# STA4030: Categorical Data Analysis
## One-way Tables

Instructor: Bojun Lu

School of Data Science
CUHK(SZ)

September 15, 2020

# Agenda

# 2.1 One-way Tables

## 2.1.1 Examples

In the first chapter, we covered

- what kinds of data we are interested in
  - Nominal, Ordinal, Interval
- what kinds of distribution will arise
  - Binomial, Multinomial, Poisson
- how to estimate and test the parameters of these distributions
  - Wald, Score, LR tests and confidence intervals

# 2.1 One-way Tables

(From Exercise 1.5 in Agresti (2013))

> *When the 2010 General Social Survey asked, "Please tell me whether or not you think it should be possible for a pregnant women to obtain a legal abortion if she is married and does not want any more children," 587 replied "yes" and 636 replied "no".*

# 2.1 One-way Tables

You are a floor manager at a large casino in Macau. You suspect one of your craps dealers is cheating, in conjunction with a customer. You test the dice they are using by rolling it 60 times. You observe the following:

| Face | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Count | 10 | 14 | 6 | 6 | 4 | 20 |

Should you punish the craps dealer and customer?

# 2.1 One-way Tables

It's 1898. Ladislaus Bortkiewicz, given the job of investigating the number of soldiers in the Prussian army killed by kicks from horses and mules, has collected the following data: by observing 14 army corps for 20 years each (for a sample of 280 corps-years), there were 144 corps-years with no deaths; 91 corps-years with one death; 32 corps-years with two deaths; 11 corps years with three deaths; 2 corps-years with four or more deaths.

What can you say about these deaths in the Prussian army?

# 2.1 One-way Tables

Each of those tables can be written as one-way frequency tables:

| Answer | Yes | No |
|---|---|---|
| Count | 587 | 636 |

| Face | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Count | 10 | 14 | 6 | 6 | 4 | 20 |

| # deaths | 0 | 1 | 2 | 3 | 4+ |
|---|---|---|---|---|---|
| Count | 144 | 91 | 32 | 11 | 2 |

# 2.1 One-way Tables

## 2.1.2 Notation

- We represent a one-way table with $c$ categories by a vector $\boldsymbol{X} = (X_1, \ldots, X_c)$, where $X_j$ is the (random variable) count or frequency in cell/category $j$.
- We then represent the observed counts or frequencies in cell $j$ by $n_j, j = 1, \ldots, c$. The total number of observations is $n = \sum_{j=1}^{c} n_j$.
- We are concerned with the joint distribution of $(X_1, \ldots, X_n)$. Let this be $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_c)$, where $\pi_j$ is the probability that a randomly selected person/item/thing from the population under study falls into category $j$ and $\sum_{j=1}^{c} \pi_j = 1$.
- We want to estimate these $\pi_j$s and their joint distribution. Think of $\pi_j$ as a population proportion with a particular characteristic. Let $p_j = n_j/n$ be the sample version of $\pi_j$.

# 2.1 One-way Tables

Our analysis begins with an assumption about how the data is generated.

One question suffices: did sampling occur with a fixed sample size or not?

- if fixed, **binomial** or **multinomial** sampling has been employed.
- if not fixed, perhaps **Poisson** sampling has been.

# 2.2 Binomial Sampling

## 2.2.1 Characterisation

Binomial sampling is characterised by

- $n$ is fixed
- each observation is a "trial" with only two possible outcomes. Nominally we can call these "success" and "failure".
- the trials are IID, meaning the probability of success $\pi$ does not change between trials and the trials do not affect each other.

# 2.2 Binomial Sampling

## 2.2.2 Inference

- Inference in binomial sampling focuses on parameter $\pi$. We have already found its MLE: $\hat{\pi} = n_s/n$, where category $s$ is the "success" category.
- In order to the use the Wald, Score and LR test for $\pi$, we need approximate normality of the sample proportion of $\hat{\pi}$. This comes with "large" sample size $n$.
- But what is large? A good rule of thumb is $n\pi \geq 5$ and $n(1-\pi) \geq 5$, so that the number of successes and failures in the sample are not small.
- With large $n$, we have the triumvirate of tests at our disposal. However, if proportions are extreme, e.g. $\pi$ or $1-\pi < 0.2$, the Score and LR approaches can be preferred to the Wald.

# 2.2 Binomial Sampling

## 2.2.3 Example - Exercise 1.5 from Agresti (2013)

We summarize the data obtained in a one-way table:

| Answer | Yes | No |
|---|---|---|
| Count | 587 | 636 |

Let "success" be "Yes". We want to know if $\pi$ is different from 0.5 at the $\alpha = 0.05$ level. The sample size $n = 1223$ is clearly large enough for asymptotic tests to be used. Applying these, we find

- The MLE is $\hat{\pi} = 587/1223 = 0.48$.
- The Wald, Score and LR 95% CIs all equal $(0.452, 0.508)$.

We conclude at the $5\%$ significance level, we cannot reject the null hypothesis that $\pi = 0.5$. That is, a random person is equally likely to answer the question "Yes" or "No".

# 2.3 Multinomial Sampling

## 2.3.1 Characterisation

Multinomial sampling is characterised by

- $n$ is fixed
- each observation is a "trial" with only $c$ possible outcomes.
- the trials are IID, meaning the probability of success $\pi$ does not change between trials and the trials do not affect each other.

We shall see later that the requirement "$n$ is fixed" is not too restrictive.

# 2.3 Multinomial Sampling

## 2.3.2 Inference

Recall the distribution of $\boldsymbol{X} \sim Mult(n, \boldsymbol{\pi})$

$$p(X_1 = n_1, \cdots, X_c = n_c) = \frac{n!}{n_1! n_2! \cdots n_c!} \pi_1^{n_1} \pi_2^{n_2} \cdots \pi_c^{n_c}$$

Hence, the loglikelihood function is

$$l(\boldsymbol{\pi}; (n_1, \ldots, n_c)) = \sum_{j=1}^{c} n_j \log(\pi_j) + Const.$$

To find the MLE of $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_c)$ is to solve the problem

Maximize: $l(\boldsymbol{\pi}; (n_1, \ldots, n_c))$

Subject to the constraint: $\pi_1 + \cdots + \pi_c = 1$

# 2.3 Multinomial Sampling

This kind of problem can be solved using Lagrange Multipliers (if you know how), or simply by taking the constraint into account straight away:

$$l(\boldsymbol{\pi}; (n_1, \ldots, n_c)) = \sum_{j=1}^{c-1} n_j \log(\pi_j) + n_c \log(1 - \sum_{j=1}^{c-1} \pi_j) + Const$$

Hence, for $j = 1, \ldots, c-1$

$$\frac{\partial l(\boldsymbol{\pi}; (n_1, \ldots, n_c))}{\partial \pi_j} = \frac{n_j}{\pi_j} - \frac{n_c}{1 - \sum_{j=1}^{c-1} \pi_j} = \frac{n_j}{\pi_j} - \frac{n_c}{\pi_c}$$

# 2.3 Multinomial Sampling

Setting these $c-1$ equations to zero and solving yields the MLEs

$$\hat{\pi}_j = \frac{n_j}{n_c}\hat{\pi}_c, \quad j = 1, \ldots, c-1$$

and summing these equations gives

$$1 - \hat{\pi}_c = \sum_{j=1}^{c-1} \hat{\pi}_j = (n - n_c)\frac{\hat{\pi}_c}{n_c}$$

telling us $\hat{\pi}_c = \frac{n_c}{n}$ and consequently $\hat{\pi}_j = \frac{n_j}{n}$ for $j = 1, \ldots, c-1$ too.

Recall that the marginal distribution for each $X_j$ is $B(n, \pi_j)$. Hence individual CIs for each $\pi_j$ can be found as before.

# 2.3 Multinomial Sampling

What kind of hypothesis we will consider with multinomial sampling? We could generalise the test we considered for binomial sampling, i.e. move from

$$H_0 : \pi = \pi_0 \quad \text{to} \quad H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0 = (\pi_{10}, \ldots, \pi_{c0})$$
$$H_1 : \pi \neq \pi_0 \quad \text{to} \quad H_1 : \boldsymbol{\pi} \neq \boldsymbol{\pi}_0$$

where $\boldsymbol{\pi}_0$ is a completely specified distribution. For example, when testing the fairness of a die, $\boldsymbol{\pi}_0 = (1/6, \ldots, 1/6)$.

Of the tests we have considered so far, the LR is best placed to test hypotheses like these.

# 2.3 Multinomial Sampling

Recall the LR test statistic, which we now denote $G^2$: if $X_1, \cdots, X_n$ are sampled from $p(x; \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \in \mathbb{R}^k$, and we consider the hypotheses

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \text{ vs } H_1 : \boldsymbol{\theta} \in \Theta - \Theta_0 \text{ where } \Theta_0 \subset \Theta$$

then we have

- Likelihood function: $L(\boldsymbol{\theta}) = \prod_i p(x_i; \boldsymbol{\theta})$
- Loglikelihood function: $l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta})$
- MLE of $\boldsymbol{\theta}$ under $\Theta_0$: $\hat{\boldsymbol{\theta}}_0$
- MLE of $\boldsymbol{\theta}$ under $\Theta$: $\hat{\boldsymbol{\theta}}$
- Likelihood-ratio test statistic:
  $G^2 = -2 \log \Lambda = -2 \log \frac{L(\hat{\boldsymbol{\theta}}_0)}{L(\hat{\boldsymbol{\theta}})} = -2[l(\hat{\boldsymbol{\theta}}_0) - l(\hat{\boldsymbol{\theta}})]$

# 2.3 Multinomial Sampling

Under $H_0$, we have:

$$G^2 = -2 \log \Lambda \to \chi^2_r$$

where the degrees of freedom $r$ is the difference between the dimensions of the parameter spaces under $H_0 \cup H_1$ (i.e. $\Theta$) and under $H_0$ (i.e. $\Theta_0$).

Alternatively, think of $r$ as:

(# parameters estimated under $H_0 \cup H_1$) minus
(# parameters estimated under $H_0$).

# 2.3 Multinomial Sampling

We use the LR test to test the hypotheses

$$H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0 = (\pi_{10}, \ldots, \pi_{c0})$$
$$H_1 : \boldsymbol{\pi} \neq \boldsymbol{\pi}_0.$$

Under $H_0$, likelihood $L(\boldsymbol{\pi}; (n_1, \ldots, n_c))$ can only take one value:

$$L(\boldsymbol{\pi}_0) = \frac{n!}{n_1! n_2! \cdots n_c!} \pi_{10}^{n_1} \pi_{20}^{n_2} \cdots \pi_{c0}^{n_c}.$$

Under $H_0 \cup H_1$, the likelihood is maximized by MLEs $\hat{\pi}_j = n_j/n$:

$$L(\hat{\boldsymbol{\pi}}) = \frac{n!}{n_1! n_2! \cdots n_c!} (n_1/n)^{n_1} (n_2/n)^{n_2} \cdots (n_c/n)^{n_c}$$

hence

$$\Lambda = \prod_{j=1}^{c} \left( \frac{n\pi_{j0}}{n_j} \right)^{n_j}$$

# 2.3 Multinomial Sampling

The test statistic $G^2$ is given by

$$G^2 = -2 \log \Lambda = 2 \sum_{j=1}^{c} n_j \log \left( \frac{n_j}{n \pi_{j0}} \right)$$

and will have the $\chi^2_{c-1}$ distribution for large $n$ because under $H_0$ no parameters were estimated (they were all given to us) and under $H_0 \cup H_1$ we needed to estimated $c - 1$ of the $\pi_j$s (the last $\pi_j$ is determined by the fact they must sum to one).

We are now in a position to test whether our craps dealer and customer were cheating.

# 2.3 Multinomial Sampling

## 2.3.3 Example: Cheating at Craps?

We want to test

$$H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0 = (1/6, \ldots, 1/6)$$
$$H_1 : \boldsymbol{\pi} \neq \boldsymbol{\pi}_0.$$

Calculate using the MLEs:

| Face | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $O_j :=$ Count $n_j$ | 10 | 14 | 6 | 6 | 4 | 20 |
| $E_j := n\pi_{j0}$ | 10 | 10 | 10 | 10 | 10 | 10 |
| $O_j \log(O_j/E_j)$ | 0 | 4.71 | -3.06 | -3.06 | -3.67 | 13.86 |

from which we find $G^2 = 17.56$. The $P$-value is 0.0036, so we reject the null hypothesis at the 1% level. The dealer and customer must be punished.

# 2.3 Multinomial Sampling

## 2.3.4 Inference: revisited

With a little thought, we realise the LR test can handle much more than a simple test of whether $\pi$ equals an *exact* value or not. It can handle null hypotheses like

- $H_0 : \pi_1 = \pi_2,\ \pi_3 = \pi_4$
- $H_0 : \pi_1 + \pi_2 = \pi_3$

in which not all (or maybe none of) the parameters are completely specified, but rather a relationship between them is.

How would the LR test work with hypotheses like these, where $\pi_0$ is unknown?

# 2.3 Multinomial Sampling

If $\pi_0$ is unknown, we need to estimate it using the data we have. This gives us estimates $\hat{\pi}_j, j = 1, \ldots, c$ for the category/cell probabilities under $H_0$. Multiply these by $n$ to have $E_j, j = 1, \ldots, c$, the expected cell values under $H_0$.

Use these $E_j$ with the observed cell values $O_j$ to compute the LR test statistic (also called the deviance test statistic) $G^2$:

$$G^2 = 2 \sum_{j=1}^{c} O_j \log \left( \frac{O_j}{E_j} \right).$$

The critical value comes from the $\chi^2_r$ distribution, where $r = c - 1 -$ # parameters estimated under $H_0$.

# 2.3 Multinomial Sampling

## 2.3.5 Example: With $\pi_0$ unspecified

We observe the following data

| Cell | 1 | 2 | 3 | 4 | Total |
|---|---|---|---|---|---|
| Frequency | 12 | 13 | 20 | 25 | 70 |

and wish to test the hypothesis $H_0 : \pi_1 = \pi_2,\ \pi_3 = \pi_4$ against the alternate hypothesis that the model does not hold.

First find the MLEs of the $\pi_j, j = 1, \ldots, 4$ under $H_0$. The loglikelihood function is

$$\begin{aligned} l(\boldsymbol{\pi}) &= n_1 \log \pi_1 + n_2 \log \pi_1 + n_3 \log \pi_3 + n_4 \log \pi_3 \\ &= (n_1 + n_2) \log \pi_1 + (n_3 + n_4) \log \pi_3 \end{aligned}$$

# 2.3 Multinomial Sampling

Now, $\pi_1 + \pi_2 + \pi_3 + \pi_4 = 2\pi_1 + 2\pi_3 = 1$, hence $\pi_3 = 1/2 - \pi_1$ and

$$l(\boldsymbol{\pi}) = (n_1 + n_2)\log \pi_1 + (n_3 + n_4)\log(1/2 - \pi_1)$$

$$\Rightarrow \frac{\partial l(\boldsymbol{\pi})}{\partial \pi_1} = \frac{n_1 + n_2}{\pi_1} - \frac{n_3 + n_4}{1/2 - \pi_1}$$

Thus the MLE for $\pi_1$ satisfies

$$\frac{n_1 + n_2}{\hat{\pi}_1} - \frac{n_3 + n_4}{1/2 - \hat{\pi}_1} = 0$$

which yields the solution $\hat{\pi}_1 = \frac{n_1+n_2}{2n}$. The rest of the MLEs quickly follow: $\hat{\pi}_2 = \hat{\pi}_1$ and $\hat{\pi}_3 = \frac{n_3+n_4}{2n} = \hat{\pi}_4$.

# 2.3 Multinomial Sampling

Plugging in the numbers, we see that, the MLEs under $H_0$ are:

$$\hat{\pi}_1 = \hat{\pi}_2 = \frac{25}{140} = \frac{5}{28} \text{ and } \hat{\pi}_3 = \hat{\pi}_4 = \frac{45}{140}$$

We then have

| Cell $j$ | 1 | 2 | 3 | 4 | Total |
|---|---|---|---|---|---|
| $O_j$ = Frequency | 12 | 13 | 20 | 25 | 70 |
| $E_j = n\hat{\pi}_{j0}$ | 12.5 | 12.5 | 22.5 | 22.5 | 70 |

Now calculate the test statistic: $G^2 = 2\sum_{j=1}^{4} O_j \log(\frac{O_j}{E_j}) = 0.597$, and the degrees of freedom is 2: three of $\pi_j, j = 1, 2, 3, 4$ were estimated under $H_0 \cup H_1$ and one $(\pi_1)$ was estimated under $H_0$. Therefore the critical value is given by $\chi^2_{2,0.05} = 5.991$ and we do not reject $H_0$.

# 2.4 Pearson's Chi-squared Test

## 2.4.1 Testing models

The last example gives some insight into how general the LR test can be. Really, what we tested was whether the model $\pi_1 = \pi_2, \pi_3 = \pi_4$ was a good fit, compared to the alternative. More generally, it looks the LR test can be used to test hypotheses like

$$H_0 : \text{Model } M_0 \text{ fits.}$$
$$H_1 : \text{Model } M_0 \text{ does not fit.}$$

There is another test which is commonly used to see if models fit table data. It is called Pearson's Chi-squared test.

# 2.4 Pearson's Chi-squared Test

## 2.4.2 Another test

Recall $O_j$ is the observed count of category $j$ and $E_j$ is the expected count of category $j$ if $H_0$ were true.

Pearson's goodness-of-fit statistic, $X^2$, is given by

$$X^2 = \sum_{j=1}^{c} \frac{(O_j - E_j)^2}{E_j}.$$

When $n$ is large, $X^2$ has the $\chi^2_r$ distribution, where $r = c - 1 -$# parameters estimated under $H_0$.

# 2.4 Pearson's Chi-squared Test

Where does $X^2$ come from?
For an illustration, consider a one-way table with $c = 2$ and $H_0 : \pi = \pi_0$. We observe a large sample of size $n$, with counts $O_1$ (in the "success" category) and $O_2$. Calculate $X^2$:

$$\begin{aligned} X^2 &= \frac{(O_1 - n\pi_0)^2}{n\pi_0} + \frac{(O_2 - n(1-\pi_0))^2}{n(1-\pi_0)} \\ &= \frac{(O_1 - n\pi_0)^2}{n\pi_0} + \frac{(n - O_1 - n(1-\pi_0))^2}{n(1-\pi_0)} \\ &= \frac{(O_1 - n\pi_0)^2}{n\pi_0(1-\pi_0)} \end{aligned}$$

Under $H_0$, since $n$ is large, $O_1 \sim N(n\pi_0, n\pi_0(1-\pi_0))$. Hence, under $H_0$, $X^2$ is the square of a standard normal random variable and therefore follows the $\chi^2_1$ distribution.

# 2.4 Pearson's Chi-squared Test

## 2.4.3 Example revisited: Cheating at Craps?

| Face | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $O_j := \text{Count } n_j$ | 10 | 14 | 6 | 6 | 4 | 20 |
| $E_j := n\pi_{j0}$ | 10 | 10 | 10 | 10 | 10 | 10 |
| $(O_j - E_j)^2/E_j$ | 0 | 1.6 | 1.6 | 1.6 | 3.6 | 10 |

from which we find $X^2 = 18.4$ and the $P$-value is 0.0025. Recall $G^2 = 17.56$ and the $P$-value was 0.0036. In good practice to perform the Pearson Chi-squared test and the LR test. If $X^2$ and $G^2$ are similar, we can be confident that the large-sample approximation to normality has worked.

# 2.4 Pearson's Chi-squared Test

## 2.4.4 Example revisited: With $\pi_0$ unspecified

From before

| Cell $j$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $O_j$ = Frequency | 12 | 13 | 20 | 25 |
| $E_j = n\hat{\pi}_{j0}$ | 12.5 | 12.5 | 22.5 | 22.5 |
| $(O_j - E_j)^2/E_j$ | 0.02 | 0.02 | 0.278 | 0.278 |

from which we find the test statistic $X^2 = 0.596$ and the $P$-value is 0.742. The $P$-value from the LR test was 0.742.

# 2.4 Pearson's Chi-squared Test

## 2.4.5 Small expected cell counts

In the binomial setting, we said that having $n\pi$ and $n(1-\pi)$ both $\geq 5$ was good enough to be confident of using the large-sample approximation employed by the likelihood based analysis.

In the multinomial setting, the rule of thumb *used* to be $E_j \geq 5$ for each category $j$. However, these days the rule is slightly modified: we can have $E_j < 5$ for at most 20% of the cells, but none of the $E_j$s can be smaller than $1$.

If some of the $E_j$s are too small, try combining categories until the rule of thumb allows you to use the large-sample approximation.

# 2.5 Poisson Sampling/Distribution

## 2.5.1 Characterisation

The Poisson distribution is used to model counts which occur randomly over a fixed time/place.

Poisson sampling is characterised by

- The total sample size $n$ is not fixed
- The counts $X_1, \ldots, X_c$ are independent Poisson variables, with rates $\mu_1, \ldots, \mu_c$.
- The Poisson distribution itself requires independence of events; homogeneity of the rate; and that the time or space in which the events are observed is fixed.

# 2.5 Poisson Sampling/Distribution

## 2.5.2 Inference

We have already shown that given a sample $y_1, \ldots, y_n$ from a $Po(\mu)$ distribution, the MLE for parameter $\mu$ is $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i$, the sample mean.
Moreover, as $n$ grows large, $\hat{\mu} \sim N(\mu, \frac{\mu}{n})$.

- The Wald test statistic is $\frac{\hat{\mu}-\mu_0}{\sqrt{\hat{\mu}/n}}$.
- The Score test statistic is $\frac{\hat{\mu}-\mu_0}{\sqrt{\mu_0/n}}$.
- The LR test statistic is $2n(\mu_0 - \hat{\mu}) + 2n\hat{\mu}\log(\hat{\mu}/\mu_0)$.

These are useful for testing whether a Poisson parameter takes a particular value. But if we want to test whether the data we observe follows some $Po(\mu)$ distribution?

# 2.5 Poisson Sampling/Distribution

## 2.5.3 Example: Death by donkey kicks

The LR and Pearson's Chi-squared test can handle a question like this, since it is asking if the Poisson distribution is a good model for the data. We demonstrate the method with the Prussian army data.

Check whether the sample looks like it could come from a Poisson distribution:

- the time/space in which the events were observed were fixed (280 corps-years)
- the total sample of deaths (196) was not fixed
- the # deaths in a corps-year does not affect the # deaths in another corps-year (seems reasonable)
- the mean # deaths is the same for all corps-years (seems reasonable).

# 2.5 Poisson Sampling/Distribution

We wish to test $H_0$ : Data follows some $Po(\mu)$ distribution.

Under $H_0$, $\pi_1$, the probability a random chosen corps-year contains zero deaths, is given by $e^{-\mu}$ for some $\mu$. Likewise, $\pi_2 = e^{-\mu}\mu$, $\pi_3 = \frac{1}{2}e^{-\mu}\mu^2$, $\pi_4 = \frac{1}{6}e^{-\mu}\mu^3$ and $\pi_5 = 1 - \sum_{j=1}^{4} \pi_j$.

Essentially, we now see that $H_0$ states the multinomial probabilities $\pi_j$ depend on some unknown parameter $\mu$ in a particular way. Let $\Theta_0$ be the parameter space of all $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_5)$ which satisfy the equations above for some $\mu$. Let $\Theta$ be the parameter space of all $\boldsymbol{\pi}$. Then our test is actually

$$H_0 : \boldsymbol{\pi} \in \Theta_0 \text{ vs } H_1 : \boldsymbol{\pi} \in \Theta - \Theta_0.$$

# 2.5 Poisson Sampling/Distribution

We begin by estimating the Poisson parameter $\mu$ via the MLE, $\hat{\mu}$, the sample mean. To do that, we need to know what observations fell in the "4+" category. It turns out they were two observation of 4 deaths a corps-year. Hence

$$\hat{\mu} = \frac{144 \times 0 + 91 \times 1 + 32 \times 2 + 11 \times 3 + 2 \times 4}{280} = 0.7.$$

Now use $\hat{\mu}$ to estimate the $\pi_j$s:

$$\hat{\pi}_1 = e^{-\hat{\mu}} = 0.497 \qquad \hat{\pi}_2 = e^{-\hat{\mu}}\hat{\mu} = 0.348$$

$$\hat{\pi}_3 = e^{-\hat{\mu}}\hat{\mu}^2/2 = 0.122 \qquad \hat{\pi}_4 = e^{-\hat{\mu}}\hat{\mu}^3/6 = 0.028$$

$$\hat{\pi}_5 = 1 - \sum_{j=1}^{4} \hat{\pi}_j = 0.006$$

# 2.5 Poisson Sampling/Distribution

| # deaths | 0 | 1 | 2 | 3 | 4+ |
|---|---|---|---|---|---|
| Count | 144 | 91 | 32 | 11 | 2 |
| $E_j$ | 139.0 | 97.3 | 34.1 | 7.9 | 1.6 |

The "4+" category has a small expected value under $H_0$, but since the other 4 are $\geq 5$, we can assume the large-sample approximation is valid.

We calculate $G^2 = 1.86$ and $X^2 = 1.98$. Under $H_0$, there was one parameter to estimate: $\mu$. Under $H_0 \cup H_1$, there were four: 4 from $\pi_j$, $j = 1, \ldots, 5$. Therefore, we calculate $P$-values from the $\chi^2_3$ distribution.

The $P$-value from the LR test is 0.603. The $P$-value from the Pearson's Chi-squared test is 0.577. Both tests suggest the Poisson distribution is a reasonable model for the data.