(ii) "<u>Accidents</u>": To test if accidents are truly distributed "randomly" over drivers, consider one specific (but generic) bus driver. <u>If accidents are really that</u> (i.e. if accidents shouldn't be more likely to happen to any driver rather than any other), then each time one occurs, there is a small π = 1/708 chance that it happens to the one driver we are considering. There are n=1623 such accidents to be 'distributed' so our driver has n "opportunities" to be dealt an accident [we ignore practical details such as whether the driver is still off work from the last one!].

We could therefore work out the binomial probability that the driver has 0, 1, 2, .. accidents. Now n is large enough and π small enough that using the Poisson formula with $\mu = n\pi$ = 1623 / 708 will be quite accurate and save a lot of calculation. We then use the probability that any one driver will have y accidents as a synonym for the proportion of all drivers that will have y accidents

We can now compare the observed distribution of accidents and see how well it agrees with this theoretical Poisson distribution. Colton (pages 16-17) examines the fit of the Poisson model to the actual data...

Colton Table 2.5 Observed and "expected" numbers of accidents during a 3-year period among 708 Ulster (Northern Ireland Transport Authority) bus drivers.

# accidents in 3-year period (y)	Number of drivers with this many accidents: Observed	Expected †	O × y
0	117	71.5	0
1	157	164.0	157
2	158	187.9	316
3	115	143.6	345
4	78	82.3	312
5	44	37.7	220
6	21	14.4	126
7	7		49
8	6		48
9	1	(7-11 combined) 6.6	9
10	3		30
11	1		11
Σ	708	708	1623

$$\hat{\mu} = \frac{\#\text{accidents}}{\#\text{drivers}} = \frac{0\times117 + 1\times157 + ..}{708} = \frac{1623}{708} = 2.29.$$

var(y) = 3.45 >> mean(y).

†Expected number = Poisson Prob(# accidents=y $|\hat{\mu}$) × 708

(e.g. prob(0) = exp[-2.29]= 0.101 ; expected # of 0's = 708 × 0.101 = 71.5)

"Comparison of observed and expected tabulations reveals more than the expected number of drivers with no accidents and with five or more accidents . <u>These data suggest that the accidents did not occur completely at random</u>; in fact it appears that there is some indication of accident proneness. From this example, what conclusions are justified concerning the random or nonrandom distribution of bus accidents?"