Lecture 11

Normal approximation to the binomial (Practice)

Go to https://gallery.shinyapps.io/dist_calc/ and choose Binomial coin experiment in the drop down menu on the left.

- Set the number of trials to 20 and the probability of success to 0.15. Describe the shape of the distribution of number of successes.
- Keeping p constant at 0.15, determine the minimum sample size required to obtain a unimodal and symmetric distribution of number of successes. Please submit only one response per team.
- Further considerations:
  - What happens to the shape of the distribution as n stays constant and p changes?
  - What happens to the shape of the distribution as p stays constant and n changes?

An analysis of Facebook users

A recent study found that "Facebook users get more than they give". For example:

- 40% of Facebook users in our sample made a friend request, but 63% received at least one request
- Users in our sample pressed the like button next to friends' content an average of 14 times, but had their content "liked" an average of 20 times
- Users sent 9 personal messages, but received 12
- 12% of users tagged a friend in a photo, but 35% were themselves tagged in a photo

Any guesses for how this pattern can be explained?

An analysis of Facebook users

A recent study found that ``Facebook users get more than they give". For example:

- 40% of Facebook users in our sample made a friend request, but 63% received at least one request
- Users in our sample pressed the like button next to friends' content an average of 14 times, but had their content "liked" an average of 20 times
- Users sent 9 personal messages, but received 12
- 12% of users tagged a friend in a photo, but 35% were themselves tagged in a photo

Any guesses for how this pattern can be explained?

Power users contribute much more content than the typical user.

This study also found that approximately 25% of Facebook users are considered power users. The same study found that the average Facebook user has 245 friends. What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users? Note any assumptions you must make.

We are given that $n = 245, p = 0.25$, and we are asked for the probability $P(K \geq 70)$. To proceed, we need independence, which we'll assume but could check if we had access to more Facebook data.

We are given that $n = 245, p = 0.25$, and we are asked for the probability $P(K \geq 70)$. To proceed, we need independence, which we'll assume but could check if we had access to more Facebook data.

$$\begin{aligned} P(X \geq 70) &= P(K = 70 \text{ or } K = 71 \text{ or } K = 72 \text{ or } \cdots \text{ or } K = 245) \\ &= P(K = 70) + P(K = 71) + P(K = 72) + \cdots + P(K = 245) \end{aligned}$$

This seems like an awful lot of work...

Could Use R ...

```
sum(dbinom(x = 70:245, size = 245, prob = 0.25))
```

```
## [1] 0.112763
```

```
pbinom(q = 69, size = 245, prob = 0.25, lower.tail = FALSE)
```

```
## [1] 0.112763
```

Normal approximation to the binomial

When the sample size is large enough, the binomial distribution with parameters n and p can be approximated by the normal model with parameters $\mu = np$ and $\sigma = \sqrt{np(1-p)}$.

- In the case of the Facebook power users, $n = 245$ and $p = 0.25$.

$$\mu = 245 \times 0.25 = 61.25 \qquad \sigma = \sqrt{245 \times 0.25 \times 0.75} = 6.778$$

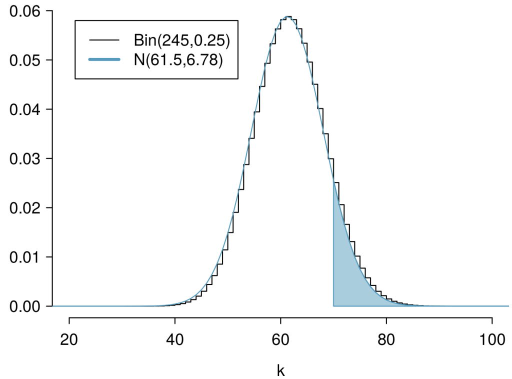
- $Bin(n = 245, p = 0.25) \approx N(\mu = 61.25, \sigma = 6.778)$.

Normal approximation to the binomial

- In the case of the Facebook power users, $n = 245$ and $p = 0.25$.

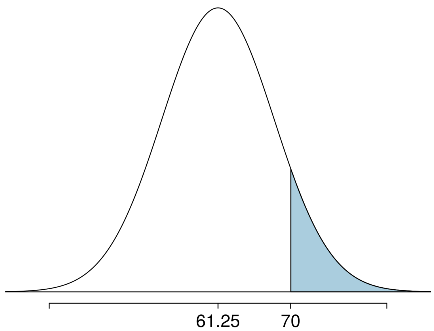
$$\mu = 245 \times 0.25 = 61.25 \qquad \sigma = \sqrt{245 \times 0.25 \times 0.75} = 6.778$$

- $Bin(n = 245, p = 0.25) \approx N(\mu = 61.25, \sigma = 6.778)$.



Computing the Approximation

What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?



$$
\begin{aligned}
Z &= \frac{obs - mean}{SD} = \frac{70 - 61.25}{6.778} \\
&= 1.29094 \\
P(Z > 1.29094) &= 1 - 0.90164 \\
&= 0.09836
\end{aligned}
$$

Computing the Approximation

What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?

But where did this $P(Z > 1.29)$ answer come from? R again!

$$
\begin{aligned}
Z &= \frac{obs - mean}{SD} = \frac{70 - 61.25}{6.778} \\
&= 1.29094 \\
P(Z > 1.29094) &= 1 - 0.90164 \\
&= 0.09836
\end{aligned}
$$

Computing Normal Probabilities

Just like we did for **pbinom()** and **dbinom()**, we can do for **pnorm()** and **dnorm()**. You saw this in workshop last week.

```
pnorm(1.290964, lower.tail = FALSE)
```

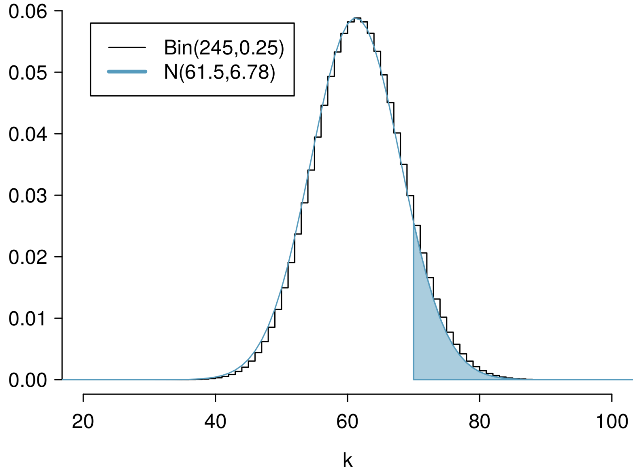
```
## [1] 0.09835808
```

This seems … bad

We know the exact probability, done using **pbinom()**, is 0.1128. So why is this "approximation" giving an answer of 0.09836?

This seems … bad

We know the exact probability, done using **pbinom()**, is 0.1128. So why is this "approximation" giving an answer of 0.09836?



"Correction for Continuity"

The normal approximation to the binomial can be a little rough. There is a **correction for continuity** which can be used instead:

- The cutoff values for the lower end of a shaded region should be reduced by 0.5
- The cutoff values for the upper end should be increased by 0.5.

Since we are doing a "greater than" probability, the lower end of the shaded region is our relevant object, so we reduce.

Computing Normal Probabilities

$$
\begin{aligned}
Z &= \frac{obs - mean}{SD} = \frac{(70 - 0.5) - 61.25}{6.778} \\
&= 1.217173 \\
P(Z > 1.217173) &= 1 - 0.8882308 \\
&= 0.1117692
\end{aligned}
$$

```
pnorm(1.217173, lower.tail = FALSE)
```

```
## [1] 0.1117692
```

That's much better!

Examples

Example 1

Suppose we toss a fair coin 20 times. What is the probability of getting between 9 and 11 heads?

Example 2

In a particular program at Trent, 60% of students are men and 40% are women. In a random sample of 50 students what is the probability that more than half are women?

Example 3

"Skip the Dishes" finds that 70% of people order through them and ask for Chinese food, and 30% for Italian food. Last week 426 orders were made through the local, Peterborough, unit. Find the probability that at least 200 orders were for Italian food.