Hypothesis Testing - First Examples

M&Ms Example

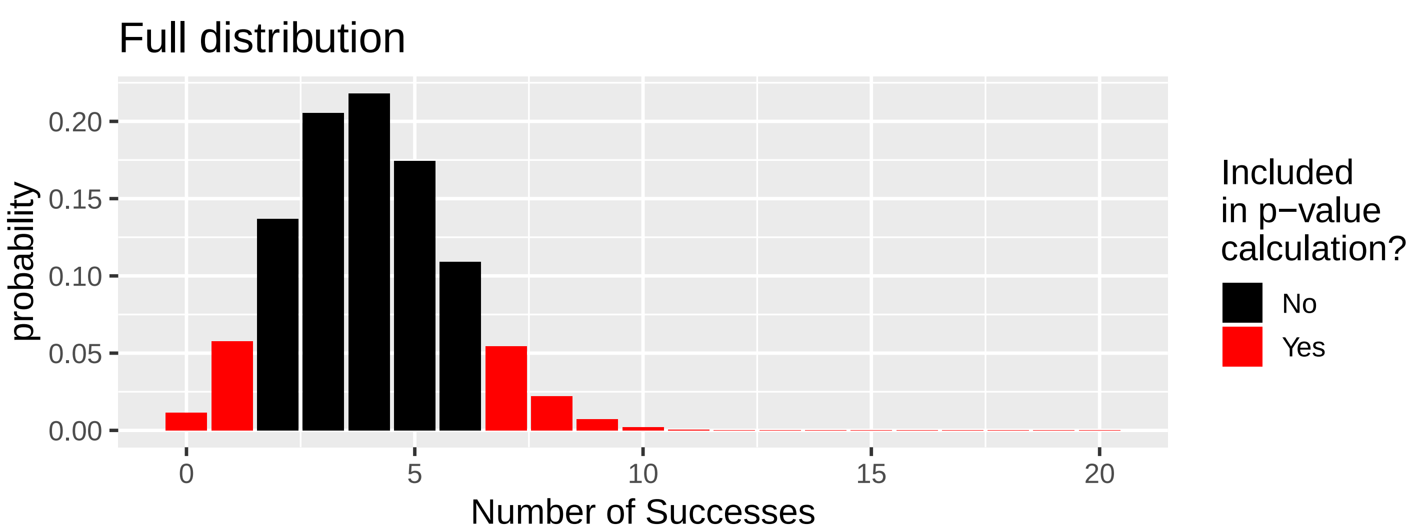
Part 1: Simple Vs. Composite Hypotheses

Sample size $n = 20$, observed $x = 7$ blue M&Ms

- Our model is $X \sim \text{Binomial}(20, \theta)$
- Our hypotheses are: $H_0 : \theta = 0.2$ and $H_A : \theta \neq 0.2$
- We can use the statistic $W = g(X) = X$ (g is the identity function) to test this hypothesis
- If H_0 is true, then $X \sim \text{Binomial}(20, 0.2)$
- The p-value is $P(X \text{ at least as extreme as } 7)$ given that $X \sim \text{Binomial}(20, 0.2)$
  - $E(X) = 20 * 0.2 = 4$
  - $7 - 4 = 3$
  - $4 - 3 = 1$
  - $P(X \text{ at least as extreme as } 7 | \theta = 0.2) = P(X \leq 1 \text{ or } X \geq 7 | \theta = 0.2)$

```
pbinom(q = 1, size = 20, prob = 0.2) +
  pbinom(q = 7, size = 20, prob = 0.2, lower.tail = FALSE)
```

```
## [1] 0.101318
```



Sample size $n = 541$, observed $x = 138$ blue M&Ms

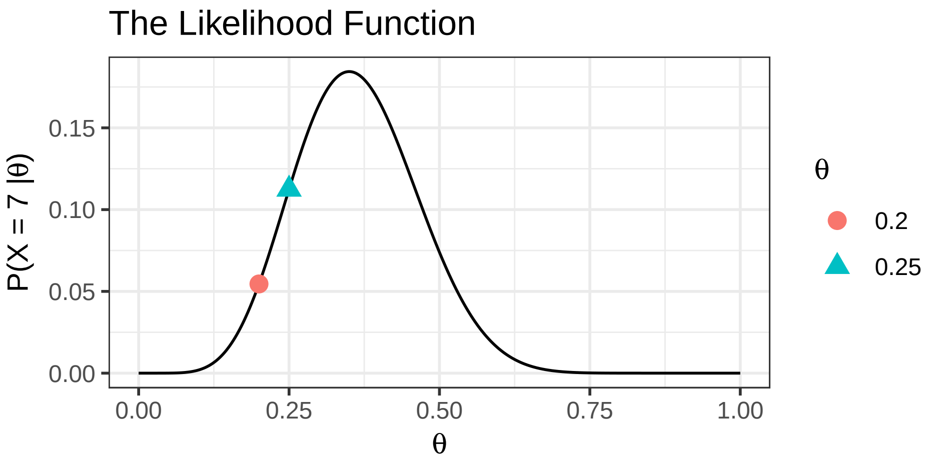
- If H_0 is true, then $X \sim \text{Binomial}(541, 0.2)$
- The p-value is $P(X \text{ at least as extreme as } 138)$ given that $X \sim \text{Binomial}(541, 0.2)$
  - $E(X) = 541 * 0.2 = 108.2$
  - $138 - 108.2 = 29.8$
  - $108.2 - 29.8 = 78.4$
  - $P(X \text{ at least as extreme as } 138) = P(X \leq 78 \text{ or } X \geq 138)$

```
pbinom(q = 78, size = 541, prob = 0.2) +
  pbinom(q = 138, size = 541, prob = 0.2, lower.tail = FALSE)
```

```
## [1] 0.001244054
```

Part 2: Simple Hypotheses

- I heard a rumor that the proportion of M&Ms that are blue was changed to 25. Is it true??
- Our hypotheses are: $H_0 : \theta = 0.2$ and $H_A : \theta = 0.25$
- We can use the statistic $W = g(X) = \frac{f_X(X|\theta=0.2)}{f_X(X|\theta=0.25)}$ (the **likelihood ratio**)




```
w_obs <- dbinom(x = 7, size = 20, prob = 0.2)/dbinom(x = 7, size = 20, prob = 0.25)
w_obs
```

```
## [1] 0.4852922
```

Questions:

1. Is a small likelihood ratio or a large likelihood ratio stronger evidence against the null hypothesis?

2. What should count as "at least as extreme" for the purpose of calculating a p-value based on the likelihood ratio, W?

- If H_0 is true, then $X \sim \text{Binomial}(20, 0.2)$
- The p-value is $P(W \leq w)$ given that $X \sim \text{Binomial}(20, 0.2)$

```
# Manual calculation of the probability distribution of W
x <- seq(from = 0, to = 20)
W_X_distn <- data.frame(
  x = x,
  probability = dbinom(x, size = 20, prob = 0.2),
  w = dbinom(x, size = 20, prob = 0.2) / dbinom(x, size = 20, prob = 0.25)
)

W_X_distn
```

```
##     x  probability          w
## 1   0 1.152922e-02 3.63558642
## 2   1 5.764608e-02 2.72668981
## 3   2 1.369094e-01 2.04501736
## 4   3 2.053641e-01 1.53376302
## 5   4 2.181994e-01 1.15032226
## 6   5 1.745595e-01 0.86274170
## 7   6 1.090997e-01 0.64705627
## 8   7 5.454985e-02 0.48529221
## 9   8 2.216088e-02 0.36396915
## 10  9 7.386959e-03 0.27297687
## 11 10 2.031414e-03 0.20473265
## 12 11 4.616849e-04 0.15354949
## 13 12 8.656592e-05 0.11516212
## 14 13 1.331783e-05 0.08637159
## 15 14 1.664729e-06 0.06477869
## 16 15 1.664729e-07 0.04858402
## 17 16 1.300570e-08 0.03643801
## 18 17 7.650410e-10 0.02732851
## 19 18 3.187671e-11 0.02049638
## 20 19 8.388608e-13 0.01537229
## 21 20 1.048576e-14 0.01152922
```

```
# Find the p-value
sum(W_X_distn$probability[W_X_distn$w <= w_obs])
```

```
## [1] 0.08669251
```

- Note: in this example, using X or $W = g(X) = \frac{f_X(X|\theta=0.2)}{f_X(X|\theta=0.25)}$ is *equivalent*
  - A larger value of X is more consistent with H_A than H_0
  - A smaller value of W is more consistent with H_0 than H_A

```
## [1] 0.08669251
```

Another example

We have 2 batches of paint, one of which is quick-dry; quick-dry paint will dry in an average of 10 minutes, and regular paint in an average of 25 minutes. The paint is unlabeled, and we forgot which was which! We paint 5 boards from batch 1 and record the drying time. We believe that the drying times are normally distributed with a standard deviation of 5 minutes for both paint types.

We have a vague feeling that batch 1 is quick-dry; we will calculate a p-value to see how strong the evidence is that the batch we used is *not* 25 minutes.

$H_0 : \theta = 25$

$H_A : \theta = 10$

Find the form of the likelihood ratio statistic in terms of $\bar{x}$, simplified as much as possible. Is a small or large value of $\bar{x}$ evidence against H_0?

Note that if $X \sim \text{Normal}(\theta, 5^2)$ then $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\frac{(x_i-\theta)^2}{\sigma^2}\right]$

Suppose we observe $\bar{x} = 20$.

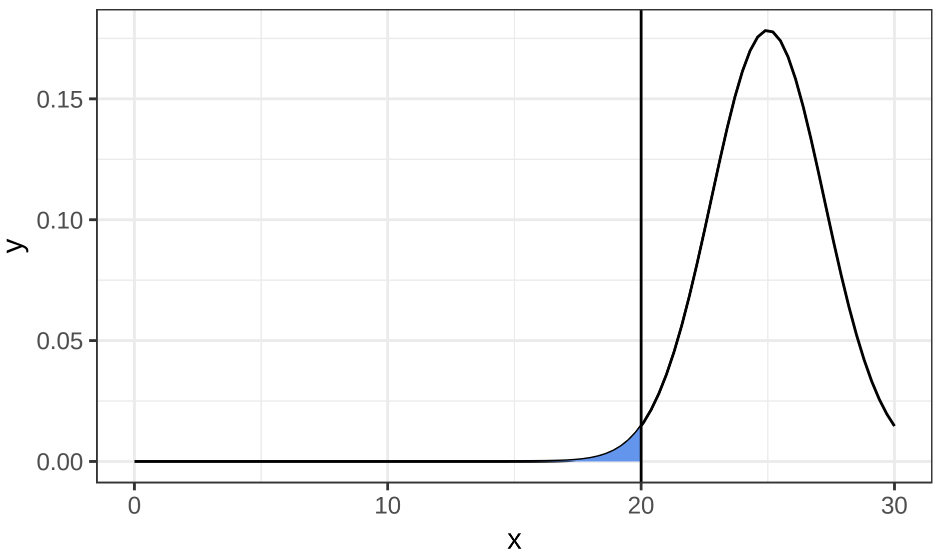
If H_0 is true, then $\bar{X} \sim \text{Normal}(25, 5^2/5)$

An "extreme" (small) value of W corresponds to a small value of $\bar{X}$.

```
x_grid_1 <- seq(from = 0, to = 20, length = 101)

region_to_shade1 <- data.frame(
  x = c(0, x_grid_1, 20),
  y = c(0, dnorm(x_grid_1, mean = 25, sd = sqrt(5), log = FALSE), 0)
)

ggplot(data = data.frame(x = c(0, 30)), mapping = aes(x = x)) +
  stat_function(fun = dnorm, args = list(mean = 25, sd = sqrt(5))) +
  geom_polygon(
    mapping = aes(x = x, y = y),
    fill = "cornflowerblue",
    data = region_to_shade1) +
  geom_vline(xintercept = 20) +
  coord_cartesian(xlim = c(0, 30), expand = TRUE) +
  theme_bw()
```



p-value:

```
pnorm(20, mean = 25, sd = 5)
```

```
## [1] 0.1586553
```