspector can count both the lawnmowers that started and those that did not start. Consequently, a p-chart is appropriate with $n = 100, \bar{p} = 0.04$. Then the standard deviation of the percent of defectives in a sample is

$$\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \sqrt{\frac{0.04(1-0.04)}{100}} = 0.020$$

Two-sigma control limits can be computed using the following:

$$UCL = \bar{p} + z\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.04 + 2 * 0.02 = 0.08$$

$$LCL = \bar{p} - z\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.04 - 2 * 0.02 = 0.00.$$

□

ANSWER for Exercise 5:

a) From samples, we obtain

$$\bar{p} = \frac{\text{Total number of defectives}}{\text{Total number of observations}} = \frac{220}{20(100)} = 0.11$$

$$\sigma_p = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \sqrt{\frac{0.11(1-0.11)}{100}} = 0.03$$

From a normal distribution table, we see that 99.74% of any normal distribution is contained within 3 standard deviations away from the mean. In other words, the probabilities that the normal variable is less than 3 standard deviation and that it is more than 3 standard deviation are both $Normdist(-3,0,1,1) = 1 - Normdist(3,0,1,1) = 0.0013 = (1-0.9974)/2$. Then, the control limits are

$$UCL_p = \bar{p} + z(\sigma_p) = 0.11 + 3(0.03) = 0.20,$$

$$LCL_p = \bar{p} - z(\sigma_p) = 0.11 - 3(0.03) = 0.02.$$

b) You can see that the process is not in control: sample 8 ($22/100 = .22$)and sample 15 ($21/100 = .21$) are above the upper control limit. The process would be stopped at those points to find and correct the causes.

c) The p-control chart does not change because $\bar{p}$ and n remain the same after correctly disributing the defective bills to the samples. However, now all the samples have fewer than 20 defectives so the process is in control. □

ANSWER for Exercise 6:

We have $\bar{\bar{x}} = 3$ mins. and $\bar{R} = 0.016$ mins. From Table 8.2, $A_2 = 0.37$ for $n = 8$.

$$UCL = \bar{\bar{x}} + A_2\bar{R} = 3 + .37(.016) = 3.006 \text{ mins},$$

$$LCL = \bar{\bar{x}} - A_2\bar{R} = 3 - .37(.016) = 2.994 \text{ mins}.$$

□

3 Exercises

1. **[C4ISN at Lennox]** In a C4ISN (http://som.utdallas.edu/c4isn) project, a UTD research team works on improving AC (air condition) demand forecasting procedure for Lennox (lennox.com). Lennox is located on the eastern side of the UTD campus. David is a graduate student at UTD and a member of the research team. He is preparing an Ishikawa (fishbone) diagram to illustrate the factors that contribute to the air conditioner demand. He identified eight factors as the demand drivers: Consumer, Government regulations, New product introductions, Macroeconomic factors, AC retailers (installers), Competition, Relevant industry indices, Environmental factors. Help David to detail his fishbone diagram: i) Choose five factors, that you know more about, out of the eight factors, ii) Provide one subcategory for each of your five factors chosen in i). If necessary, explain what your subfactor means. For example, if you choose Relevant industry indices and say that ROI is an important index, you must explain what ROI stands for. You might mean Real-estate Outlook Index by ROI while many people would interpret that as Return On Investment.

2. **[Photovoltaic Dishes at Lennox]** Given high prices and environmental concerns for conventional energy sources, solar energy can become a viable alternative in the U.S. as it has in other countries; see www.pv-ec.net and www.epia.org. A method of collecting sunbeams is by using *photovoltaic dishes*. These dishes look like satellite dishes. They collect the sun beam waves at their receiver as opposed to electromagnetic waves carrying broadcast signals collected by satellite dishes. Production and especially installation of a photovoltaic dish is similar to those of an air conditioner and a furnace which are the speciality of Lennox. Moreover, Lennox has expertise in connecting air conditioners and furnaces to the house power supply. This expertise can be used to directly connect photovoltaic dishes to existing heating/cooling systems to avoid energy losses. Suppose that Lennox is considering to produce and install photovoltaic dishes for residential customers and they employ a UTD Cohort student Li Argy[2] to forecast the potential demand for the dishes before making substantial investment into this product. Help Li to make an Ishikawa (fishbone) diagram to illustrate the factors that will contribute to the photovoltaic dish demand. Also comment in 1-2 sentences about the differences between forecasting for air conditioner demands and for photovoltaic dish demands.

3. **[Expensive - Ineffective Healthcare]** Let us first read excerpts from the following article: S. Jauhar. *Many Doctors, Many Tests, No Rhyme or Reason.* New York Times, March 11, 2008 issue. The author, Dr. Sandeep Jauhar, is a cardiologist on Long Island. He says:

   > ...a 50-year-old man who had been admitted to the hospital short of breath. During his month-long stay he was seen by a hematologist, an endocrinologist, a kidney specialist, a podiatrist, two cardiologists, a cardiac electrophysiologist, an infectious-diseases specialist, a pulmonologist, an ear-nose-throat specialist, a urologist, a gastroenterologist, a neurologist, a nutritionist, a general surgeon, a thoracic surgeon and a pain specialist. He underwent 12 procedures, including cardiac catheterization, a pacemaker implant and a bone-marrow biopsy. ...When he was discharged, follow-up visits were scheduled ...with 7 specialists. This ...case, in which expert consultations sprouted with little rhyme, reason or coordination, reinforced a lesson I have learned many times since entering practice: In our health care system, where doctors are paid piecework for their services, if you have a slew of physicians and a willing patient, almost any sort of terrible excess can occur.
   >
   > ...
   >
   > Are we getting our money's worth? ...The *U.S.* ranks ...in [the] *last* place ...among major industrialized countries in *health-care quality* .... And in the U.S., regions that spend the most on health care appear to have higher mortality rates than regions that spend the least, perhaps because of increased hospitalization rates that result in more life-threatening errors and infections. It has been

[2] This question does not reflect any of Lennox decisions or strategies. It is inspired by discussions with Gary Li, Cohort 2008.

estimated that if the entire country spent the same as the lowest spending regions, the Medicare program alone could save about $40 billion a year.

...

Overutilization is driven by many factors "defensive" medicine by doctors trying to avoid lawsuits; patients demands; a pervading belief among doctors and patients that newer, more expensive technology is better. The most important factor, however, may be the perverse financial incentives of our current system.

Doctors are usually reimbursed for whatever they bill. As reimbursement rates have declined in recent years, most doctors have adapted by increasing the quantity of services. ... Overconsultation and overtesting have now become facts of the medical profession.

Consider medical imaging. ... from 1999 to 2004 the growth in the volume of imaging services per Medicare patient far outstripped the growth of all other physician services. In 2004, the cost of imaging services was close to $100 B, or an average of roughly $350 per person ....

...

Not long ago, I visited a friend a cardiologist ... at his office ... to ask him about imaging in private practices.

> When I started in practice, I wanted to do the right thing. ... A young woman would come in with palpitations. I would tell her she was fine. But then I realized that she would just go down the street to another physician and he would order all the tests anyway: echocardiogram, stress test, Holter monitor stuff she did not really need. Then she would go around and tell her friends what a great doctor a thorough doctor the other cardiologist was. I tried to practice ethical medicine, but it did not help. It did not pay ... from a financial ... [or] a reputation standpoint.

His nuclear imaging camera was in an adjoining "procedure" room. He broke down the monthly costs ...: camera lease, $4,500; treadmill lease, $400; office space, $1,000; technician fee, $1,800; nurse fee, $1,000; and miscellaneous expenses of $200. "... say I get ... $850 per ... stress test, ... Then I have to do ... 10 ... tests a month ... to cover the costs .... "

...

Last year, Congress approved steep reductions in Medicare payments for certain imaging services. Deeper cuts will almost certainly be forthcoming. This is good; unnecessary imaging is almost certainly taking place, leading to false-positive results, unnecessary invasive procedures, more complications .... But the problem ... is much larger than imaging. *Doctors* are doing too much testing and too many procedures, often for *the sake of business.*

...

You can also read the entire article from nytimes.com. In light of this article and other information you have, make an Ishikawa (fishbone) diagram to list the factors that contribute to expensive and ineffective healthcare.[3]

4. **[Explaining your grades]** Make an Ishikawa diagram to illustrate the variables and the factors because of which you did not get 100% from the quiz.

5. **[Truffle Quality]** A chocolate factory has weighed 4 chocolate truffles coming out of production over 5 working days to obtain the following data, weights are in grams

	Mon	Tue	Wed	Thu	Fri
10am	30.50	30.30	30.15	30.60	30.15
11am	29.75	31.00	29.50	32.00	30.25
2pm	29.90	30.20	29.75	31.00	30.50
3pm	30.25	30.50	30.00	30.00	29.70

[3]This question is partly inspired by Blue Cross Blue Shield's Texas President Marty Foster's presentation at Richardason Chamber of Commerce on February 27, 2008.

a) Group the data by day into five samples and find the sample averages.
b) Compute the range for each sample.
c) Use the ranges and control chart parameter (n, A_2, D_3, D_4, d_2) of the textbook to make an X-bar chart.

6. **[TPS in Tennessee]** A boiler plant in Tennessee wanted to implement TPS in the early 1990s. The plant produces replacement boilers for nuclear reactors in the power plants or navy ships. The unionized employees believed that efficiency improvements would cut down the required work force and hence eventually resulting in lay offs. Employees refused to implement the improvements.
a) In view of the employee tenure in the USA shown in Figure 1, in what aspect the employees are right?
b) Do you think the employee tenure in Japan show the same trends as that in the USA? If not, how does this help companies like Toyota to implement improvements in Japan?

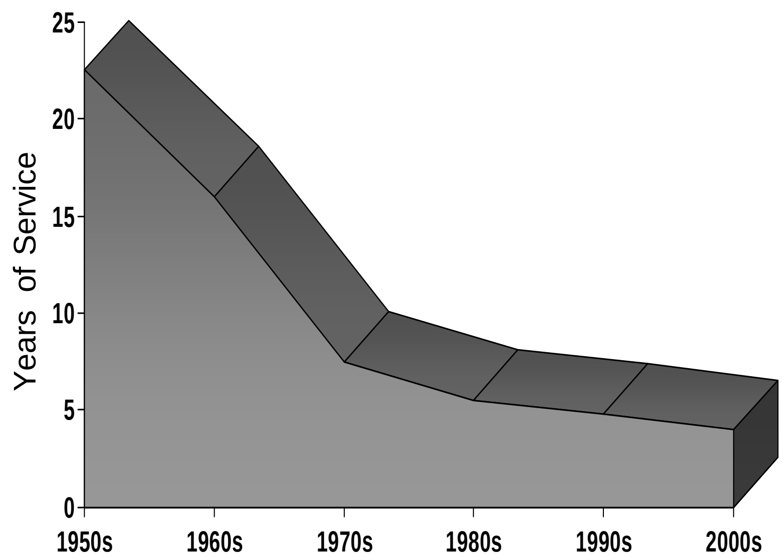

Figure 1: Average employee tenure in the U.S. over 1950-2000.

7. **[Complaints against Toyota]** Table below has the actual number of complaints against Toyota logged by NHTSA, National Highway Traffic Safety Administration.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2006	183	162	209	159	194	229	223	244	271	217	224	235
2007	296	207	266	238	264	268	265	296	258	287	276	226
2008	304	232	264	271	310	299	341	322	290	278	276	312
2009	361	288	346	344	380	372	396	459	448	510	517	469

Using the appropriate quality control chart, find the upper and lower control limits by using the data in the table above.
a) There were 354 complaints in January 2010. Is this number within your control limits?
b) Drop the month of December 2009 from your data set, compute the control limits as if we are at the begining of December 2009. Does 469 complaints in December indicate that the process is in control in December?

c) Drop the months of November and December 2009 from your data set, compute the control limits as if we are at the begining of November 2009. Does 517 complaints in November indicate that the process is in control in November?

4 Exercises from BA3352

1. Every year about 500 people apply for UTD's full time MBA program. Over the years it has been observed that GMAT score of each of these people are distributed normally with mean 650 and variance 300.
a) If UTD decides to accept all applicants whose GMAT score is above 670, on average how many people will be accepted per year?
b) If UTD decides to accept 50 students with highest GMAT scores every year, what should be the cut off GMAT score (lowest score among the 50 accepted students).

...
4. The DFW airport passenger data below pertains to the first eight months of 2001. Suppose that every month has 30 days. Number of passengers flying out of DFW airport per day and the number of passengers who are searched per day are:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
	$\bar{y}_{Jan}$	$\bar{y}_{Feb}$	$\bar{y}_{Mar}$	$\bar{y}_{Apr}$	$\bar{y}_{May}$	$\bar{y}_{Jun}$	$\bar{y}_{Jul}$	$\bar{y}_{Aug}$
Average # of passengers/day	15000	14000	12600	13300	14700	14100	16800	17500
	$\bar{z}_{Jan}$	$\bar{z}_{Feb}$	$\bar{z}_{Mar}$	$\bar{z}_{Apr}$	$\bar{z}_{May}$	$\bar{z}_{Jun}$	$\bar{z}_{Jul}$	$\bar{z}_{Aug}$
Average # of searched passengers/day	47	53	61	41	42	44	51	43

The average number of passengers per day is computed as follows. Let $y_{i,j}$ be the number of the passengers on the ith day of month j. The average number of passengers per day for month j is $\bar{y}_j$ defined as

$$\bar{y}_j = \frac{\sum_{i=1}^{30} y_{i,j}}{30} \qquad \text{for } j \in \{Jan, Feb, Mar, Apr, May, Jun, Jul, Aug\}.$$

The average number of passengers searched per day is computed similarly. Let $z_{i,j}$ be the number of the passengers searched on the ith day of month j. The average number of passengers searched per day for month j is $\bar{z}_j$ defined as

$$\bar{z}_j = \frac{\sum_{i=1}^{30} z_{i,j}}{30} \qquad \text{for } j \in \{Jan, Feb, Mar, Apr, May, Jun, Jul, Aug\}.$$

a) What is the sample size n for computing averages in the table?
b) Suppose that the standard deviation of the number of passengers ($y_{i,j}$) flying out of DFW every day is 3000, what is the standard deviation of the average number of passengers ($\bar{y}_j$) flying out of DFW per day?
c) Assuming a Normal distribution for the number of passengers, how many sigmas (σ) will give you a Type I error of 20% for an $\bar{x}$-chart on the average number of passengers flying out of DFW per day?

5. Refer to question 4.
a) Find out 3-sigma UCL and LCL for an $\bar{x}$ chart on the average number of passengers flying out of DFW per day.
b) Is the process in control during the first eight months? Explain.

6. Refer to question 4.
a) Compute the variance of the average number of passengers searched ($\bar{z}_j$) per day during the first eight

months. In other words, find the variance of the population $\{\bar{z}_{Jan}, \bar{z}_{Feb}, \bar{z}_{Mar}, \bar{z}_{Apr}, \bar{z}_{May}, \bar{z}_{Jun}, \bar{z}_{Jul}, \bar{z}_{Aug}\}$ by using the data in the table. Let us call this variance $\sigma_{\bar{z}}^2$.
b) Compute the ratio of $\sigma_{\bar{z}}^2$ to the grand mean of the averages of the passengers searched per day during the first eight months. Looking at this ratio and considering the fact that the number of searches per day is an integer number, what distribution would be appropriate to study the number of searches?
c) What are UCL and LCL for a 2.5-sigma c-control chart for the number of passengers searched per day?

7. Refer to question 4.
a) Obtain the proportion $\bar{r}_j$ of passengers searched per day for each month. In other words, construct the population $\{\bar{r}_{Jan}, \bar{r}_{Feb}, \bar{r}_{Mar}, \bar{r}_{Apr}, \bar{r}_{May}, \bar{r}_{Jun}, \bar{r}_{Jul}, \bar{r}_{Aug}\}$ by using the data in the table.
b) Compute the grand mean and the variance $\sigma_{\bar{r}}^2$ of the population in a).
c) What are UCL and LCL for a 2.5-sigma p-control chart for the proportion of passengers searched?

8. Refer to questions 4,6 and 7. Below are average number of passengers and average the number of passengers searched in September and October 2001.

	Sep	Oct
Average number of passengers/day	9100	6200
Average number of searched passengers/day	57	63

Using c- and p-control charts obtained in questions 6 and 7 and the recent numbers above determine if
a) The number of passengers searched per day is in control?
b) The proportion of passengers searched per day is in control?
c) How can you reconcile your answers if you say "yes" to either a) or b) above, and "no" to the other?