Statistical Analysis in Support of Mains Replacement Programme 2013 to 2021

For Jonathan Dennett, National Grid

10th October 2011

Iron mains replacement programme 2013 to 2021

Background

National Grid has provided data for the number of fatalities in 86 incidents occurring from 1990 to 2010 as shown in Table 1.

Number of fatalities	Number of incidents
0	73
1	9
2	2
3	0
4	2
Total number of incidents	86
Mean fatalities per incident	21/86 = 0.244
Standard deviation	0.718
Standard error	$0.718/\sqrt{86} = 0.077$

Table 1 Data and summary statistics

It is required to determine a cut-off level for MRPS risk scores above which gas mains are to be replaced during the 2013 to 2021 replacement programme.

MRPS x 10^{-6} is the probability of an incident caused by a gas main per km.

MRPS x expected number of fatalities per incident/number of premises < 1

MRPS < (1/ expected number of fatalities per incident) x number of premises.

The larger the expected number of fatalities per incident, the lower the cut-off for MRPS.

Expected number of fatalities

The expected number of fatalities is based on previous experience of fatalities in incidents averaged over the available data from 1990 to 2010. The expected value is used to determine the cut-off for MRPS rather than basing the analysis on the outcomes of single incidents.

The mean value of the data in Table 1 can be used to estimate the expected number of fatalities. However, the observed mean value of 21/86 is a point estimate and is subject to uncertainty. In the calculation of a cut-off value for MRPS, it is sensible to use an upper bound for the expected number of fatalities per incident associated with a measure of confidence.

Confidence intervals and statistical models

It is necessary to evaluate the uncertainty around the observed mean number of fatalities per incident to construct confidence intervals.

Count data are often modelled by the Poisson distribution, however, the data in Table 1 are not well modelled by the Poisson distribution[1] because the numbers of fatalities are more variable (over-dispersed). The over-dispersion could be caused by too many zeros (see Appendix 1) and one approach is to consider the data to arise from a mixture of distributions where zero-fatality incidents occur with a certain probability and non-zero-fatality incidents occur according to Poisson probabilities. This is called a Zero Inflated Poisson model. This model is not adopted here because it implies that there is something intrinsically different about incidents producing zero fatalities. A note has been provided by Jonathan Dennett that is included within this paper (Appendix 2), which discusses the many factors that can affect the outcome of an incident.

An alternative to the Poisson distribution for count data is the Negative Binomial model which allows for data that is more variable than expected in a Poisson model. This model is widely used in transport statistics and ecology where counts of various items are regularly analysed. The Negative Binomial model is a good fit to the data[2], because the observed frequencies of different counts of fatalities are very similar to the expected frequencies (see Appendix 1). However, using this model, the fit is less good for 4 fatalities and the expected number of incidents resulting in 5 or more fatalities is extremely low, whereas it is known that such incidents can occur. Therefore, although the Negative Binomial model fits the bulk of the observed data well, it is unwise to rely on it for the confidence interval.

The mean number of fatalities per incident can be modelled by the Normal distribution. The Normal model for the mean can be used even though the individual values are counts as in Table 1.The Normal model fits better for mean values from larger sample sizes, the sample size being particularly important when the individual data are skewed (a result due to the central limit theorem). The data in Table 1 are skewed as most incidents have zero fatalities. The occurrence of incidents with higher fatalities than those observed would make the observed data distribution even more skewed. However, the sample size is fairly large and it was found in simulation studies in [1] that the Normal distribution gave reliable confidence intervals for the mean with data of the form in Table 1 when the sample size was 86.

The Binomial distribution is a familiar statistical model and is also considered here for completeness. However, the Binomial distribution is only suitable if the outcome of each incident can take one of two values. The Binomial distribution could therefore be used to model the expected proportion of incidents with one or more fatalities, but that is not what is required here.

The conclusion is that the Normal distribution provides a reasonable model for the mean but confidence intervals calculated using statistical models should be taken as guidelines rather

[1] A goodness-of-fit test for Poisson has chi-square=5.36 with 1 degree of freedom giving a p value of 0.021 which shows a significant difference from Poisson

[2] A goodness-of-fit test for Negative Binomial has chi-square=0.05 with 1 degree of freedom giving a p value of 0.823 which shows no significant difference from Negative Binomial

than fact. The cut-off level chosen for MRPS should correspond to *at least* the value given by the confidence interval.

Sample size

The sample size is 86. In [1] it was found that sample sizes of 30 to 40 were sufficient to give robust estimates for the parameters of the Negative Binomial distribution and that the Normal distribution gave reliable confidence intervals for the mean when the sample size was 86. It is reasonable to conclude that the sample size in Table 1 is sufficient to estimate expected fatalities.

Confidence levels

Confidence intervals are commonly calculated at 95% and 99% levels. A two-sided confidence interval gives an upper and a lower bound, a one-sided confidence interval gives an upper bound only. The interpretation for a 99% one-sided confidence interval is that the true value will be less than the upper bound found in 99% of confidence intervals constructed across many separate samples. The numerical value of a 99% one-sided upper confidence bound is the same as the numerical value of the upper value of a 98% two-sided confidence interval. Similarly the numerical value of a 95% one-sided upper confidence bound is the same as the numerical value of the upper value of a 90% two-sided confidence interval.

Confidence intervals

Confidence intervals for different models and confidence levels are shown in Table 2 for comparison. The Poisson and Negative Binomial confidence intervals are calculated as in [2] and [1] respectively. The Normal confidence intervals are calculated as the mean (0.244) plus or minus a multiple of the standard error (0.077).[3]

Model	Upper of 1-sided 90%	Upper of 2-sided 95%	Upper of 1-sided 99%	Upper of 2-sided 99%
Poisson	0.352	0.373	0.399	0.418
Negative Binomial	0.343	0.366	0.393	0.413
Normal	0.372	0.396	0.424	0.444

Table 2 Confidence intervals for mean fatality per incident

Conclusion and recommendations

The summary statistics for the data in Table 1 depend on the sample. The outcomes of additional incidents that may follow will change the values. The sample values are used to provide the estimates of mean and variance and models of the underlying random variation are used to give the confidence intervals. The upper 99% two-sided confidence interval is a conservative choice and I would recommend that this confidence level is used.

[3] The multiples are 1.64 for 90%, 1.96 for 95%, 2.33 for 99% and 2.58 for 99.5% in Table 2

The Poisson model does not fit the data very well. The Negative Binomial model fits the observed data very well but additional data may well change the pattern and upset the fit. The Normal model has been found to give reliable confidence intervals for the mean of medium to large sample sizes. I would therefore recommend using *at least* the upper Normal confidence bound which is 0.444 in Table 2.

The confidence bound should be at least 0.444 because higher numbers of fatalities have occurred in the past. It should be noted that the occurrence of an incident with a high number of fatalities would increase the expected number of fatalities. For example, if another incident caused 22 fatalities as happened in the past, then the mean would increase to 0.494.

On the basis of the 86 incidents from 1990 to 2010, I would recommend a cut-off based on expected fatality of *at least* 0.444 per incident.

Shirley Coleman

References

[1] David Shilane, Steven N. Evans, Alan E. Hubbard (2010)

Confidence Intervals for Negative Binomial Random Variables of High Dispersion

The International Journal of Biostatistics, 6(1).

[2] http://www.graphpad.com/quickcalcs/Confinterval2.cfm

Appendix 1

Count	**Observed**	Poisson probabilities	**Poisson expected**	Negative Binomial probabilities	**Negative Binomial expected**
0	**73**	0.78349	**67.3799**	0.84873	**72.9908**
1	**9**	0.19117	**16.4407**	0.09802	**8.4294**
2	**2**	0.02332	**2.0058**	0.03147	**2.7066**
3	**0**	0.00190	**0.1632**	0.01226	**1.0546**
4	**2**	0.00012	**0.0100**	0.00520	**0.4470**

Table A1 Observed and expected frequencies using Poisson and Negative Binomial models

Note that the Negative Binomial probability of 5 or more fatalities is 0.00432, expected number of incidents in a sample of 86 is 0.37 which less than 1 incident.

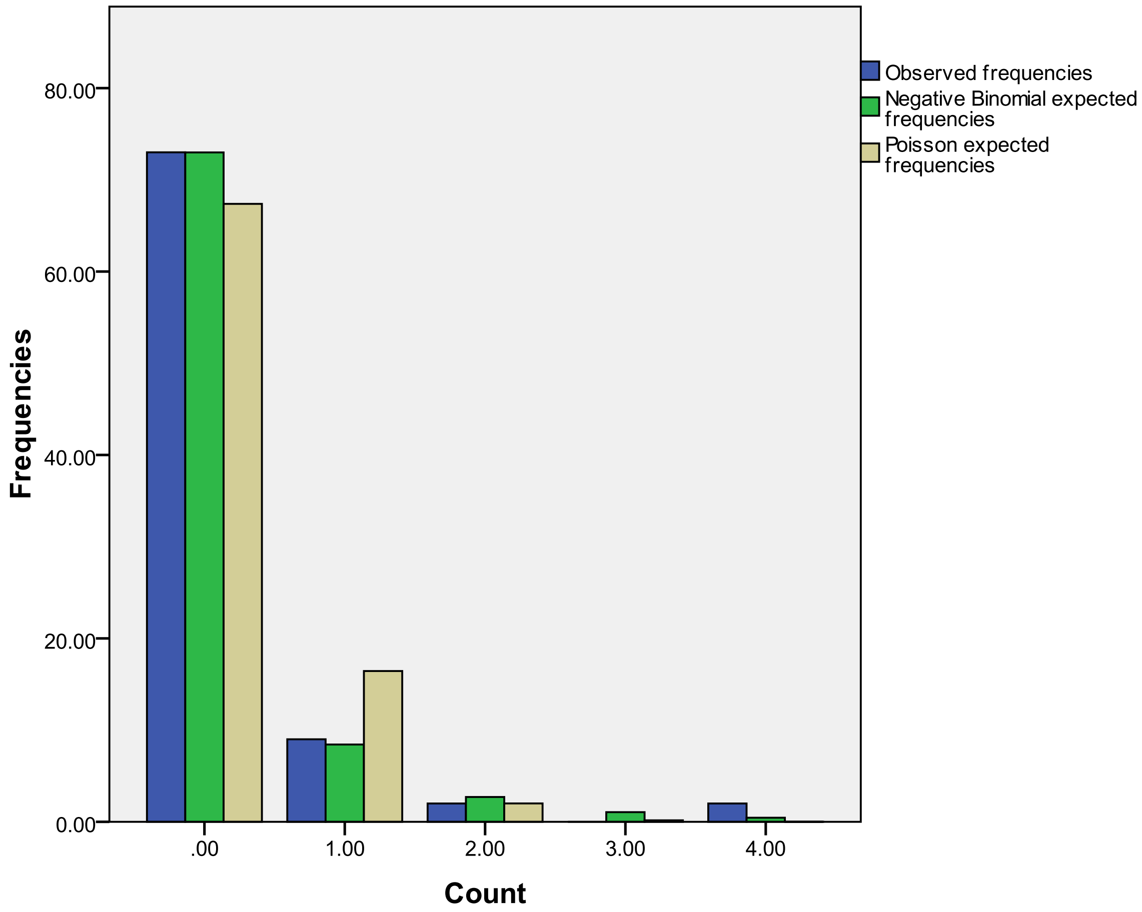

Figure A1 Observed and expected frequencies

The Normal model is used for the mean value and cannot be used to find expected frequencies of individual counts.

Appendix 2 – A note provided by Jonathan Dennett, National Grid, which discusses the factors which affect the outcome of an explosion.

Factors affecting the outcome of an explosion

For a property to be blown up by a gas explosion a flammable gas air mixture must first be achieved within some part of the property and this mixture must subsequently be ignited.

Once the flammable volume is ignited the outcome the effect of the explosion on human life depends on the following factors.

1. Number of people present
2. Where those people are within the building
3. The nature of the building
4. The location and power of the explosion
5. Subsequent fire
6. The health of the people
7. The response of the emergency services

The more people present within or close to a building the more people may be killed or injured when it experiences an explosion. Clearly where nobody is present nobody will be harmed.

The location of persons within a building may be significant. Sometimes the building only partly collapses and if the people are within a part of the building that survives they stand a higher chance of survival themselves.

The nature of the building appears to be very important however the position is not as simple as it might be thought. Very light and badly built buildings are not necessarily as unsafe during an explosion as seemingly stronger buildings because often the explosion is vented through the failure of a wall or window before the building as a whole experiences excess overpressure that causes it to collapse. In general steel or concrete framed modern buildings with curtain wall construction are likely to be safer than masonry buildings with load bearing walls because curtain walls can fail without the frame of the building failing; in this eventuality the building as a whole will remain standing even though there is significant damage. This increases survivability.

The power of an explosion is controlled by the gas / air concentration (the closer to a stoichiometric concentration the more powerful the blast), the amount of the building that contains the flammable volume and the pressure at which the building fails. Strong buildings with little venting achieve high overpressure that can result in the explosion transitioning from deflagration to a detonation with resulting great devastation. (Deflagrating explosions can vent limiting overpressure however as soon as detonation is initiated venting is not important because the detonation proceeds at a velocity that is too high for such venting to be significant in controlling the peak pressure achieved.)

An explosion can be confined to a small part of a property; in this case little damage may be done. However, when a flammable volume is present in more than one room the resulting blast can be significantly worse than if the volume had been present in a single room of equivalent total size. This is because as an explosion vents from one area into another, e.g. through a doorway, the flame velocity accelerates and detonation becomes more likely, particularly if a door between the two rooms had been closed and fails at some point leading to the propagation of the explosion. Cellars are particularly risky because they tend not to vent an explosion and as a result huge pressures can be generated – enough to blow apart all but the strongest structures. In addition to killing persons within the structure such a blast would imperil all those within a significant distance as large items such as roofing joists and lumps of masonry are typically thrown tens of metres.

Sometimes an explosion is followed by fire; the escaping gas may or may not be a significant factor in this. In the event that people are trapped the probability that they survive will be lessened by the presence of fire.

The probability that somebody survives an explosion event is influenced by their health. Vulnerable people such as young children, the elderly, the sick and disabled are more likely to be killed and may not be able to escape any subsequent fire even if they are not trapped.

The response time and effectiveness of the emergency services and the success of any subsequent medical treatment will influence the probability that an incident produces fatalities.

It is clear that most of these factors are not linked with the type of gas main involved. This is because in respect of every incident that occurs a flammable volume is ignited within the building. Initiating events, such as type of main, its location etc. may be ignored because these factors contribute towards the probability that a flammable volume exists – it has no influence on what happens after it ignites.

In fact it might be possible to argue that Tier 2 Iron mains may systematically be located closer to more or less vulnerable buildings (in fact they are probably located to more vulnerable buildings because they were laid in the period 1860 to 1960 and buildings that existed prior to the main being laid, which will tend to be located in such areas are more likely to be vulnerable to explosion) however at this point such evidence in anecdotal.