*Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.*
*https://simfit.org.uk*
*https://simfit.silverfrost.com*

The situation envisaged is where a user has a set counts of nonnegative integers and wishes to see if the frequencies observed are consistent with a Poisson distribution with parameter $\lambda$. In SIMFIT this can be done using a chi-square test, a dispersion test, a Fisher exact test, or a Kolmogorov-Smirnov test.

From the SIMFIT main menu choose `[A/Z]`, then program **binomial**, and select the option to test for a Poisson distribution. The most widely used test is a chi-square test so this is done first, choosing to use the sample estimate of $\hat{\lambda} = 1.1$ instead of the current fixed value of 2, and opting for a minimum of 5 counts per bin, leading to the next results.

**Chi-square test for $P(\lambda)$ with $\lambda = 1.1$**

| | | |
|---|---|---|
| $H_0$: Poisson distribution for data with title: | | |
| Test file `Poisson.tf1`: 40 random numbers | | |
| Sample estimate used in chi-square test | | |
| Sample estimate for $\lambda$ | 1.100 | |
| Lower 95% confidence limit | 0.7993 | |
| Upper 95% confidence limit | 1.477 | |
| Mean of $x$-values | 1.100 | |
| Variance of $x$-values | 0.8103 | |
| Standard deviation of $x$ | 0.9001 | |
| Mean using fixed $\lambda$ | 2.000 | |
| Poisson dispersion value $D$ | 28.73 | |
| $P(\chi^2 \geq D)$ | 0.8863 | |
| Number of partitions (bins) used | 3 | |
| Number of degrees of freedom | 1 | |
| Chi-square test statistic $C$ | 5.857 | |
| $P(\chi^2 \geq C)$ | 0.0155 | *Reject $H_0$ at 5% level* |
| Upper tail 5% critical point | 3.841 | |
| Upper tail 1% critical point | 6.635 | |

Now it should be emphasized that the chi-square test is an approximate test, as the test statistic only becomes asymptotic to a chi-square variable with large samples. Further, if any cells have a small frequency, say $< 5$, it is usually recommended to combine adjacent bins until this is the case. That is why choosing a minimum frequency of 5 resulted in only 3 bins. If the test is now repeated on the same data but choosing a minimum frequency of 3 the next results are obtained.

| | | |
|---|---|---|
| Number of partitions (bins) used | 4 | |
| Number of degrees of freedom | 2 | |
| Chi-square test statistic $C$ | 5.858 | |
| $P(\chi^2 \geq C)$ | 0.0535 | *Consider accepting $H_0$* |
| Upper tail 5% critical point | 5.991 | |
| Upper tail 1% critical point | 9.210 | |

It should be noted that by simply changing the number of bins from 3 to 4 a rejection ($p = 0.0155$) becomes an acceptance ($p = 0.053$), which should serve to emphasize how the outcome of such chi-square tests depend on the number of bins.

A classical example is the famous data collected by von Bortkiewicz for 0, 1, 2, 3, or 4 deaths per year by horse kick in the Prussian cavalry during the period 1875 - 1894 for 10 groups, as this is a typical example of using the Poisson distribution to model rare events.

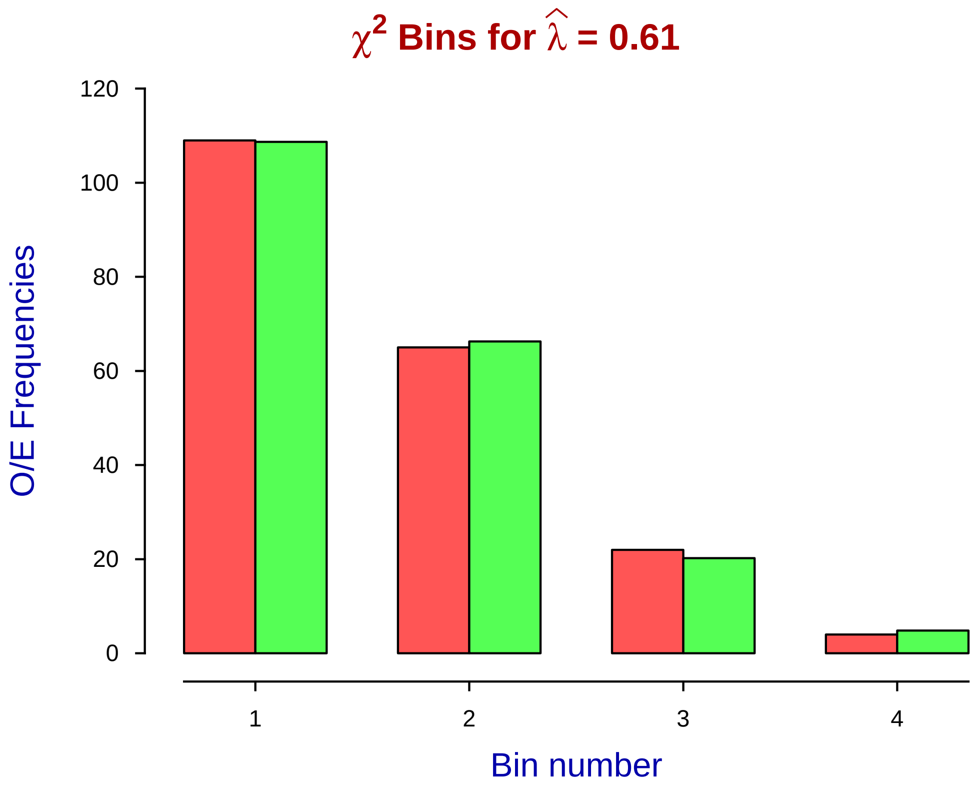
The 200 frequencies recorded are contained in the test file `poisson.tf2` as follows

| Deaths | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Frequency | 109 | 65 | 22 | 3 | 1 |

leading to the result, shown below, that $p = 0.8506$, so that the null hypothesis of a Poisson distribution with $\hat{\lambda} = 0.61$ cannot be rejected.

| **Chi-square test for** $P(\lambda)$ **with** $\lambda = 0.61$ | | |
|---|---|---|
| $H_0$: Poisson distribution for data with title: | | |
| Death from horse kicks in Prussian cavalry 1875-1894 | | |
| Sample estimate used in chi-square test | | |
| Sample estimate for $\lambda$ | 0.6100 | |
| Lower 95% confidence limit | 0.5066 | |
| Upper 95% confidence limit | 0.7283 | |
| Mean of $x$-values | 0.6100 | |
| Variance of $x$-values | 0.6110 | |
| Standard deviation of $x$ | 0.7816 | |
| Number of partitions (bins) used | 4 | |
| Number of degrees of freedom | 2 | |
| Chi-square test statistic $C$ | 0.3235 | |
| $P(\chi^2 \geq C)$ | 0.8506 | *Consider accepting* $H_0$ |
| Upper tail 5% critical point | 5.991 | |
| Upper tail 1% critical point | 9.210 | |

The observed and expected values used in this test are displayed below.



The Fisher Exact test will frequently fail with large or aberrant data sets and this will be indicated when it happens, but the dispersion test can always be used to test for data that are too uniform or too clustered to be from a Poisson distribution. This is used to study clumping of objects studied microscopically, and similar situations concerning spatial or temporal distributions of counts. Consider, for instance, the analysis of this data set contained in the test file `poisson.tf3`.

| Counts | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 0 | 2 | 10 | 0 | 0 | 0 | 1 | 2 | 9 | 1 | 5 |

**Dispersion and Fisher-exact Poisson tests**

| | | |
|---|---|---|
| Bonferroni $n = 2$ | | |
| Data: Poisson clumping data | | |
| Sample size | 30 | |
| Sample total | 173 | |
| Sample sum of squares | 1333 | |
| Sample mean | 5.767 | |
| Lower 95% confidence limit | 4.939 | |
| Upper 95% confidence limit | 6.693 | |
| Sample variance | 11.56 | *Too large ?* |
| Dispersion ($D$) | 58.16 | |
| $P(\chi^2 \geq D)$ | 0.00104 | *Reject $H_0$ at 1% sig.level* |
| Number of degrees of freedom | 29 | |
| Fisher exact probability | 1.00000 | |
| IFAIL = 1: Fisher $p$ is only an upper bound | | |

In a Poisson distribution the mean is equal to the variance, and a variance much less than the mean suggests a distribution that is too uniform, while a variance exceeding the mean could indicate clustering, as will be clear from the results above and the following plot.