

Normal Approximation to the Binomial

e.g. Estimate the probability that a school of 1200 students contains more than 150 left-handed students

Q: How would you solve such a problem?

A: One approach would be to take a large sample, say 50 students, and count the number of left handed students. From that information you could estimate the probability.

Say the sample contained 8 left handed students, we would estimate the probability of being left handed as;

$$P(\text{left handed}) = \frac{8}{50} = 0.16$$

Using binomial probability;

Let X = number of left handed students

$$X \sim \text{Bin}(1200, 0.16)$$

$$P(X > 150) = 1 - P(X \leq 150)$$

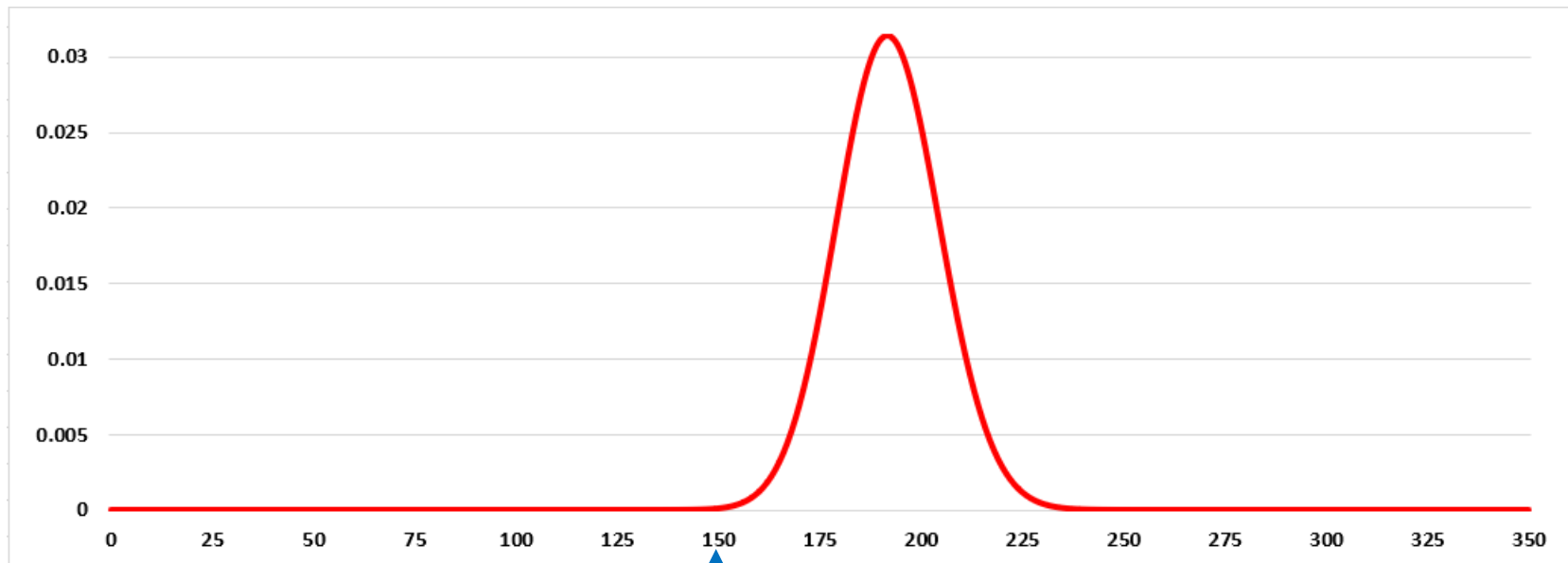
$$\begin{aligned} = 1 - & \binom{1200}{0} (0.84)^{1200} (0.16)^0 - \binom{1200}{1} (0.84)^{1199} (0.16)^1 - \dots \\ & \dots - \binom{1200}{149} (0.84)^{1051} (0.16)^{149} - \binom{1200}{150} (0.84)^{1050} (0.16)^{150} \end{aligned}$$

even using the idea of complimentary events, it still involves 151 calculations

$$= 1 - 0.0003838$$

$$= \underline{0.9996 \text{ (to 4 dp)}}$$

The distribution's polygon would look like



we just found the area to the right of

- The distribution has a modal class somewhere in the middle of the range of values
- The distribution is symmetrical
- The frequency density tails off fairly rapidly as values move further away from the modal class

These are the features of a normal distribution

So what would have happened if we assumed our distribution was normal;

first we need to find the mean and variation

$$\begin{aligned} X \sim \text{Bin}(1200, 0.16) \quad \mu &= np & \sigma^2 &= np(1-p) \\ &= (1200)(0.16) & &= (1200)(0.16)(0.84) \\ &= 192 & &= 161.28 \end{aligned}$$

$$X \sim N(192, 161.28) \Leftrightarrow Z \sim N(0, 1) \text{ where } Z = \frac{X - \mu}{\sigma}$$

we need to
standardise the
data

$$\begin{aligned} P(X > 150) &= 1 - P(X \leq 150) \\ &= 1 - P\left(Z \leq \frac{150 - 192}{\sqrt{161.28}}\right) \\ &= 1 - \Phi(-3.31) \\ &= 1 - 0.0046654 \\ &= \underline{0.9995} \text{ (to 4 dp)} \end{aligned}$$

% difference to
binomial answer

$$\frac{0.999616326 - 0.99953346}{0.999616326} \times \frac{100}{1}$$

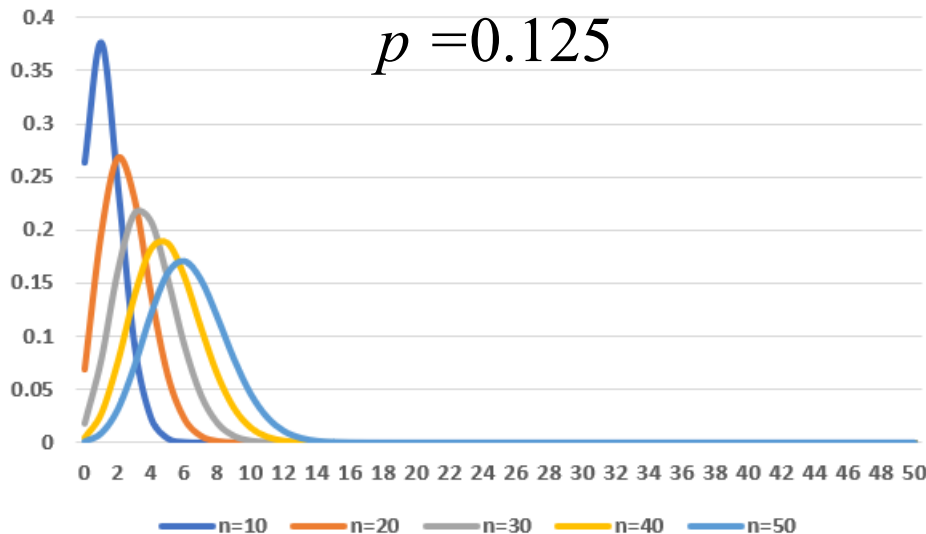
= 0.008%

When is it okay to approximate a binomial distribution with a normal distribution?

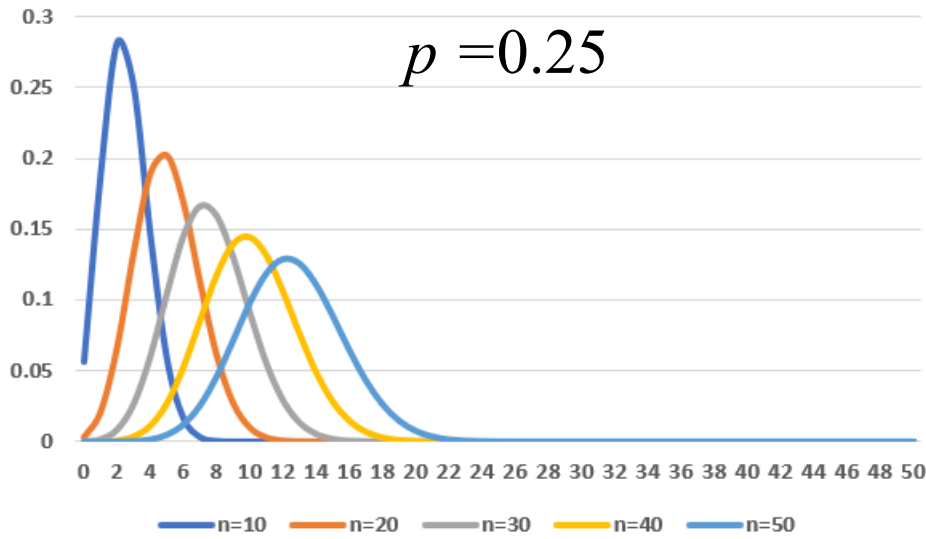
Keep in mind the three key features of a normal distribution

- The distribution has a modal class somewhere in the middle of the range of values
- The distribution is symmetrical
- The frequency density tails off fairly rapidly as values move further away from the modal class

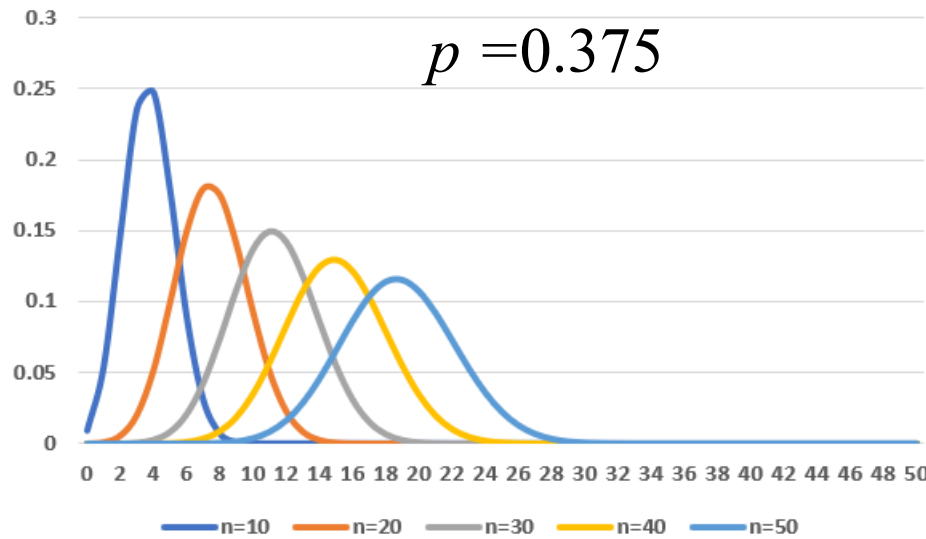
Let's take a look at some binomial distributions;



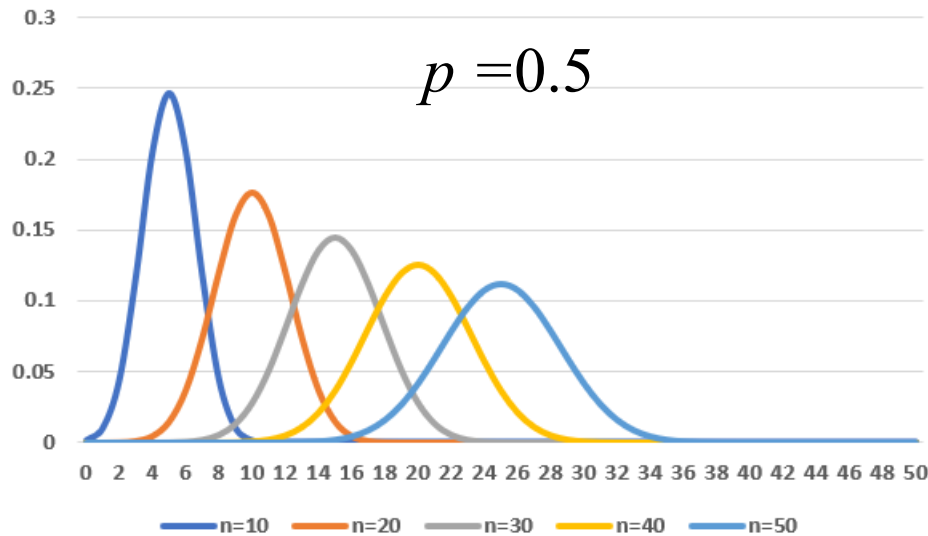
n	p	q	np	nq	✓ or ✗
10	0.125	0.875	1.25	8.75	✗
20	0.125	0.875	2.50	17.50	✗
30	0.125	0.875	3.75	26.25	✗
40	0.125	0.875	5.00	35.00	✓ ✗
50	0.125	0.875	6.25	43.75	✓



n	p	q	np	nq	<input checked="" type="checkbox"/> or <input checked="" type="checkbox"/>
10	0.25	0.75	2.5	7.5	<input checked="" type="checkbox"/>
20	0.25	0.75	5.0	15.0	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
30	0.25	0.75	7.5	22.5	<input checked="" type="checkbox"/>
40	0.25	0.75	10.0	30.0	<input checked="" type="checkbox"/>
50	0.25	0.75	12.5	37.5	<input checked="" type="checkbox"/>



n	p	q	np	nq	<input checked="" type="checkbox"/> or <input checked="" type="checkbox"/>
10	0.375	0.625	3.75	6.25	<input checked="" type="checkbox"/>
20	0.375	0.625	7.50	12.50	<input checked="" type="checkbox"/>
30	0.375	0.625	11.25	18.75	<input checked="" type="checkbox"/>
40	0.375	0.625	15.00	25.00	<input checked="" type="checkbox"/>
50	0.375	0.625	18.75	31.25	<input checked="" type="checkbox"/>



n	p	q	np	nq	☑ or ☒
10	0.5	0.5	5.00	5.00	☒ ☑
20	0.5	0.5	10.00	10.00	☑
30	0.5	0.5	15.00	15.00	☑
40	0.5	0.5	20.00	20.00	☑
50	0.5	0.5	25.00	25.00	☑

As a general rule;

for a satisfactory approximation
 $np > 5$ and $nq > 5$

We are approximating the area under the polygon with bins 1 unit apart, so for the interval $a \leq x \leq b$, the area under the polygon is actually

$$\int_{a-0.5}^{b+0.5} f(x) dx$$

(this might be more obvious if you think of the histogram)

continuity correction for small samples

for small n ; $P(a \leq X \leq b)$ use $P(a - 0.5 \leq X \leq b + 0.5)$

Exercise 17C;

1, 2adg, 3, 4, 5, 7, 9, 11, 12