R TUTORIAL, #13: NORMAL APPROXIMATIONS TO BINOMIAL DISTRIBUTIONS

The (>) symbol indicates something that you will type in.
A bullet (•) indicates what the R program should output (and other comments).

FAIR COIN EXAMPLE (COUNT HEADS IN 100 FLIPS)

• We will obtain the table for $\text{Bin}\left(n = 100,\ p = \frac{1}{2}\right)$.

> Type: probs = dbinom(0:100, size=100, prob=1/2)
> Type: round(data.frame(0:100, probs), digits=5)
- We get the probabilities rounded to five decimal places.

• We can do a spike plot for the distribution.
> Type: plot(0:100, probs, type="h", xlim=c(0,100), ylim=c(0,.1))
> Type: points(0:100, probs, pch=16, cex=.5)

> Type: curve(dnorm(x, mean=50, sd=5), from=0, to=100, xlim = c(0, 100), ylim = c(0, 0.5), xlab = "x", add=T, col="blue")
- We superimpose the density curve of the normal approximation.
- 'add=T' adds the curve to the current plot.

• Find $P(45 \text{ heads} \le X \le 55 \text{ heads})$.
> Type: sum(dbinom(45:55, size=100, prob=1/2))

• What does the normal approximation (with continuity corrections) give us?
> Type: pnorm(55.5, mean=50, sd=5) - pnorm(44.5, mean=50, sd=5)

• What if we hadn't used continuity corrections?
> Type: pnorm(55, mean=50, sd=5) - pnorm(45, mean=50, sd=5)

• Find $P(X \le 45 \text{ heads})$.
> Type: sum(dbinom(0:45, size=100, prob=1/2))

• What does the normal approximation (with continuity corrections) give us?
> Type: pnorm(45.5, mean=50, sd=5)

• Find $P(X > 55 \text{ heads})$.
> Type: sum(dbinom(56:100, size=100, prob=1/2))

• What does the normal approximation (with continuity corrections) give us?
> Type: 1 - pnorm(55.5, mean=50, sd=5)

<u>WHY SHOULD WE USE CONTINUITY CORRECTIONS?</u>

• This is best illustrated by the distribution $\text{Bin}\left(n = 10, p = \frac{1}{2}\right)$, which is the "simplest" binomial distribution that is eligible for a normal approximation.
> Type: probs2 = dbinom(0:10, size=10, prob=1/2)

• Let's do a probability histogram for this distribution.
• The 'barplot' command better suits our purposes.
> Type: barplot(probs2, names.arg=c(0:10), space=0, xlim=c(0,10), ylim=c(0,0.3))
    • 'space=0' ensures that we do not have any spaces between our bars. Otherwise, our coordinate system get messed up.

• Let's superimpose the approximating normal density curve.
> Type: curve(dnorm((x-0.5), mean=5, sd=sqrt(2.5)), from=0, to=11, xlim = c(0, 11), add=T, col="blue")
    • The '(x-0.5)' indicates that we have to shift the curve horizontally by 0.5 units to accommodate the barplot.

• Find $P(3 \le X \le 7)$.
> Type: sum(dbinom(3:7, size=10, prob=1/2))

• What does the normal approximation (with continuity corrections) give us?
> Type: pnorm(7.5, mean=5, sd=sqrt(2.5)) - pnorm(2.5, mean=5, sd=sqrt(2.5))

• What if we hadn't used continuity corrections?
> Type: pnorm(7, mean=5, sd=sqrt(2.5)) - pnorm(3, mean=5, sd=sqrt(2.5))
    • Do you see from the plot why we are getting an underestimate?