

# Conceptueel inzicht in hypothese- toetsen opbouwen met simulaties

Aan de slag met het specifieke minimumdoel rond  
hypothesetoetsen

Kortrijk, 18/11/23

Johan Deprez

slides op mijn website van KU Leuven (zoek via wie-is-wie)

# De beginsituatie

# Jullie (!) beginsituatie

- Wat weten jullie zelf (nog) over het toetsen van hypothesen?  
wij zullen het vandaag beperken tot hypothese over één proportie  
voorlopig enkel terminologie!
- Wat moet in deze brainstorm zeker aan bod komen?
  - nulhypothese en alternatieve hypothese
  - populatie en steekproef
  - populatieparameter en steekproefstatistiek
  - significantieniveau  $\alpha$  (vaak 0,05)
  - eventueel betrouwbaarheidsniveau (is dan 95%)
  - éénzijdig of tweezijdig toetsen
  - al dan niet verwerpen van een hypothese
  - type-I-fout en type-II-fout

# De specifieke minimumdoelen

- De leerlingen toetsen hypothesen.
  - Onderliggende (kennis)elementen: Nulhypothese, alternatieve hypothese, p-waarde, significantieniveau, steekproevenverdeling
  - Statistiek, Gevorderde wiskunde
  - voor studierichtingen Humane wetenschappen, Wiskunde-wetenschappen, Technologische wetenschappen en engineering, ...
- De leerlingen leggen in betekenisvolle situaties de betekenis uit van nulhypothese, alternatieve hypothese, significantieniveau en p-waarde.
  - Onderliggende (kennis)elementen: Steekproevenverdeling
  - Pakket uit statistiek, Uitgebreide wiskunde in functie van wetenschappen/economie, Uitgebreide wiskunde
  - voor studierichtingen Welzijnswetenschappen, Latijn-wetenschappen, Mechatronica, Economie-moderne talen, Bedrijfswetenschappen, ...

# Ons voorbeeld


# Ons voorbeeld

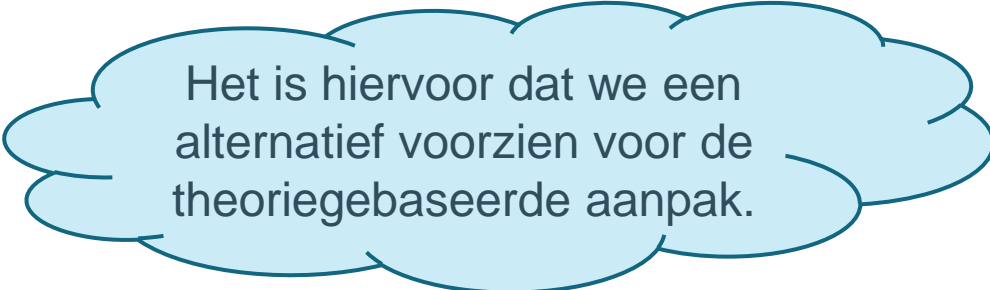
Ik vermoed van een dobbelsteen dat hij onzuiver is, meer bepaald dat de kans op een 6 kleiner is dan  $1/6$ . Hoe kan ik onderzoeken of ik gelijk heb?

- populatie = alle mogelijke worpen met deze dobbelsteen
- nulhypothese
  - dobbelsteen is zuiver (kans op een 6 is gelijk aan  $1/6$ )  
dus het tegendeel van mijn vermoeden!
  - nulhypothese: het 'effect' dat we willen aantonen, is niet aanwezig  
we proberen aan te tonen dat deze hypothese niet houdbaar is
  - hypothese heeft betrekking op een populatieparameter (proportie is  $1/6$ )
- alternatieve hypothese
  - éénzijdig! kans op een 6 is strikt kleiner dan  $1/6$

# Ons voorbeeld

Ik vermoed van een dobbelsteen dat hij onzuiver is, meer bepaald dat de kans op een 6 kleiner is dan  $1/6$ . Hoe kan ik onderzoeken of ik gelijk heb?

- populatie = alle mogelijke worpen met deze dobbelsteen
- nulhypothese: dobbelsteen is zuiver (kans op 6 is gelijk aan  $1/6$ )
- alternatieve hypothese: kans op een 6 is strikt kleiner dan  $1/6$
- werkwijze:
  - neem een (één!) steekproef  
bv. gooi 120 keer, 15 keer 6  
steekproefstatistiek:  $15/120 = 1/8$
  - bereken p-waarde 
  - vergelijk met significantieniveau om te besluiten of je de nulhypothese al dan niet zult verwerpen



Het is hiervoor dat we een alternatief voorzien voor de theoriegebaseerde aanpak.

# Kennismaking met de applet

# Samenvatting

Ik vermoed van een dobbelsteen dat hij onzuiver is, meer bepaald dat de kans op een 6 kleiner is dan  $1/6$ . Hoe kan ik onderzoeken of ik gelijk heb?

- populatie = alle mogelijke worpen met deze dobbelsteen
- nulhypothese: dobbelsteen is zuiver (kans op 6 is gelijk aan  $1/6$ )
- één steekproef: gooi 120 keer, 15 keer 6, steekproefstatistiek:  $15/120 = 1/8$

# De applet

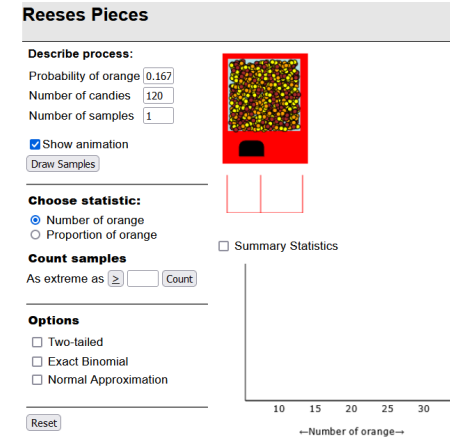
- [www.rossmanchance.com](http://www.rossmanchance.com)

> Applet Collection > Reeses Pieces

- ik bedien de applet terwijl we een onderwijsleergesprek houden, maar je mag meedoen als je wil

valkuil: als er iets misloopt, ben je daar misschien teveel op gefocust en let je niet meer op wat er gezegd wordt

- we bootsen het gooien van de dobbelsteen na
  - op het eerste gezicht: door het draaien van een snoepje
  - oranje snoepje draaien = een 6 gooien
  - in feite: via een (pseudo)randomgenerator op een computer
  - nabootsen = simuleren



# De applet

Je kan instellen:

- percentage oranje snoepjes ( $p$ ,  $1/6 \approx 0.1667$ )
- aantal snoepjes dat je draait (number of candies, 120)
- aantal keer dat je een aantal snoepjes draait (number of samples, voorlopig 1)
- animatie aan/uit
- aantal of proportie
- (en nog wat andere dingen, die we pas later nodig hebben)

## Reeses Pieces

### Describe process:

Probability of orange

Number of candies

Number of samples

Show animation

### Choose statistic:

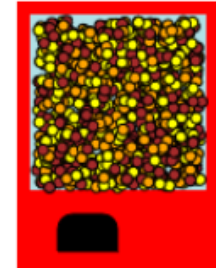
- Number of orange  
 Proportion of orange

### Count samples

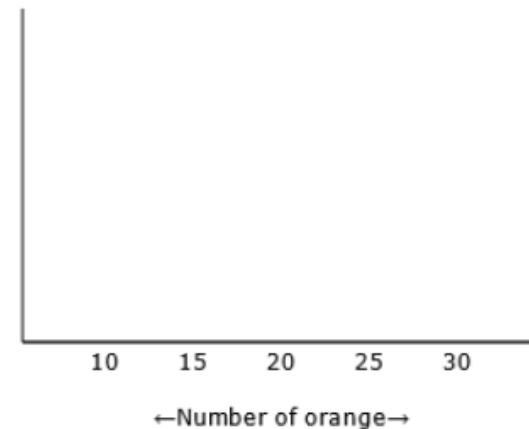
As extreme as

### Options

- Two-tailed  
 Exact Binomial  
 Normal Approximation



Summary Statistics



# De applet

## De applet

- geeft het aantal oranje snoepjes in de steekproef (hier: 16, tel in de figuur)
- geeft het percentage oranje snoepjes in de steekproef ( $p^\wedge$ , hier 0.133)
- duidt dit aan onder de vorm van een puntje op de grafiek
- je kunt voor de tweede en derde bullet ook kiezen voor 'aantal' i.p.v. 'proportie'

## Reeses Pieces

### Describe process:

Probability of orange

Number of candies

Number of samples

Show animation

Total Samples = 1

### Choose statistic:

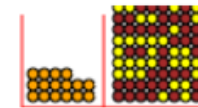
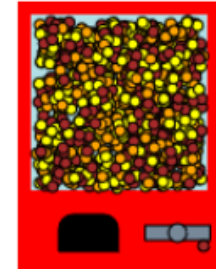
- Number of orange  
 Proportion of orange

### Count samples

As extreme as

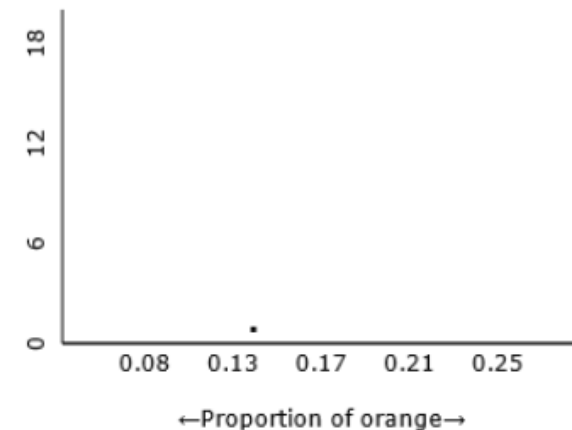
### Options

- Two-tailed  
 Exact Binomial  
 Normal Approximation



Most recent  $\hat{p} = 0.133$

Summary Statistics



De applet

We simuleren gedrag van  
een

**zuivere**

**dobbelsteen!**

(probability of orange =  $1/6$ )

# De applet

- Simulatie
  - we onderzoeken wat er gebeurt met de proportie 6-en als we een reeks van 120 worpen doen met een dobbelsteen die aan de nulhypothese voldoet
  - we zullen hiervoor heel veel reeksen van 120 worpen simuleren
    - in de applet: 'number of samples' (sample = steekproef)
    - in het Engels soms 'trial' i.p.v. 'sample', hier gebruiken we: 'reeks'
  - (als het ware) gedachtenexperiment
- Ons onderzoek
  - de dobbelsteen voldoet wel of niet aan de nulhypothese
  - we voeren maar één reeks van 120 worpen uit (steekproef)
  - het gaat over de echte dobbelsteen en over een echt uitgevoerde reeks van 120 worpen

# Simuleren van reeksen van 120 werpen met een zuivere dobbelsteen

# Niet één, maar enkele reeksen

Wat verwacht je te zien als we **50** reeksen van 120 worpen met een zuivere dobbelsteen simuleren?

- Elke reeks levert een aantal of proportie op. Hierbij treden verschillende aantallen en proporties op.
- **Steekproevenvariabiliteit**
- Let op: het gaat niet over variabiliteit van de kleuren binnen één reeks (bv. oranje, geel, geel, bruin, ...), maar over variabiliteit van de 50 proporties (één per reeks) over de reeksen heen (bv. 21/120, 16/120, 18/120, ...)

## Reeses Pieces

### Describe process:

Probability of orange

Number of candies

Number of samples

Show animation

Total Samples = 50

### Choose statistic:

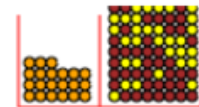
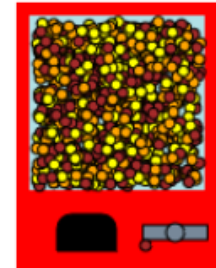
- Number of orange  
 Proportion of orange

### Count samples

As extreme as

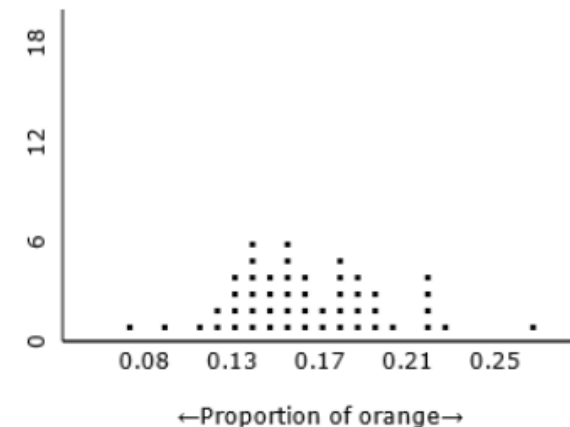
### Options

- Two-tailed  
 Exact Binomial  
 Normal Approximation



Most recent number of orange candies = 21

Summary Statistics



# Niet enkele, maar heel veel reeksen

Wat verwacht je te zien als we **5000** reeksen van 120 worpen met een zuivere dobbelsteen simuleren?

- Nu niet enkel 'chaos' (steekproevenvariabiliteit), maar ook regelmaat!
- veel proporties liggen vrij dicht bij  $1/6$
- waarden die verder van  $1/6$  afwijken komen ook voor maar minder vaak
- gemiddelde van de 5000 proporties is bij benadering  $1/6$
- ...
- **Steekproevenverdeling**

## Reeses Pieces

### Describe process:

Probability of orange

Number of candies

Number of samples

Show animation

Total Samples = 5000

### Choose statistic:

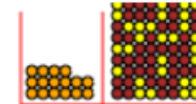
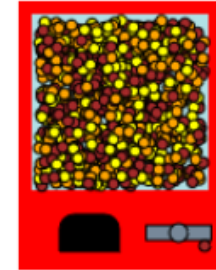
- Number of orange  
 Proportion of orange

### Count samples

As extreme as

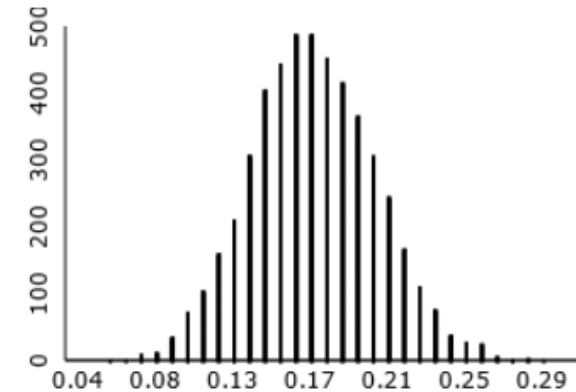
### Options

- Two-tailed  
 Exact Binomial  
 Normal Approximation



Most recent  $\hat{p} = 0.133$

Summary Statistics



←Proportion of orange→

# Niet enkele, maar heel veel reeksen

Wat verwacht je te zien als we **5000** reeksen van 120 worpen met een zuivere dobbelsteen simuleren?

- **Steekproevenverdeling**
- wij weten, maar voor leerlingen nu nog niet belangrijk: voor 'oneindig veel' reeksen
  - binomiale verdeling met  $n = 120$  en  $p = 1/6$
  - bij benadering normale verdeling met verwachtingswaarde  $1/6$  en standaardafwijking  $\sqrt{\frac{1}{6} \cdot \frac{5}{6} / 120}$

## Reeses Pieces

### Describe process:

Probability of orange

Number of candies

Number of samples

Show animation

Total Samples = 5000

### Choose statistic:

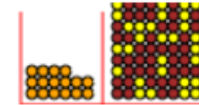
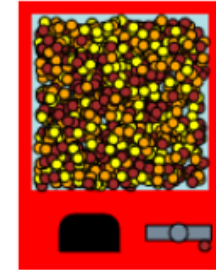
- Number of orange  
 Proportion of orange

### Count samples

As extreme as

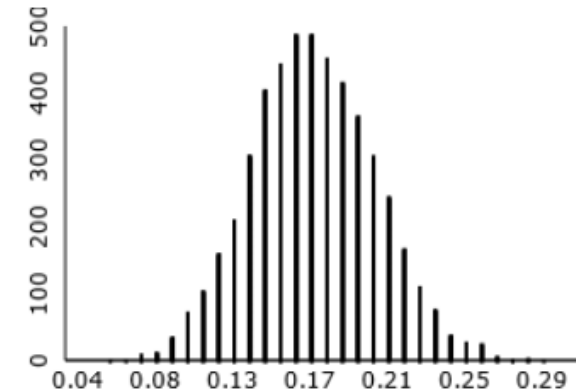
### Options

- Two-tailed  
 Exact Binomial  
 Normal Approximation



Most recent  $\hat{p} = 0.133$

Summary Statistics



←Proportion of orange→

# Niet enkele, maar heel veel reeksen

Wat verwacht je te zien als we **5000** reeksen van 120 worpen met een zuivere dobbelsteen simuleren?

- **Steekproevenverdeling**
- Let op: het gaat niet over verdeling van de kleuren binnen één reeks (bv. oranje, geel, geel, bruin, ...), maar over verdeling van de 5000 (of eigenlijk: 'oneindig veel') proporties (één per reeks) over de reeksen heen (bv. 21/120, 16/120, 18/120, ...)
- o.a. daarom gebruik je beter 'steekproevenverdeling' dan 'steekproeffverdeling'

## Reeses Pieces

### Describe process:

Probability of orange

Number of candies

Number of samples

Show animation

Total Samples = 5000

### Choose statistic:

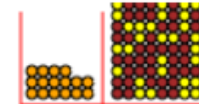
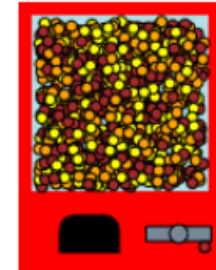
- Number of orange  
 Proportion of orange

### Count samples

As extreme as

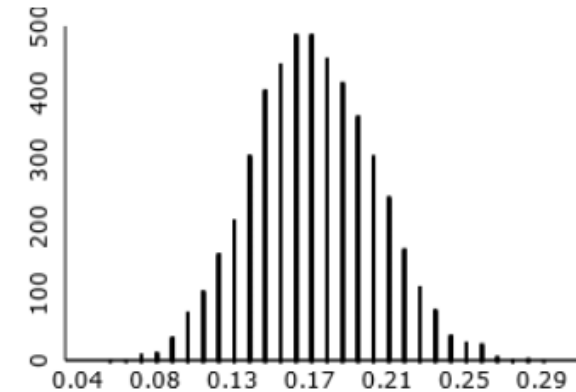
### Options

- Two-tailed  
 Exact Binomial  
 Normal Approximation



Most recent  $\hat{p} = 0.133$

Summary Statistics



←Proportion of orange→

# Niet enkele, maar heel veel reeksen

Wat verwacht je te zien als we **5000** reeksen van 120 worpen met een zuivere dobbelsteen simuleren?

- **Steekproevenverdeling**
- Let op: het gaat **niet** over het **herhalen van de steekproeftrekking in de echte wereld** (met een dobbelsteen waarvan we niet weten of die al dan niet zuiver is), maar **wel** over ons **gedachtenexperiment** waarmee we zicht willen krijgen op wat we kunnen verwachten voor de proportie 6-en in een reeks van 120 worpen met een zuivere dobbelsteen.

## Reeses Pieces

### Describe process:

Probability of orange

Number of candies

Number of samples

Show animation

Total Samples = 5000

### Choose statistic:

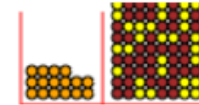
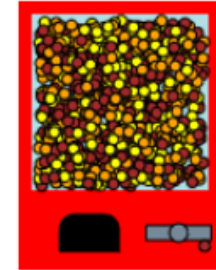
- Number of orange  
 Proportion of orange

### Count samples

As extreme as

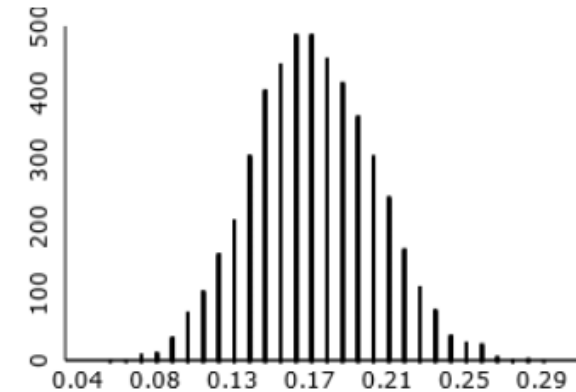
### Options

- Two-tailed  
 Exact Binomial  
 Normal Approximation



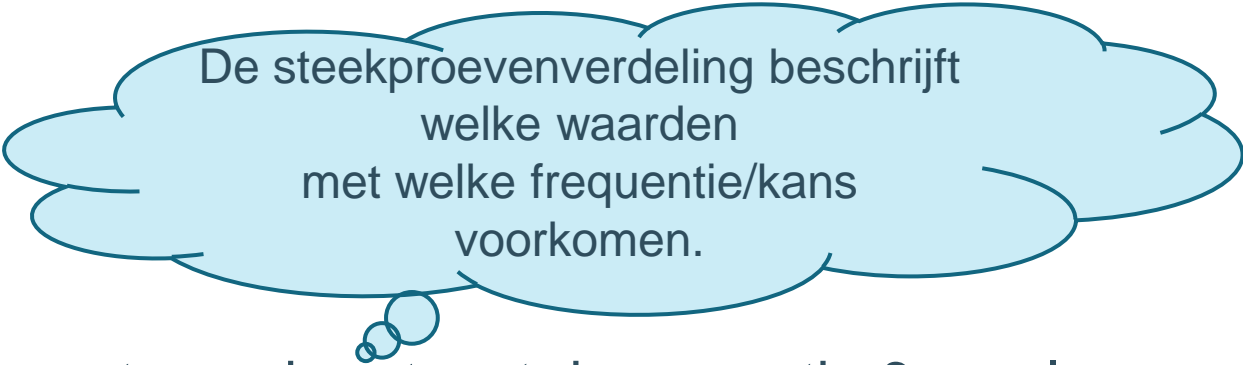
Most recent  $\hat{p} = 0.133$

Summary Statistics



←Proportion of orange→

# De applet



De steekproevenverdeling beschrijft welke waarden met welke frequentie/kans voorkomen.

- Simulatie

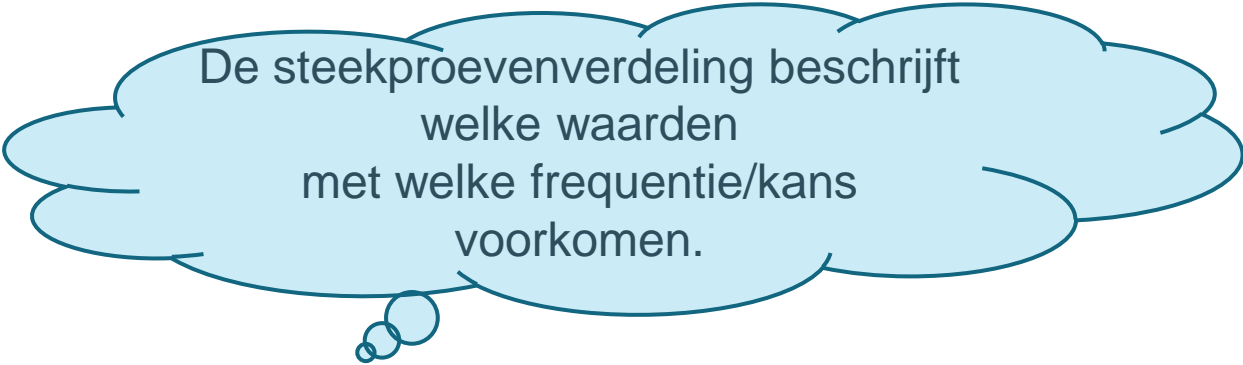
- we onderzoeken wat er gebeurt met de proportie 6-en als we een reeks van 120 worpen doen met een dobbelsteen die aan de nulhypothese voldoet
- we zullen hiervoor heel veel reeksen van 120 worpen simuleren in de applet: ‘number of samples’ (sample = steekproef) in het Engels soms ‘trial’ i.p.v. ‘sample’, hier gebruiken we: ‘reeks’
- (als het ware) gedachtenexperiment

- Ons onderzoek

- de dobbelsteen voldoet wel of niet aan de nulhypothese
- we voeren maar één reeks van 120 worpen uit (steekproef)
- het gaat over de echte dobbelsteen en over een echt uitgevoerde reeks van 120 worpen

Is de (echte) dobbelsteen zuiver of onzuiver?

# De applet



De steekproevenverdeling beschrijft welke waarden met welke frequentie/kans voorkomen.

- Simulatie
  - we onderzoeken wat er gebeurt met de proportie 6-en als we een reeks van 120 worpen doen met een dobbelsteen die aan de nulhypothese voldoet
- Ons onderzoek
  - de dobbelsteen voldoet wel of niet aan de nulhypothese
  - we voeren maar één reeks van 120 worpen uit (steekproef)
  - we vergelijken de proportie die uit onze ene steekproef komt met de steekproevenverdeling van een nulhypothese-dobbelsteen

# Steekproefproportie interpreteren

We vergelijken onze steekproefproportie (15 op 120,  $1/8$ ) met de steekproeven-verdeling van de nulhypothese-dobbelsteen.

- zie 'Count samples'
- let op overgang: hoogstens  $1/8$
- van de 5000 reeksen van 120 worpen met een nulhypothese-dobbelsteen geven er 611 (12,22%) een proportie 6-en die  $\leq 1/8$  is
- te weinig reden voor twijfel aan de zuiverheid v.d. dobbelsteen
- 12,22% is (via simulatie bepaalde) p-waarde

## Reeses Pieces

### Describe process:

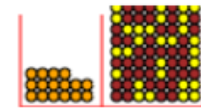
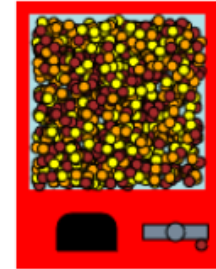
Probability of orange

Number of candies

Number of samples

Show animation

Total Samples = 5000



Most recent  $\hat{p} = 0.133$

### Choose statistic:

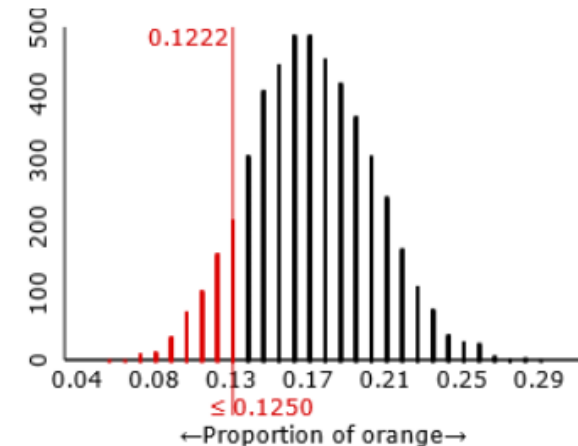
- Number of orange  
 Proportion of orange

### Count samples

As extreme as

Proportion of samples:  
 $611 / 5000 = 0.1222$

Summary Statistics



### Options

- Two-tailed  
 Exact Binomial  
 Normal Approximation

# Steekproefproportie interpreteren

Stel dat de steekproef niet **15**, maar slechts **12** keer 6 opleverde...

- steekproefproportie = 0,1
- 'experimenteel bepaalde p-waarde' is nu 2,68%
- grens tussen al dan niet verwerpen van de nulhypothese = significantieniveau  $\alpha$
- wordt vooraf bepaald, vaak 5%
- met dit significantieniveau nu wel voldoende evidentie om de nulhypothese te verwerpen

## Reeses Pieces

### Describe process:

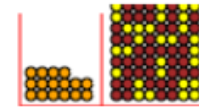
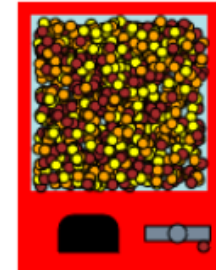
Probability of orange

Number of candies

Number of samples

Show animation

Total Samples = 5000



Most recent  $\hat{p} = 0.133$

### Choose statistic:

- Number of orange  
 Proportion of orange

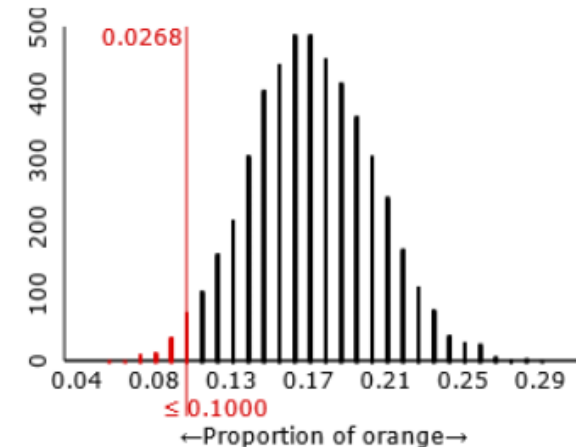
### Count samples

As extreme as

Proportion of samples:

134 / 5000 = 0.0268

Summary Statistics



### Options

- Two-tailed  
 Exact Binomial  
 Normal Approximation

# Zijn we zeker van onze conclusie?

- 12 keer 6 (proportie 0,1) in een steekproef van 120 worpen
  - we verwerpen de nulhypothese
  - nochtans zien we dat zo'n reeks van 120 worpen ook bij een zuivere dobbelsteen kan optreden, zij het met een kleine kans
  - we zijn dus niet zeker!
  - we kunnen een type-I-fout maken met kans =  $\alpha$
  - een kleine kans! → we geven de nulhypothese het voordeel van de twijfel
- 15 keer 6 (proportie 1/8) in een steekproef van 120 worpen
  - we verwerpen de nulhypothese niet
  - maar dit is geen hard bewijs dat de nulhypothese waar is
  - we zijn dus niet zeker!
  - we kunnen een type-II-fout maken
  - kans hierop is niet zonder meer te bepalen

# Zijn we zeker van onze conclusie?

- 12 keer 6 (proportie 0,1) in een steekproef van 120 worpen
  - we verwerpen de nulhypothese
  - nochtans zien we dat zo'n reeks van 120 worpen ook bij een zuivere dobbelsteen kan optreden, zij het met een kleine kans
  - we zijn dus niet zeker!
  - we kunnen een type-I-fout maken met kans = p-waarde
  - een kleine kans! → we geven de nulhypothese het voordeel van de twijfel
- 15 keer 6 (proportie 1/8) in een steekproef van 120 worpen
  - we verwerpen de nulhypothese niet
  - maar dit is geen hard bewijs dat de nulhypothese waar is
  - we zijn dus niet zeker!
  - we kunnen een type-II-fout maken
  - kans hierop is niet zonder meer te bepalen

# Slot

# De twee aanpakken naast elkaar

## Theoretische benadering

## Simulatiegebaseerde aanpak

Stap 1: Gegevens kan je zien als een steekproef van 40 willekeurige observaties. We observeren een steekproefproportie  $\hat{p} = 0,70$ .

Stap 2:

$$H_0: p = 0,50$$

$$H_a: p > 0,50$$

Stap 3: CLS + binomiaalverdeling

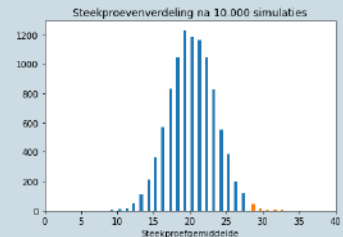
Indien  $H_0$  waar:  $\hat{p} \sim N\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$

$$\text{Of } z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$$

Stap 4: Wij observeren:  $z = \frac{0,75 - 0,50}{\sqrt{\frac{0,50 \cdot 0,50}{40}}} = 2,53$

$$P(z > 2,53) = 0,0057. \text{ (Tabel, GRM of software)}$$

Stap 3: Simuleer het 40 keer gooien van een eerlijk muntstuk met een script, applet, ... en tel aantal keer kop



Stap 4: empirische p-waarde

$$\text{p-waarde} = \frac{\#\text{simulaties} > 28}{10.000} = \frac{77}{10.000} = 0,0077$$

Stap 5: Op basis van zo'n kleine p-waarde hebben we evidentie om de nulhypothese te verwerpen ten voordele van de alternatieve hypothese. Als het muntstuk eerlijk was, zou 28 keer kop in een reeks van 40 zeer onwaarschijnlijk zijn.

slide overgenomen van studenten educatieve master die hun onderzoek hierrond uitvoeren (met een ander voorbeeld)

# Waarom nood aan een andere aanpak?

- In vakdidactisch onderzoek werden heel wat problemen vastgesteld bij studenten die enkel theoriegebaseerde aanpak kregen
- zie hiernaast + volgende slide
- weer overgenomen van mijn studenten!



## Multi-level perspectief

Perspectiefwissel: overgang van één steekproef naar steekproevenverdeling

Hypothetische wereld onder de nulhypothese:  
de echte wereld vs  
de wereld onder de nulhypothese



## Veelheid aan concepten nodig

Steekproevenverdeling, significantieniveau, nulhypothese (Castro-Sotos et al 2007; Makar & Rubin, 2018)

Nulhypothese en alternatieve hypothese verwarren (Castro-Sotos et al 2007)

P-waarde verwarren met kans dat nulhypothese waar is

Mensen hebben in het algemeen moeite met kansen en probabilistisch redeneren (Kahneman & Tversky)



## Proces-object-dualiteit

Procedureel denken aanvullen met conceptueel begrip

P-waarde wel kunnen gebruiken om beslissing te nemen maar het niet als object beheersen (Cobb, 2007)



## Inferentiële logica

Modus tollens karakter van statistisch redeneren

Als de nulhypothese waar is zouden we x observeren. We observeren niet x, dus is de nulhypothese niet waar (del Mas, 2004; Falk & Greenbaum, 1995)

Komt bij theoretische benadering pas op het einde

# Maar een simulatiegebaseerde aanpak lost niet vanzelf alle problemen op!



## **Specifieke simulatieproblemen: risico's bij het visualiseren van steekproevenverdeling**

Vb. De nulhypothese is gecentreerd rond nul waardoor leerlingen denken dat die klopt (Case & Jacobbe, 2018)

Eenheden in steekproef vs in steekproevenverdeling verwarren (replicatie vs simulatie)

Rol van de nulhypothese in de simulatie uitleggen

# Deze aanpak is breder toepasbaar

- andere hypothesetoetsen
  - gemiddelde i.p.v. proportie
  - verschil tussen twee gemiddeldes of twee proporties
  - bij de dobbelsteen niet enkel naar '6 ogen' kijken maar naar alle ogen aantallen
    - zelf simulatietool gemaakt met Excel
    - zie mijn website > 2016 > Groningen
- betrouwbaarheidsintervallen
  - zie ook website die we vandaag gebruikt hebben

Mijn dobbelsteen (probleem 1)

# ogen	mijn
1	23
2	26
3	19
4	18
5	25
6	9
TOT	120

Is mijn dobbelsteen onzuiver?

Bedankt voor je aandacht!