Understanding Microbial Count Distributions: Choosing the Right Model for Control Charts

1. INTRODUCTION

Microbial counts are critical indicators in pharmaceutical environments, cleanrooms, and other controlled settings.

Microbial counts are an example of **count data**, meaning they represent the number of occurrences of an event within a fixed unit of measurement (*e.g.*, time, area, or volume). A **count variable** is a discrete variable that can only take non-negative integer values (0, 1, 2, ...), making it fundamentally different from continuous data. Importantly, **microbiological data are continuous in the sense that they accumulate over time and across different conditions, but they remain fundamentally count-based, discrete variables**. This distinction is crucial because **the statistical models and control charts used for analysis must be appropriate for discrete count data, rather than continuous measurements**.

In statistical modeling, count data are often assumed to follow a Poisson distribution, where the variance equals the mean.

Mathematically, a discrete random variable, *X*, has the Poisson distribution if its probability function is:

$$p(x) \begin{cases} = \frac{(np)^x}{x!} e^{-np} = \frac{\lambda^x}{x!} e^{-\lambda} & x = 0,1,2,\ldots \\ = 0 & \text{elsewhere} \end{cases}$$

and its variance is equal to the mean and the parameter λ:

$$\sigma^2 = \mu = \lambda$$

Because of this there are «different» Poisson Distributions for different values of the mean, μ.

Introduced by Siméon Denis Poisson in a book he wrote regarding the application of probability theory to lawsuits (1837), the Poisson distribution is commonly used to model the number of occurrences of « rare events » within a fixed interval of time, volume, or space such as:

- number of misprints on a page (or number of pages) in a book,
- number of people in a community reaching the age of 100,
- number of wrong phone numbers dialed in a day,
- number of equipment failures occurring in a given time period, *etc*.

This model assumes that events occur independently, the average rate of occurrence is constant, and the probability of multiple events occurring simultaneously is negligible. These conditions align well with microbial contamination scenarios in controlled environments, where counts are typically low (*e.g.*, 0–3 CFU) and measured over fixed intervals (*e.g.*, per m^3 of air or per 100 mL of water).

For practical purposes, a dataset of 100 simulated microbial counts was generated using a Poisson distribution with a mean of $\lambda \approx 2$. The simulated values are representative of possible microbial counts from air or water samples and are summarized in Table 1 below.

Table 1

Sample ID	Count	Sample ID	Count	Sample ID	Count	Sample ID	Count	Sample ID	Count
1	4	21	4	41	1	61	2	81	2
2	4	22	1	42	2	62	5	82	1
3	1	23	6	43	0	63	3	83	1
4	3	24	4	44	5	64	2	84	2
5	2	25	0	45	2	65	3	85	3
6	2	26	2	46	5	66	1	86	2
7	3	27	1	47	4	67	1	87	1
8	0	28	4	48	2	68	3	88	0
9	2	29	2	49	5	69	3	89	0
10	3	30	3	50	2	70	1	90	1
11	2	31	3	51	1	71	0	91	2
12	3	32	3	52	1	72	1	92	0
13	4	33	1	53	1	73	1	93	1
14	1	34	3	54	3	74	2	94	4
15	2	35	0	55	0	75	1	95	4
16	4	36	3	56	3	76	3	96	3
17	5	37	0	57	3	77	0	97	1
18	0	38	1	58	1	78	1	98	2
19	2	39	4	59	1	79	2	99	3
20	2	40	2	60	2	80	0	100	2

For the data set in Table 1:

- Mean = 2.1
- Variance = 2.0
- Overdispersion Factor = 0.97
- Percentage of zeros = 13%

The Overdispersion Factor is a numerical value that quantifies how much extra variability is present in a dataset compared to what we would expect under a Poisson distribution. It is defined as:

$$Overdispersion\ Factor\ = \frac{Variance}{Mean}$$

where:

- Variance measures the spread of the data.
- Mean is the expected count of occurrences.

In Poisson distribution, the mean and variance are equal:

$$\text{Variance=Mean}$$

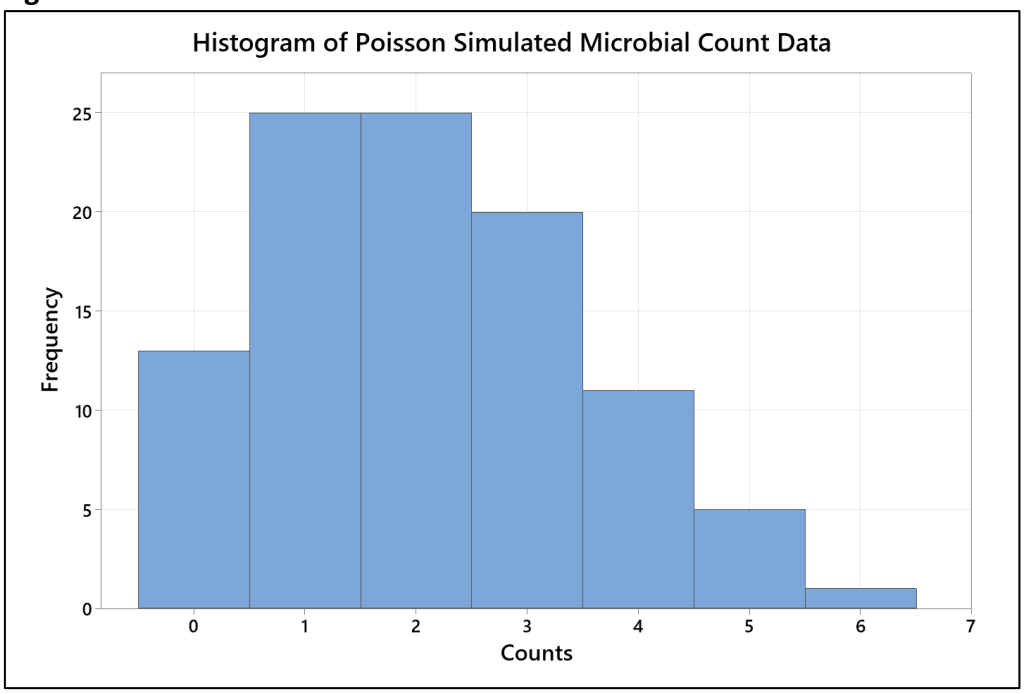
which means the Overdispersion Factor should be close to 1.
Since, in this case, the Overdispersion Factor = 0.97 ≈ 1, the Poisson model is appropriate.

The Overdispersion Factor has important practical consequences. In fact, if there is overdispersion (*i.e.*, Overdispersion Factor > 1), the use of a Poisson model can lead to incorrect conclusions (for example, too many false alarms in control charts). Furthermore, traditional control charts assume Poisson-distributed data.

Graphically, the data in Table 1 are presented as shown in Figure 1 below.

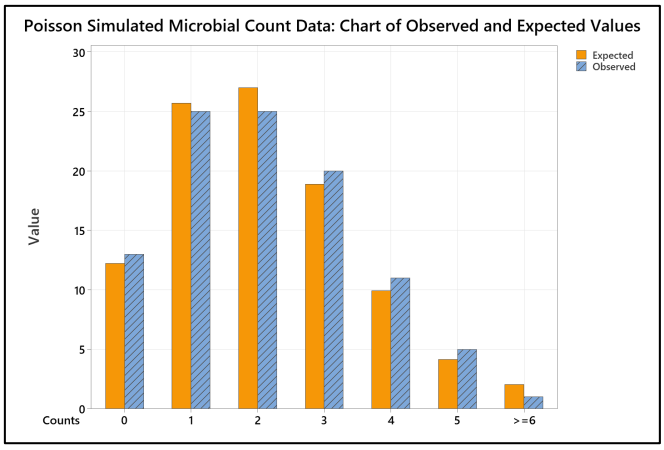
All graphs in this article were generated using Minitab® 22.1 and R. The R scripts used for the analyses and control charts are available in my GitHub repository at https://github.com/rbonfichi/microbial-counts.

Figure 1



As shown in Figure 2, using the Poisson goodness-of-fit test, it is possible to determine that the data in Table 1 follows a Poisson distribution.

Figure 2



Chi-Square Test

Null hypothesis H_0: Data follow a Poisson distribution
Alternative hypothesis H_1: Data do not follow a Poisson distribution

DF	Chi-Square	P-Value
5	1,09567	0,954

As mentioned above, identifying the most correct distribution to represent the experimental data is fundamental for choosing the most suitable control chart to monitor the process from which they come.

If the data follows a Poisson distribution, the appropriate control chart to use is the *c*-chart (count chart) as it is designed for count data (*e.g.*, microbial counts, defect counts).

The *c*- chart should be used to monitor the number of defects when each item can have multiple defects, and the so-called "subgroups" are all the same size. A typical example is the monitoring of the total number of CFU on an agar plate where 10 mL of sample are always plated.
The control limits are based on the Poisson mean (λ)
Since Poisson-distributed data has variance equal to its mean, the standard deviation is:

$$\sigma_c = \sqrt{\lambda}$$

Following the classical:

$$\text{UCL} = \mu+3\sigma, \text{LCL} = \mu-3\sigma$$

control limits for a *c*-chart are calculated as:

$$UCL = \bar{c} + 3\sqrt{\bar{c}}$$

$$LCL = \bar{c} - 3\sqrt{\bar{c}}$$

where:

- $\bar{c}$ = average count per sample (Poisson mean λ).
- UCL = Upper Control Limit
- LCL = Lower Control Limit (set to zero if negative, since counts cannot be negative).

In the case of data in Table 1 where we have microbial counts with a mean count of λ = 2.1:

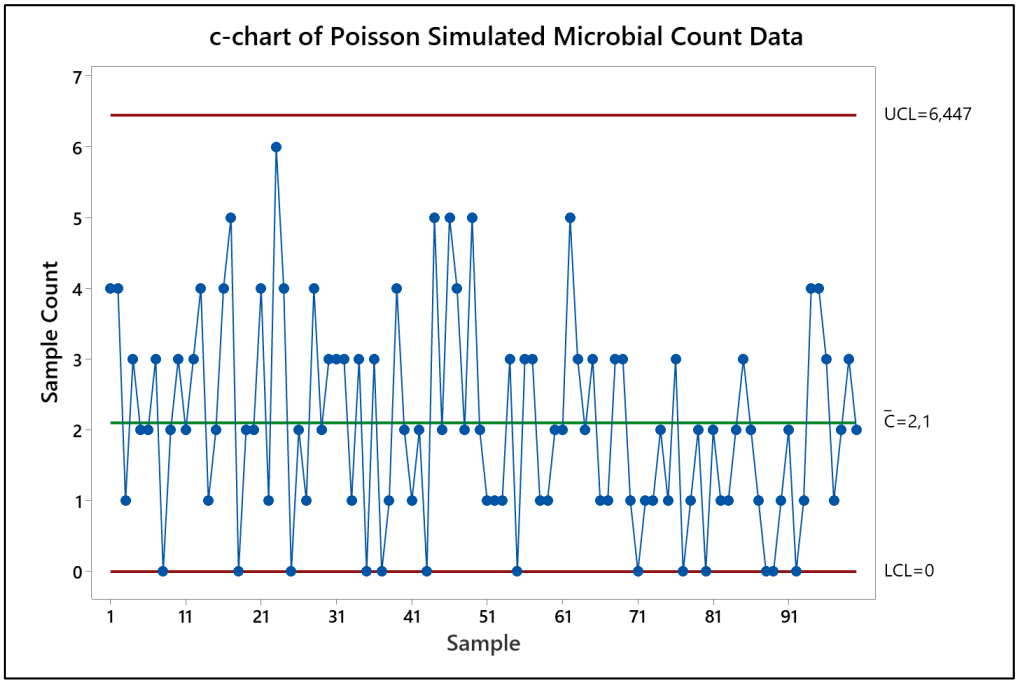
- $\sigma_c = \sqrt{\lambda} = \sqrt{2.1} = 1.45$
- UCL = 2.1 + 3(1.45) = 6.45
- LCL = 2.1 - 3(1.45) = -2.25 = 0

So, the parameters characterizing the control chart are, in this case:

$$\text{UCL= 6.45, LCL=0, } \bar{c} \text{ =2.1}$$

and the corresponding control chart is shown in Figure 3 below.

Figure 3

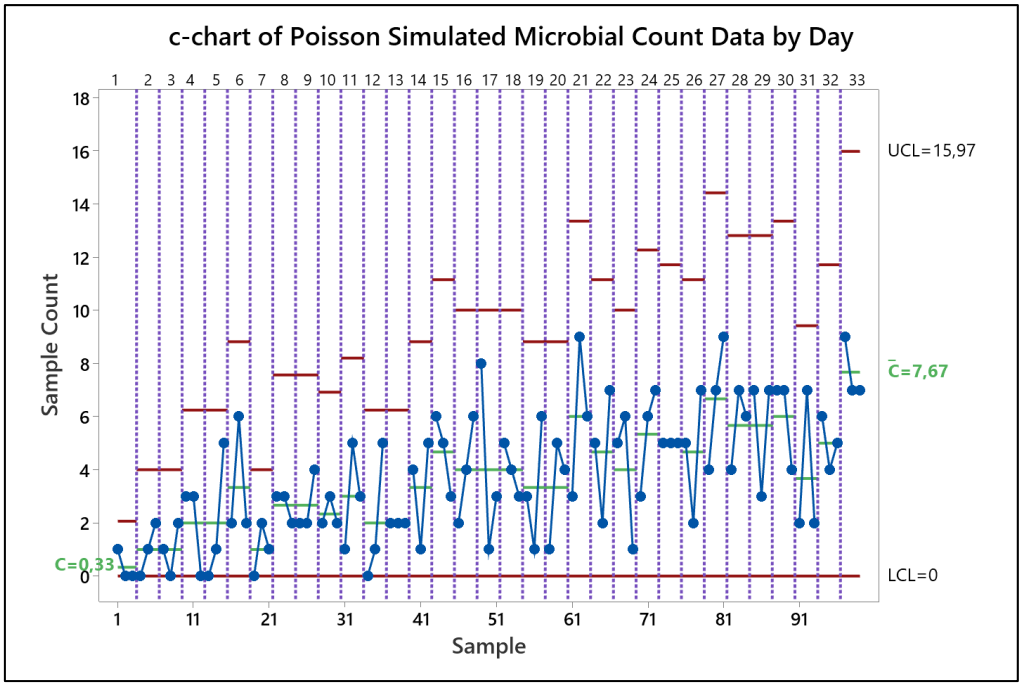


For many reasons such as:

- contamination accumulation because of ineffective cleaning,
- temperature and environmental changes (*e.g.*, warmer conditions during later shifts)
- sampling or equipment contamination (*e.g.*, residual contamination from previous samplings)
- operator influence (*e.g.*, differences in handling or sampling procedures across shifts)

it is possible that microbial counts may gradually increase showing a growth trend over time. Let's consider, for example, the case of monitoring a given sampling point conducted over the course of a month and, every day, at each work shift. Figure 4 shows how a *c*-control chart could, for example, show an upwards shift.

Figure 4



However, in practice, microbial counts often deviate from the Poisson assumptions. Overdispersion—where the variance exceeds the mean—is frequently observed due to environmental variability, sampling issues, and other factors. Additionally, many cleanroom datasets contain excess zeros, requiring the consideration of zero-inflated models.

Let us begin by examining the phenomenon of overdispersion.

2. WHY OVERDISPERSION OCCURS?

Overdispersion in microbial counts arises due to several factors that violate the assumptions of the Poisson distribution:

- Variability in environmental conditions: Fluctuations in airflow, humidity, and cleanliness can cause inconsistent microbial counts, leading to greater variability than expected under a Poisson model. Instead, according to the Poissonian model, rare occurrences happen at a constant rate.
- Non-independence of contamination events: A single contamination source can result in clusters of high counts, violating the independence assumption of the Poisson distribution.
- Sampling variability: Differences in sample collection methods, sample sizes, and handling introduce additional variability not accounted for by the Poisson model.

- Heterogeneous microbial distribution: Uneven distribution of microbes in controlled environments contributes to variability, conflicting with the Poisson assumption of a constant average rate of occurrence.

3. WHY OVERDISPERSION MATTERS?

Let's imagine two cleanrooms with the same average microbial count. If one has very variable counts (*e.g.*, 0, 0, 0, 10, 2) while the other has consistent counts (*e.g.*, 2, 3, 2, 3, 1), the first case shows overdispersion. Traditional Poisson-based control charts may trigger false alarms in such cases, necessitating models like the Negative Binomial.

4. FEATURES OF OVERDISPERSED DATA

As mentioned above, overdispersed data are mainly characterized by a variance which significantly exceeds the mean, contrary to the Poisson model where mean equals variance.
Additionally, a high proportion of zero counts, interspersed with occasional large counts, creating a distribution with long tails.
Graphically, overdispersed data may appear as a right-skewed distribution, highlighting the limitations of the Poisson model in accurately describing such data.
The comparison shown in Figure 5 between overdispersed simulated data (Figure 5a) and Poisson-type simulated data (Figure 5b) is illustrative.

Figure 5a

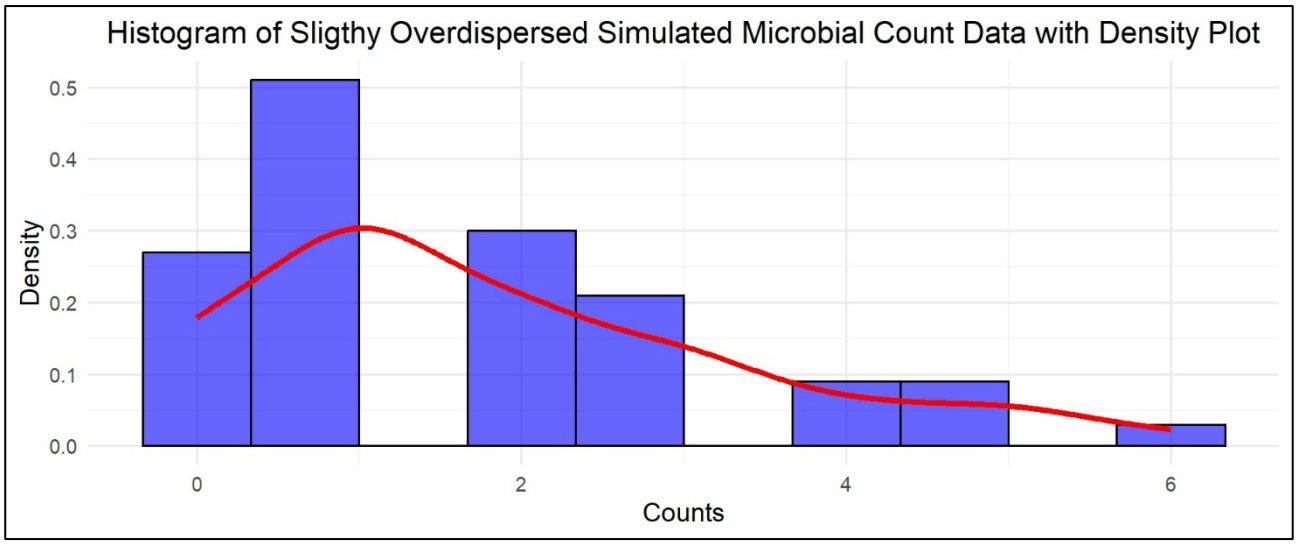
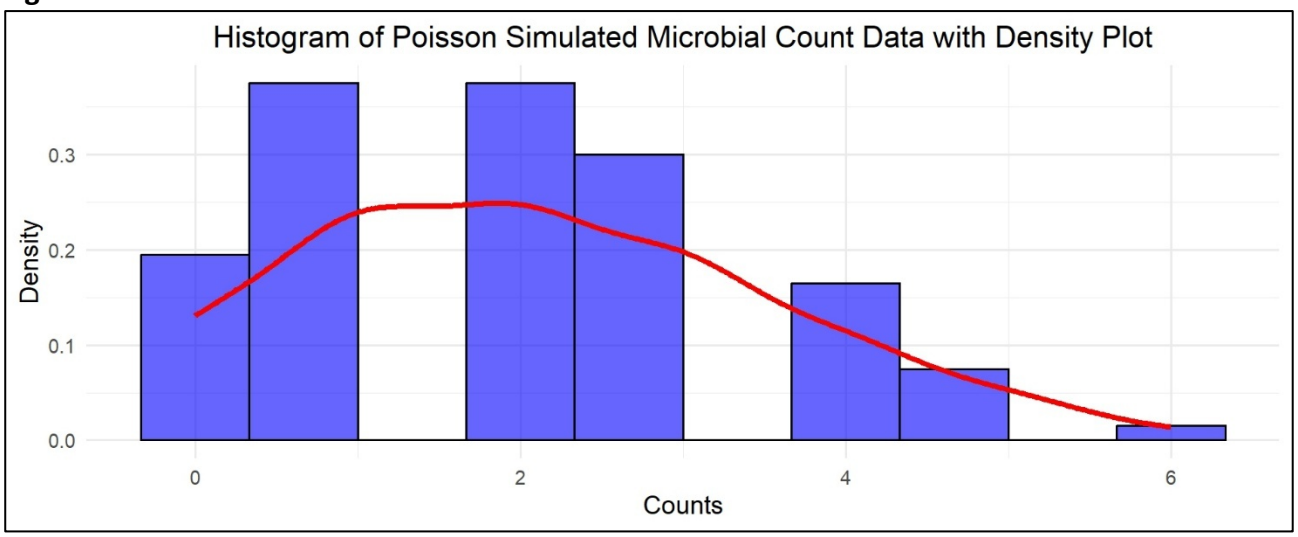


Figure 5b



The data graphically represented in Figure 5a are those collected in Table 2 below.

Table 2

Sample ID	Count	Sample ID	Count	Sample ID	Count	Sample ID	Count	Sample ID	Count
1	1	21	1	41	6	61	3	81	3
2	2	22	3	42	3	62	2	82	2
3	0	23	3	43	1	63	5	83	1
4	3	24	2	44	4	64	5	84	5
5	0	25	3	45	1	65	3	85	0
6	1	26	2	46	1	66	1	86	3
7	2	27	0	47	5	67	0	87	2
8	1	28	0	48	4	68	1	88	1
9	3	29	4	49	1	69	3	89	4
10	4	30	1	50	1	70	1	90	1
11	0	31	1	51	5	71	2	91	2
12	1	32	2	52	0	72	1	92	0
13	1	33	3	53	2	73	4	93	1
14	5	34	2	54	0	74	1	94	1
15	1	35	1	55	1	75	2	95	2
16	3	36	1	56	0	76	0	96	0
17	0	37	1	57	1	77	1	97	2
18	1	38	0	58	0	78	2	98	2
19	2	39	2	59	1	79	6	99	2
20	0	40	0	60	1	80	1	100	3

For the dataset in Table 2:

- Mean = 1.8
- Variance = 2.3
- Overdispersion Factor = 1.26
- Percentage of zeros = 18%

In this case the overdispersion factor is slightly greater than 1.

The simulated Overdispersed data displayed in Figure 5a, and collected in Table 2, were obtained using the so-called **Negative Binomial distribution**, a more versatile alternative to the Poissonian model.

5. NEGATIVE BINOMIAL DISTRIBUTION

Negative Binomial is an extension of the Poisson distribution that introduces an extra dispersion parameter, allowing the variance to be greater than the mean.
It is used when overdispersion is present, meaning the variance of the data is significantly larger than the mean.
While Poisson assumes that the variance = mean, the Negative Binomial assumes:

$$Variance \ = \mu + \frac{\mu^2}{r}$$

where:

- μ is the mean count,
- r is the dispersion parameter (lower values indicate greater overdispersion).

Negative Binomial distribution should be used when:

- count data exhibit overdispersion (*i.e.*, variance > mean).
- the excess variability arises from unobserved heterogeneity (*e.g.*, different environmental conditions leading to fluctuating microbial counts).

Even in this case, identifying the correct distribution to represent the experimental data is fundamental for choosing the most suitable control chart to monitor the process from which they come.
To handle overdispersed counts it can be used:

- Negative Binomial Control Chart (*NB*-chart)
- Laney *u'* and *p'*-Charts (for overdispersion adjustments).

6. NEGATIVE BINOMIAL AND LANEY CONTROL CHARTS

It is important to note that Negative Binomial Control Charts (*NB*-charts) and Laney charts (*u'*- and *p'*-charts) are related but not the same. They are both used for overdispersed data, but they handle overdispersion in different ways. In particular:

- Negative Binomial Control Charts (*NB*-charts) model the data using a Negative Binomial distribution

- Laney *u'* and *p'*-charts apply a statistical correction (sigma-z scaling) to adjust control limits.

So, even if Negative Binomial control chart is not a Laney chart, they are both used to address overdispersion as summarized in Table 3 below:

Table 3

Feature	Negative Binomial Control Chart (*NB*-chart)	Laney *u'* & *p'*-Charts
Distribution Used	**Negative Binomial** (variance > mean)	**Poisson or Binomial** (adjusted for overdispersion)
How It Works	Uses a **Negative Binomial model** with an extra dispersion parameter (r)	Applies a **sigma-z correction** to adjust control limits
Best For	Data with **excess variance** due to **clustering of events** (*e.g.*, microbial counts)	**Slight to moderate overdispersion** in Poisson or Binomial data
Advantages	More flexible, **models high variability accurately**	Easy to implement in existing **Poisson/Binomial charts**
Limitations	Requires estimating an extra **dispersion parameter (r)**	May not fully correct extreme overdispersion

As is once again evident, it is essential, first, to establish how the data is distributed.
To sum it all up in a few words:

- if the variance is only slightly greater than the mean, use Laney charts.
- if the variance is significantly greater than the mean, use a Negative Binomial Control Chart.

Both Laney *u'*-charts and *p'*-charts adjust for overdispersion, but they are used for different types of data. In particular:

Table 4

Chart Type	Used For	Formula Adjusted
Laney *u'*-Chart	Rates (per unit defects or microbial counts per sample)	$\sigma' = \sigma \times Z$ (sigma-z correction)
Laney *p'*-Chart	Proportions (defectives per batch, pass/fail data)	$\sigma' = \sigma \times Z$ (sigma-z correction)

The key differences between these two types of control charts are as follows:

- ***u'*-chart** → Used for **count-based data**, like microbial counts per sample.
- ***p'*-chart** → Used for **proportions**, like % of samples exceeding a microbial limit.

The formula for Overdispersion Correction or **Laney's Z-Score Adjustment**, is:

$$\sigma' = \sigma \times Z$$

where *Z* is a **correction factor** that adjusts for **excess variance**.

Everything we have seen so far is summarized in the summary Table 4 below.

Table 5

Scenario	**Best Control Chart**
Poisson-distributed microbial counts	Traditional c-chart
Poisson-distributed defect rates (per unit)	*u*-chart
Binomial-distributed defect proportions	*p*-chart
Overdispersed Poisson or Binomial data (mild/moderate overdispersion)	Laney *u'* or *p'*- charts
Highly overdispersed data (Variance >> Mean)	Negative Binomial Control Chart (*NB*-chart)

In light of the above, the data in Table 2 showing mild/moderate overdispersion (the overdispersion factor is in fact slightly higher than 1) can therefore be treated equally well using both Negative Binomial and Laney *u'*-control charts (see Figures 6 and 7).

Figure 6

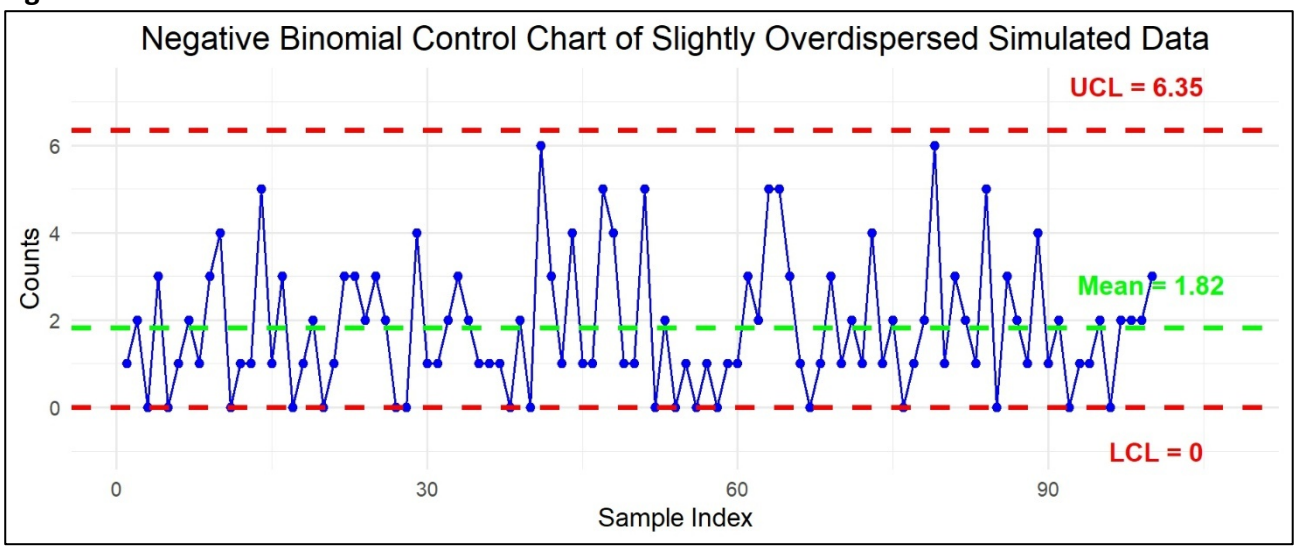
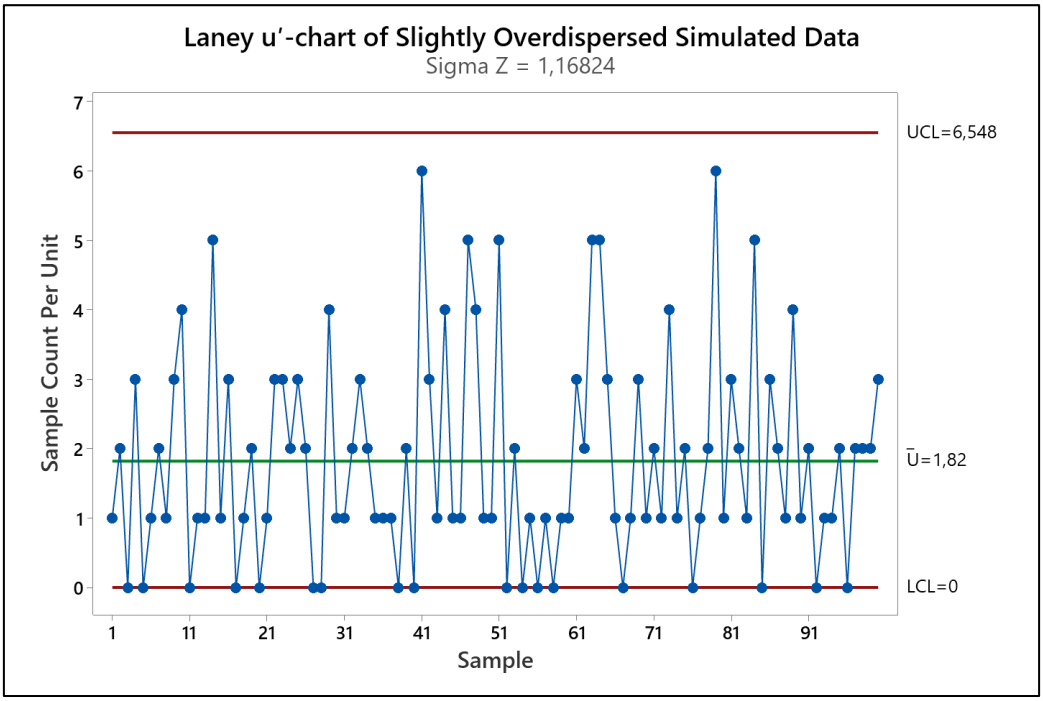


Figure 7



Let us now consider, as an example, a series of simulated data characterized by **high overdispersion** such as those reported in Table 6 below.

Table 6

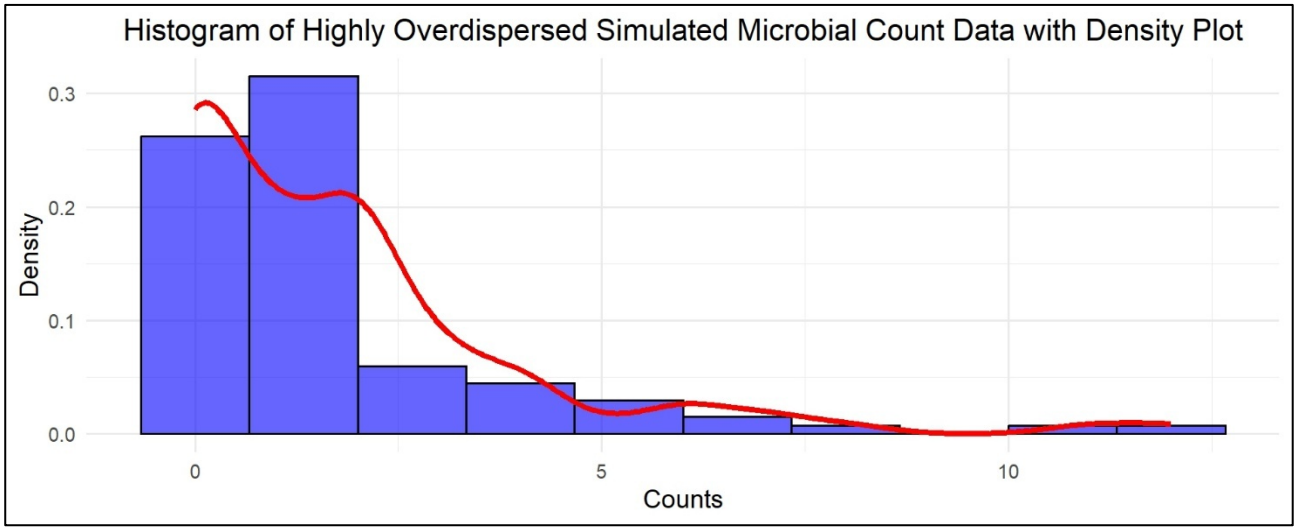
Sample ID	Count	Sample ID	Count	Sample ID	Count	Sample ID	Count	Sample ID	Count
1	4	21	0	41	0	61	0	81	1
2	1	22	0	42	4	62	2	82	1
3	0	23	1	43	2	63	0	83	2
4	1	24	2	44	6	64	2	84	0
5	1	25	4	45	2	65	1	85	1
6	0	26	5	46	2	66	3	86	0
7	2	27	8	47	2	67	7	87	0
8	7	28	1	48	1	68	0	88	0
9	0	29	6	49	1	69	0	89	1
10	0	30	0	50	2	70	0	90	2
11	1	31	3	51	2	71	0	91	2
12	1	32	0	52	3	72	4	92	0
13	1	33	2	53	3	73	0	93	1
14	0	34	0	54	0	74	0	94	2
15	1	35	2	55	3	75	0	95	0
16	2	36	0	56	11	76	0	96	0
17	0	37	2	57	6	77	1	97	0
18	3	38	12	58	2	78	4	98	1
19	2	39	0	59	4	79	2	99	2
20	3	40	0	60	2	80	0	100	3

For the dataset in Table 6:

- Mean = 1.8
- Variance = 5.2
- Overdispersion Factor = 2.85
- Percentage of zeros = 35%

The overdispersion factor is now significantly greater than 1.
In this case, the data show a much more pronounced right-skewed distribution than that shown by the data in Table 2 and visualized in Figure 5a.

Figure 8



As can be seen from the control charts shown below in Figures 9 and 10, in the case of highly overdispersed data, the choice of the correct reference distribution makes the difference for the purposes of correct monitoring of the process which, otherwise, would be full of alarms.
In Figure 9 (Laney *u'*-chart) we observe at least five data points that exceed or are at the level of the Upper Control Limit (UCL) while in Figure 10 (Negative Binomial control chart) this number is reduced to three.

Figure 9

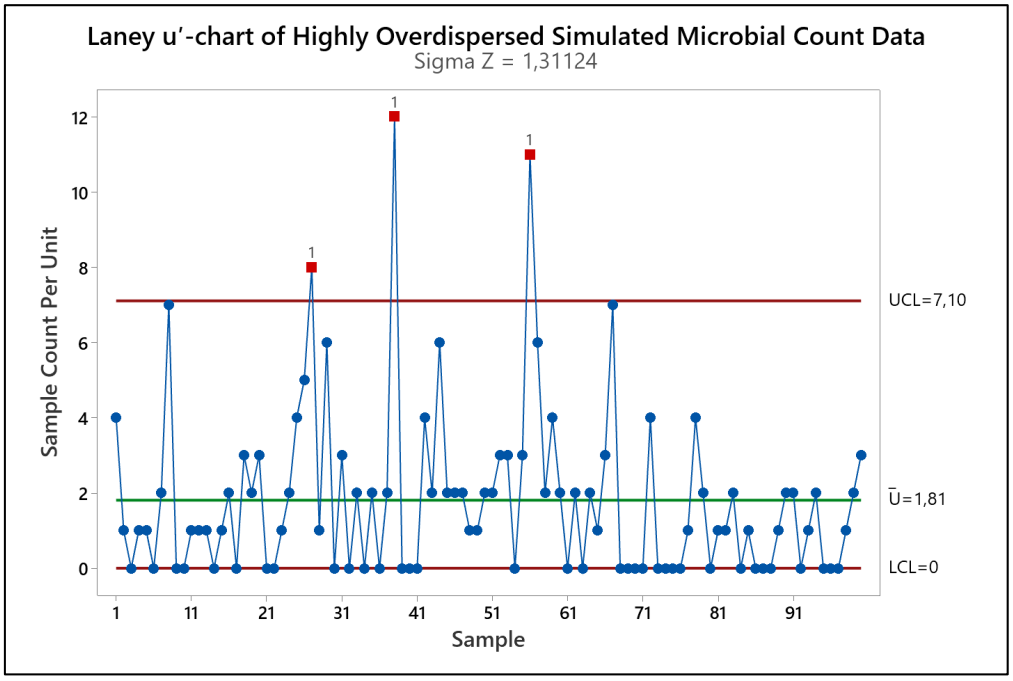
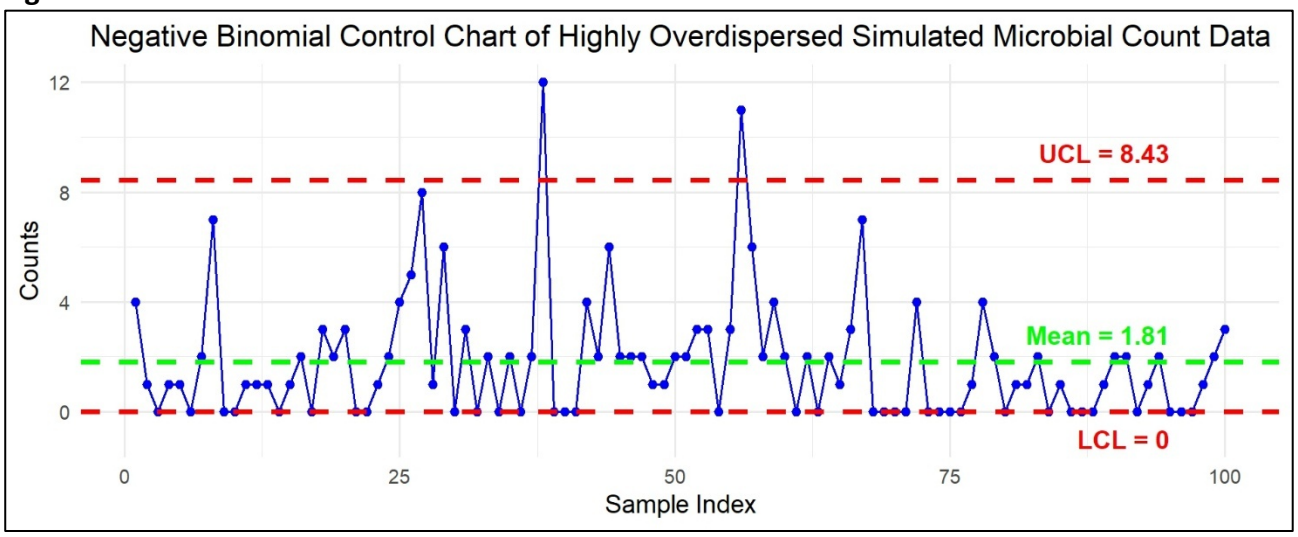


Figure 10



While Negative Binomial distribution handles high overdispersion as in the case of Table 6 data, it does not explicitly model excess zeros. If data have more zeros than expected, a **Zero-Inflated Model (ZIM)** might be better.

7. ZERO-INFLATED MODELS (ZIMS)

Real-world microbial data often show greater variability than that observed until now.
In fact, in controlled environments such as cleanrooms, microbial counts often include excess zeros. These zeros may arise from two different processes:

- true structural zeros (no contamination possible in certain conditions)
- random zeros (low but nonzero microbial presence)

In this case, the **Zero-Inflated Poisson (ZIP)** and **Zero-Inflated Negative Binomial (ZINB)** models account for these two sources of zeros by combining a zero-generating process with either a Poisson or Negative Binomial count process.
In practice a Zero-Inflated Model consists of:

- a binary process that decides if the observation is in the "always zero" group.
- a count process (Poisson or Negative Binomial) for cases where counts occur.

Mathematically, the probability mass function (PMF) is:

$$P(Y = y) \begin{cases} = p + (1-p)\, f(0,\ \theta) & if\ y = 0 \\ = (1-p)\, f(y,\ \theta) & if\ y > 0 \end{cases}$$

where:

- p is the probability of being in the **structural zero** state,
- $f(y, \theta)$ is the Poisson or Negative Binomial probability mass function.

There are basically two types of Zero-Inflated Models:

- **Zero-Inflated Poisson (ZIP)**: when the count data follows a Poisson distribution.
- **Zero-Inflated Negative Binomial (ZINB)**: when the count data is overdispersed and follows a Negative Binomial distribution.

They should be used when:

- there are more zeros than a standard Poisson or Negative Binomial model would predict.
- the data suggest two sources of zeros (*e.g.*, some locations always have no microbes, while others follow a count distribution).

In short, the main differences between Negative Binomial and Zero-Inflated Models can be summarized as shown in Table 7 below:

Table 7

Feature	Negative Binomial (NB)	Zero-Inflated Models (ZIP/ZINB)
Handles Overdispersion?	Yes	Yes (ZINB)
Handles Excess Zeros?	No	Yes
Uses Two Processes?	No	Yes (binary + count)
Best When?	Overdispersion is the main issue	There are excess zeros from two sources

In summary we should therefore use:

- **Negative Binomial (NB)** if variance > mean but zeros follow the expected pattern.
- **Zero-Inflated Poisson (ZIP)** if there are more zeros than expected.
- **Zero-Inflated Negative Binomial (ZINB)** if there are both excess zeros and overdispersion.

To simplify and clarify better, let's now look at two explanatory examples.

8. EXAMPLE 1

Table 8 below summarizes a series of simulated microbial count data characterized by **excess zeros**, that is, the percentage of zeros relative to the total observations is greater than 30%. This, in general, is a good sign of **zero inflation**.

Table 8

Sample ID	Count	Sample ID	Count	Sample ID	Count	Sample ID	Count	Sample ID	Count
1	0	21	0	41	0	61	2	81	0
2	0	22	0	42	0	62	3	82	2
3	2	23	0	43	1	63	0	83	1
4	0	24	0	44	2	64	1	84	3
5	4	25	0	45	2	65	0	85	0
6	2	26	3	46	1	66	2	86	0
7	0	27	1	47	0	67	0	87	0
8	1	28	0	48	1	68	2	88	0
9	2	29	2	49	3	69	3	89	0
10	6	30	3	50	0	70	2	90	2
11	0	31	0	51	2	71	0	91	0
12	0	32	2	52	3	72	0	92	3
13	0	33	0	53	0	73	0	93	1
14	1	34	0	54	1	74	1	94	2
15	3	35	2	55	0	75	0	95	0
16	2	36	0	56	1	76	4	96	0
17	7	37	0	57	1	77	0	97	0
18	3	38	3	58	2	78	0	98	0
19	0	39	0	59	0	79	0	99	0
20	2	40	0	60	0	80	1	100	0

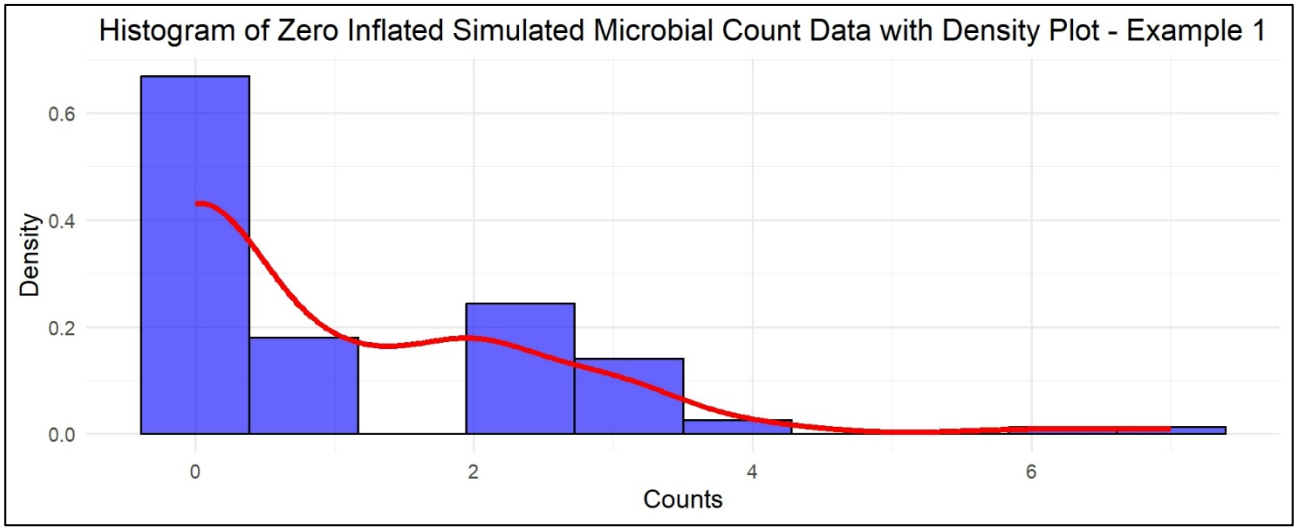
For the dataset in Table 8:

- Mean = 1.1
- Variance = 2.0
- Overdispersion Factor = 1.85
- Percentage of zeros = 52%

In this case the overdispersion factor is certainly greater than 1 (→ **overdispersion**), but the dominant feature of the dataset is the percentage of zeros present which is equal to 52% (→ **zero inflation**).

Again, the data show a pronounced right-skewed distribution as in Figure 8 but, unlike Figure 8, here the zero frequency dominates the histogram.

Figure 11



Since a comparison based on **Akaike Information Criterion (AIC)** of which of the two models ZIP or ZINB was more suitable for the data in Table 8 did not reveal any significant differences, I decided to use ZIP (Zero-Inflated Poisson) as it is simpler. Based on this, the control limits were calculated for the chart shown in Figure 12 as follows:

$$UCL = \lambda + 3\sqrt{\lambda\,(1-p)}$$

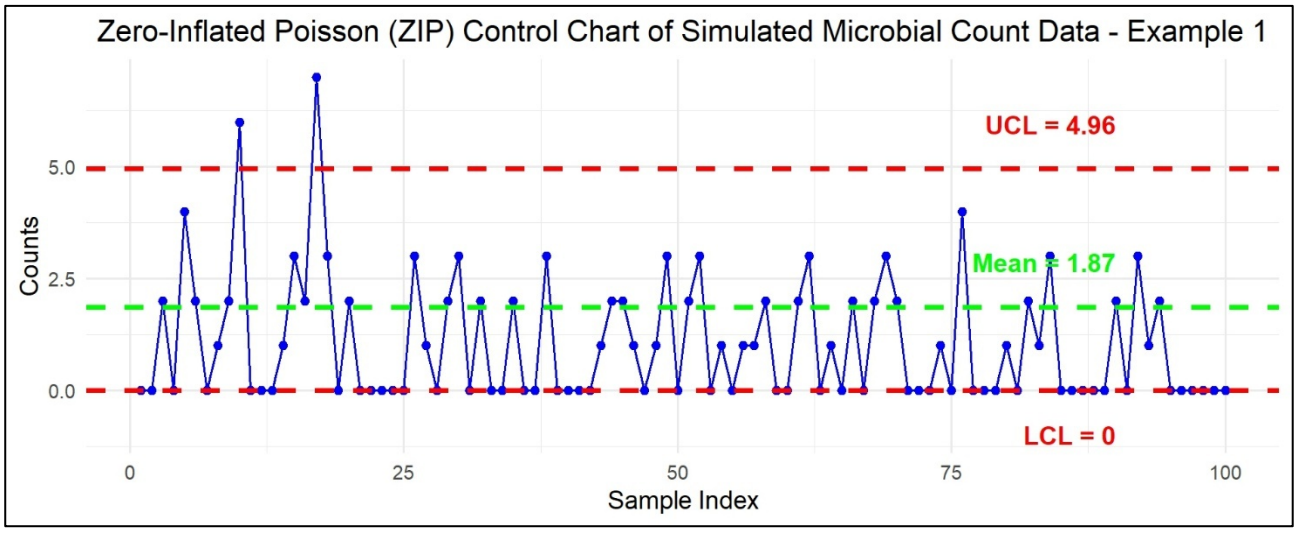
$$LCL = max\left(0, \lambda - 3\sqrt{\lambda\,(1-p)}\right)$$

where:

λ = Poisson mean

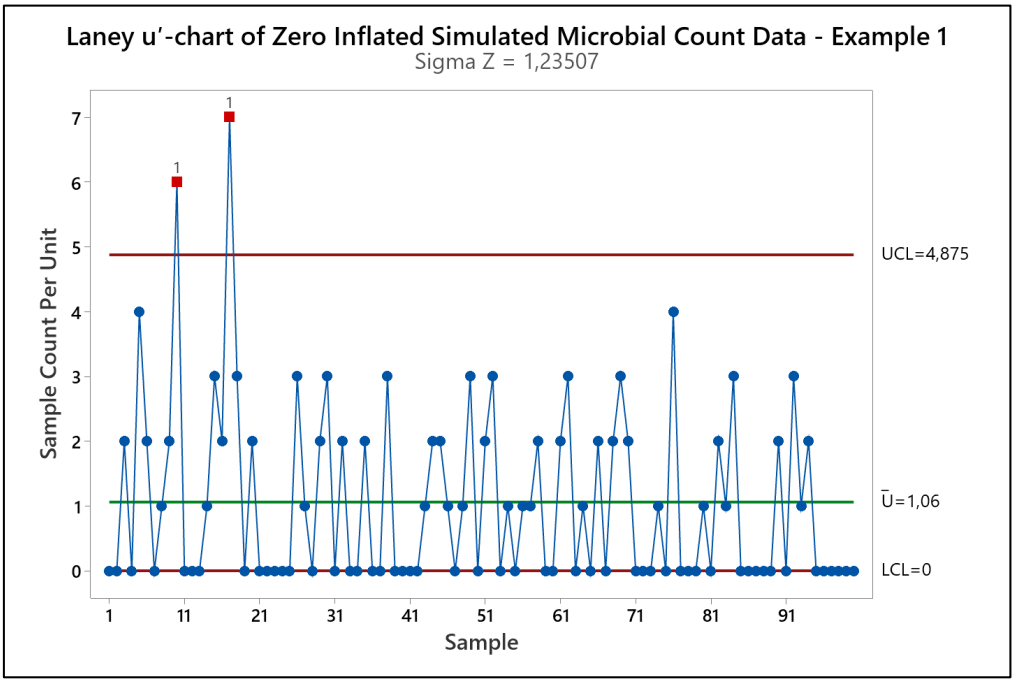
p = Zero-Inflation probability.

Figure 12



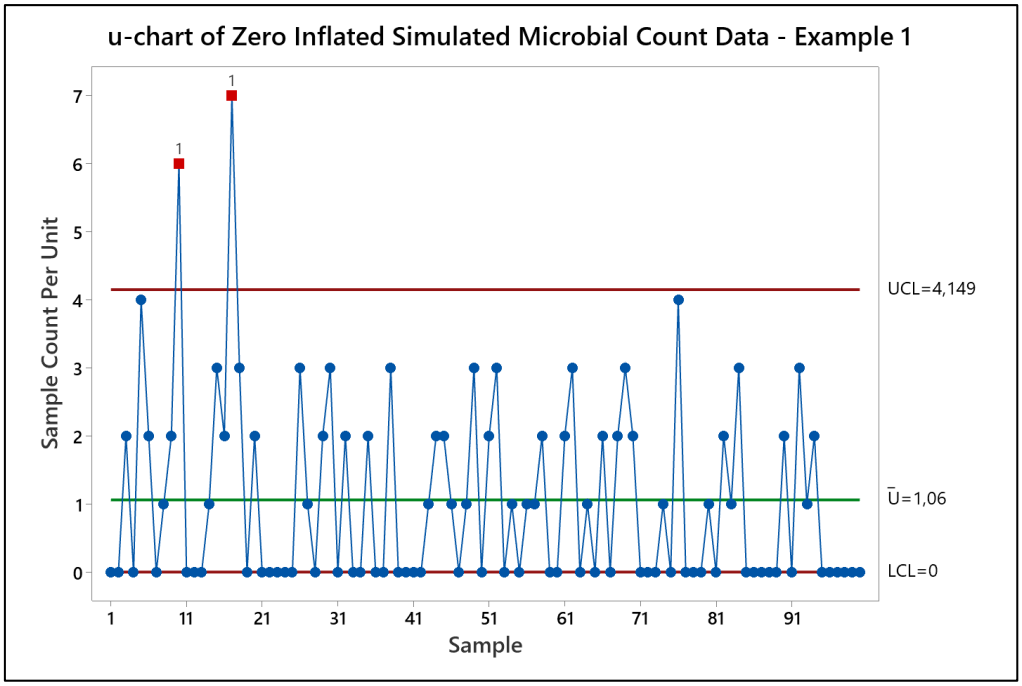
Apart from the differences in the mean and UCL values, the Laney *u*′-control chart in Figure 13 is quite similar to the one obtained using ZIP.

Figure 13



Conversely, as shown in Figure 14, using a simple *u*-control chart would have instead led to a lower UCL and thus increased the possibility of alarms.

Figure 14



9. EXAMPLE 2

In Table 9 below, a second set of simulated microbial count data is collected, characterized by an excess of zero values, that is the percentage of zeros with respect to the total observations is greater than 30%. This, in general, is a good sign of zero inflation.

Table 9

Sample ID	Count	Sample ID	Count	Sample ID	Count	Sample ID	Count	Sample ID	Count
1	0	21	2	41	0	61	3	81	0
2	0	22	6	42	0	62	2	82	0
3	0	23	3	43	0	63	0	83	1
4	1	24	0	44	0	64	2	84	0
5	5	25	2	45	1	65	1	85	1
6	3	26	0	46	0	66	3	86	0
7	2	27	1	47	0	67	1	87	0
8	0	28	6	48	0	68	2	88	0
9	4	29	4	49	3	69	1	89	5
10	0	30	0	50	0	70	0	90	3
11	0	31	0	51	3	71	8	91	4
12	0	32	0	52	1	72	1	92	0
13	0	33	1	53	0	73	0	93	0
14	1	34	0	54	2	74	0	94	4
15	0	35	2	55	4	75	0	95	0
16	0	36	0	56	1	76	0	96	0
17	0	37	4	57	0	77	0	97	0
18	0	38	3	58	0	78	0	98	5
19	0	39	3	59	4	79	0	99	0
20	1	40	0	60	0	80	4	100	1

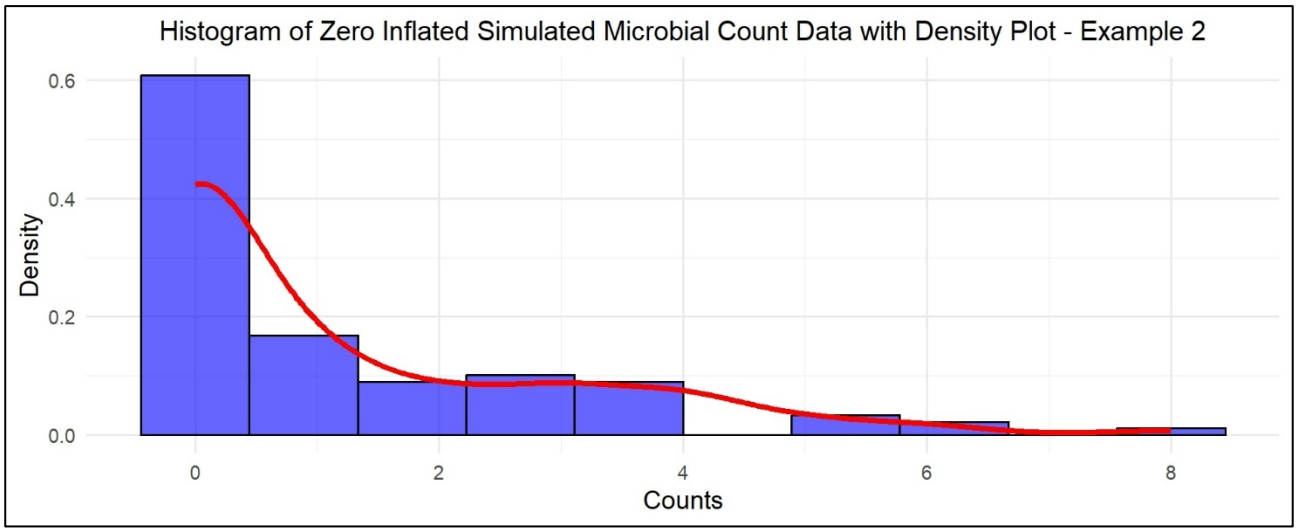
For the dataset in Table 9:

- Mean = 1.3
- Variance = 3.1
- Overdispersion Factor = 2.51
- Percentage of zeros = 54%

In this case the percentage of zeros in the dataset is comparable to that of Example 1 (54% *vs.* 52%) while the overdispersion factor is significantly higher (2.51 *vs.* 1.85).

The appearance of the data (see Figure 15 below) shows a markedly right-skewed distribution with zero frequency dominating the histogram.

Figure 15



A comparison based on Akaike Information Criterion (AIC) and Overdispersion factor has shown that ZINB model was more suitable for the data in Table 9. Based on this, the control limits were calculated for the chart shown in Figure 16 as follows:

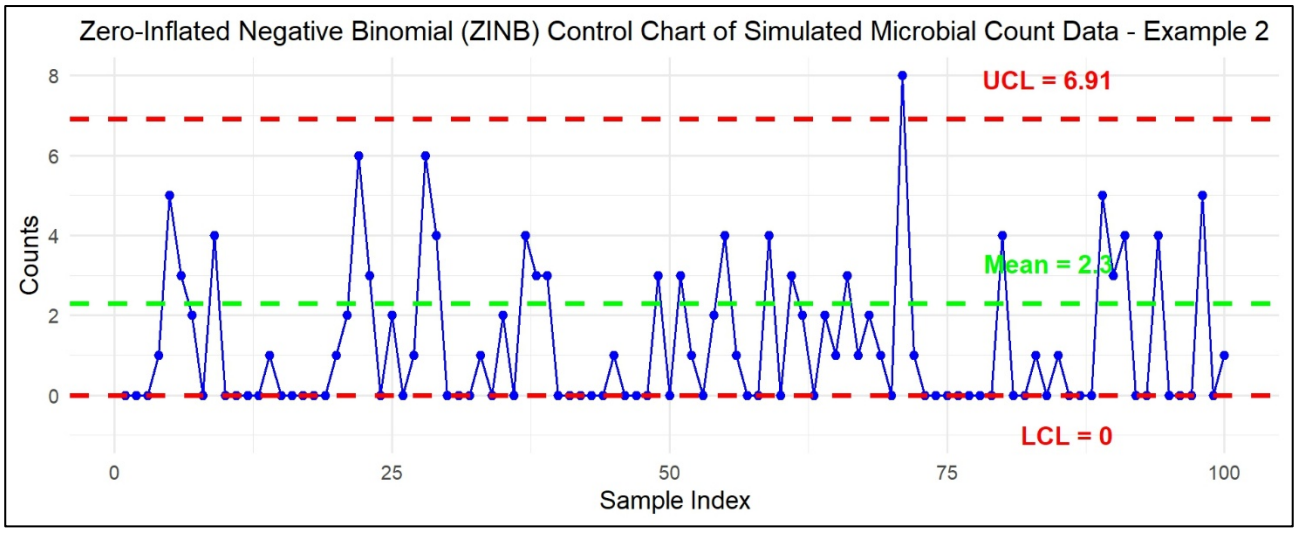
$$UCL = \lambda + 3 \times \sqrt{\lambda\,(1-p) + \frac{\lambda^2}{\theta}}$$

$$LCL = max\left(0, \lambda - 3 \times \sqrt{\lambda\,(1-p) + \frac{\lambda^2}{\theta}}\right)$$

where:

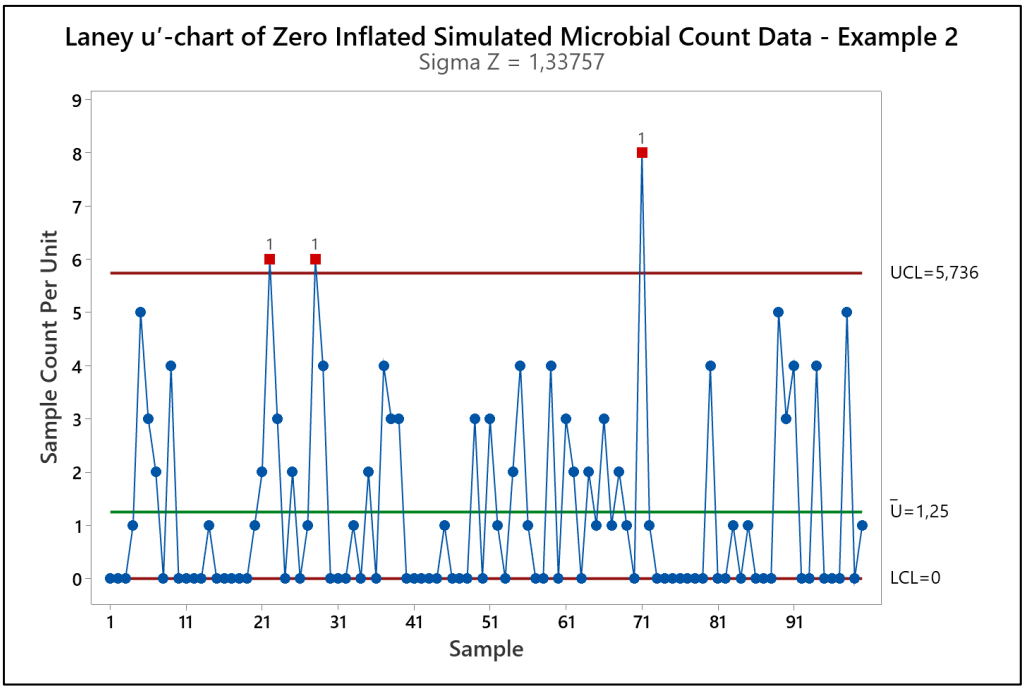
λ	=	Estimated mean count (fitted by ZINB model)
p	=	Zero-inflation probability (excess zeros adjustment)
θ	=	Dispersion parameter (controls overdispersion in NB)
$\sqrt{\lambda\,(1-p)+\frac{\lambda^2}{\theta}}$	=	Standard deviation, considering overdispersion and zero-inflation
$3\times$	=	Multiplies by 3 standard deviations to set control limits
max(0, LCL)	=	Ensures LCL is not negative

Figure 16



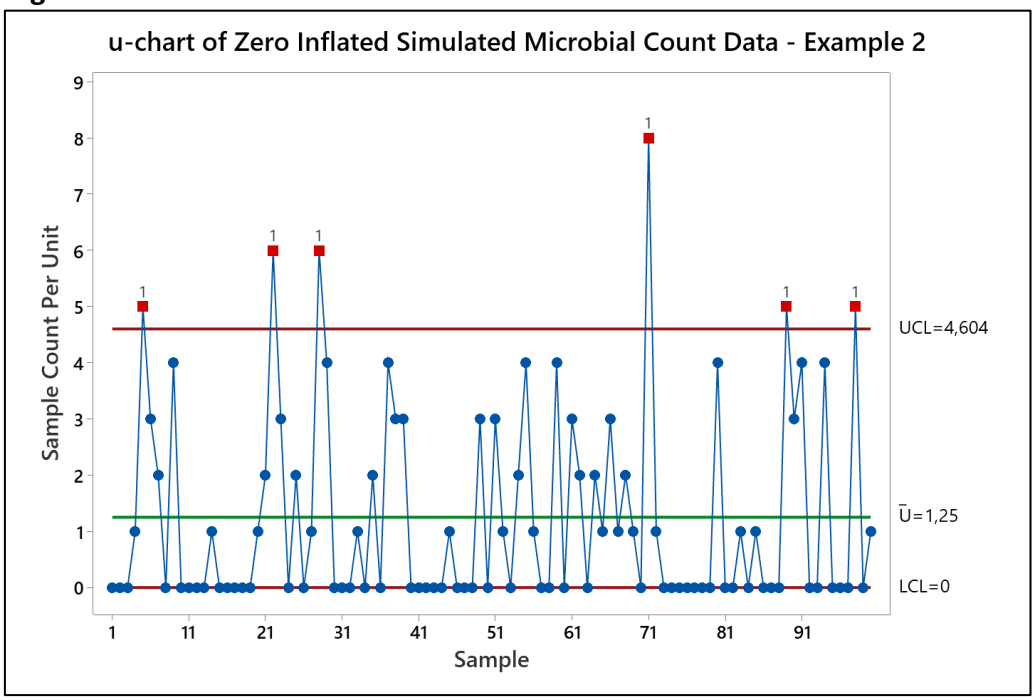
In this case, unlike Example 1 where the Overdispersion Factor was less different from 1, the Laney u' control chart in Figure 17 looks quite different from the one obtained by ZINB.

Figure 17



If only a simple *u*-control chart had been used, the resulting situation would have been even worse as clearly highlighted in Figure 18.

Figure 18



10. CONCLUSIONS: AVOIDING FALSE ALARMS WITH THE RIGHT MODEL

Selecting the correct statistical model for microbial count data is **critical** in pharmaceutical and controlled environments. While the **Poisson distribution is foundational**, it often does **not** adequately describe real-world data due to **overdispersion** or **excess zeros**.

This study highlights that:

- When **variance is significantly greater than the mean**, **Negative Binomial models** provide a **better fit** than Poisson.
- When **excess zeros** are present, **Zero-Inflated Models (ZIP, ZINB)** are often necessary to **avoid false alarms** in control charts.
- **Control charts must be adapted** to the data distribution to prevent **incorrect process monitoring decisions**.

If the data distribution is **misidentified**, traditional control charts may lead to **frequent unnecessary interventions (false alarms) or failure to detect real process deviations**. **There is no "one-size-fits-all" solution**—choosing the right statistical model is essential for ensuring **reliable monitoring and decision-making**.

Furthermore, an improper selection of control charts results in **incorrect control limits**, which not only affect current process monitoring but also have long-term consequences. These control limits are often used in **Continued Process Verification (CPV)**, meaning that an initial misclassification of data distribution could lead to **inadequate long-term process control strategies**. Overly tight limits may cause **excessive false alarms**, while excessively wide limits could **fail to detect real deviations** in microbial trends.

Crucially, **microbial count data are discrete by nature**, meaning they require **statistical models that appropriately reflect their properties**. Misapplying continuous-variable models to discrete count data can result in **misleading control limits and incorrect process evaluations**. Recognizing this fundamental characteristic is essential for both **short-term monitoring and long-term process verification**.

By integrating **data analysis tools**, practitioners can perform **goodness-of-fit tests**, assess overdispersion, and select the **most appropriate control chart**, ensuring that microbial monitoring remains **both sensitive and reliable** not only today but also in the long-term context of process validation and Continued Process Verification.

11. REFERENCES

- D.D. Wackerly, W. Mendenhall, R.L. Schaeffer, *Mathematical Statistics with Applications*, CPI Group (UK) Ltd., 2021
- J.M. Hilbe, *Modeling Count Data*, Cambridge University Press, 2014
- A.F. Zuur. E.N. Ieno, *Beginner's Guide to Zero-inflated Models with R*, Highland Statistics Ltd., 2016
- H. Wickham, *ggplot2 - Elegant Graphics for Data Analysis*, 2nd Ed., Springer, 2016
- H. Wickham, M. Çetinkaya-Rundel, G. Grolemund, *R for Data Science – Import, Tidy, Transform, Visualize, and Model Data*, 2nd Ed., O'Reilly, 2023