SPC

Topic: c charts

Fill in the following blanks:

1. Let the random variable X represent the **number of defects on an item** and let the **mean number of defects on such an item be 4**. Thus, X is a Poisson random variable with parameter *c* = 4. Determine the following.

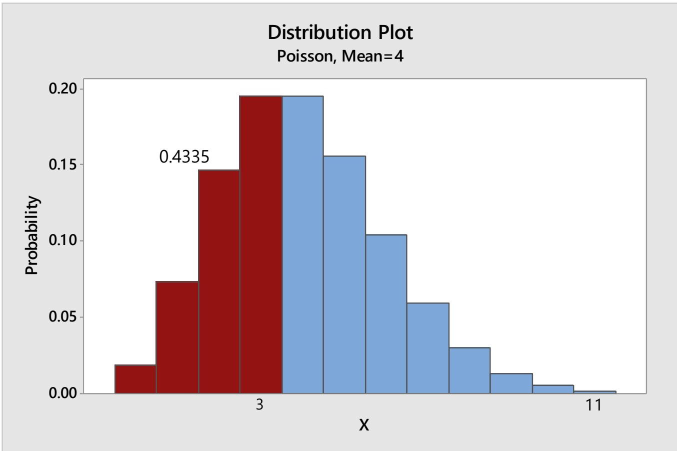
$E(X) = \text{Mean}(X) =$ ______4_________, $\text{Variance}(X) =$ ______4_______, $\text{StdDev}(X) =$ _______2_______.

2. What is the probability that there will be *at most* 3 (meaning 0, 1, 2, or 3) defects on an item with mean *c* = 4? You can use Minitab or your calculator to determine this. Remember that Poisson is a discrete random variable (i.e. determine probabilities with sums and not integrals). Report your answer correct to 4 decimal places.

Graph > **Probability Distribution Chart** > **View Probability** > **Poisson with Mean 4**
Shaded Area > **X Value** > **Left Tail** > **X Value: 3**

Solution: The probability mass function of X is:

$$p(x) = \frac{e^{-4} \cdot 4^x}{x!} \text{ for } x = 0, 1, 2, 3, \ldots; \; P(X \leq 3) = \sum_{x=0}^{3} \frac{e^{-4} \cdot 4^x}{x!} \cong \mathbf{0.4335}$$



3. A basic c chart example: Snags per 1 square foot block of carpeting for k = 10 samples:

2, 3, 3, 2, 1, 0, 4, 2, 1, 0

Determine the **center line, upper control limit**, and **lower control limit** for this c chart by-hand or in Minitab. Report decimal values to 3 decimal places.

$CL_c = (2 + 3 + 3 + 2 + ... + 0) / 10 =$ **1.8 blemishes**

$UCL_c = 1.8+3 \bullet \sqrt{1.8} \cong \mathbf{5.825}$ $\quad LCL_c = 1.8-3 \bullet \sqrt{1.8} \cong -2.225$, which doesn't make sense, so **0**

4. Data for $k = 25$ samples of fabric from a textile mill, each 100 m^2, are selected and the number of flaws per sample is counted. Data for the 25 samples are in the worksheet **Hmwk7DATA_AttributeCharts**.

(a) Determine the center line and control chart limits for the number of nonconformities per 100 m^2 of the fabric. Report decimal values to 2 decimal places.

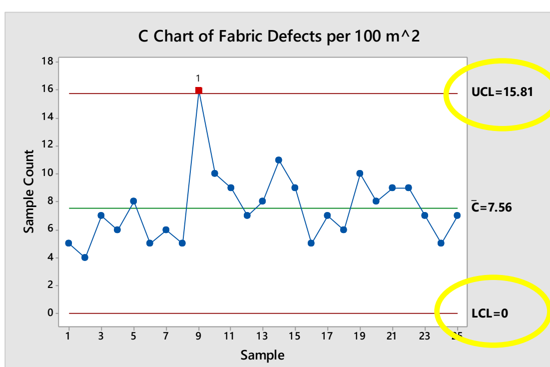
Solution: The total number of flaws in the 25 samples is 189. Thus, the average number of defects per sample is 189 / 25 = **7.56.**

$CL_c = 7.56$ defects

$UCL_c = 7.56 + 3 \bullet \sqrt{7.56} \cong \mathbf{15.81}$ $\quad LCL_c = 7.56 - 3 \bullet \sqrt{7.56} \cong -0.689$, which doesn't make sense, so **0**

Construct a control chart in Minitab for the defects. Make sure your center lines and limits match up with what you computed above.

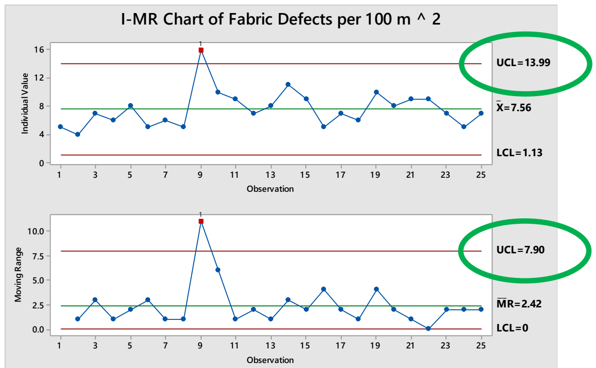
Stat > Control Charts > Attribute Charts > C



(b) As a comparison, imagine that you thought this data was measurement data and constructed an I-MR chart. What's the only difference between the c chart and I chart?

Solution: The only difference is the values of the control limits (UCL and LCL) for the two different graphs (*c and I*). You will notice that the I chart has tighter control limits and the c chart has wider control limits

- For the c chart, the standard deviation is computed as $\sqrt{c} = \sqrt{7.56} \cong \mathbf{2.75}$. The *c* chart's LCL and UCL are **0** and **15.81**, respectively.
- For the ***I* chart,** the standard deviation is computed as $\hat{\sigma} = \overline{MR}/d2 \cong 2.42/1.128 \cong \mathbf{2.15}$. The I chart has LCL, UCL as: **1.13** and **13.99**, respectively Using an *I* chart implies that ***the order in which the data is collected matters***, which technically it isn't – the defects are likely independent of order number.



(c) True or False. In general, using an *I* chart instead of a c chart could affect the number of Type I or Type II Errors reported. Note: Since Poisson charts ***aren't normally distributed***, Type I Error is no longer 0.0027, though it could be close. **Grading:** All or nothing.

True False

5. [+0.5] Go to the following website, and answer as many questions as you can before you reach 3 strikes. You can have only 1 practice round. On the front board, record your number of correct answers ***on your second try***. **Construct and attach a c chart** of the number of correct answers per person (after everyone has recorded their answers on the board). ***Make sure to title your graph***. If you've taken this quiz before, your score will most likely be an out-ofcontrol point!

http://www.notdoppler.com/theimpossiblequiz.php

Comments about Problem 5. Now if we did have data for the "number of correct answers" for the Impossible Quiz, what kind of chart would you make?

If we just had a column of values for the # **of correct answer**s, then we would make a ***c* chart**. We would use this chart since *we don't know the total number of questions associated with this quiz*. The data in class was very bimodal – pretty much – people had taken the quiz before or they had not. Here's what the data looked like:

Number of correct answers: 3, 4, 26, 3, 29, 28, 5, ...

We don't want to use an *I*-MR chart since the data is *not measurement data* – it is definitely **count data**. Also, for an I-MR chart, order matters. MRbar would be very large because of the difference between consecutive "number of correct answers" values. The c chart has centerline cbar (which is the average of all the correct answers) and the UCL would be: cbar + 3 * sqrt(cbar), while the LCL would be: cbar – 3 * sqrt(cbar).

6. The number of typographical errors is counted over chapters in a textbook. The data for $k = 25$ chapters is in the file Hmwk7DATA_AttributeCharts.

(a) Determine the center line and control chart limits for the number of errors per chapter.

$CL_c = 155$ errors / 25 chapters = $\mathbf{6.2}$ errors per chapter

$UCL_c = 6.2+3\bullet\sqrt{6.2} \cong \mathbf{13.670}$ $LCLc = 6.2-3\bullet\sqrt{6.2} \cong -1.270$, which doesn't make sense, so $\mathbf{0}$

Grading: +0.5 for CL, +0.5 for UCL, deduct 0.25 for negative LCL

(b) There is a more appropriate chart for this data – it's called a u chart. To construct a u chart, we need a piece of additional information. For this example, what additional piece of information would help to build a more useful chart? That is, what information should we take into consideration (if it is available)?

Solution: An answer that indicates that the chapter sizes are different, such as:

- Number of pages per chapter,
- Number of characters per chapter,
- Number of lines per chapter,
- Basically – chapter sizes are different.

7. In a Health Information Center, telephone counselors are available to provide up-to-date information about health-related issues to callers. When calls are received at the main switchboard, they are transferred to one of four groups of counselors, based on the classification of the caller's question. The **four classifications for callers' questions** are:

(1) General Health, (2) Pediatrics, (3) Infectious Diseases, and (4) Oncology/Hematology.

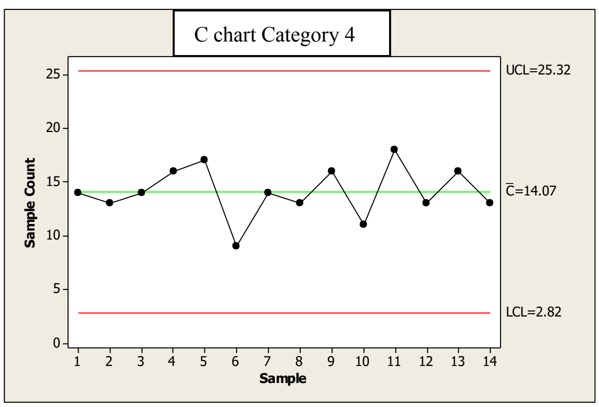
Even though the counselors have extensive training as well as many sources of medical information, questions do arise for which adequate information or knowledge is not readily available. In such cases, the counselors research the question and, later, telephone the original caller with accurate and detailed information concerning the caller's questions. Because of the amount of time lost for each caller's question that must be researched, it is desired to **determine the health area where these questions most frequently occur.** The data in columns C9-C13 in Hmwk7DATA_AttributeCharts contains the number of calls received per day for which the answer is not readily available (assume a good operational definition exists for this concept!).

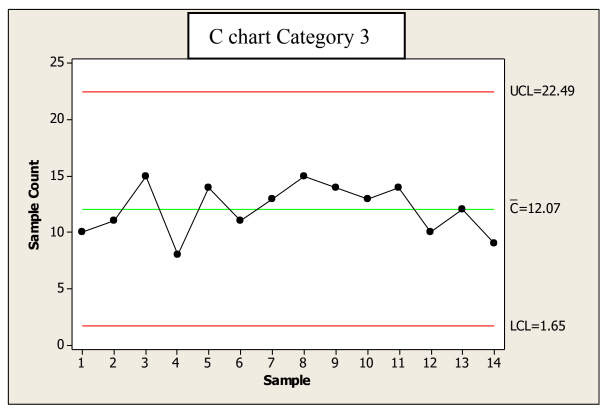
Number of "Unanswerable Calls"
Classification of Questions Asked

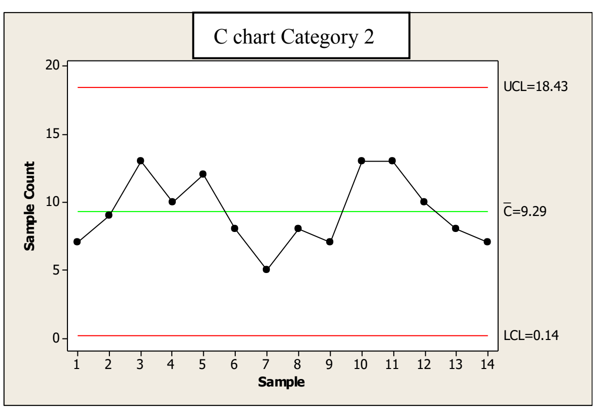
Days	General Health	Pediatrics	Infectious Diseases	Oncology
1	5	7	10	14
2	7	9	11	13
3	6	13	15	14
4	9	10	8	16
5	4	12	14	17
6	7	8	11	9
7	9	5	13	14
8	3	8	15	13
9	7	7	14	16
10	8	13	13	11
11	9	13	14	18
12	9	10	10	13
13	3	8	12	16
14	6	7	9	13

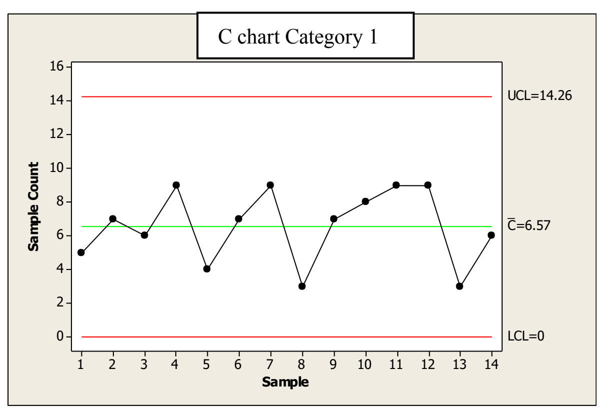
7.(a) In Minitab, plot the appropriate control charts for classifications 1, 2, 3, and 4. Record the center lines for each plot. **You do not need to attach the plots.**

Classification 1: 6.57 Classification 2: 9.29 Classification 3: 12.07 Classification 4: 14.07









Some students may have made p charts using the ***total number of unanswerable calls in a day as the sample size;*** e.g. 36, 40, 48, 43, 47, 35, 41, 39, 44, 45, 54, 42, 39, 35. These are also correct.

Classification 1: 0.1565 Classification 2: 0.2211 Classification 3: 0.2874 Classification 4: 0.3355

(b) Let's look at the **Classification 4: Oncology** control chart. If the process is in control, what's the probability of committing a Type I Error, Rule #1 (point beyond LCL or UCL) on this chart? Note: It is appropriate to use a Poisson distribution to calculate this probability, although you can *approximate* this probability with a normal distribution.

Also, you can only input positive integers into a Poisson distribution. Give your answer correct to 4 decimal places.

Solution (detailed):

In this problem, we are tracking the number of "unanswerable calls" in different departments of a hospital. We are only given the number of unanswerable calls each day. In part (a), you determined that the **number of unanswerable calls** in the **Oncology department** has a **Poisson distribution** with **mean 14.07** (since this is cbar). The variance of a Poisson distribution is cbar and the **standard deviation is sqrt(cbar).**

So, the centerline is 14.07. The **UCL** is **14.07 + 3 * sqrt(14.07) = 25.32** and the **LCL is 14.07 – 3 *sqrt(14.07) = 2.82.** You can check your control chart limits for the Oncology department control chart – and you'll see these limits agree with those. If not, then you didn't make a *c* chart.

Here's the subtle point. The data that you're tracking is the **number of unanswerable calls**. Every day we obtain an integer data value – a count. That number could be 0, 1, 2, 3, ... The **LCL is 2.82** and the **UCL is 25.32**.

- If the number of unanswerable calls is **0, 1, or 2,** then they'll be **BELOW the LCL**.
- If the number of unanswerable calls is **26, 27, 28, ...** then they'll be **ABOVE the UCL**.
- A **Type I error** occurs if the **number of unanswerable calls is 0, 1, 2 or 26, 27, 28, ...**
- A **Type I error** does **NOT** occur if the number of unanswerable calls is **3, 4, 5, ... 25**.
- Whether you count 3 and 25 as "in" or "out" makes a difference in your Type I and Type II probability calculations. Here is the **correct probability** of a **Type I Error** assuming the **mean number of calls is 14.07.**

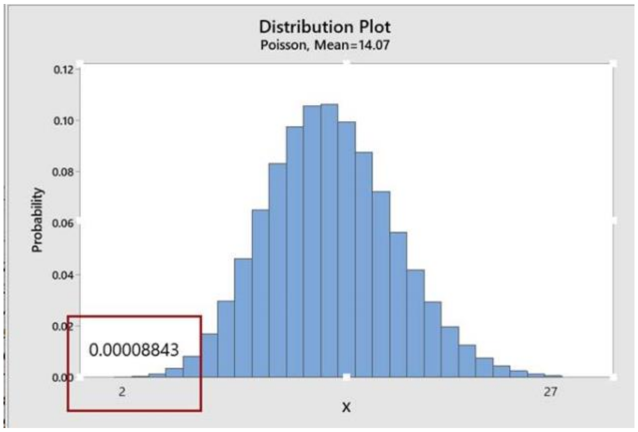
$$P(\text{Type I Error}) = 1 - P(\text{a point plots inside the LCL and UCL})$$

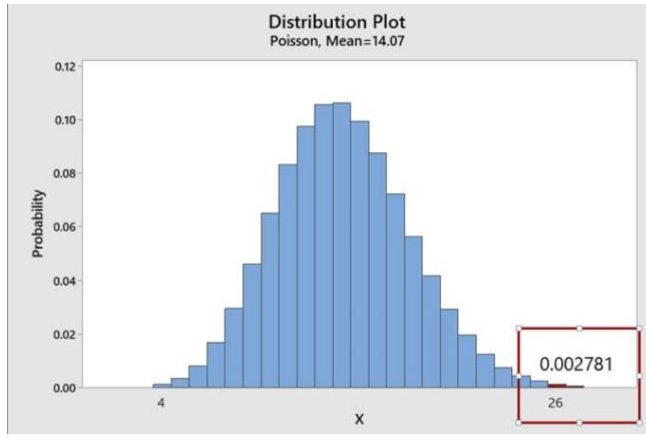
$$= 1 - \sum_{x=3}^{25} \frac{e^{-14.07} \cdot 14.07^x}{x!} \cong \mathbf{0.0029}$$

$$P(\text{Type I Error}) = P(\text{a point plots outside LCL or UCL when the process is in control})$$

$$= \sum_{x=0}^{2} \frac{e^{-14.07} \cdot 14.07^x}{x!} + \sum_{x=26}^{\infty} \frac{e^{-14.07} \cdot 14.07^x}{x!} \cong \mathbf{0.0029}$$

Here are those calculations done in Minitab:





When you sum the left and right tail probabilities, you get 0.0029.

If you didn't use a Poisson distribution to determine this value, but used a normal approximation as below, you will get full credit if you used the **CORRECT normal approximation**. If you estimated this probability with a normal distribution, then X ~ Normal (14.07, 14.07).

$$P(\text{Type I Error}) = 1 - P(\text{a point plots inside the LCL and UCL})$$

$$= 1 - \sum_{x=3}^{25} \frac{e^{-14.07} \cdot 14.07^x}{x!} \cong \mathbf{0.0029}$$

$$P(\text{Type I Error}) = P(\text{a point plots outside LCL or UCL when the process is in control})$$

$$= \sum_{x=0}^{2} \frac{e^{-14.07} \cdot 14.07^x}{x!} + \sum_{x=26}^{\infty} \frac{e^{-14.07} \cdot 14.07^x}{x!} \cong \mathbf{0.0029}$$

$$= 1 - P\left(\frac{2.82 - 14.07}{\sqrt{14.07}} < z < \frac{25.32 - 14.07}{\sqrt{14.07}}\right)$$

$$\cong P(-3 < Z < 3)$$

$$\cong \mathbf{0.0027}$$

(c) Assume that the number of unanswerable calls per day in Oncology shifts to mean 20. What is the probability that you will commit a Type II Error given the mean has shifted to 20? Note: It is appropriate to use a Poisson distribution, although you can *approximate* this probability with a normal distribution. Give your answer correct to 4 decimal places.

Solution: Again, the problem is that we're working with discrete (count) data instead of variable (measurement) data – and so the upper and lower control limits make a big difference as to whether they are included or not in determining Type I and Type II Errors.

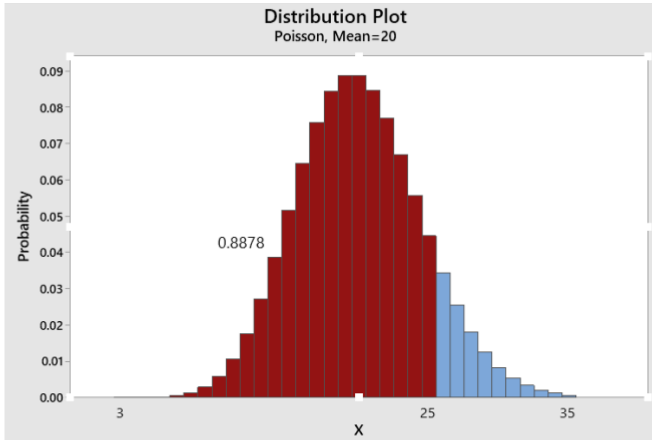
Using the same logic as in part (b), I got 0.8878 in part (c). If you used Minitab's Probability Distribution Plot with middle and used the left as 2.82 and the right as 25.32, you got "lucky" with 0.8878 since Minitab is ROUNDING 2.82 up to 3 and 25.32 down to 25. In general, if you want to compute the **exact probability of a Type II error for a Poisson**, you'll need have the **left x value** be the **ceiling of the LCL** and the **right x value** be the **floor of the UCL**.

This is only be a problem with attribute charts (count data). When you integrate, this matter doesn't come up!

$P(\text{Type II Error}) = P(\text{determine process in control when the mean has shifted})$

$= P(\text{obtain between 3 and 25 defects/day when the mean number has shifted to } \lambda = 20)$

$= \sum_{x=3}^{25} \frac{e^{-20} \cdot 20^x}{x!} \cong \mathbf{0.8878}$



If you didn't use a Poisson distribution to determine this value, but used a normal approximation as below, you will get full credit if you used the CORRECT normal approximation. If you estimated this probability with a normal distribution, then $X \sim \text{Normal}(20, \sqrt{20})$. If you used a normal approximation, but you used a different standard deviation, e.g., 2.40 from the control chart or $\sqrt{14.07} \cong 3.75$, then points will be deducted. Since the distribution is Poisson, when the mean changes to c, then standard deviation changes to $\sqrt{c}$

$$
\begin{aligned}
P(\text{Type II Error}) &= P(2.82 < X < 25.32) \\
&= P\left(\frac{2.82-20}{\sqrt{20}} < z < \frac{25.32-20}{\sqrt{20}}\right) \\
&\cong P(-3.84 < z < 1.19) \\
&\cong \mathbf{0.8829}
\end{aligned}
$$

(d) In order to directly compare the number of unanswerable calls for each classification, what additional data would be helpful to have? If you had this information, what type of control chart would be preferable for monitoring the number of unanswerable per day?

Solution: Here's one example:

- If we could track the **total number of calls in each category per day,** then we could track the proportion of defectives per day. We'd have to use a ***p* chart (and not *np*)** since the sample size would most likely vary from day to day.
- We could also track the length of each call. We could study the lengths with Individual-Moving Range charts or possibly Xbar-R charts (if we group our data).

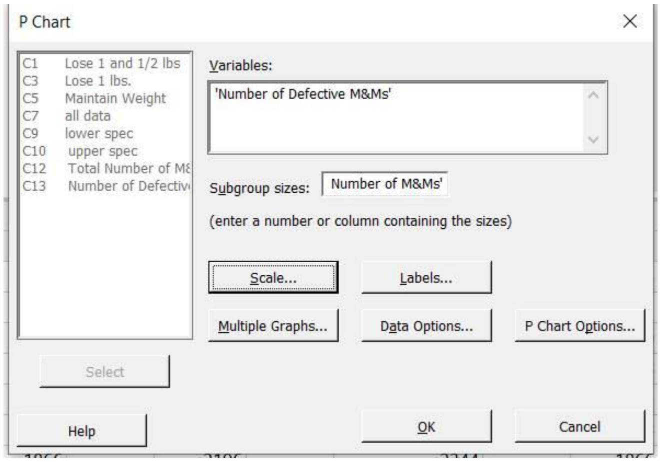
8. At the front board, write down your total number of M&M's and the number of defective ones. Construct the appropriate control chart for tracking the ***proportion of defective M&M's per bag***. Attach the control chart, ***appropriately labeled***, at the end of this assignment.

Solution: I'm going to copy and paste the data that is currently in a Google spreadsheet below.

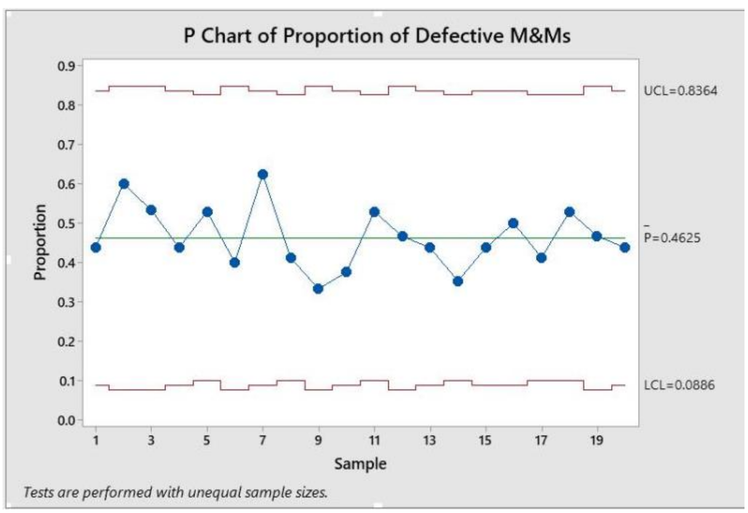
Total Number of M&Ms	Number of Defective M&Ms
16	7
15	9
15	8
16	7
17	9
15	6
16	10
17	7
15	5
16	6
17	9
15	7
16	7
17	6
16	7
16	8
17	7
17	9
15	7
16	7

The control chart is for the **"proportion of defective M&M's per bag."** We **KNOW** the number of M&M's per bag, so we can make a proportion of defectives by taking the number of defectives and divide by the total number of M&M's per bag. For example, the first several proportions are: 7/16, 9/15, 15/18, ...

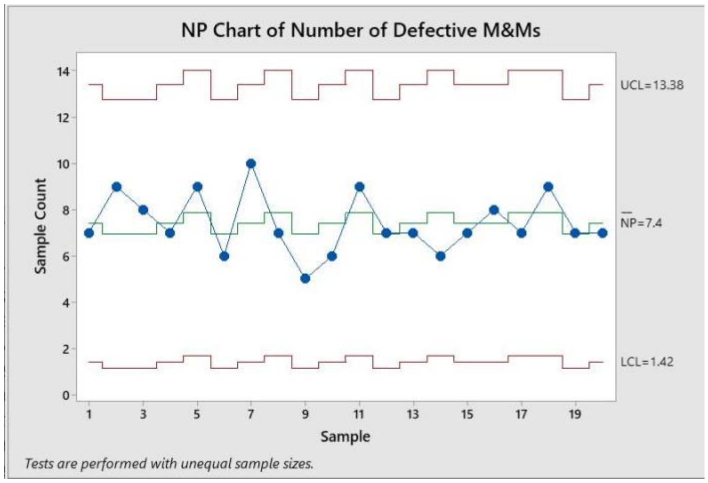
The **MOST CORRECT chart to use is a p-chart**. Why? You know the finite # of M&M's per bag and you know the # of defectives per that finite number. Don't make the proportions; let Minitab make them for you, which is what a p chart does. Here is the Minitab screen shot:



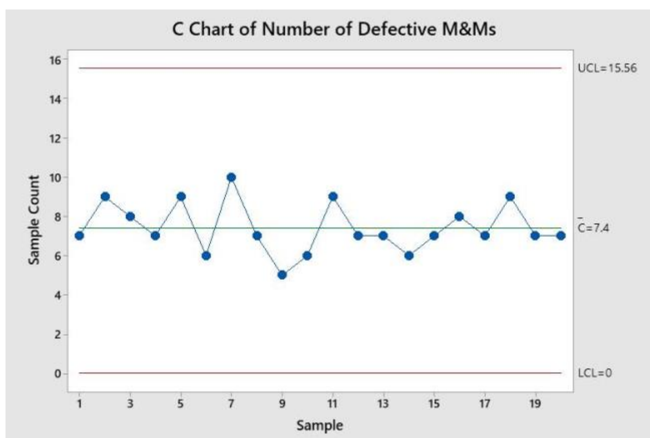
The chart will look like the one below. I changed the title of the graph to say "proportion" of defective M&M's. Note that I said that the *p* chart was the MOST CORRECT choice. There are other choices, but not as good.



Why not an np chart? It's ok – but we don't typically use the np chart when the "subgroup size" is changing. Note that the number of M&M's per bag is fluctuating – 16, 15, 15, 16 ... In our notes (and yellow reference sheet), it says to use a *p* chart over an np chart when the sample size is not constant. Why? The np chart is confusing, especially the centerline, when n varies. You can use it, but you can **ALWAYS use a p chart instead of an np chart.** You don't need to worry about sample sizes changing with a *p* chart. So, when you know the total number of something when making a proportion, use a *p* chart. Here is the *np* chart:



Why not a c chart? A c chart is used when we **DON'T know the size of the total population**. Since we **KNOW** the number of M&M's per bag, then we want to USE this information to make a more detailed plot than a c chart. If I didn't have you count the total number of M&M's per bag, then we would not have that total number. Then a c chart would be appropriate with our "area of interest" as 1 bag. So, then we'd choose a c chart and just use the # of defectives per bag, where we are considering a bag as our item. On the c chart, we'd just use the column of number of defectives. We wouldn't have the total, so we couldn't use the total. This c chart is less informative than the p chart since we're not including the # in each bag.



If you understand everything that I've said above, then you have a good handle on the difference between *p, np,* and *c* charts.

Also, I **didn't make an I-MR chart**, though you may be wondering why not. It's because we counted the number of defective M*M's. There are **no measurements involved** – this is not variable data – so we don't want to use a variable chart. It will plot the same points as the other charts, but it will use MRbar for the standard deviation, and this makes the order of data collection matter. An I-MR is wrong. The control chart limits will be totally wrong.