Approximation of Binomial with Normal Model

- Remarkably, when n, np and nq are large, or p is close to ½, the Binomial distribution is well approximated by Normal distribution.

- A Normal model works pretty well if we expect to see at least 10 successes and 10 failures. Use Success/Failure Condition($np \geq 10$ and $n(1 - p) \geq 10$).

- Recall that X_i is a Bernoulli random variable with mean:
  $$\mu = E(X) = (0)(1-p) + (1)(p) = p$$
  and variance:
  $$\sigma^2 = Var(X) = E[(X-p)^2] = (0-p)^2(1-p) + (1-p)^2(p) = p(1-p)[p+1-p] = p(1-p)$$
  With a sample of n Bernoulli distribution, which is a Binomial (n, p),
  Mean = np
  Sd = $\sqrt{npq}$
- For example, consider 4000 coin-flips with probability of head equal to 50%. Is possible to calculate the probability of getting at least 3000 heads?

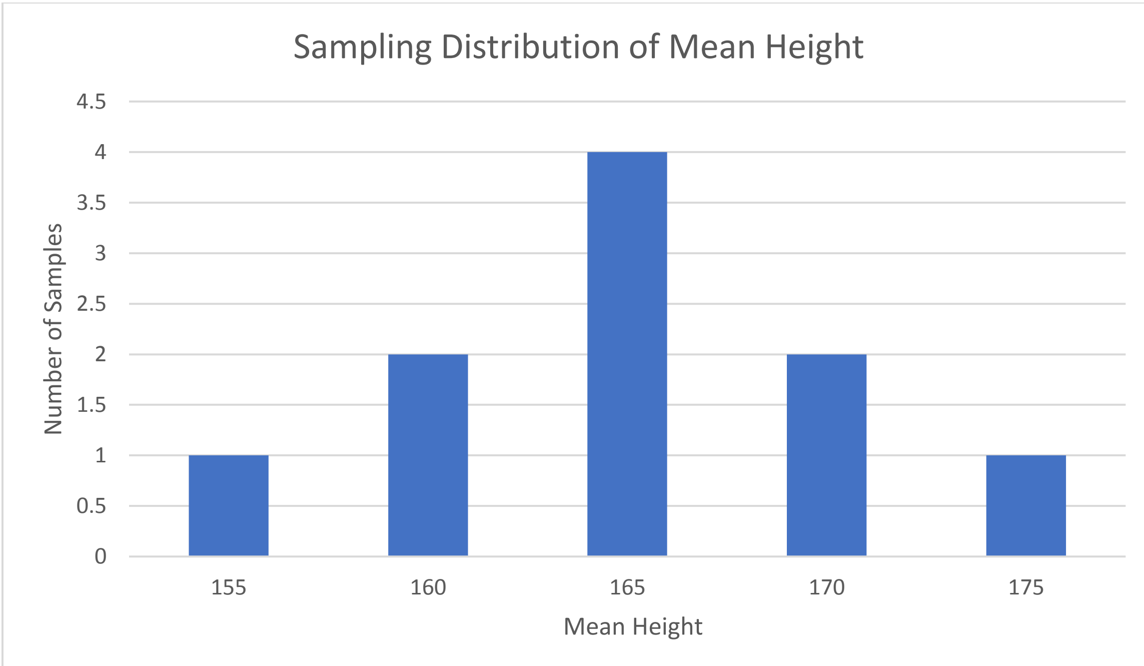
Sampling Distribution Model

- Assume we like to measure the mean height of the U of T students(Large population). In the population, there are 4 students A(170cm), B(180cm), C(160), and D(150cm). We are interested in the sampling distribution of the mean height.
- Parameter = true mean height (usually impossible to compute) = $\frac{170+180+160+150}{4}$ = 165cm
- Statistic = sampling mean height (we want to find the distribution of this!)
- Now we are taking a random sample of size 2, and here are the results

Sample	Mean height
A,B	175cm
B,C	170cm
A,C	165cm
B,D	165cm
A,D	160cm
D,C	155cm
B,C	170cm
D,A	1cm
A,C	165cm

D,B	165cm

- Sampling Distribution of Sample Mean Height for Sample Size of 2



Central Limit Theorem

- The sampling distribution of any mean becomes more nearly Normal as the sample size grows. The only assumption is independence and randomization.
- Mean μ = $\frac{sum\ of\ all\ data\ points}{n}$
- Standard deviation s = $\frac{\sigma}{\sqrt{n}}$

Question 1 (67 And More Tennis, Chapter 13 Page 484)
Suppose the tennis player in Exercise 65 serves 80 times in a match. p = 0.7.

(a) What are the mean and standard deviation of the number of good first serves expected?
(b) Verify that you can use a Normal model to approximate the distribution of the number of good first serves.
(c) Use the 68-95-99.7 Rule to describe this distribution.
(d) What is the probability that she makes at least 65 first serves?

Question 2(47 Pregnancy, Chapter 14 Page 526)

Assume that the duration of human pregnancies can be described by a Normal model with mean 266 days and standard deviation 16 days

(a) What percentages of pregnancies should last between 270 days and 280 days?

$Z_1 = \frac{y-\mu}{\sigma} = \frac{270-266}{16} = 0.25$

$Z_2 = \frac{y-\mu}{\sigma} = \frac{280-266}{16} = 0.875$

P (0.25 < Z < 0.875) = 21.1%

(b) At least how many days should the longest 25% of all pregnancies last?

$0.674 = \frac{y-266}{16}$

y = 276.8 days

(c) Supposed there are 60 pregnant women getting examined. According to CLT, what the distribution of the mean length of their pregnancies?

(d) What is the probability that the mean duration of these pregnancies will be less than 260 days?

Question 3(25 Teenage Drivers, Chapter 15 Page 557)

An Insurance company checks police records on 582 accidents selected at random and notes that teenagers were at the wheel in 91 of them.

(a) Create a 95% confidence interval for the percentage of all auto accidents that involve teenage drivers.

(b) Explain what your interval means.

(c) Explain wat “95% confidence” means.

(d) A politician urging tighter restrictions on drivers’ license issued to teen says, “In one of every five auto accidents, a teenager is behind the wheel.” Does your confidence interval support or contradict this statement? Explain.