DISCOVERY SUMMIT

Introducing Limits of Detection in the Distribution Platform

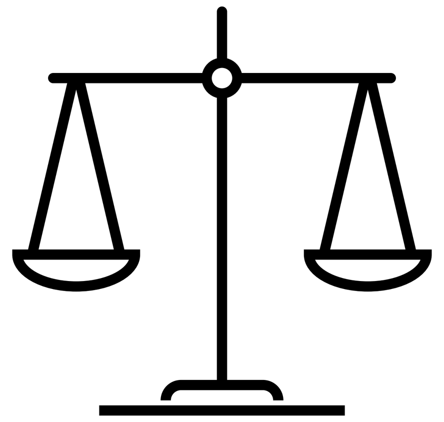
Clay Barker and Laura Lancaster, JMP Principal Research Statistician Developers

JMP STATISTICAL DISCOVERY

What is a limit of detection problem?

When our measurement device is not able to provide reliable measurements below a fixed threshold, this is limit of detection problem.

For example:
Imagine that our data is made up of weight measurements, but our scale is not able to reliably measure anything below 1 gram. Those rows are recorded as 1.0.



Because of the limitations of our scale, our data may look something like on the right.

	weight
1	1.0
2	5.5
3	3.2
4	1.0
5	1.0
6	6.0

We don't really believe those values of 1.0, we just know that they are at most 1.0.

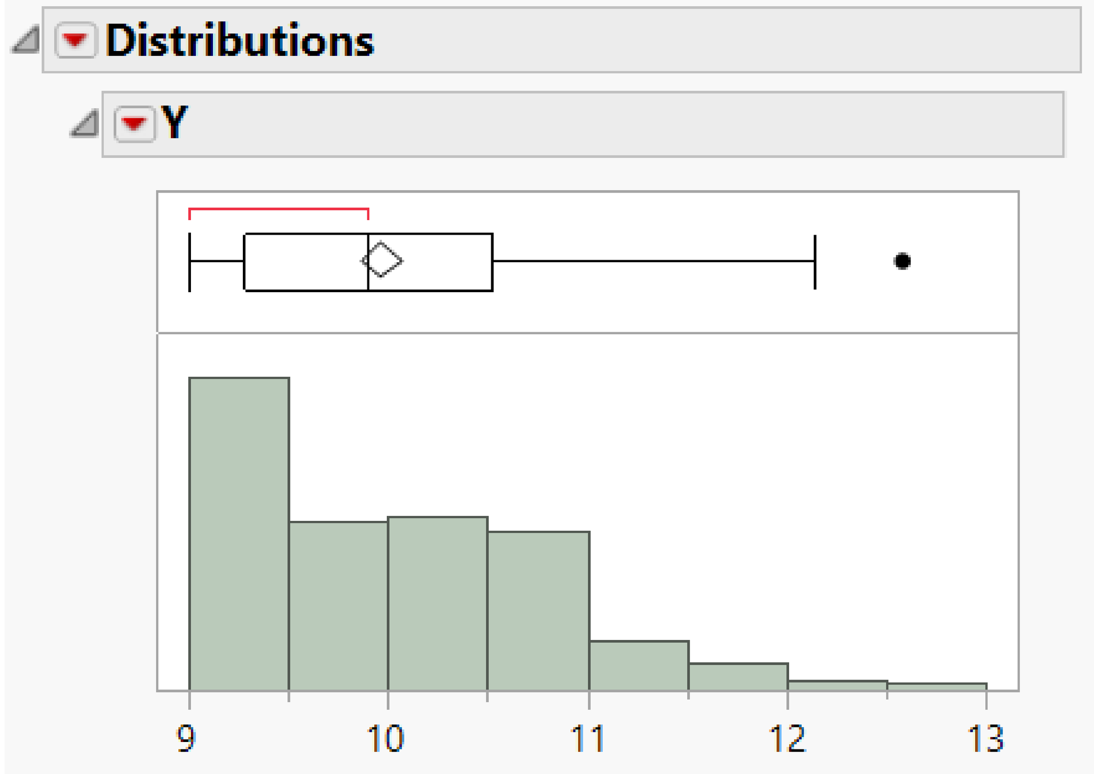
This type of data is very common in practice.

What should we do about it?

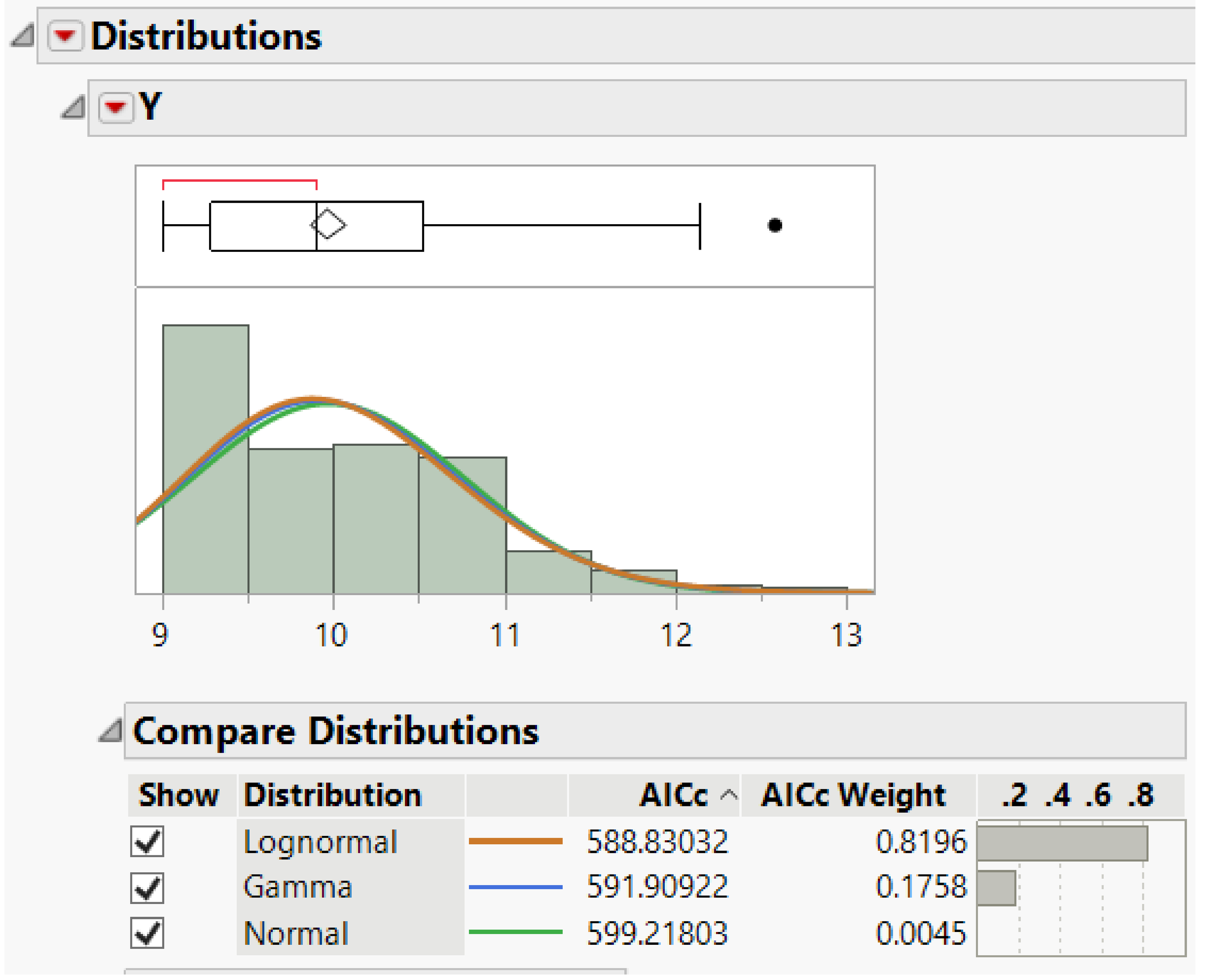
Consider a simulated example.

Suppose $Y_i \sim$ Normal(10,1), but we have a lower detection limit of 9. Our data will look like below.

1/0 Cols	Y
1	9.65
2	10.82
3	9.51
4	10.45
5	9.37
6	9.00



What happens if we model Y as-is? The results are not great.



Fitted Normal Distribution

Parameter		Estimate	Lower 95%	Upper 95%
Location	μ	9.9705033	9.8711955	10.069811
Dispersion	σ	0.7972399	0.7329456	0.873995

DISCOVERY SUMMIT

Introducing Limits of Detection in the Distribution Platform

Clay Barker and Laura Lancaster, JMP Principal Research Statistician Developers

JMP STATISTICAL DISCOVERY

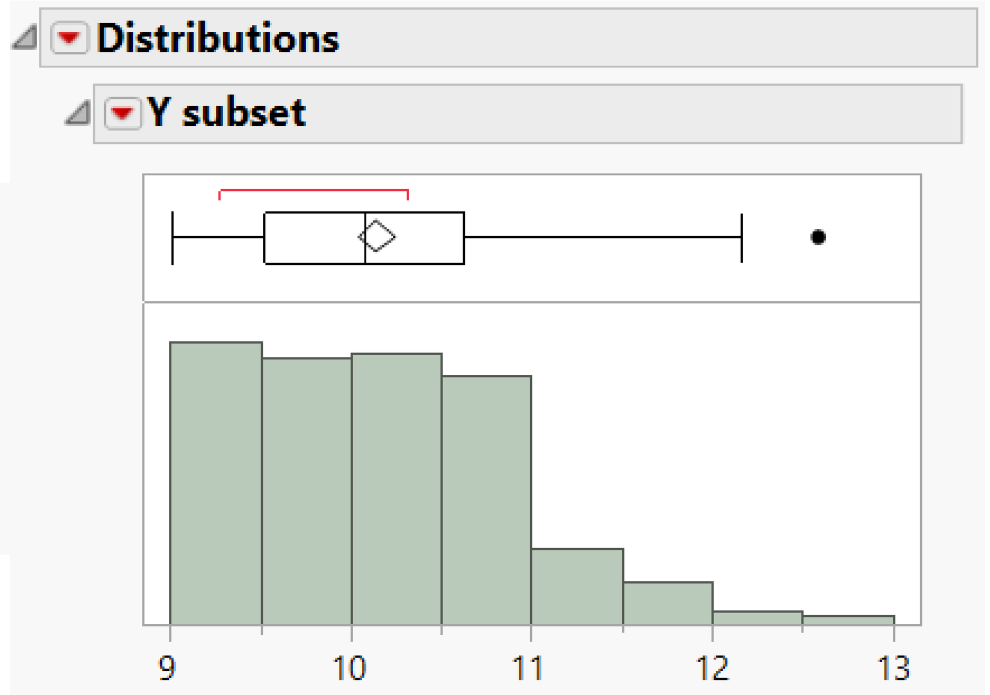
What should we do about it?

Ignoring the values at the limit didn't work out well.

What happens if we just get rid of them?
Now we have a smaller (and biased) sample.
We overestimate the mean and underestimate the variation.

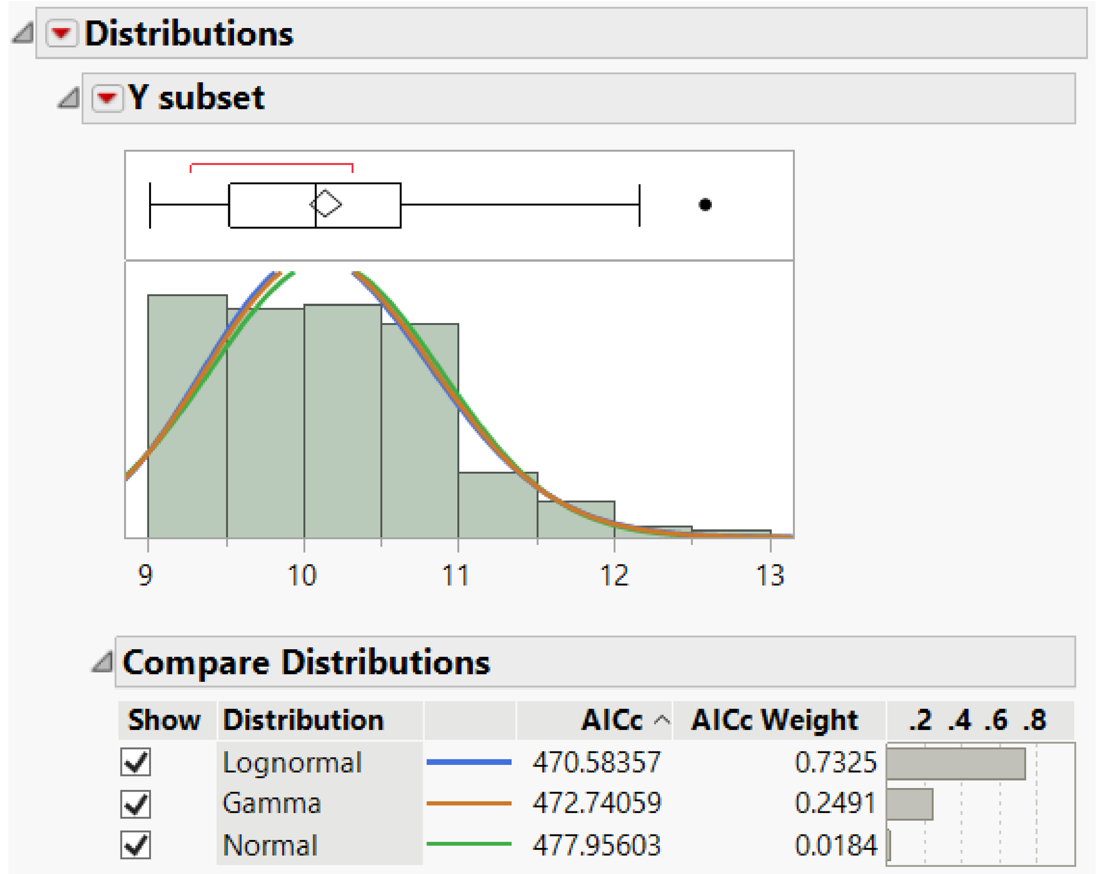
Fitted Normal Distribution

Parameter		Estimate	Lower 95%	Upper 95%
Location	μ	10.144461	10.044053	10.24487
Dispersion	σ	0.7416376	0.6771231	0.8198465



And we still would not pick the normal distribution…

So we can't ignore the limit and we can't throw those observations away.



What should we do about it?

How do we properly handle observations at our limit of detection? We need to treat such observations as censored and use appropriate methods.

Censoring is a situation where the value of a measurement/observation is only partially known.

An observation at the lower detection limit d_l is left censored,

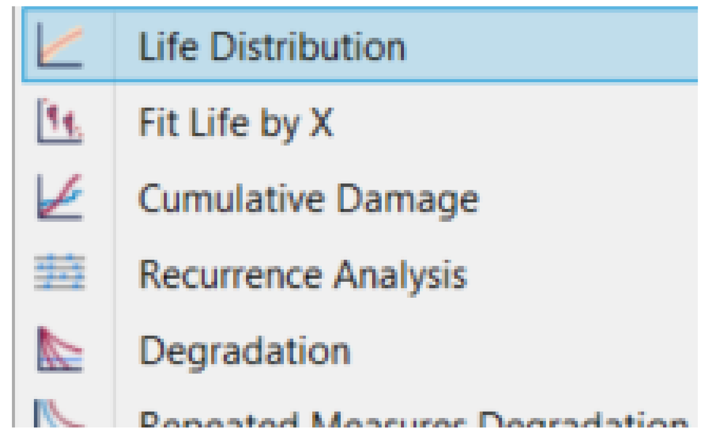
$$y = d_l \Rightarrow y \leq d_l$$

An observation at the upper detection limit d_u is right censored,

$$y = d_u \Rightarrow y \geq d_u$$

Meeker and Escobar's *Statistical Methods for Reliability Data* is an excellent resource for handling censored data.

A variety of JMP platforms handle censoring.



Introducing Limits of Detection in the Distribution Platform

Clay Barker and Laura Lancaster, JMP Principal Research Statistician Developers

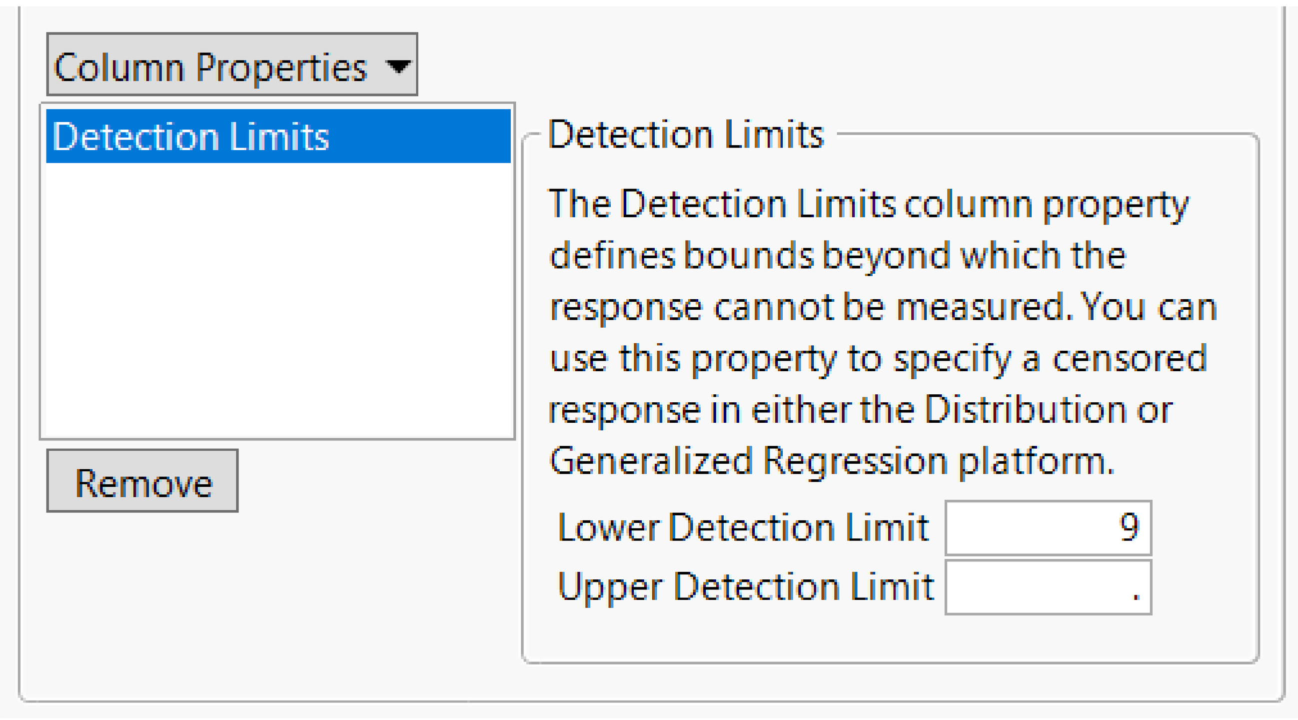
DISCOVERY SUMMIT

jmp STATISTICAL DISCOVERY

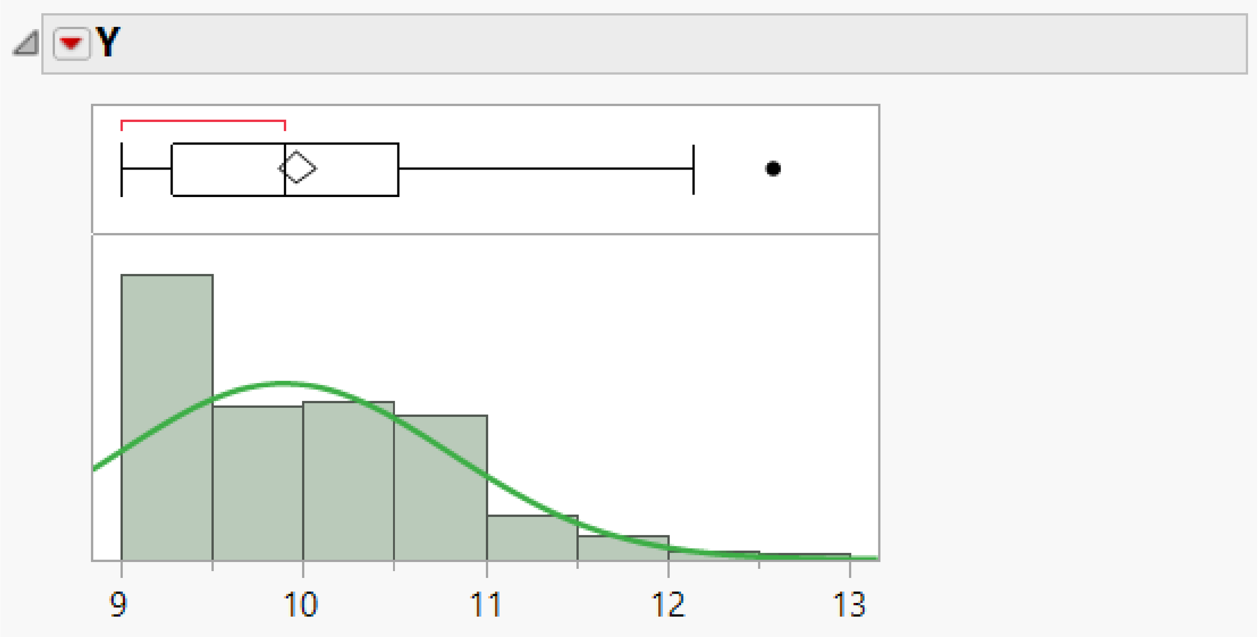
Limits of Detection in Distribution

New in JMP 17, the Distribution platform naturally handles Limit of Detection problems.

Specify them using the "Detection Limits" column property.



The Detection Limits column property ensures that the Distribution platform does estimation properly.

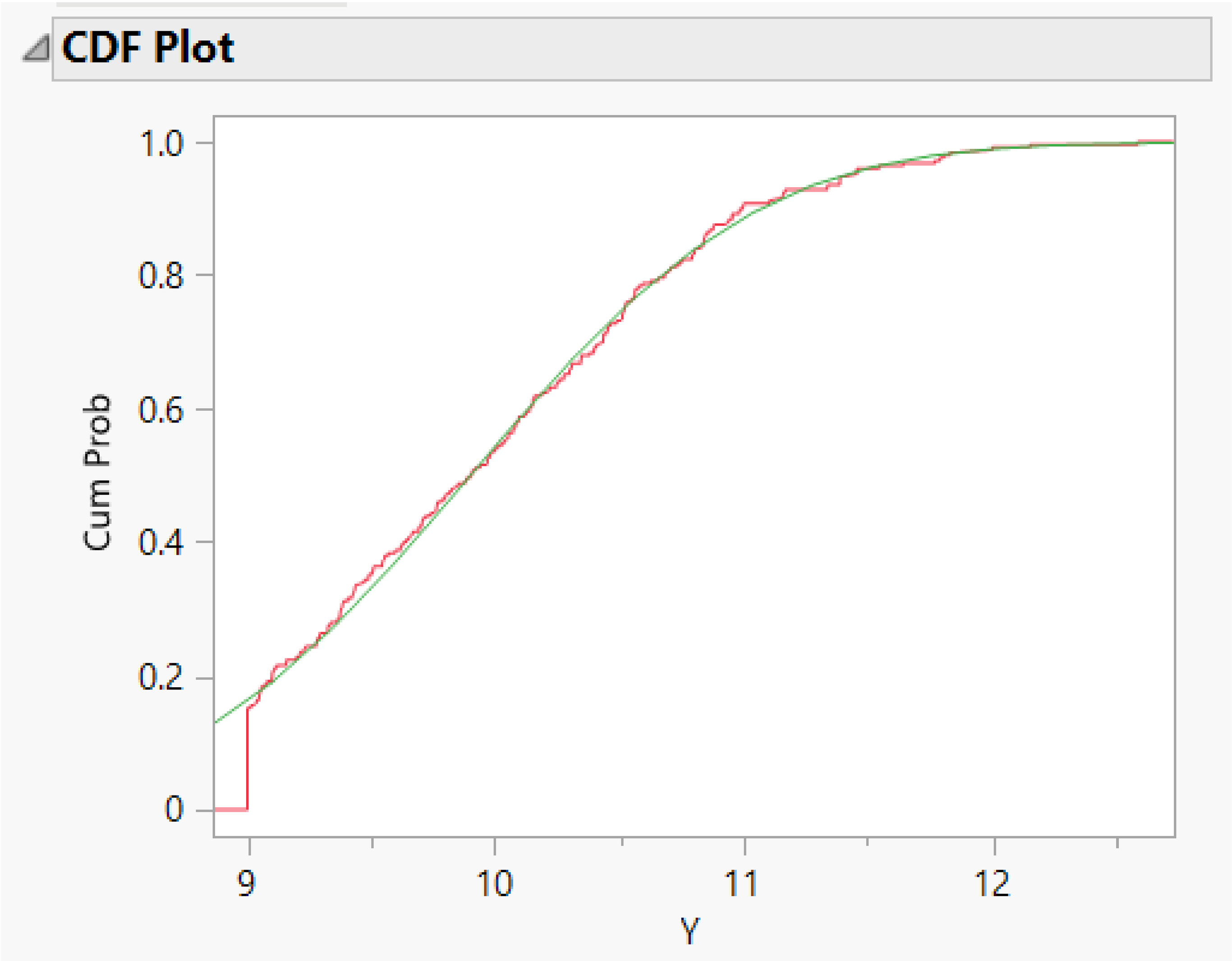


Compare Distributions

Show	Distribution	AICc	BIC	-2*LogLikelihood
☑	Normal	647.78201	654.77635	643.73343

Fitted Normal Distribution with Detection Limits (9, .)

Parameter		Estimate	Std Error	Lower 95%	Upper 95%
Location	μ	9.8973527	0.0587611	9.7805334	10.012028
Dispersion	σ	0.9111741	0.0459224	0.827998	1.0090508



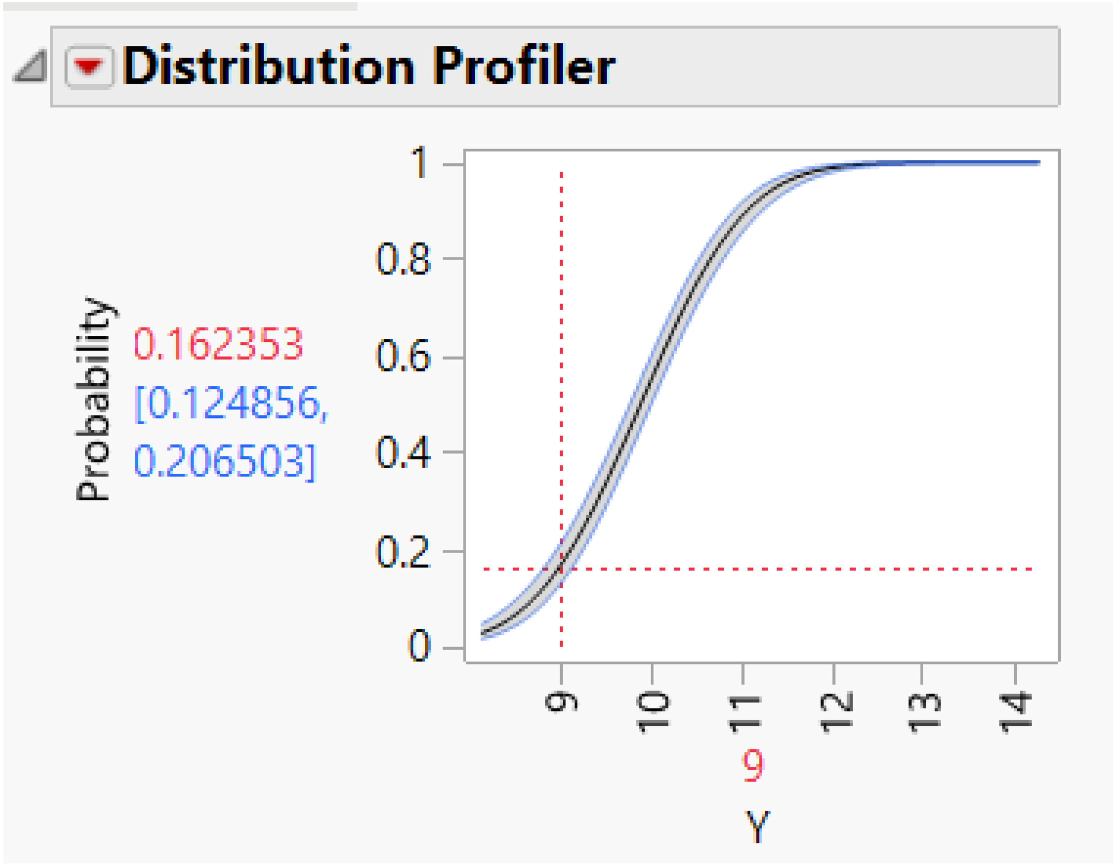
Limits of Detection in Distribution

The report makes it clear that we used detection limits.

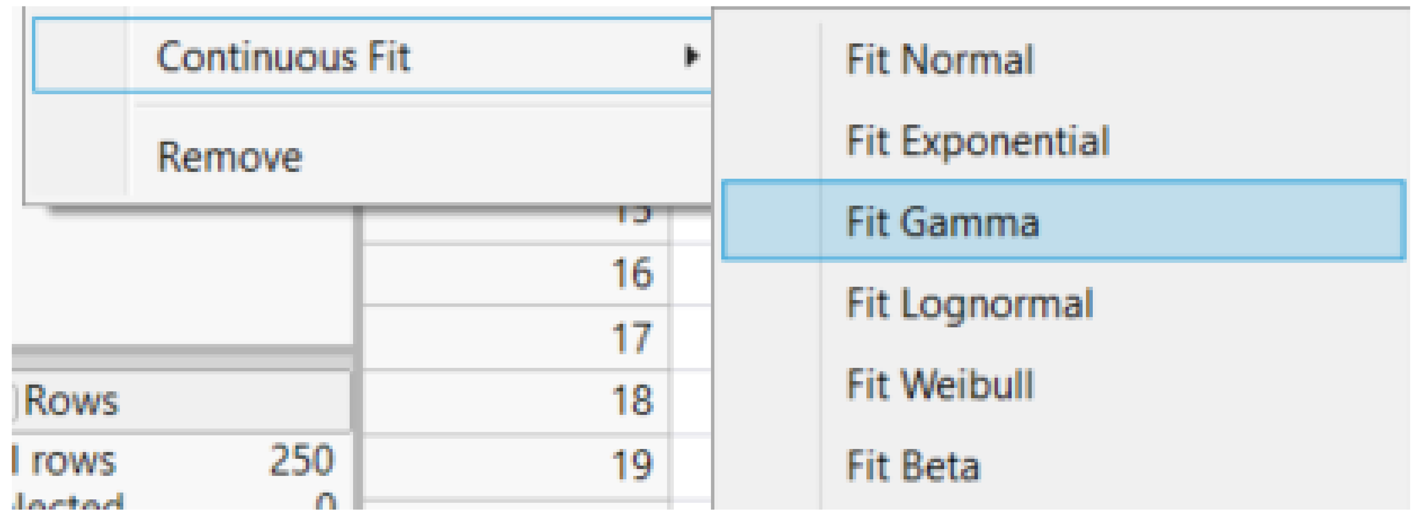
Fitted Normal Distribution with Detection Limits (9, .)

Since we are doing estimation properly, our inferences and estimates will be much more trustworthy.

Parameter		Estimate	Lower 95%	Upper 95%
Location	μ	9.8973527	9.7805334	10.012028
Dispersion	σ	0.9111741	0.827998	1.0090508



Note: not all distributions in the platform can accommodate limits of detection.



If you have JMP Pro, the Generalized Regression platform also recognizes the Detection Limits column property.

DISCOVERY SUMMIT

Introducing Limits of Detection in the Distribution Platform

Clay Barker and Laura Lancaster, JMP Principal Research Statistician Developers

JMP STATISTICAL DISCOVERY

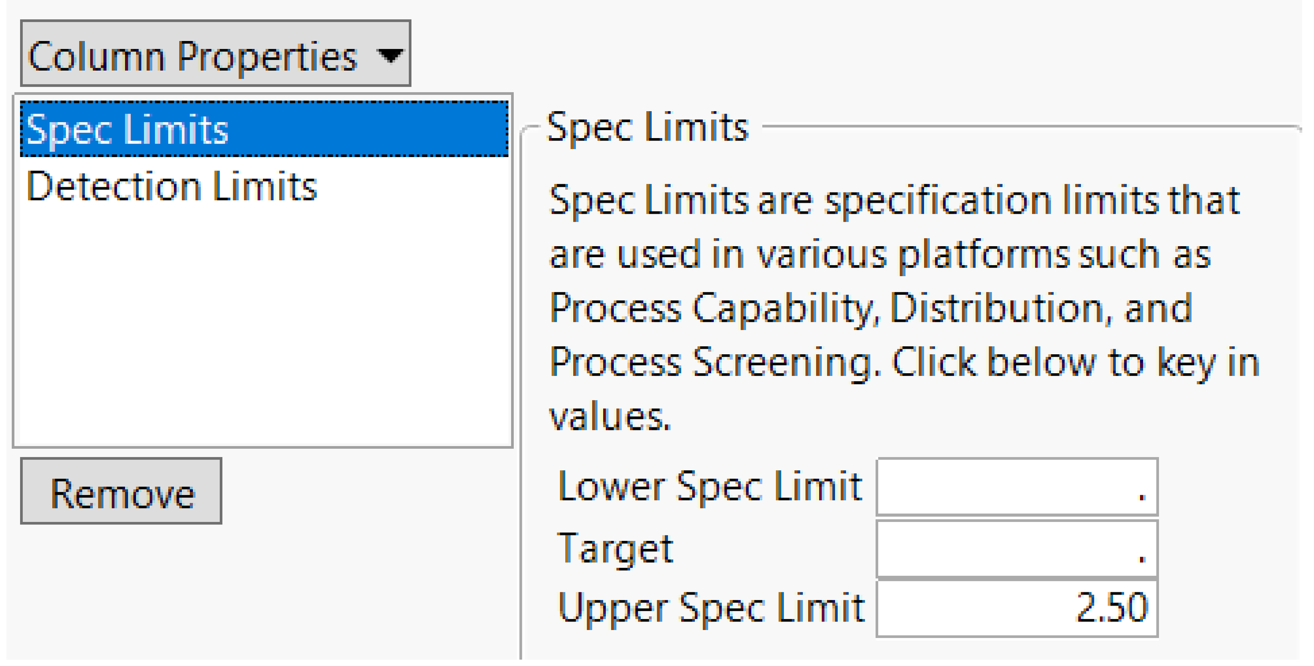
Limits of Detection and Capability

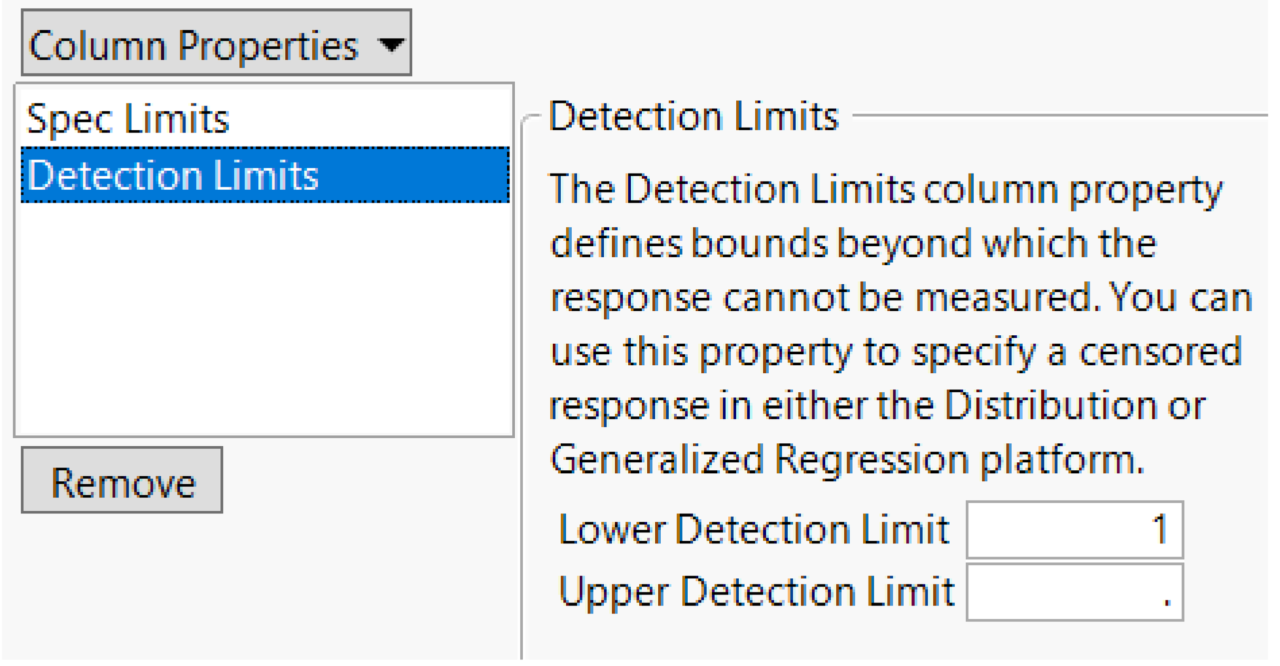
Let's say that we are manufacturing a new drug and need to measure the amount of an impurity.

	Impurity
10	1.34
11	1.83
12	1.59
13	1.00
14	1.27
15	1.00
16	1.57

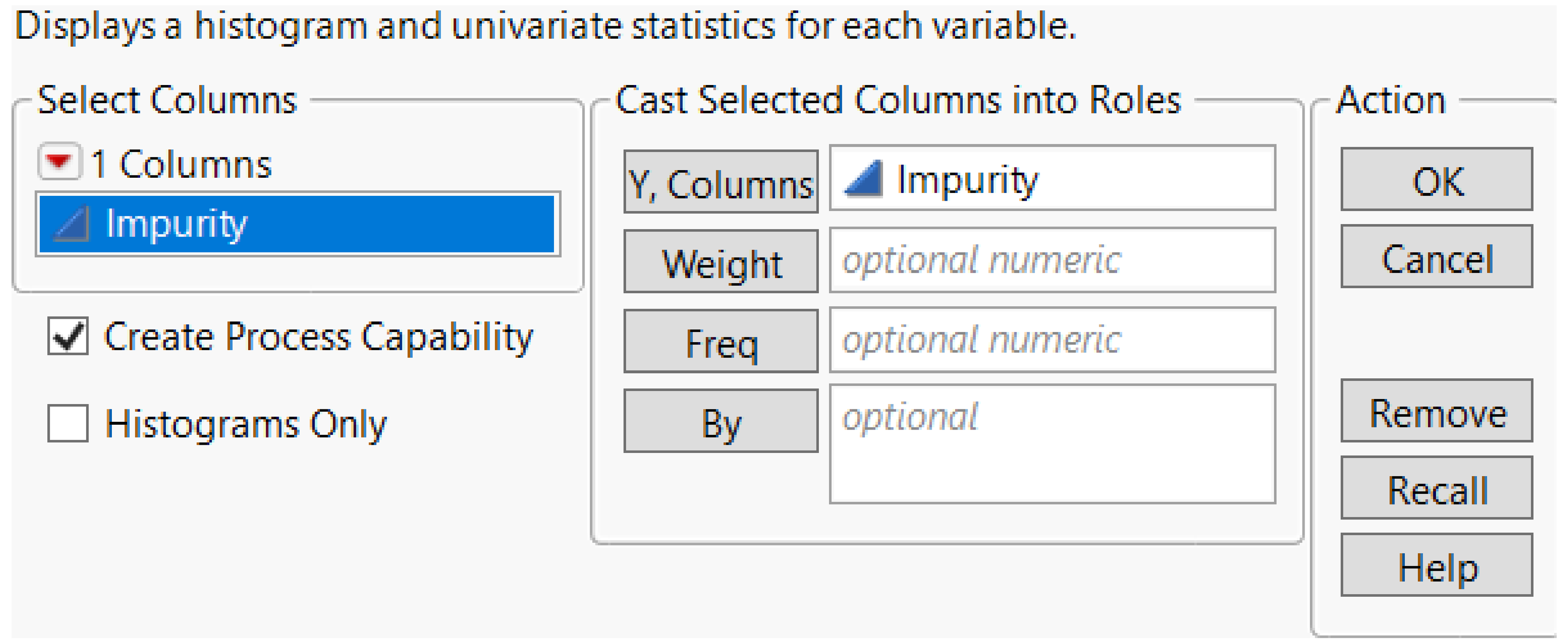
We have an Upper Spec Limit of 2.5 mg and a Lower Detection Limit of 1.0 mg.

Specify these limits with column properties.





Now we can launch Distribution like usual and it will know what to do.



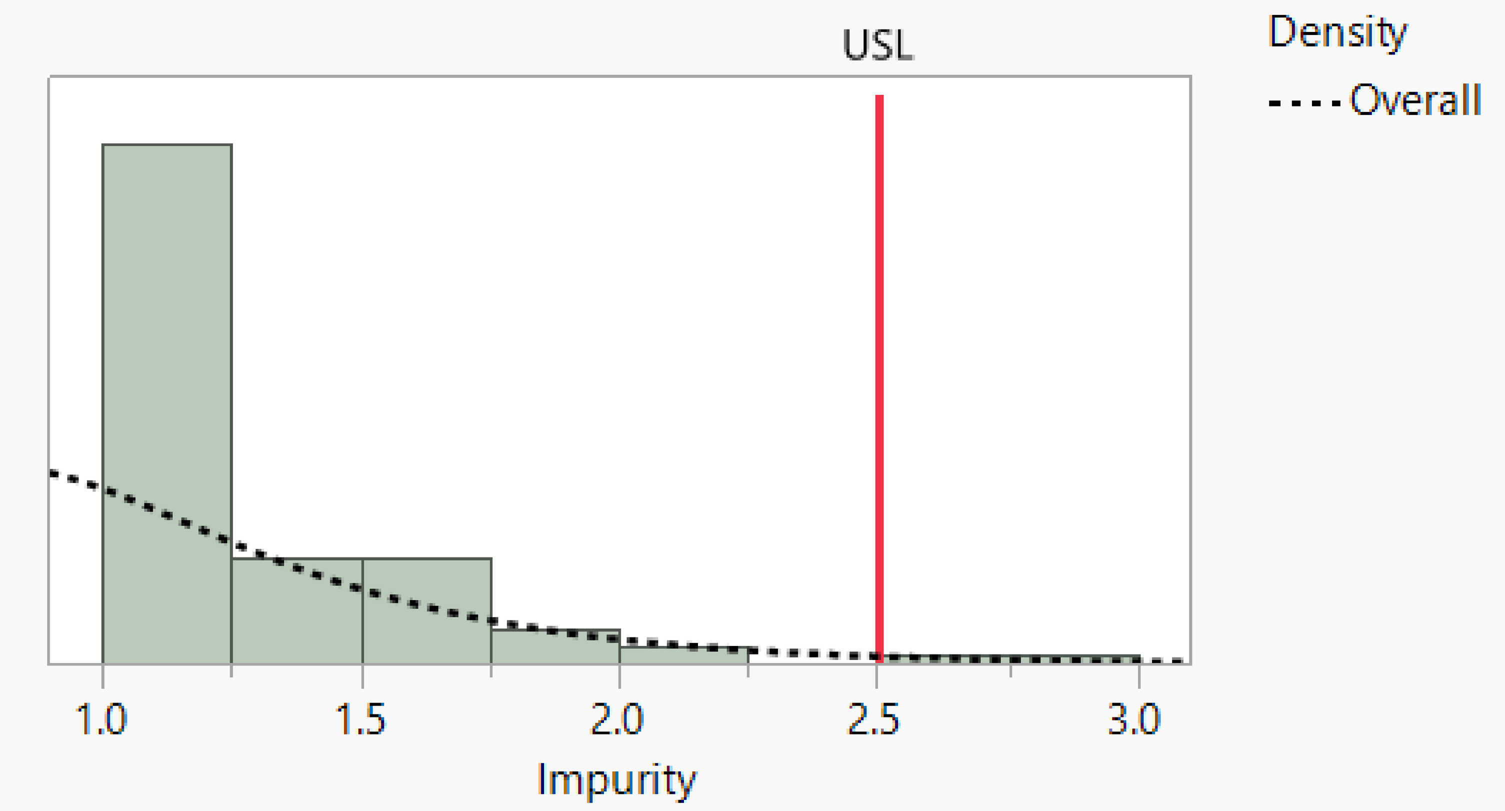
Limits of Detection and Capability

Process Capability

Impurity(Lognormal with Detection Limits (1, .)) Capability

Nonnormal capability indices calculated with the Percentiles method.

Histogram



Overall Sigma Capability

Index	Estimate
Ppk	0.546
Ppu	0.546

Parameter Estimates

Parameter	Estimate
Scale	-0.015051
Shape	0.4466259

Nonconformance

Portion	Observed %	Expected Overall %
Above USL	2.0000	1.8522
Total Outside	2.0000	1.8522

Because of the column property, we automatically do estimation respecting the LOD.

Most important: our Ppk is .546. Not too great.

Introducing Limits of Detection in the Distribution Platform

Clay Barker and Laura Lancaster, JMP Principal Research Statistician Developers

Limits of Detection and Capability

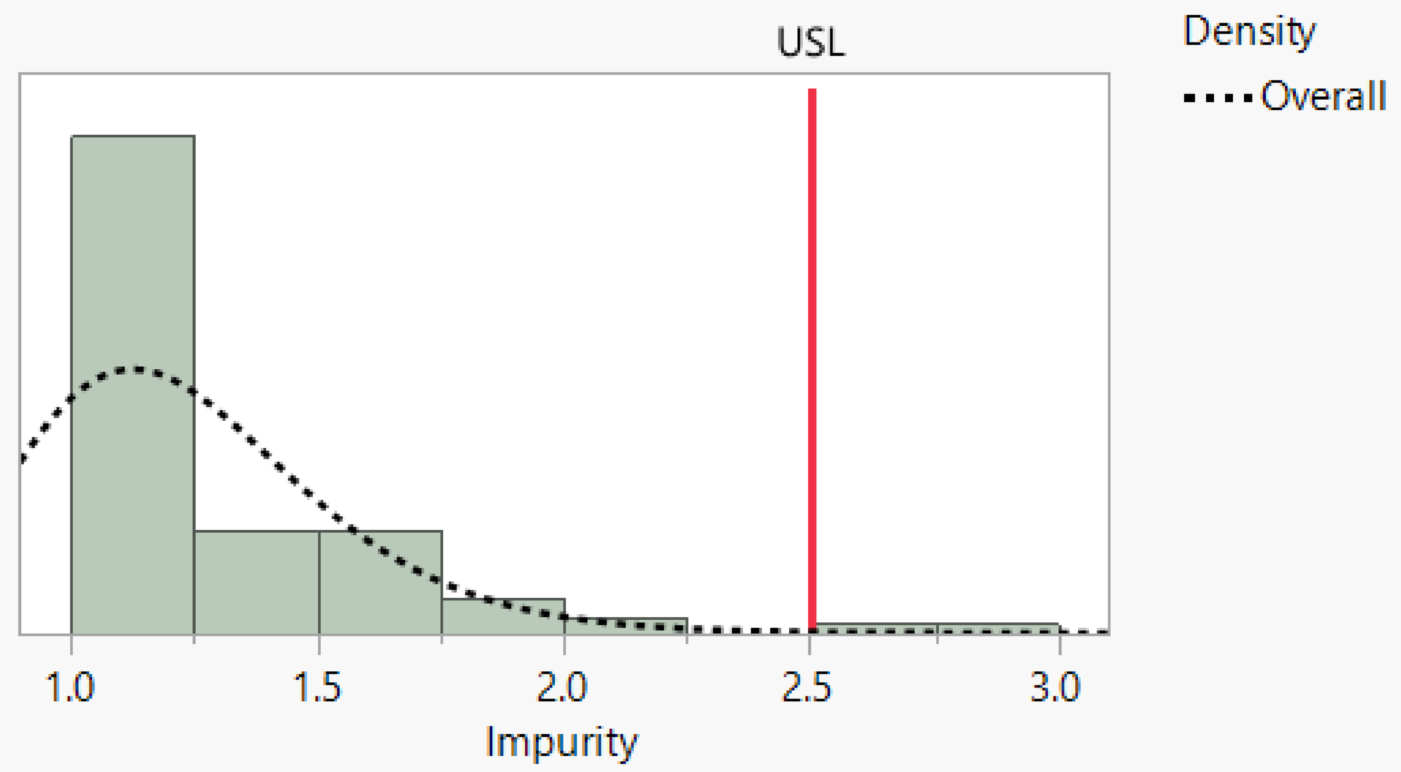
What if we had done our capability analysis ignoring the LOD?

Process Capability

Impurity(Lognormal) Capability

Nonnormal capability indices calculated with the Percentiles method.

Histogram



Overall Sigma Capability

Index	Estimate
Ppk	1.011
Ppu	1.011

Parameter Estimates

Parameter	Estimate
Scale	0.176746
Shape	0.2446504

Nonconformance

Portion	Observed %	Expected Overall %
Above USL	2.0000	0.1252
Total Outside	2.0000	0.1252

Now our Ppk is 1.011!
We don't have good estimates of our distribution parameters, so we cannot trust calculations like Ppk.

Summary

It is important to recognize when our data feature a Limit of Detection.

Ignoring these limits leads to misleading fits and bad decisions.

In our example, ignoring a lower detection limit led to

1. Biased estimates for our lognormal fit
2. Ppk estimate that was twice as big as it should be

But in JMP 17, Distribution makes it easy to avoid these pitfalls.

- Specifying the Detection Limit column property tells Distribution everything that it needs to know.
- Works for Upper and Lower limits, as well as both.
- Six different distributions support Limit of Detection.
- The same column property works with the Generalized Regression platform.