CA200 – Quantitative Analysis for Business Decisions

File name: CA200_Section_04B_StatisticsIntroductionII

Table of Contents

4. Introduction to Statistics II

4.6 Standard Discrete Distributions continued
Further Examples on Use

Example **5**:
The probability of a *good* component in inspecting assembly line output is known to be 0.8 ; probability of a *bad* component is 0.2. Inspecting components randomly from the line, a sample of size 12 is checked

What is the Expected No. of good components and what is the Standard Deviation?
Note: Assumes each inspected item picked *independently*.

Solution:
Binomial distribution ; i.e. one of two outcomes. Parameters n=12, p=0.8

X = random variable = No. components. So, from basic principles – (see previous examples), or from tables:

Expected No. (i.e. Mean No.) of Good Components = $E(X) = np = 12 \times 0.8 = \underline{9.6}$

Standard Deviation (X)= $\sqrt{\text{Var}(X)} = \sqrt{(np(1-p))}$

$$= \sqrt{(npq)} = \sqrt{(12 \times 0.8 \times 0.2)} = \underline{1.386}$$

Example 6:

Suppose components are placed into bins containing **100** each. After inspection of a large number of bins, *average no.* defective components found to be 10, with standard deviation = 3.
Assuming that same production conditions are maintained for larger bins, containing **300** components each

(a) What would be the average no. (expected no.) defective components per larger bin?
(b) What would be the S.D. of the No. defectives per larger bin?
(c) How many components must each bin hold so that S.D. of No. defective components =1% of the Total no. components in the bin?

<u>Solution</u>

Proportion defective = 0.1 (from the inspection phase). So, proportion good = 0.9.

We are given what the Mean and S.D. are for inspection phase, but as a check on what we have learned about distributions, these clearly come from:

Mean = $E(X) = np$ = 100 × 0.1 = 10 components defective on average

S.D. (X) = $\sqrt{(npq)} = \sqrt{(100 \times 0.1 \times 0.9)} = 3$

(a) Sample size 'n' now = 300

Production is assumed to be continuing as before, so proportion defective = 0.1 hence E(X) for larger bin size = 300 × 0.1 = <u>30</u>

(b) S.D. (X) = $\sqrt{(npq)}$ = now $\sqrt{(300 \times 0.1 \times 0.9)}$ = <u>5.2</u>

(c) For the S.D. (standard deviation) to be 1% of Total No. (this implies trying to impose quality control at a desirable level on the production), then would expect to have to look critically at e.g. whether sampling at sufficient intervals, suitable sampling process and so on. Here

$$\sqrt{(npq)} = n / 100$$

$$\sqrt{(n \times 0.1 \times 0.9)} = n / 100$$

$$900n = n^2$$

$n = 900$. i.e. bins must hold 900 components

<u>Note</u> :

If interested in *proportions*, rather than *Counts* (i.e. number of) then divide by n.

Hence

$$E\{Y\} = {np}/{n} = p$$

$$S.D\{Y\} = {\sqrt{npq}}/{n} = \sqrt{{pq}/{n}}$$

<u>Further</u>**: Normal Approximation to Binomial**

Binomial calculations can get quite cumbersome as we have seen in the previous section. Less work is involved for Normal, so try to use this when appropriate.

In general, applies if np > 5, when p < 0.5

If nq > 5, when p > 0.5

Approximation requires *re-writing* original variable = count (or proportion) in terms of Standardised Normal variable, **U**, as usual.

Example 7.

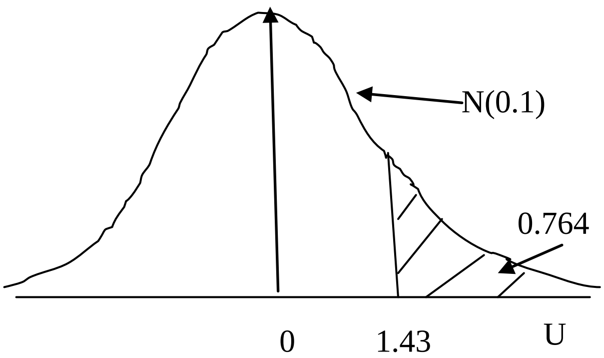
Records show that 60% of students pass their exams. at the first attempt. Using the Normal Approximation to the Binomial, calculate the probability that at least 65% of a group of 200 students will pass at the first attempt.

Solution We have p = 0.6, q = 0.4, n=200

$$\sigma = \sqrt{pq/n} = \sqrt{0.6 \times 0.4/200} = 0.035$$

$$U = \frac{(Observed\,Value - Expected\ Value)}{S.D.} = \frac{X-\mu}{\sigma} = \frac{0.65-0.60}{0.035} = 1.43$$

So want probability that 65% or more pass at first attempt.
The value , U =1.43, divides up the Normal distribution, s.t. 0 to 64.999% distribution below and 65% to 100% above this value.



So, the probability is 0.764 of 65% or more passing at the first attempt.

Recall: The **Poisson Distribution** describes No. events occurring within a given interval: Useful to work with this, instead of Binomial, when sample size (n) *large*, when we have random *independent* events and p small: recall *'rare'* event distribution.

$$P\{X = k\} = e^{-\lambda}\lambda^{k}/k! \qquad k = 0,1,2,......$$

Where k is some discrete value as shown and λ is the Poisson parameter and equal to both the Expectation and Variance of the random variable X.

Example 8:

In a transport fleet, there is, on average, one breakdown a week, which requires a recovery operation. What is the expected pattern of recoveries over 100 weeks?

Solution : the long way

No. Recoveries	Probability	Recovery Pattern for100 weeks
0	$P\{Y=0\} = e^{-1}1^0 / 0! = 0.3679$	0.3679 x 100 = 37 weeks
1	$P\{Y=1\} = e^{-1}1^1 / 1! = 0.3679$	0.3679 x 100 = 37 weeks
2	$P\{Y=2\} = e^{-1}1^2 / 2! = 0.1840$	0.1840 x 100 = 18 weeks
3	$P\{Y=3\} = e^{-1}1^3 / 3! = 0.06130$	0.0613 x 100 = 6 weeks
4	$P\{Y=4\} = e^{-1}1^4 / 4! = 0.01530$	0.0153 x 100 = 2 weeks
5	$P\{Y=5\} = e^{-1}1^5 / 5! = 0.00360$	0.0036 x 100 = 0 weeks
	Total probability = 1.00	Total weeks = 100

Alternatively: Use (Cumulative) Poisson Tables.
White, Yeats & Skipworth, "Tables for Statisticians" page 7

Solution

For Poisson parameter (= mean = variance = λ = 1 here), Cum. Poisson has values:

No. random events **(r or more)**	For λ=1.0, probabilities are:		Recovery Pattern 100 weeks
	Prob{No. events ≥ r}	Prob{No. events = r}	Prob. × 100
0	1.00000	1.0000-0.62312 = 0.36788	37 weeks
1	0.63212	0.62312-0.26424 = 0.36788	37 weeks
2	0.26424	0.26424-0.08030 = 0.18394	18 weeks
3	0.08030	0.08030-0.01899 = 0.06131	6 weeks
4	0.01899	0.01899-0.00366 = 0.01533	2 weeks
5	0.00366	0.00366-0.00059 = 0.00307	0 weeks
6	0.00059	0.00059-0.00008 = 0.00051	
7	0.00008	0.00008-0.00001 = 0.00007	
8	0.00001		
		Total Prob. =1.00	Total =100 weeks

Example 9: Customers arrive randomly at a service point at an average rate of *30 per hour*. Assuming a **Poisson distribution**, calculate the probability that:

(i) No customer arrives in any particular minute

(ii) Exactly one arrives ..

(iii) Two *or more* arrive ..

(iv) Three *or fewer* arrive ..

Solution

The time interval requested is a **minute** (not an hour). So, Mean (Poisson parameter λ) is 30/60 =0.5.

Using Tables p. 7 for λ =0.5, gives:

(i) P{No Customer} = (1.00000-0.39347) = 0.60653

(ii) P{1 customer} = (0.39347-0.09020) = 0.30327

(iii) P{2 *or more* customers} = 0.09020, (reading directly from tables)

(iv) P{ 3 *or fewer*} = 1- P{ 4 *or more*}

= 1 – 0.00175

= 0.99825

Further**: Normal Approximating Poisson**

Again, if it is possible to use less lengthy calculations, then we do so. In general, when the sample size is large, the Poisson can be approximated by the Normal. From the statistical tables, this says any value of the Poisson parameter $\lambda > 20$ can be thought of as large. In practice, the Normal approximation works reasonably well for values of $\lambda \geq 12$, provided that sample size very large and events not too 'rare, i.e. probability of an individual event not *too* small.

Example 10:

Suppose work stoppages per day (X) in a particular factory, due to faulty machines, is 12 on average.

What is the probability of *15 or fewer* work stoppages due to machine faults on any given day?

Solution

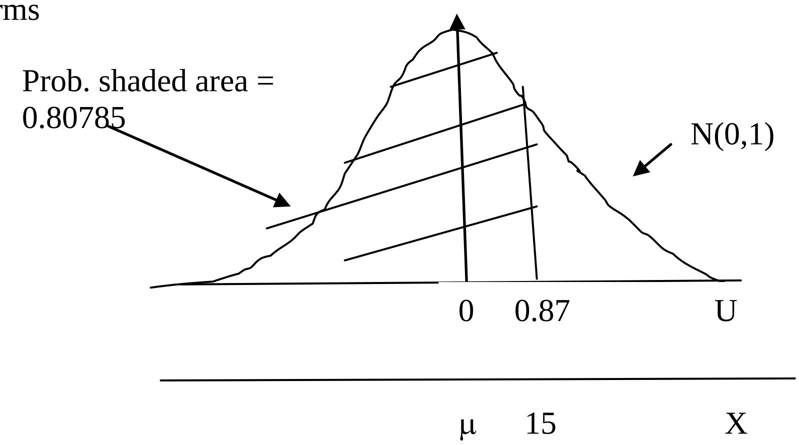
λ= 12, assumed 'large' here. Poisson distribution, so Mean = λ, and S.D. (i.e. σ) = $\sqrt{\lambda}$
Could use cumulative Poisson tables (as before) and calculate the probability as P{15 or fewer} = 1 – P{16 or more} (p.10 tables)

Alternatively: Transform to Standardised Normal variable, using information on the value of interest (observed), the mean (or expected) value and S.D., to give

$$U = \frac{(Observed Value - Expected\ Value)}{S.D.} = \frac{X-\mu}{\sigma} = \frac{15-12}{\sqrt{12}} = +0.87$$

Interested in *15 or fewer* work stoppages , so Probability of everything *below* 0.87

Mean (X) = μ = 12, transforms to U = 0 as usual



So, the probability is <u>0.80785</u> of *15 or fewer* stoppages on any given day.

Further : **Poisson Approximating Binomial**

Sample size , (n) 'large', but probability of 'success' (p) small (i.e. rare events)

For X = No. Successes, denote Mean (μ) = E(X) = np

S.D. (σ) = $\sqrt{}$(Var(X)= $\sqrt{}$(np)

[For Binomial, this is usually written as $\sqrt{}$(npq) of course, where q = 1-p as usual. Now p is very small, (rare event) so q $\cong$ 1 and n is large, so the product np is essentially a constant].

Example 11: Compute the probability of obtaining exactly 1 tyre from a sample of 20 if 8% of tyres manufactured at a particular plant are known to be defective.

For **Binomial**: Could calculate from first principles, but this is a lot of work. Also, calculation at limit of tables.

<u>Solution:</u>

$$P\{X = 1 \mid n = 20, p = 0.08\} = \binom{20}{1}(0.08)^1(0.92)^{19} = 0.3282$$

where $\left[{}^{20}C_1 \quad \textit{also} \quad \textit{written} \quad \binom{20}{1} \right]$

For **Poisson**:

$$P\{X = 1 \mid \mu = np = 1.6\} \approx \frac{e^{(-20)(0.08)}[(20)(0.08)]^1}{1!}$$

Still some work, but from Tables directly for a mean of 1.6, gives

Probability = 0.79810-0.47507 = <u>0.32303</u>

Summary Hints on which distribution applies

Binomial:

Involves discrete outcomes (counts or No. of occcurrences) with 2 outcomes per event e.g. M/F, Yes/ No or similarly; outcome probabilities p and q= 1-p respectively.

<u>Note</u>: When no. of items, **n**, is large and **p** is not close to 0 or 1, (i.e. distribution can be taken to be approximately *symmetric*), then Binomial probabilities can be *approximated* using a **Normal distribution** with mean (μ) = np, SD (σ) = √(npq).

Poisson:

Similar to Binomial, discrete distribution, but used for *rare* events

<u>Note</u>: can use *instead of Binomial* if no. items 'large', say **n** ≥ 20, when probability of event **p** ≤0.05. However, approximation still quite good for n around 20 and p up to about 0.1, so more general rule is that the mean of the Poisson (λ= np) should be < 5.

Can approximate Poisson itself by the Normal Distribution when **n** very large and **p** not too small or *more specifically* (see statistical tables), when λ > 20.

Normal:

Most commonly applied distribution and used for variables with continuous range of possible values.

Also applies when a large no. items / *large group or sample size n* is considered; in which case can also be used to approximate quantities following discrete distributions.

4.7 Exercises

These will be put up separately for students to try themselves.