

Unit 8: Final Review

1. Bayesian vs. frequentist inference

Sta 101 - Spring 2016

Duke University, Department of Statistical Science

Dr. Çetinkaya-Rundel

Slides posted at http://bit.ly/sta101_s16

► Project questions?

1

M&Ms

- We have a population of M&Ms. The percentage of yellow M&Ms is either 10% or 20%.
- You have been hired as a statistical consultant to decide whether the true percentage of yellow M&Ms is 10%. You are being asked to make a decision, and there are associated payoff/losses that you should consider.

2

Decision table

Decision	True state of the population	
	% yellow = 10%	% yellow = 20%
% yellow = 10%	<i>Your boss gives you a bonus, and I bring you candy on Monday</i>	<i>You lose your job, and no candy for you</i>
%yellow = 20%	<i>You lose your job, and no candy for you</i>	<i>Your boss gives you a bonus, and I bring you candy on Monday</i>

3

- ▶ I will show you a random sample from the population, but you pay \$200 for each M&M, and you must buy in \$1000 increments.
- ▶ That is, you may buy 5, 10, 15, or 20 M&Ms.

4

Application exercise: Set up -- data

How many M&Ms would you buy? Decide as a team and vote.

- (a) 5 (b) 10 (c) 15 (d) 20

Application exercise: Set up -- significance level

Then, discuss at what significance level you will reject the null hypothesis.

6

- ▶ Hypotheses:
 - H_0 : 10% yellow M&Ms
 - H_A : more than 10% yellow M&Ms
- ▶ Your test statistic is the number of yellow M&Ms you observe in the sample.
- ▶ The p-value will be the probability of observing this many or more yellow M&Ms given the null hypothesis is true.

5

Now we will take a sequence of M&Ms, and you record the number of yellows in the first n draws.

RGYBO BBGOY YRBRR GORBY

Application exercise: Frequentist inference

- ▶ Number of yellows in the first n draws = _____ = k
- ▶ Calculate the p-value using the Binomial distribution:
p-value = $P(k \text{ or more yellows} \mid n, \% \text{yellow is } 10\%) = \text{_____}$
- ▶ Do you reject the null hypothesis? _____

See next slide for hints...

7

Remember $Binomial(n, k) = \binom{n}{k} p^k (1-p)^{(n-k)}$

Say you picked $n = 5$, and hence $k = 1$.

$$\begin{aligned} & P(1 \text{ or more yellows} \mid n = 5, \% \text{yellow is } 10\%) \\ &= P(K = 1) + P(K = 2) + P(K = 3) + P(K = 4) + P(K = 5) \\ &= \left[\binom{5}{1} 0.1^1 \times 0.9^4 \right] + \left[\binom{5}{2} 0.1^2 \times 0.9^3 \right] + \dots + \left[\binom{5}{5} 0.1^5 \times 0.9^0 \right] \\ &= 0.32805 + 0.0729 + 0.0081 + 0.00045 + 0.00001 \\ &\approx 0.41 \end{aligned}$$

Alternatively:

```
# P(K = 1, n = 5, p = 0.1)
dbinom(1, 5, 0.1)
```

0.32805

... and

```
# P(K >= 1, n = 5, p = 0.1)
sum(dbinom(1:5, 5, 0.1))
```

0.40951

8

9

Bayesian inference

Now we will start over. Start with 1:1 odds that the percentage of yellows is 10%:20%.

- ▶ H_1 : 10% yellow M&Ms $\rightarrow P(10\% \text{ yellow}) = 0.5$
- ▶ H_2 : 20% yellow M&Ms $\rightarrow P(20\% \text{ yellow}) = 0.5$

Application exercise: Bayesian inference

Using the same data and Bayes' theorem to calculate the probability the percentage of yellow is 10% and 20% given the observed data.

See next slide for hints...

$P(10\% \text{ yellow} \mid \text{data})$:

$$\begin{aligned} P(10\% \text{ yellow} \mid \text{data}) &= \frac{P(\text{data} \mid 10\% \text{ yellow}) \times P(10\% \text{ yellow})}{P(\text{data})} \\ &= \frac{P(\text{data} \mid 10\% \text{ yellow}) \times P(10\% \text{ yellow})}{P(\text{data} \mid 10\% \text{ yellow}) \times P(10\% \text{ yellow}) + P(\text{data} \mid 20\% \text{ yellow}) \times P(20\% \text{ yellow})} \\ &= \frac{0.33 \times 0.5}{0.33 \times 0.5 + 0.41 \times 0.5} \\ &= 0.44 \end{aligned}$$

10

11

Application exercise: Bayesian vs. Frequentist inference

Regardless of the choices you made earlier about n , fill out the table below for all possible choices of n and the resulting k .

Number of yellow M&Ms in first	Frequentist: p-value		Bayesian: Posterior	
	$P(K \geq k \mid 10\% \text{ yellow})$	Decision	$P(10\% \text{ yellow} \mid n, k)$	$P(20\% \text{ yellow} \mid n, k)$
$n = 5 : k = 1$	0.41	Fail to reject H_0	0.44	
$n = 10 : k = 2$				
$n = 15 : k = 3$				
$n = 20 : k = 4$				

- We know that the true % yellow in these data is 20%:

RGYBO BBGOY YRBRR GORBY

- However the Frequentist approach (using p-values) would not allow us to reject the null hypothesis of 10% yellow.
- On the other hand, the Bayesian approach yields a higher posterior probability for 20% yellow.