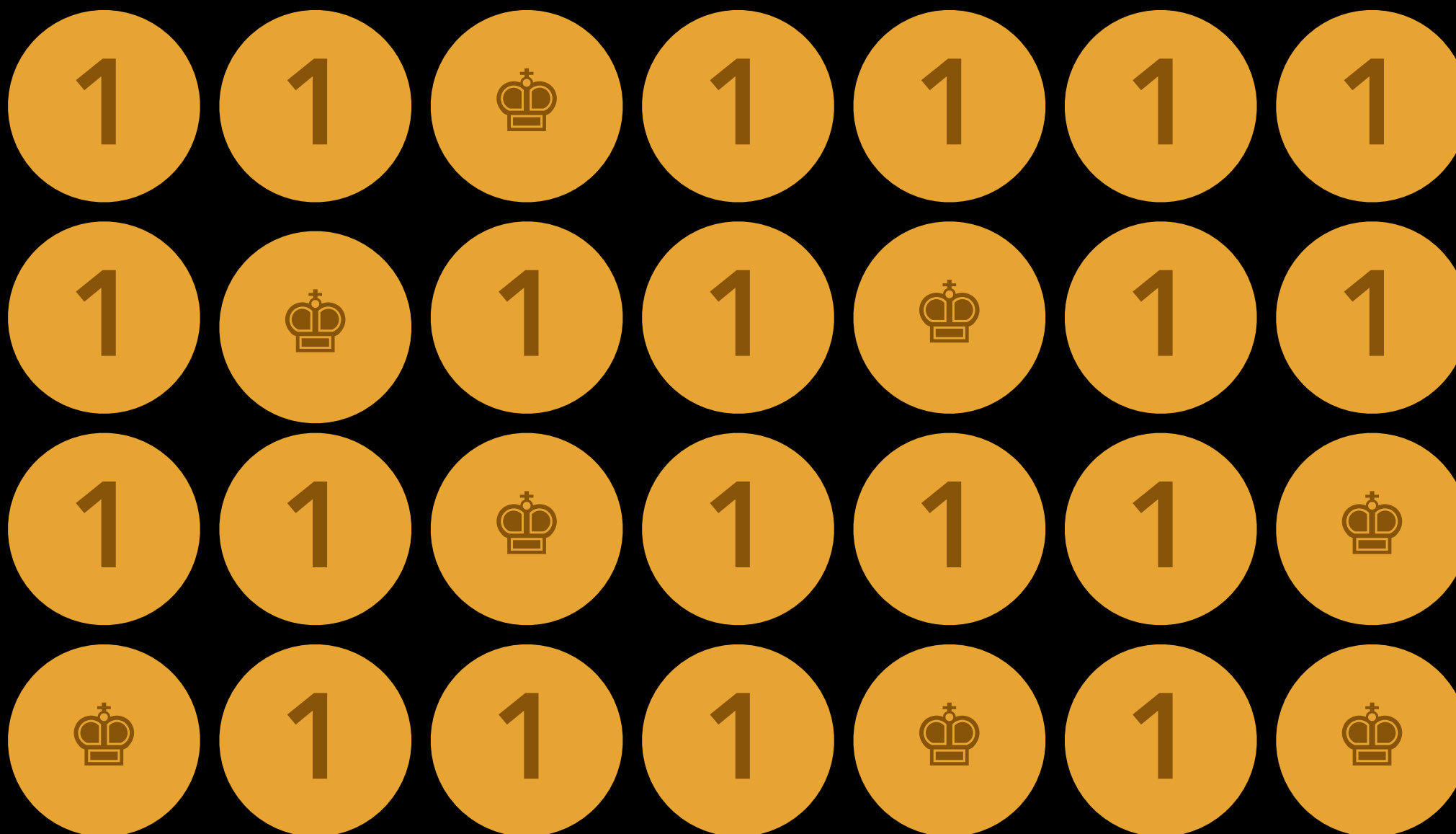


# Statistika za hekerje

# Statistics for Hackers



*Jake VanderPlas*  
*PyCon 2016*



20×1

8×Crown



20 × 1

8 ×



57	71	74	35	60	83	74	62
43	65	63	30	51	62	57	72
64	51	43	55	85	54	77	45
69	46	97	59	73	65	57	73

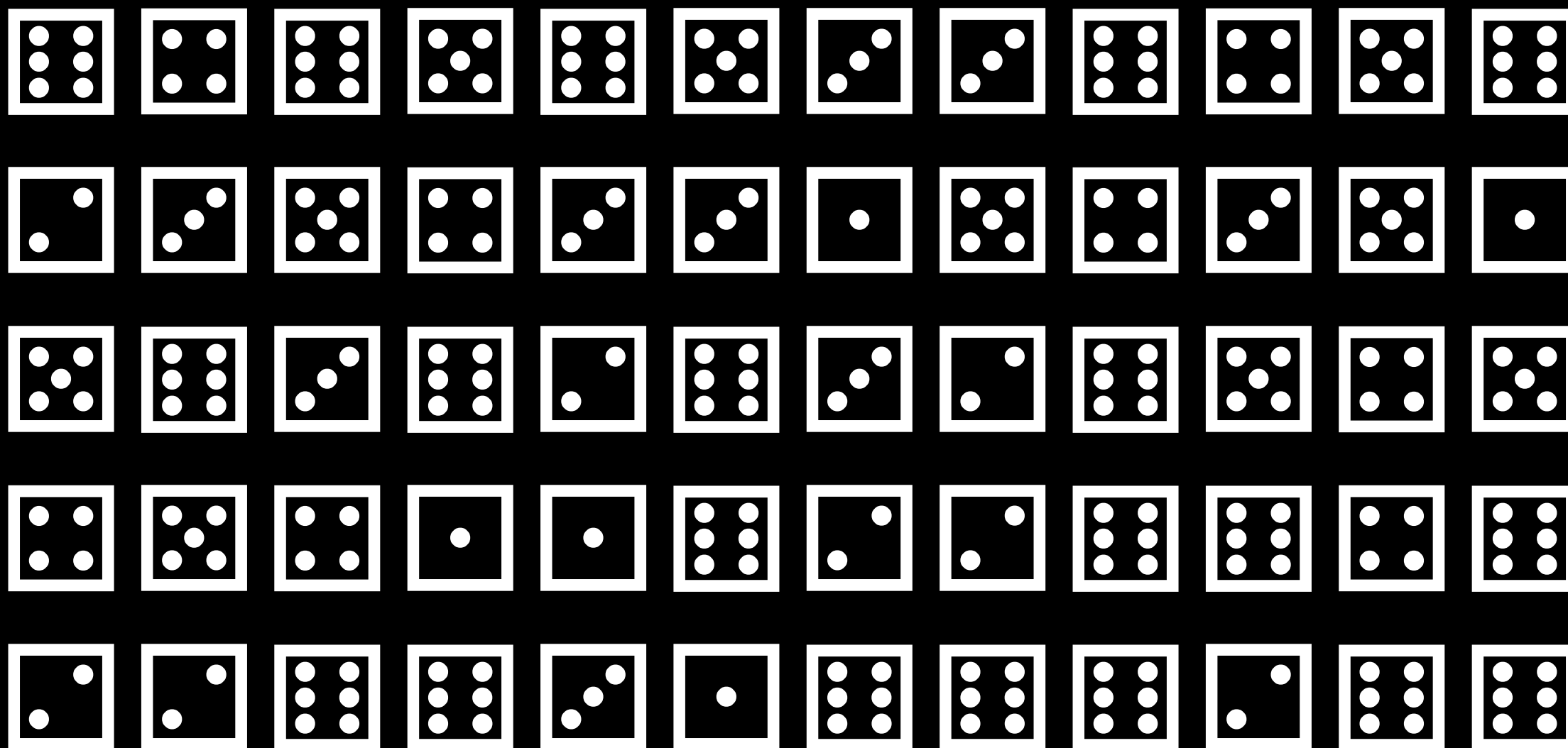
**fantje: 59,13 dekleeta: 63,82**

**razlika: 4,69**

57	71	74	35	60	83	74	62
43	65	63	30	51	62	57	72
64	5	So dekleta znala več?					45
69	46	97	59	70	7	7	73

**fantje: 59,13 dekleta: 63,82**

**razlika: 4,69**



5× 8× 9×

8× 10× 20×



5 ×  8 ×  9 × 

8 ×  10 ×  20 × 



13

12

5



9

5

9





*Sta izbiri hrane in  
pijače povezani?*

9

5

5

9



20×

1

8×





20×

1

8×



“Vse srečne družine so si podobne,  
vsaka nesrečna družina  
pa je nesrečna po svoje.”

Lev Nikolajevič Tolstoj: *Ana Karenina*

**ničelna hipoteza:**  
kovanec je pošten



Verjetnost, da v 28 metih  
poštenega kovanca pade 20 cifer

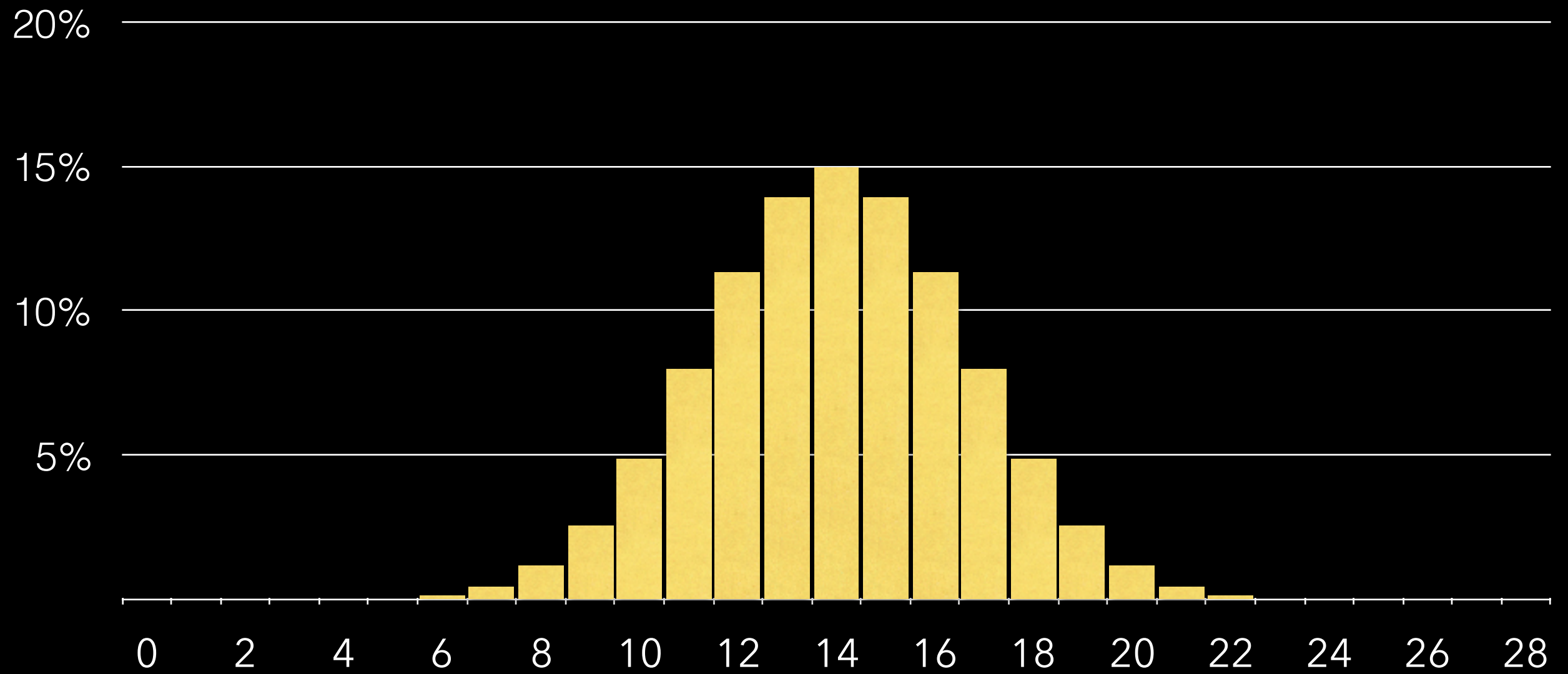
$$P(C = 20) = \frac{\binom{28}{20}}{2^{28}} \approx 1,16\%$$

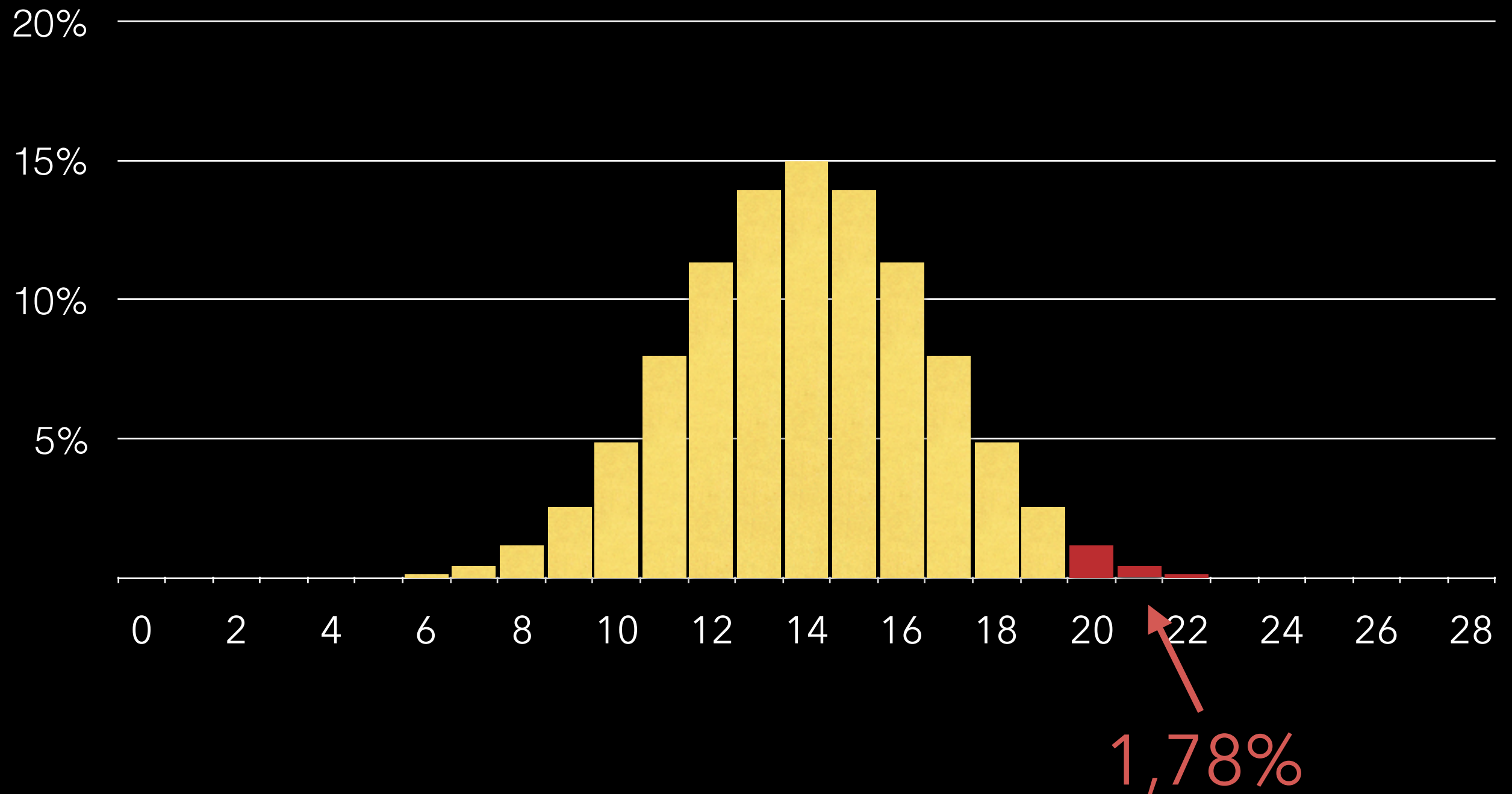
Verjetnost, da v 10000 metih  
poštenega kovanca pade 5000 cifer

$$P(C = 5000) = \frac{\binom{10000}{5000}}{2^{5000}} \approx 0,80\%$$

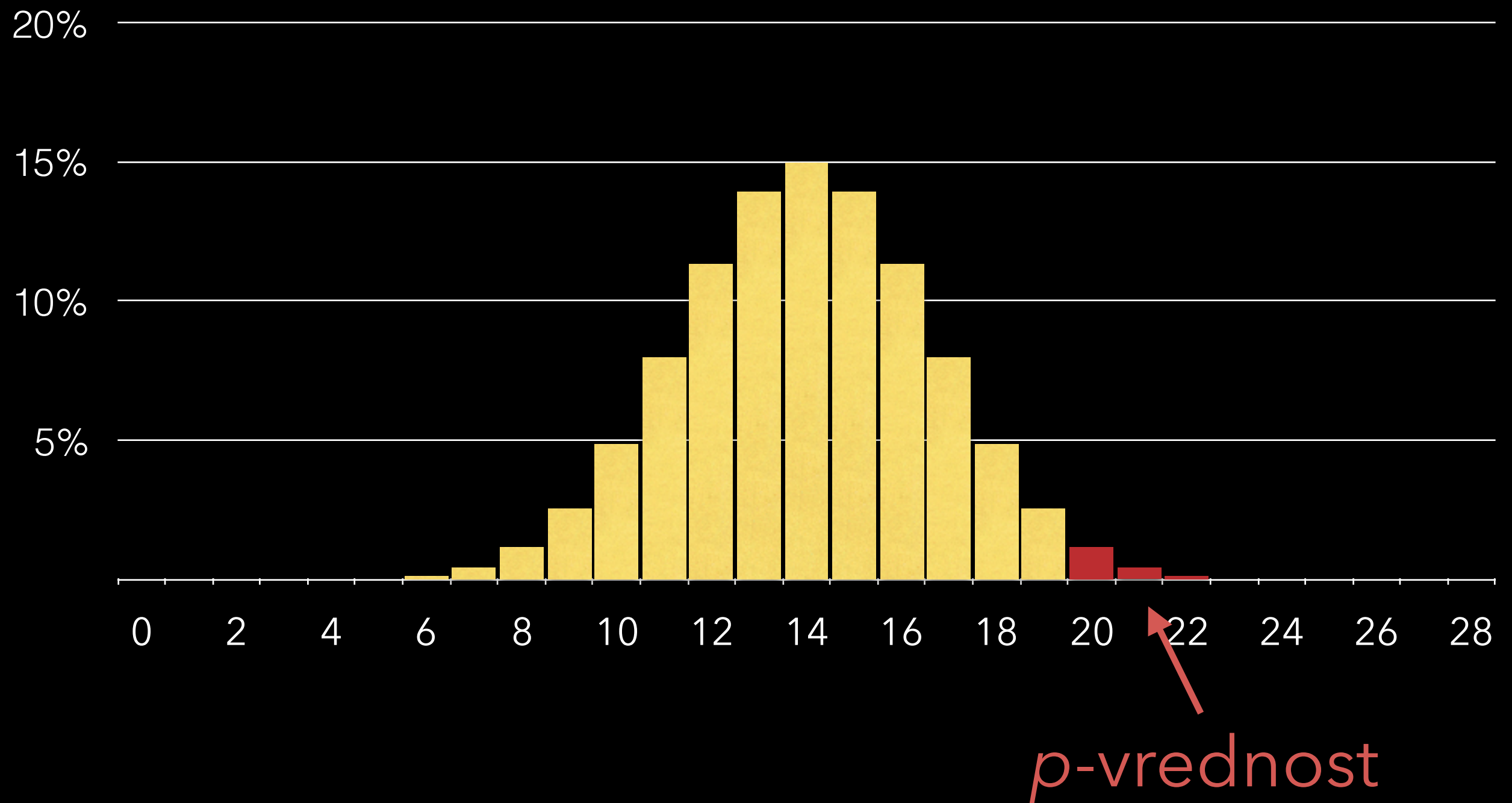
Verjetnost, da v 28 metih  
poštenega kovanca pade **vsaj** 20 cifer

$$P(C \geq 20) = \sum_{k=20}^{28} P(C = k) \approx 1,78\%$$









# $p$ -vrednost

Verjetnost, da ob ničelni hipotezi  
dobimo vsaj tolikšno odstopanje  
kot pri izmerjenem rezultatu.

# $p$ -vrednost

Verjetnost, da  
ničelna hipoteza drži.



# $p$ -vrednost

Verjetnost, da ob ničelni hipotezi  
dobimo vsaj tolikšno odstopanje  
kot pri izmerjenem rezultatu.



# Preverjanje hipotez

1. postavimo ničelno hipotezo
2. določimo porazdelitev dogodkov
3. izračunamo  $p$ -vrednost

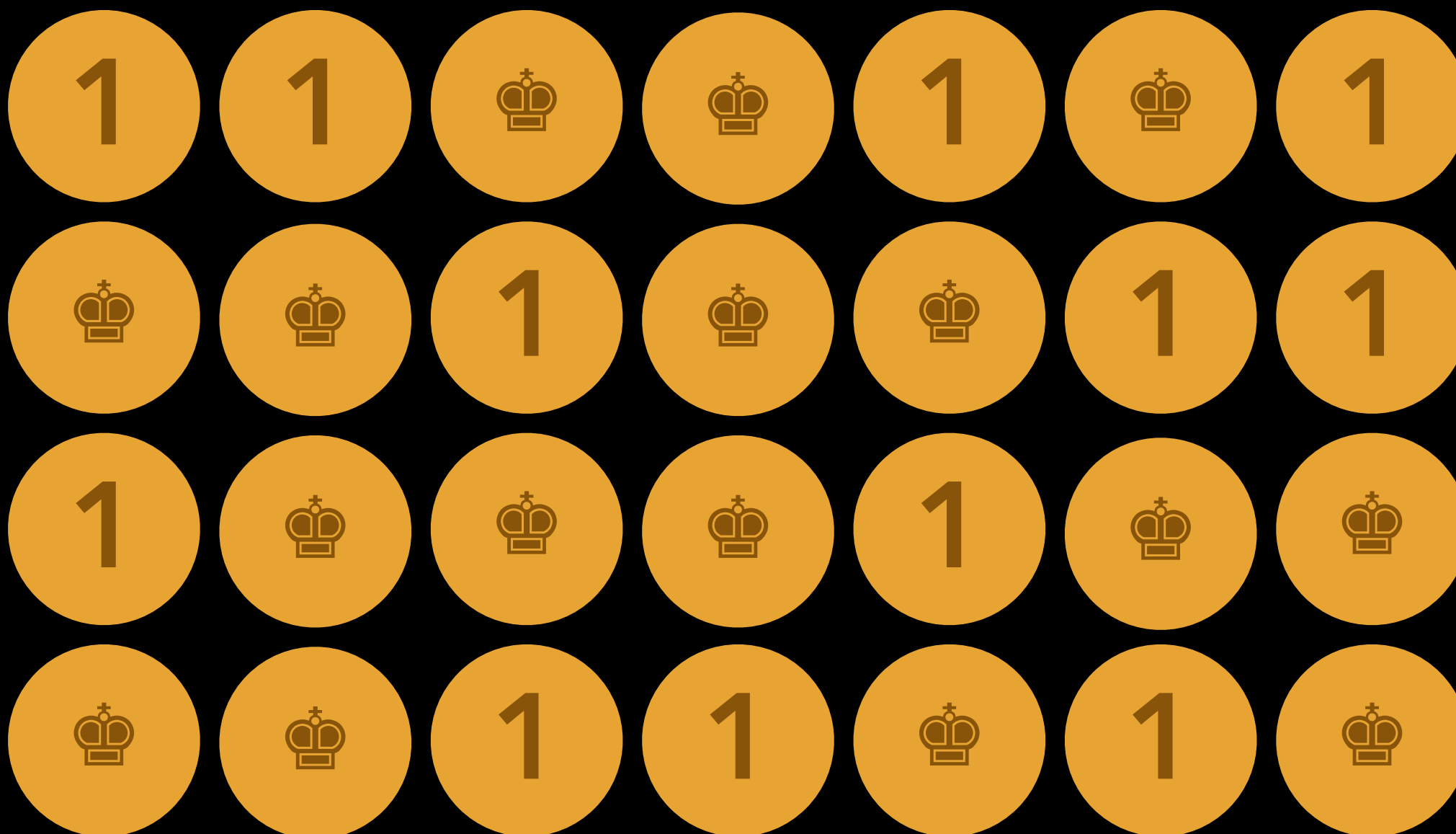


# Hekersko preverjanje hipotez

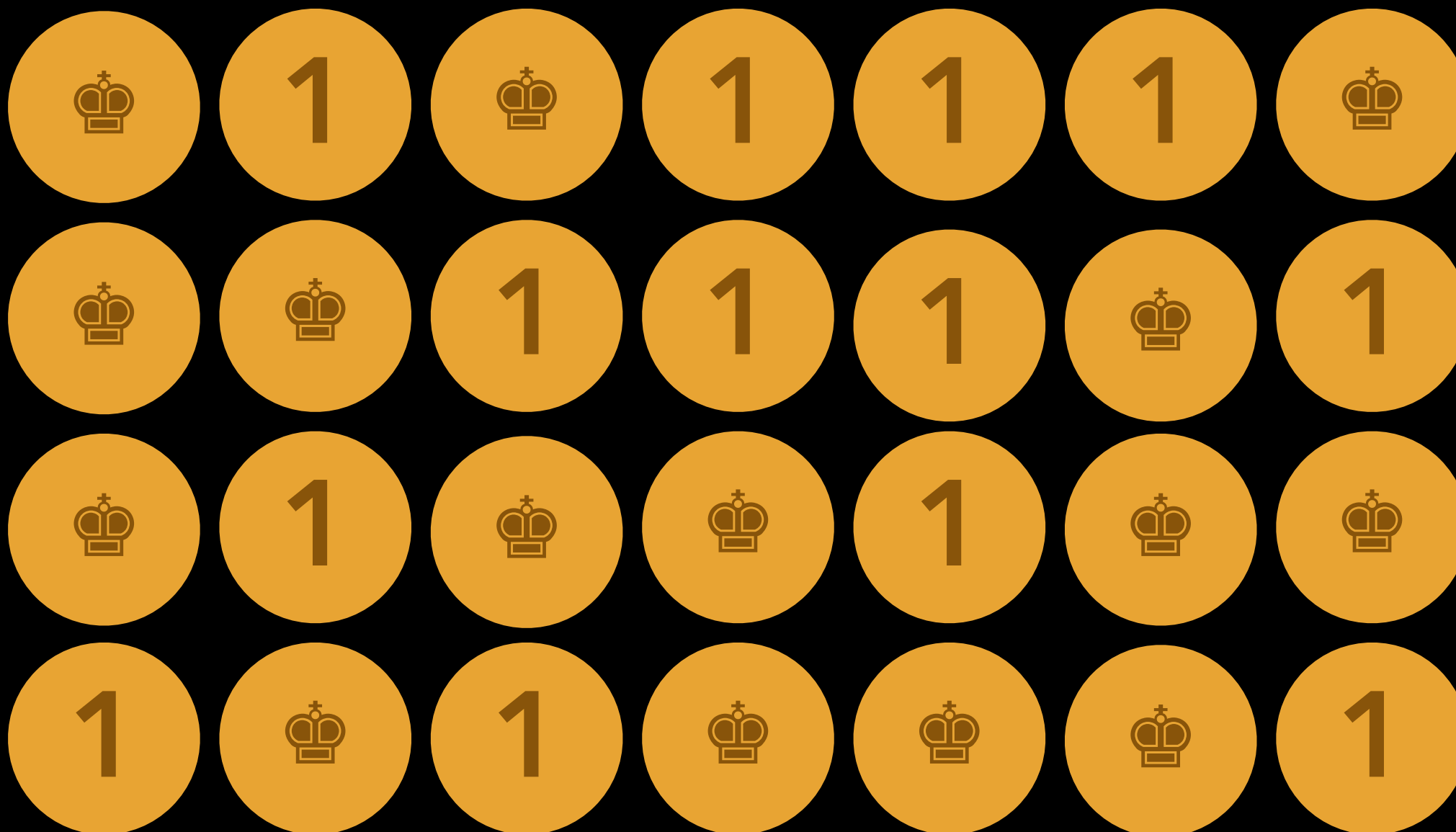
1. postavimo ničelno hipotezo
2. generiramo veliko dogodkov
3. ocenimo  $p$ -vrednost

# Hekersko preverjanje hipotez

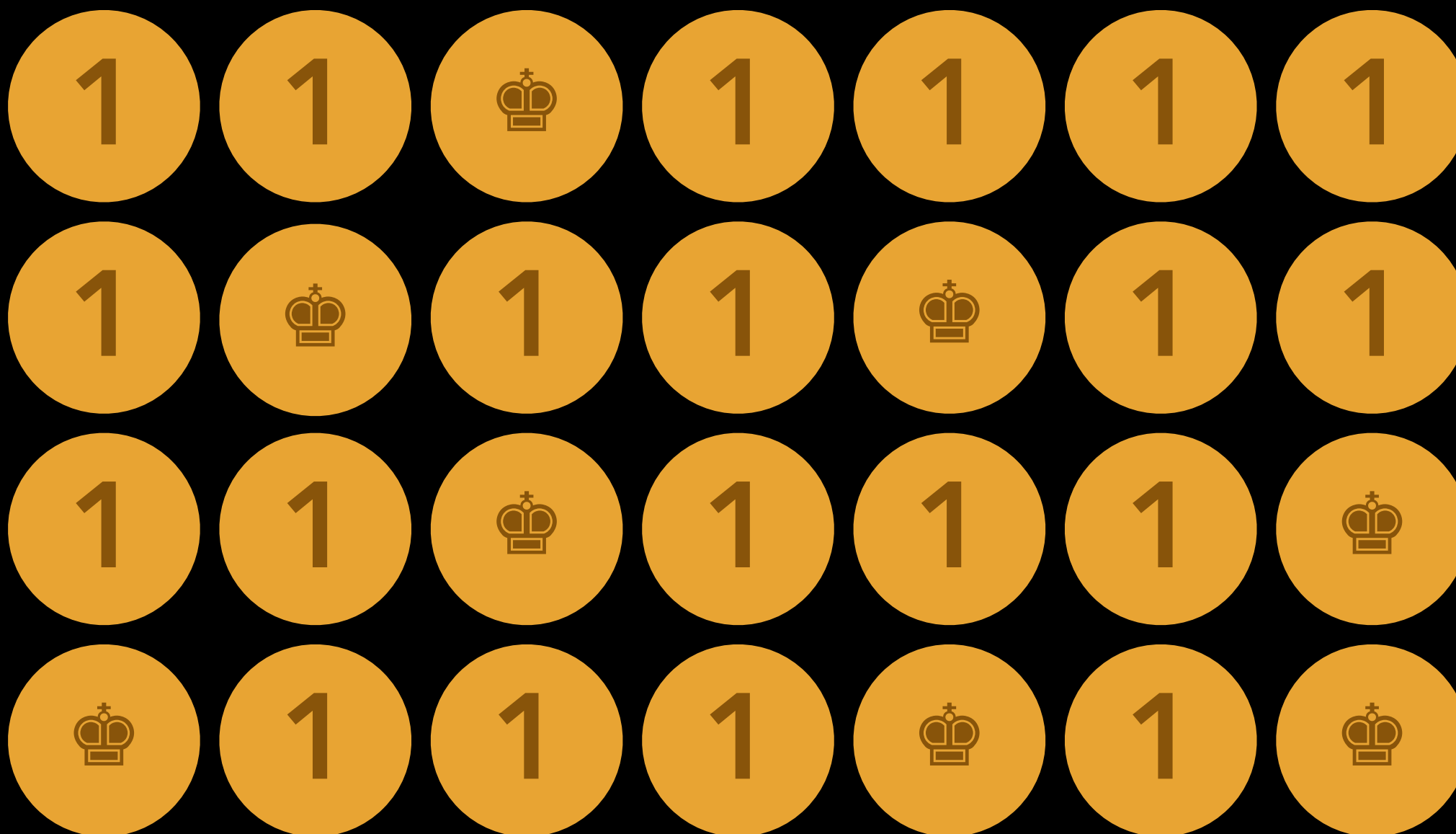
1. postavimo ničelno hipotezo ✓
2. generiramo veliko dogodkov
3. ocenimo  $p$ -vrednost




12x  16x 

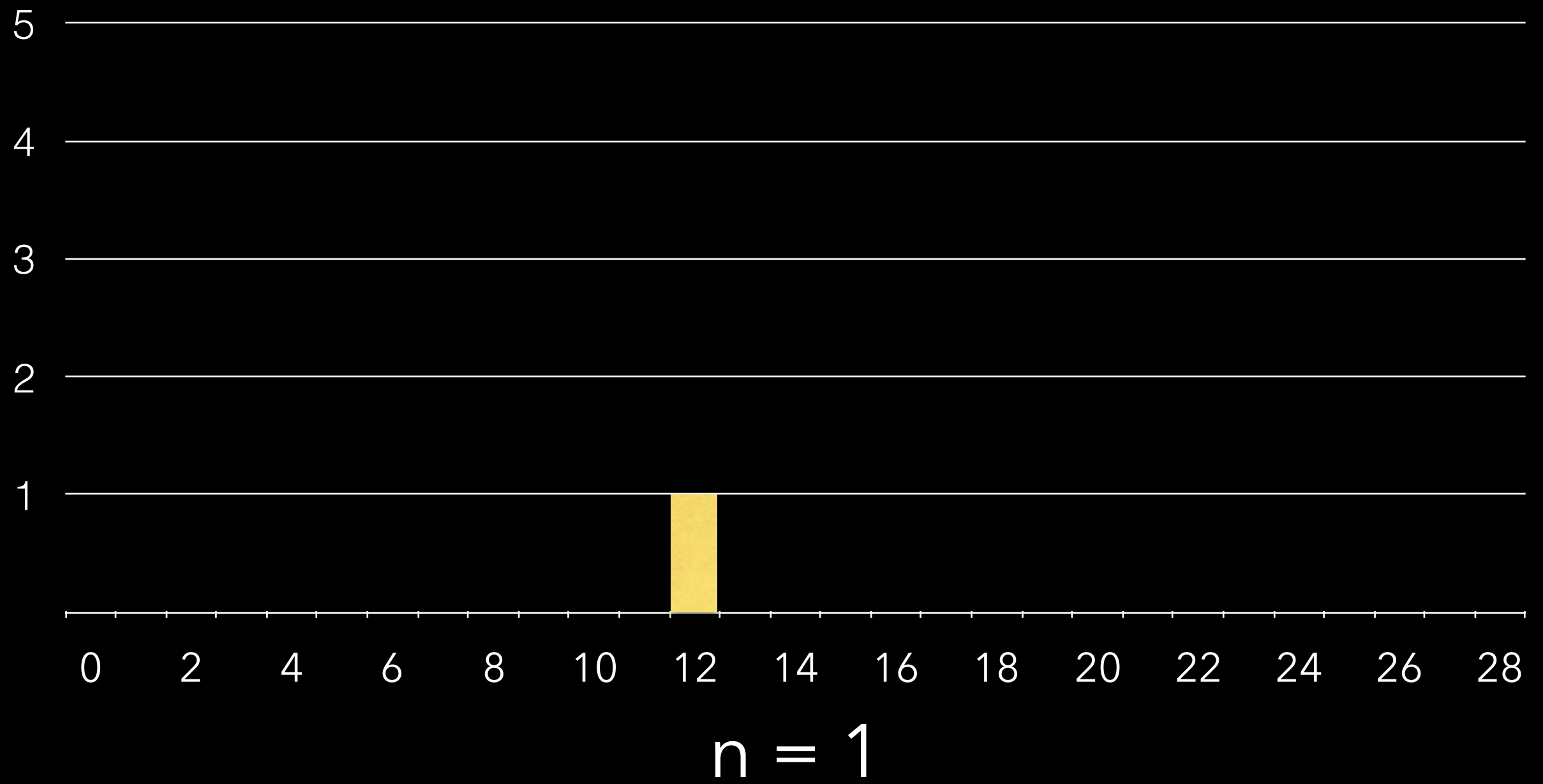


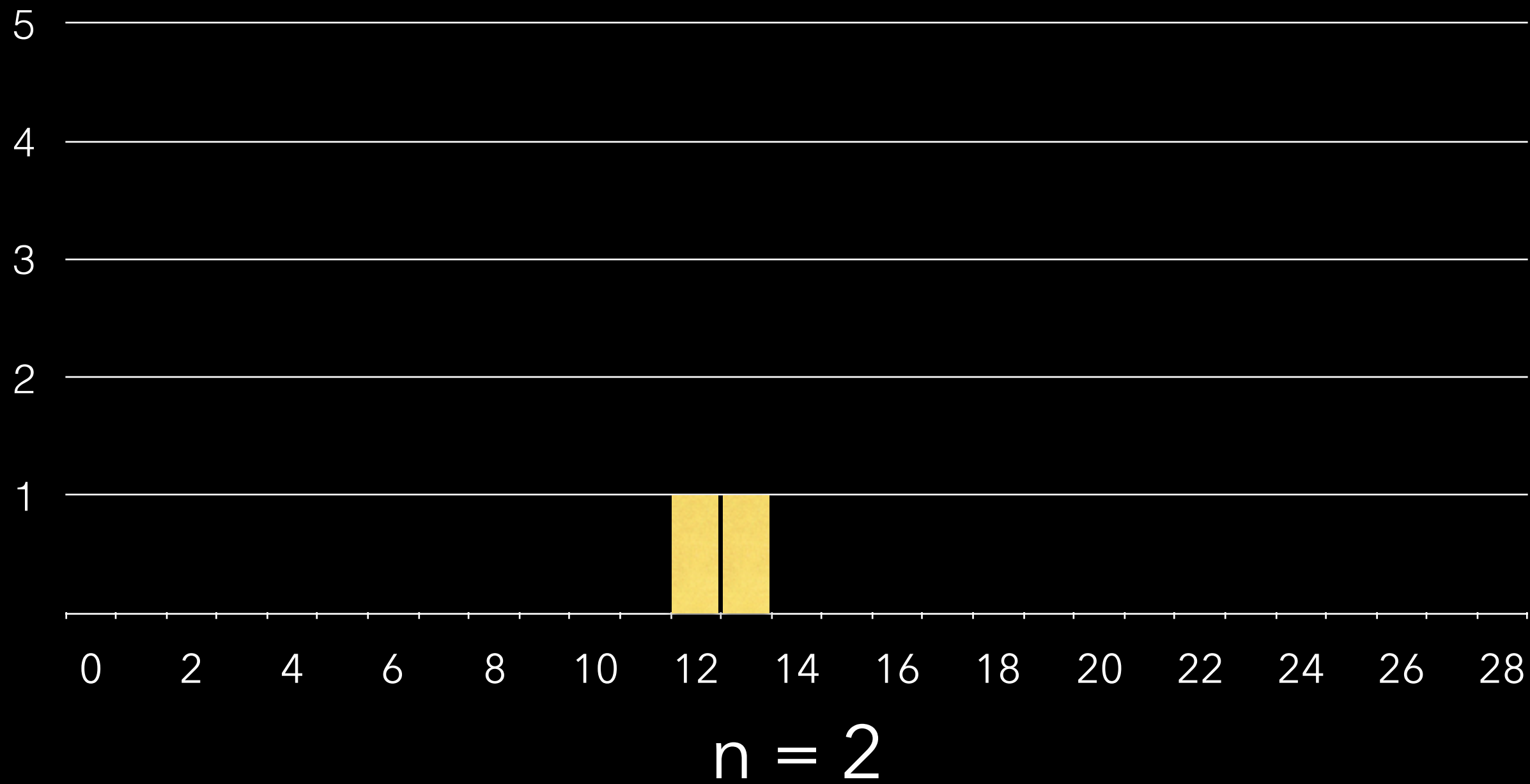
13x 1 15x Crown

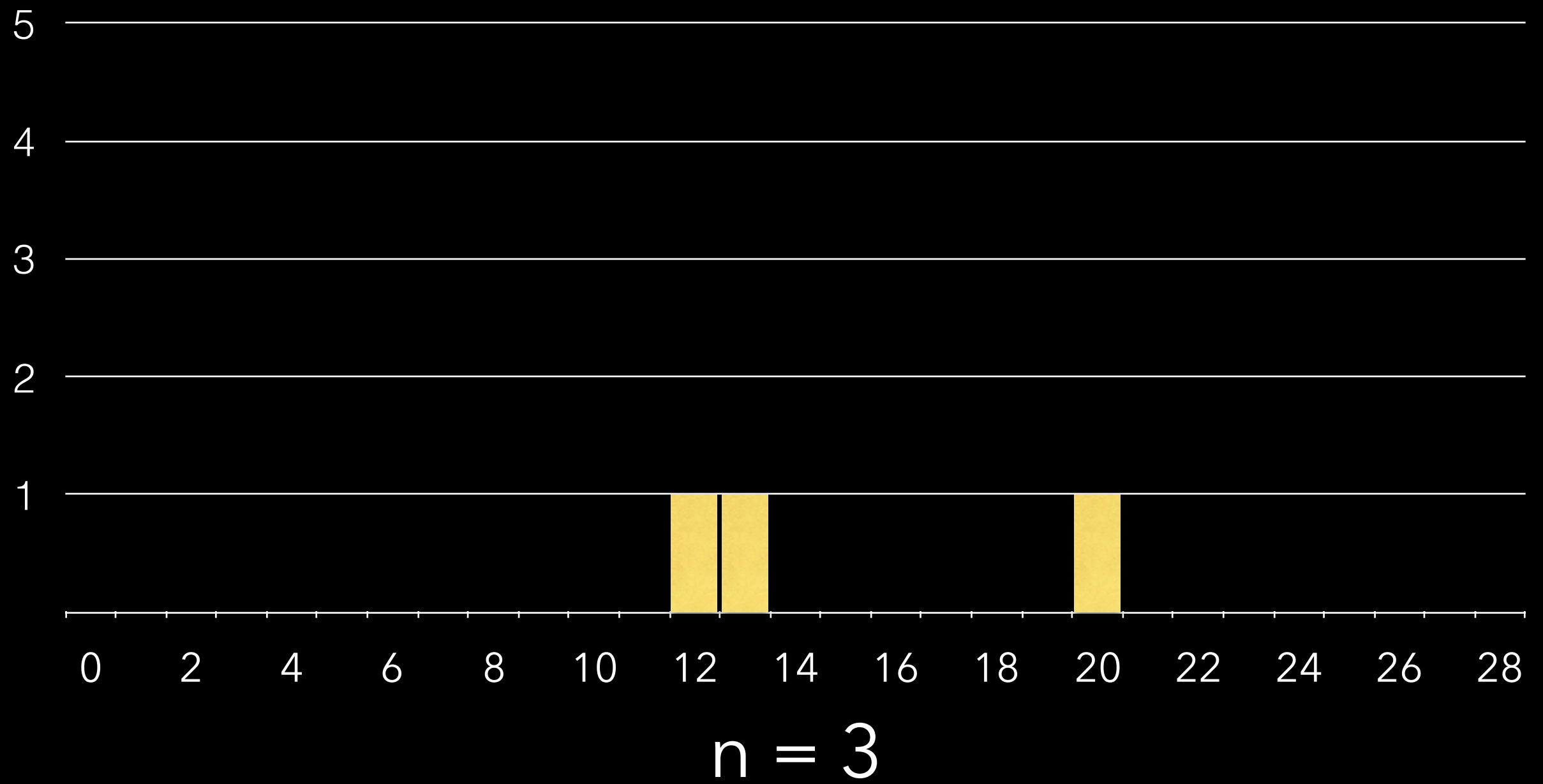


20× 

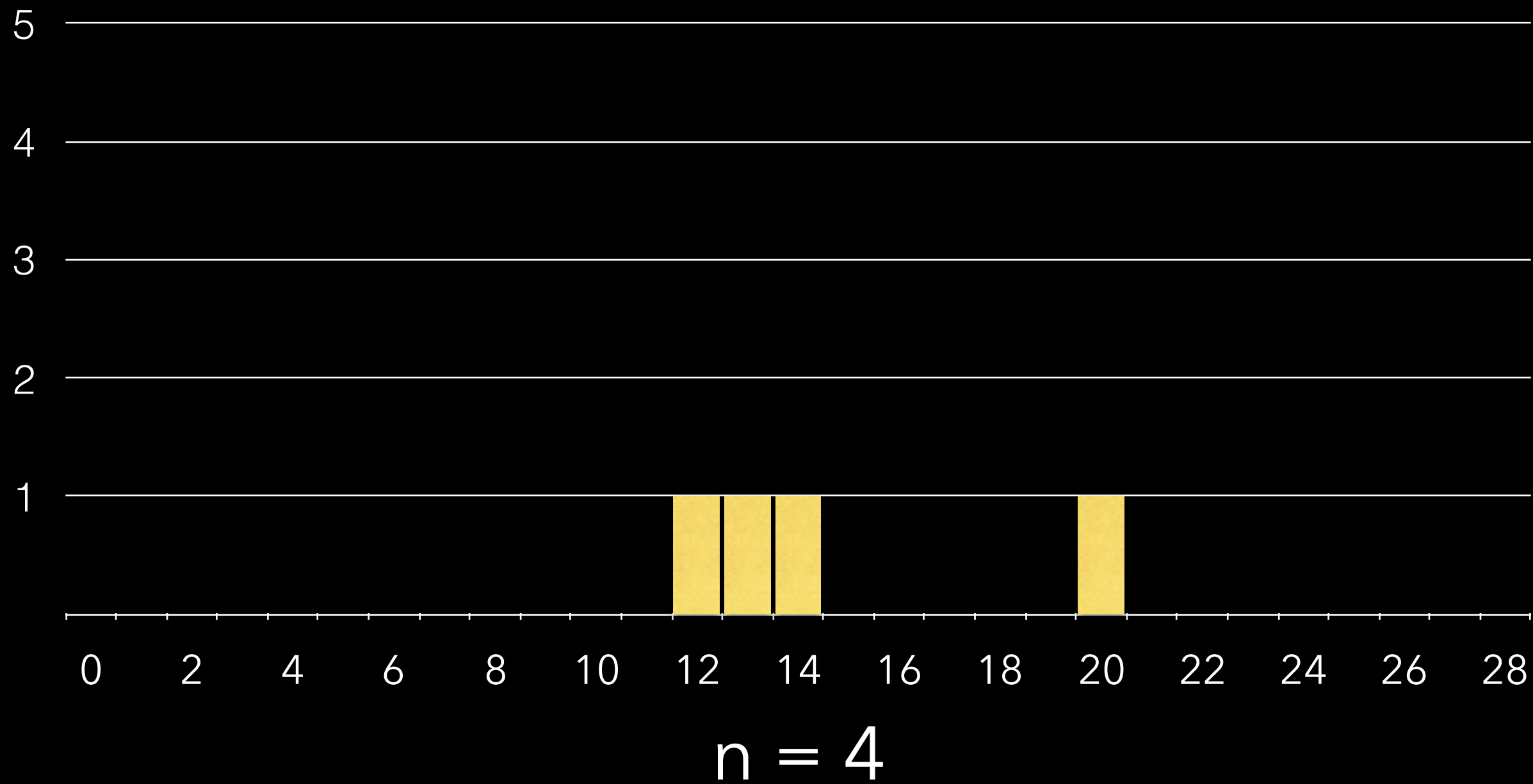
8× 

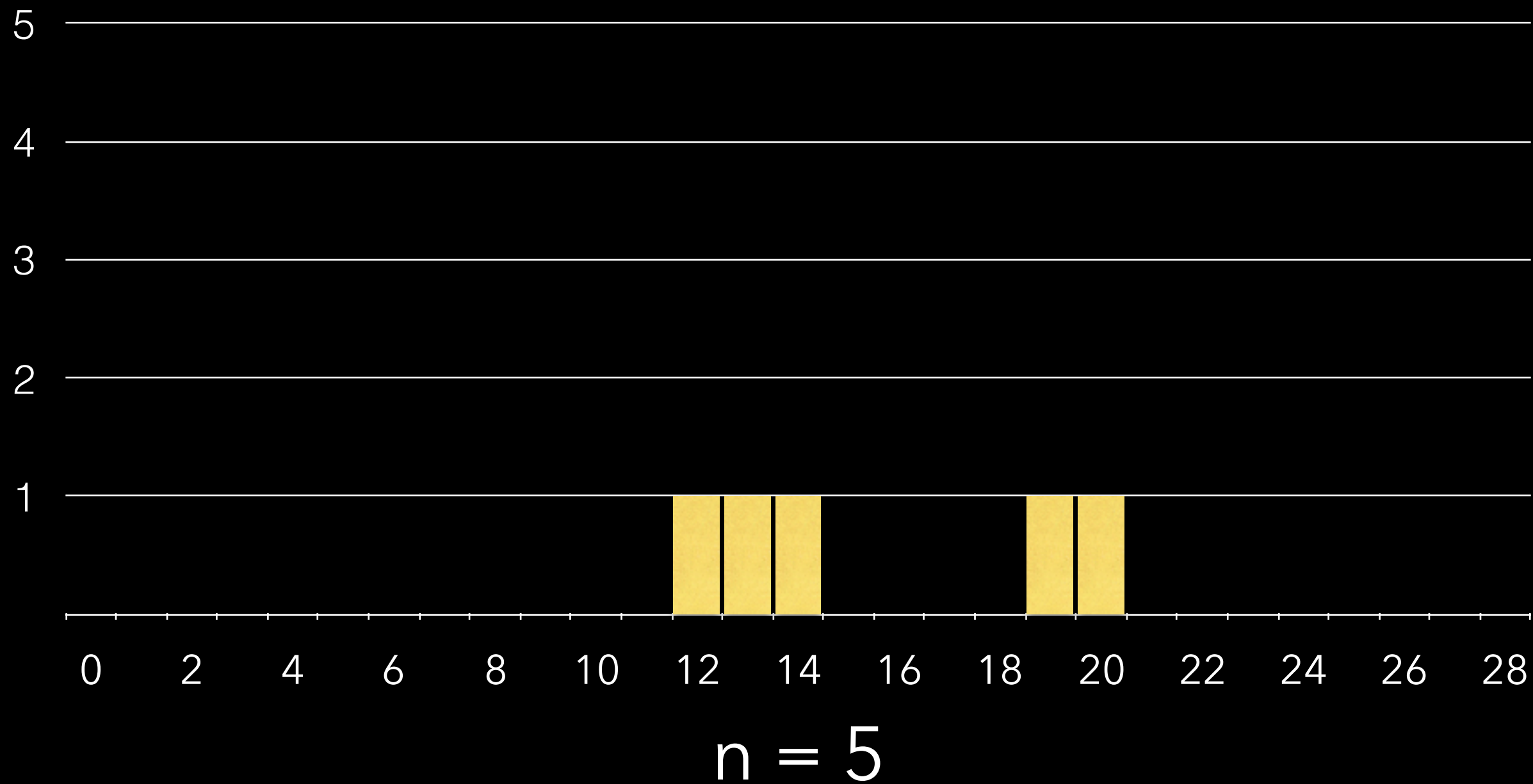


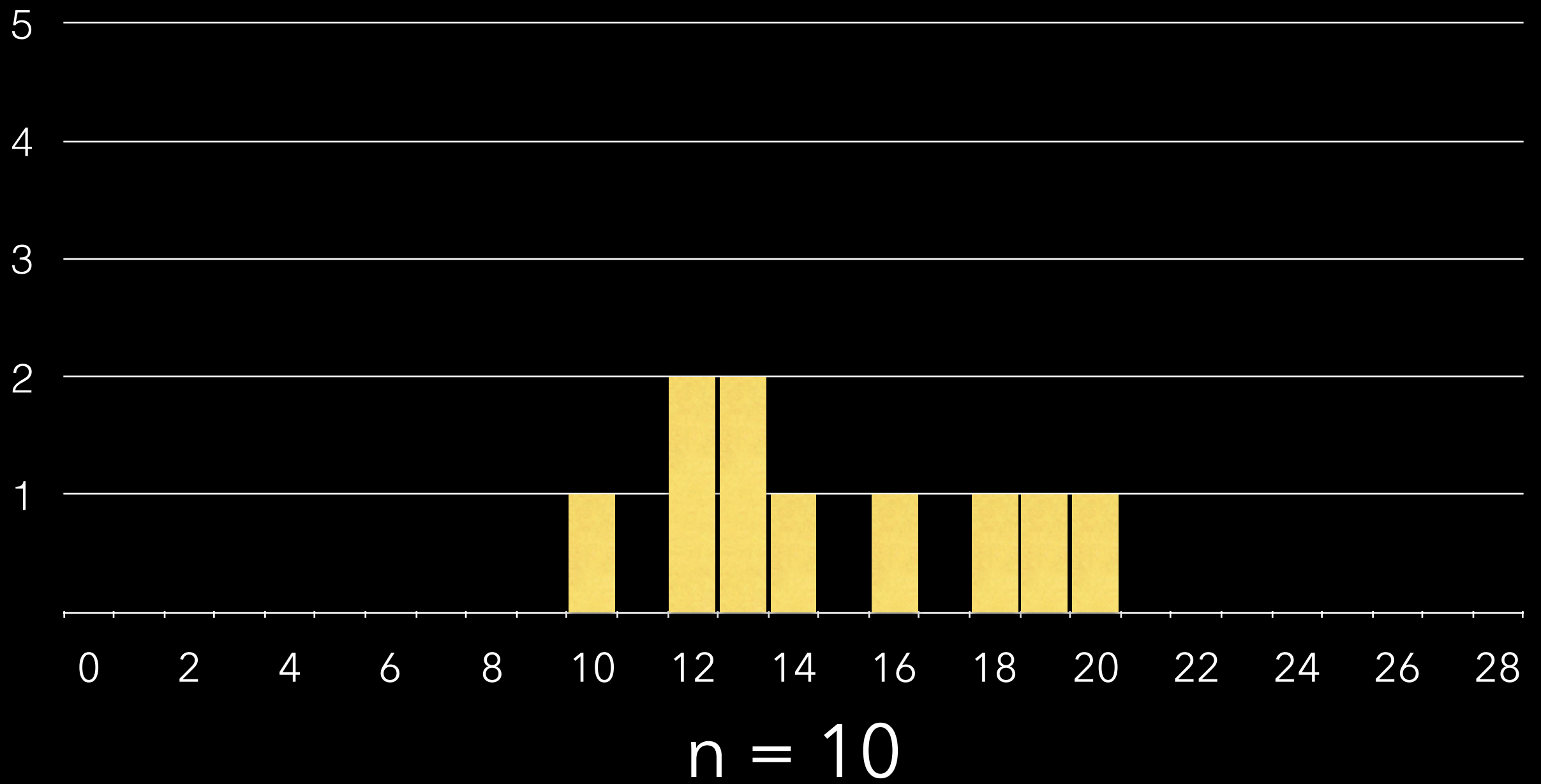


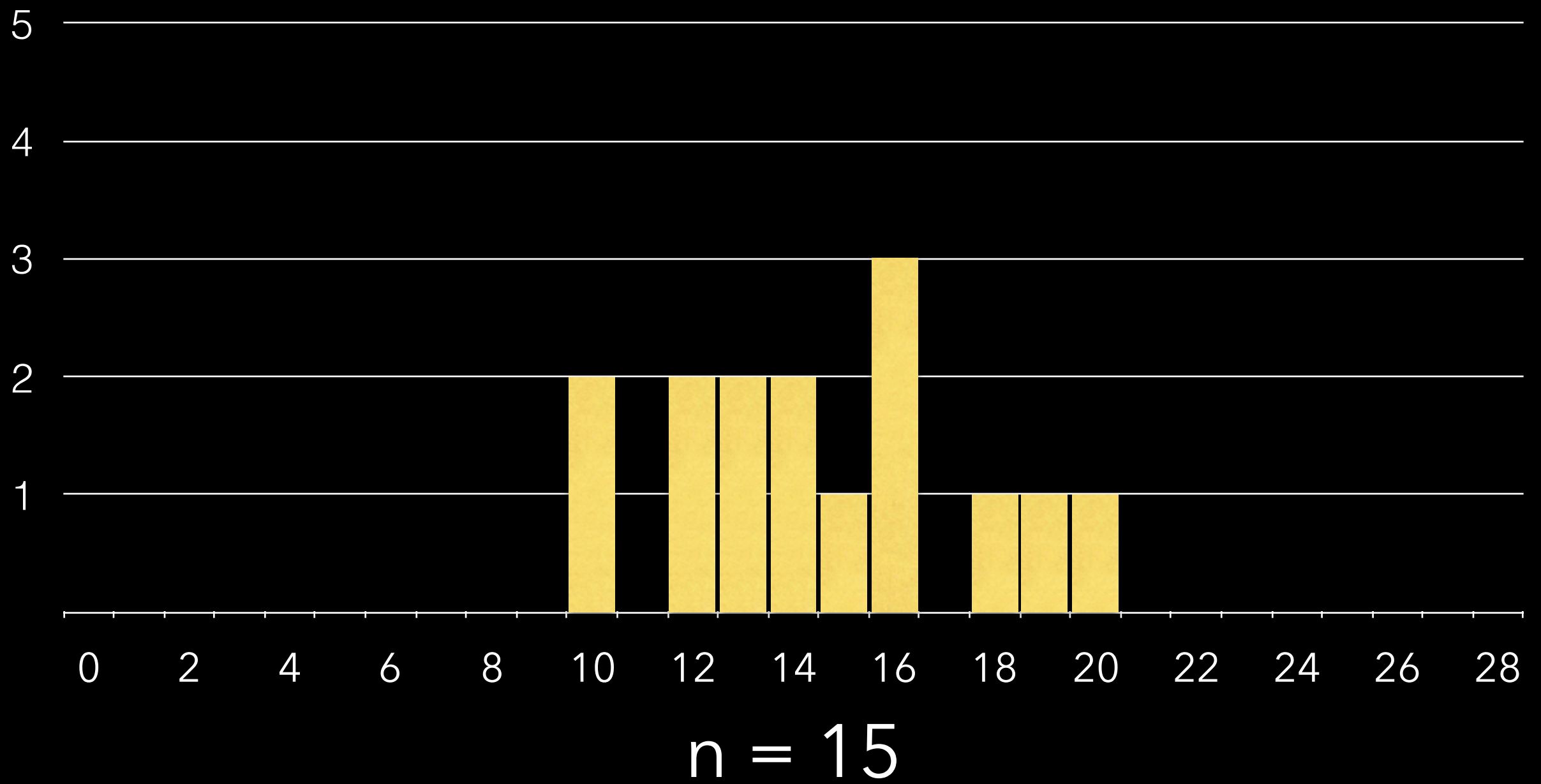


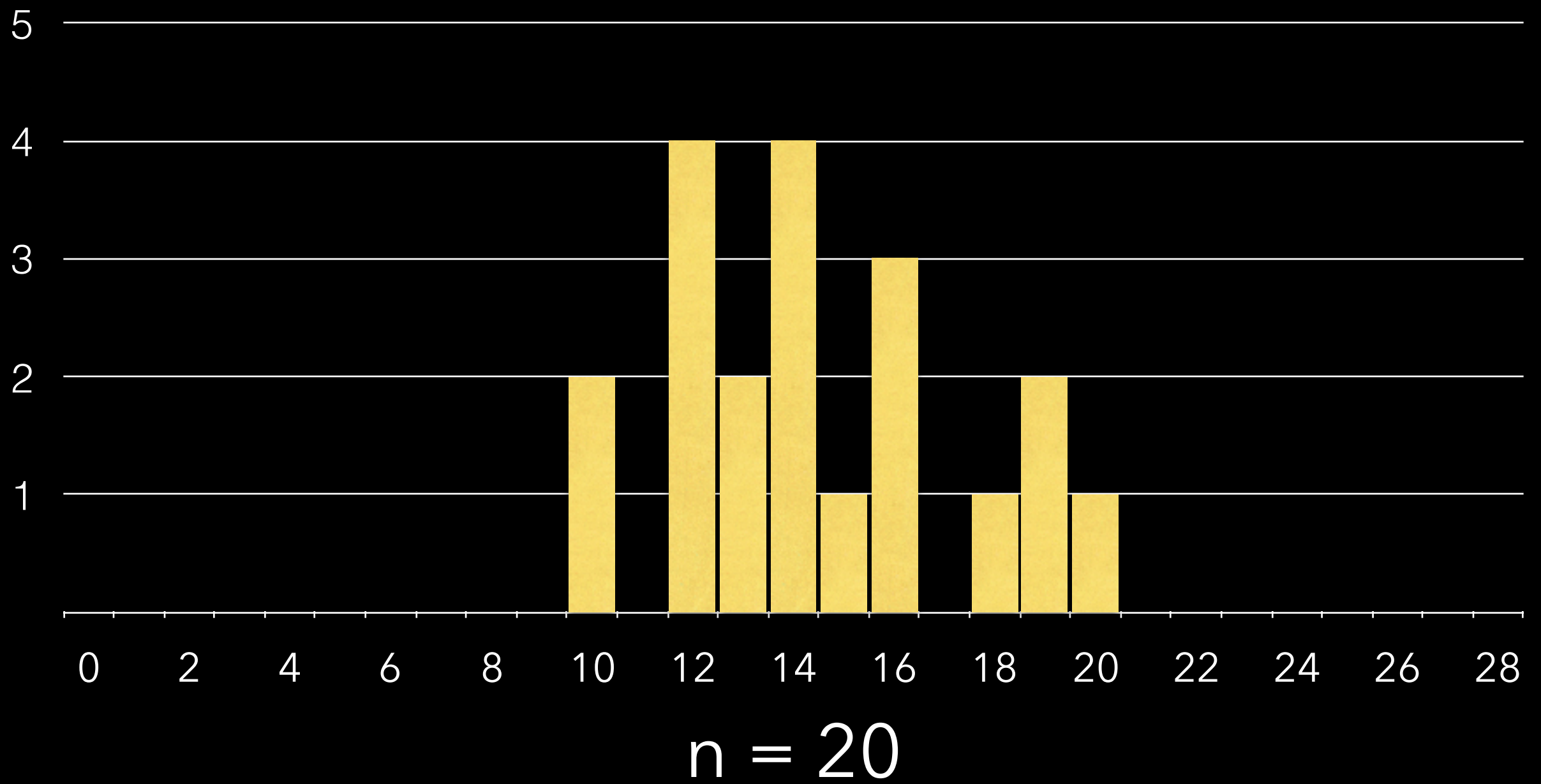


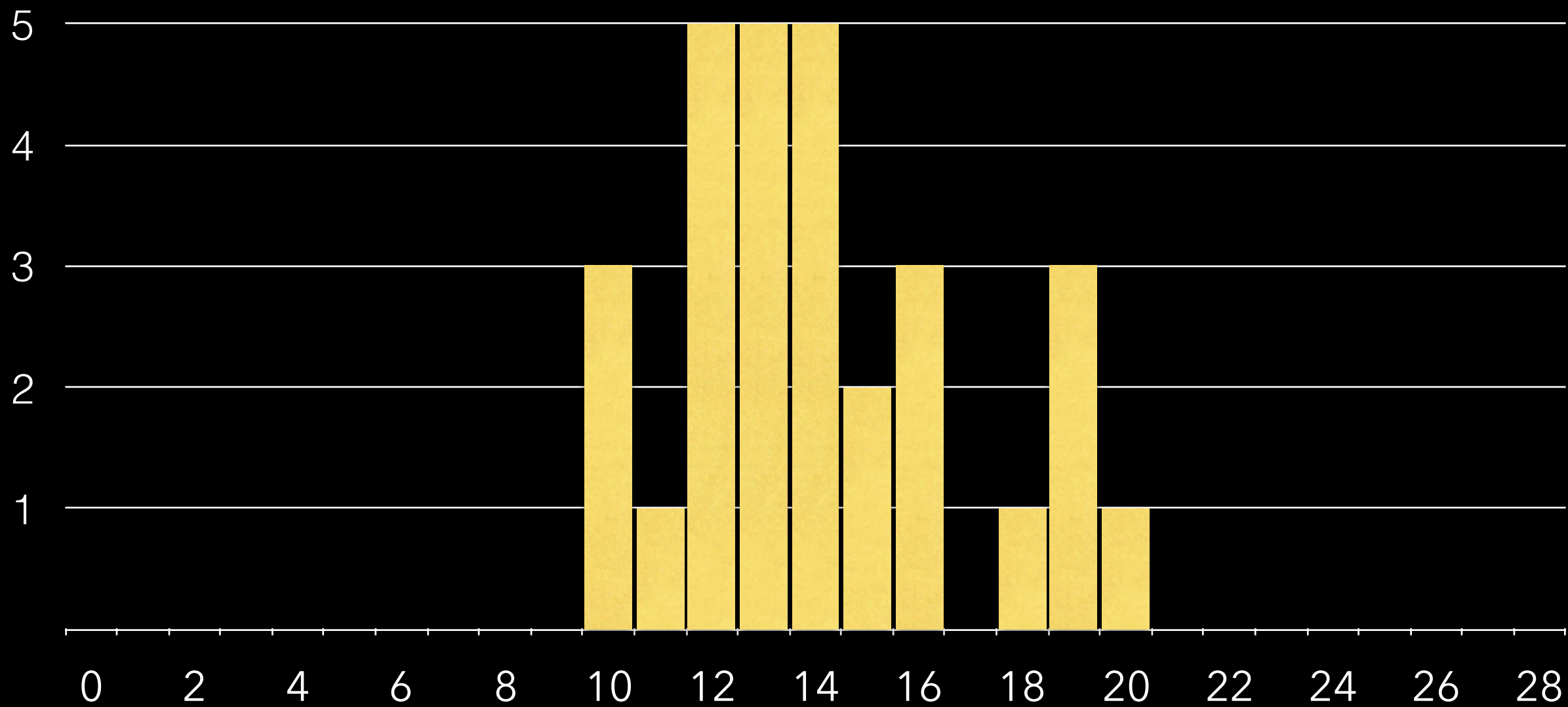






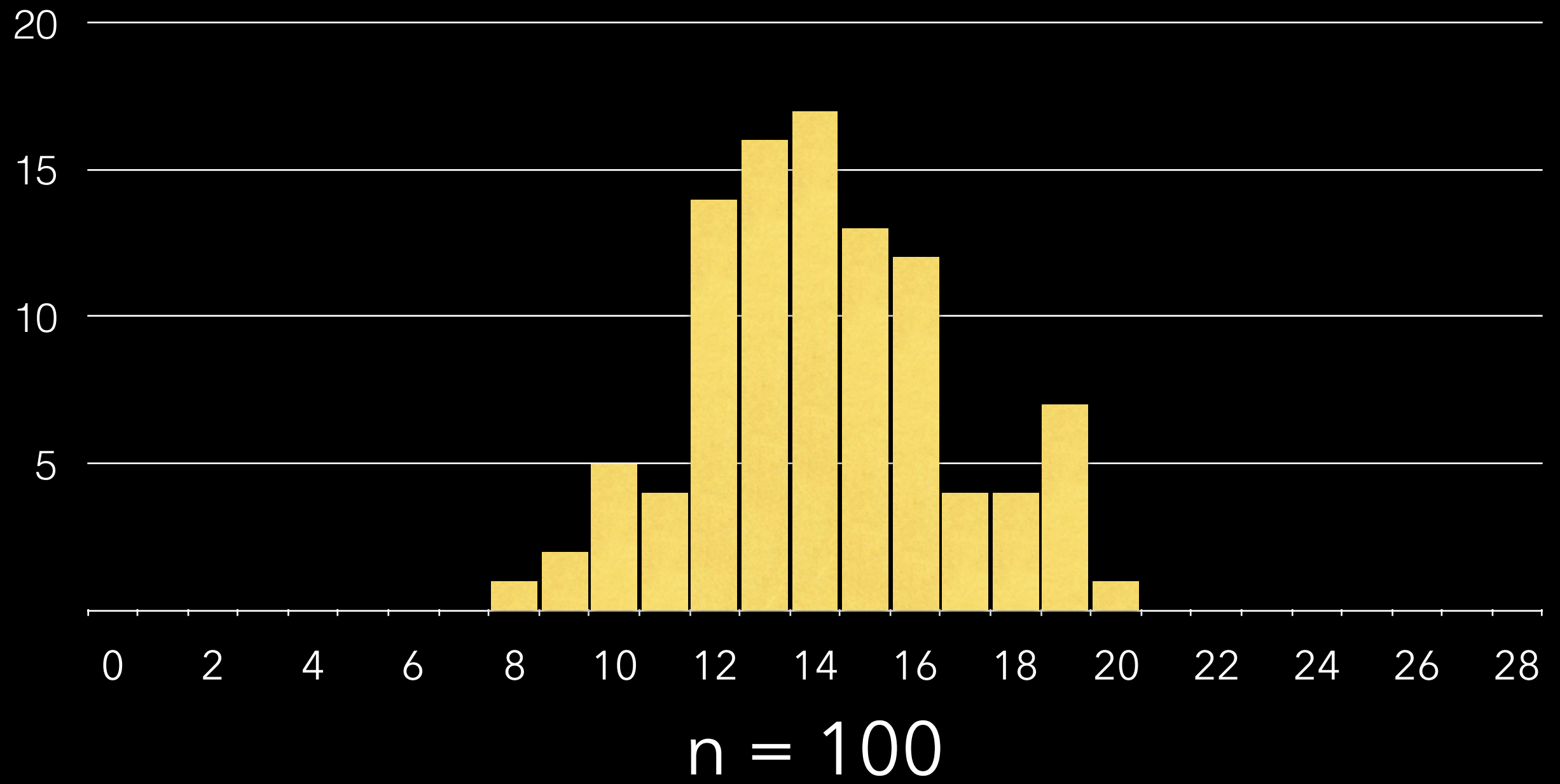




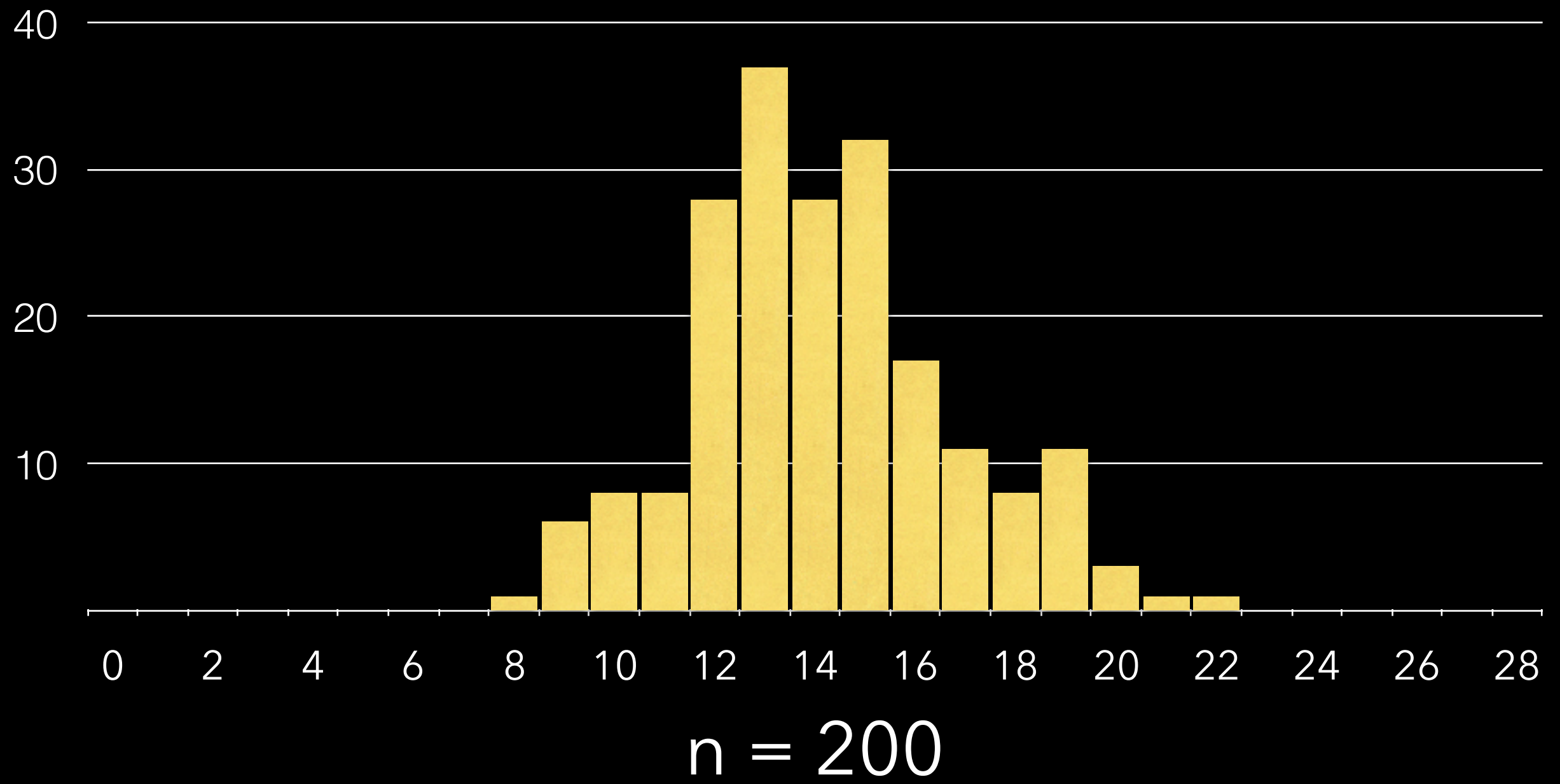


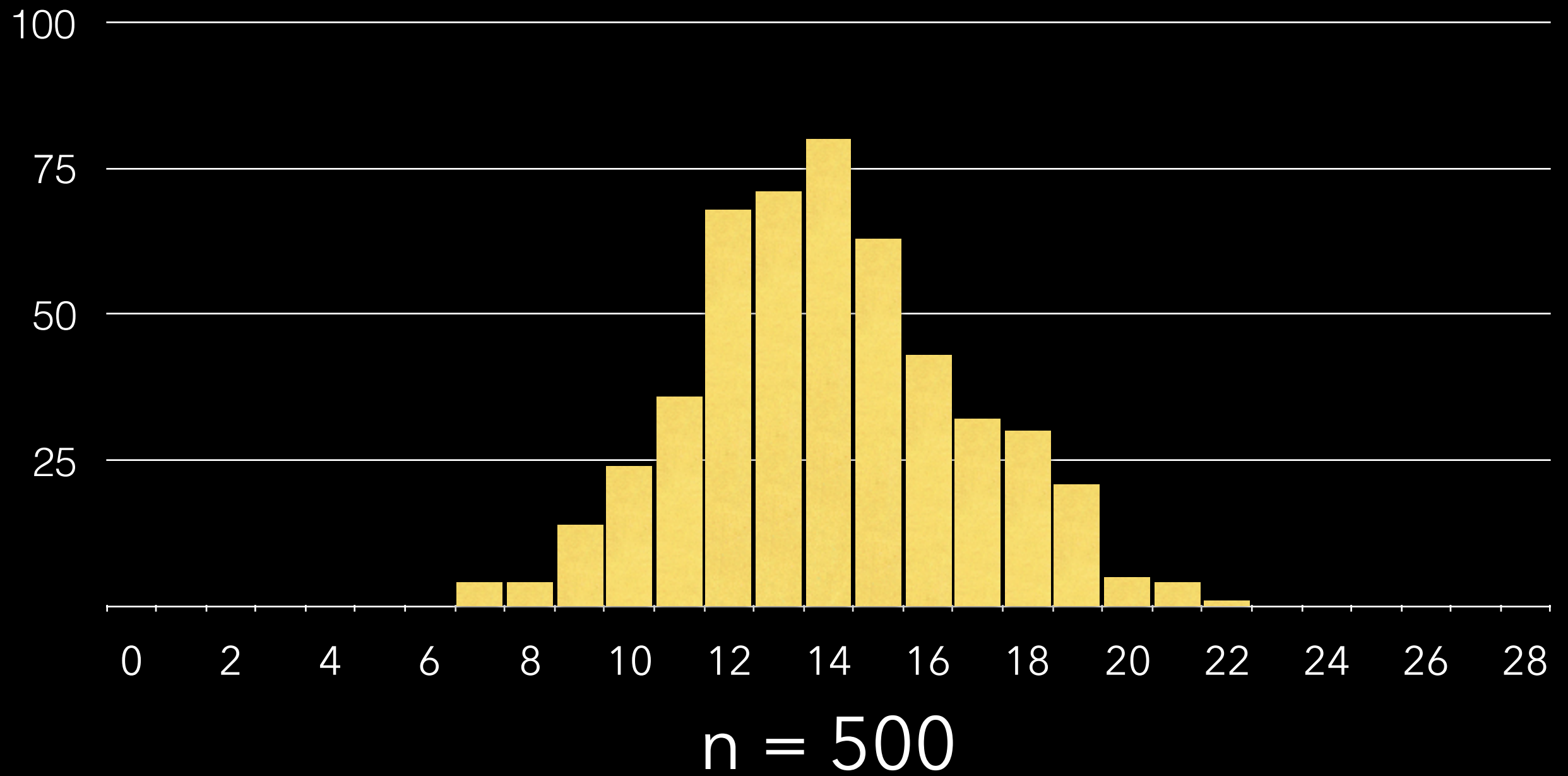
$n = 25$

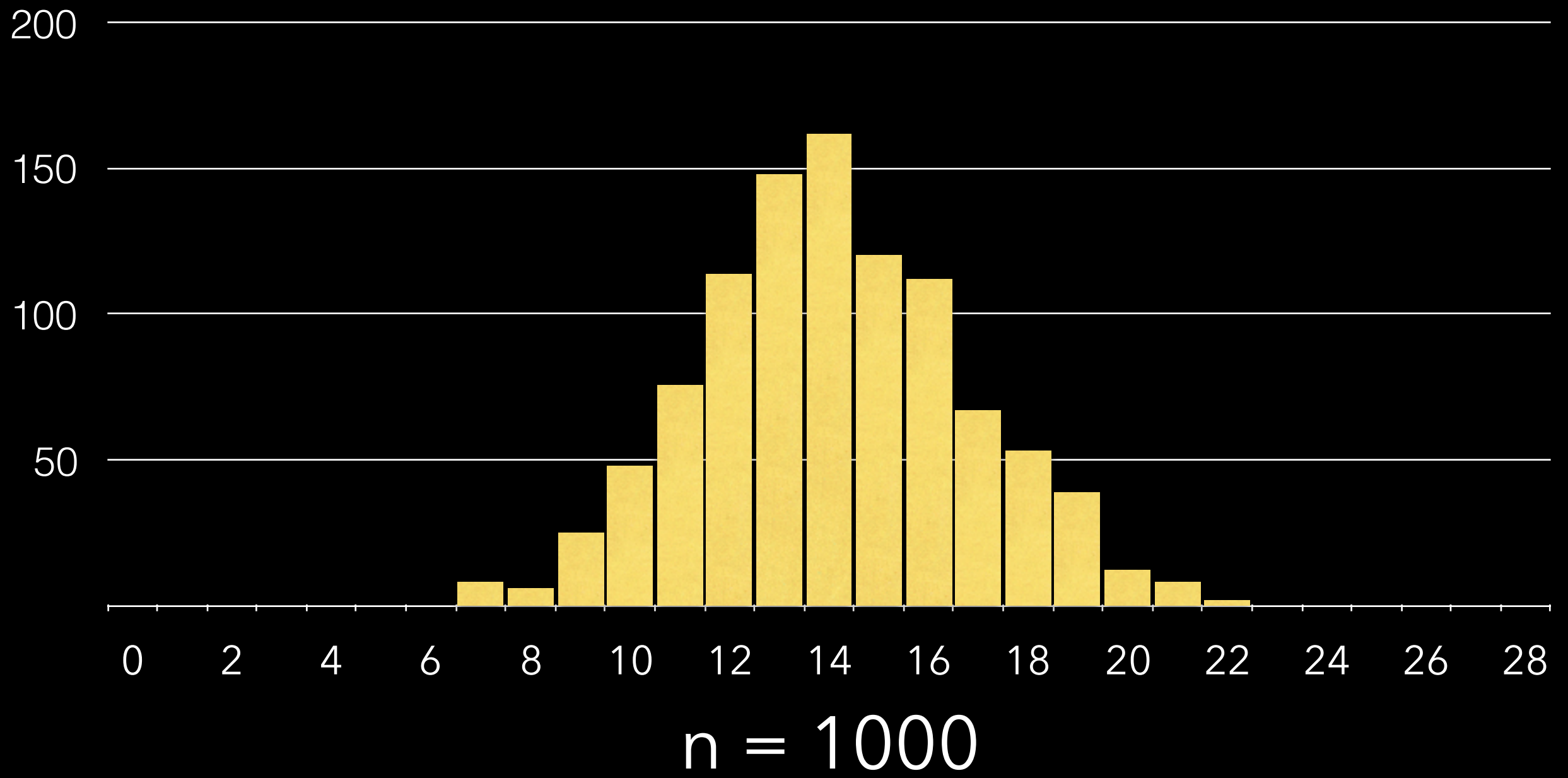


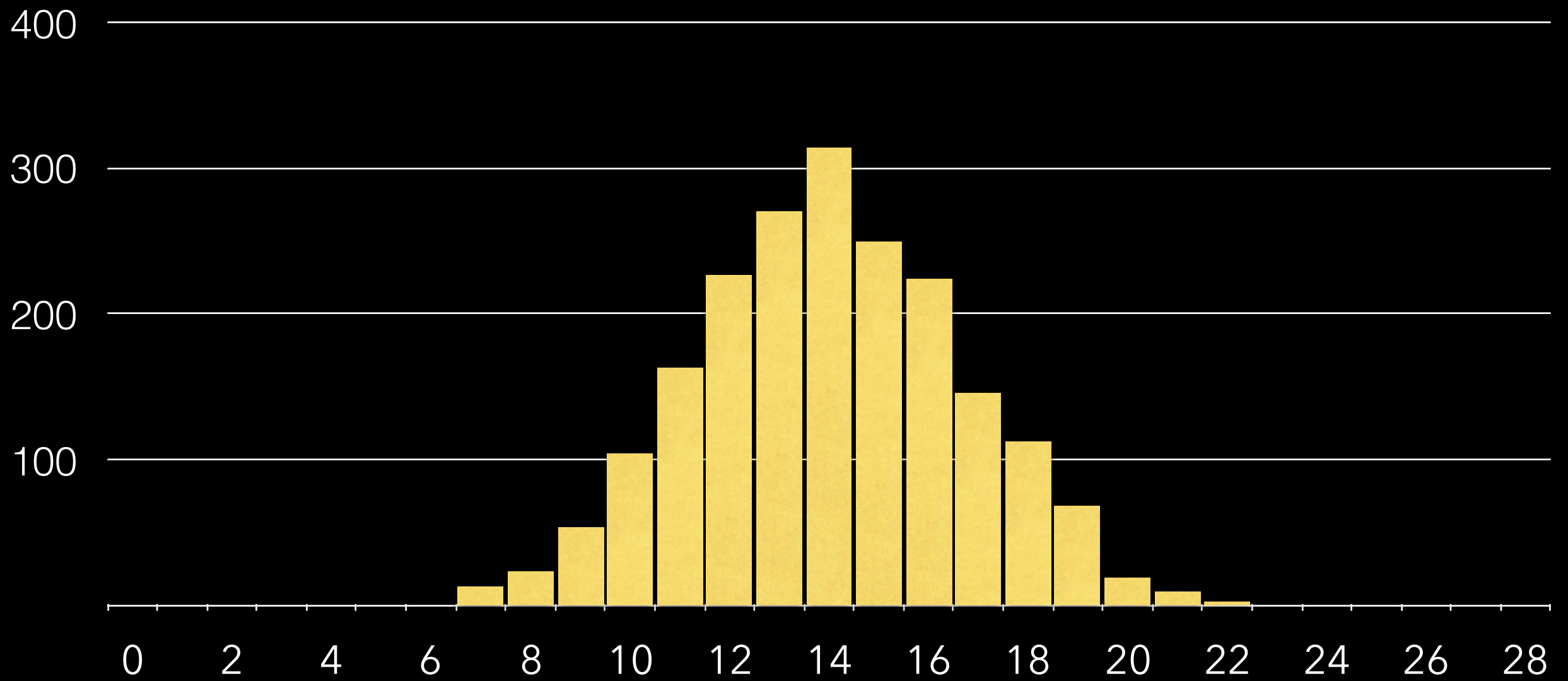




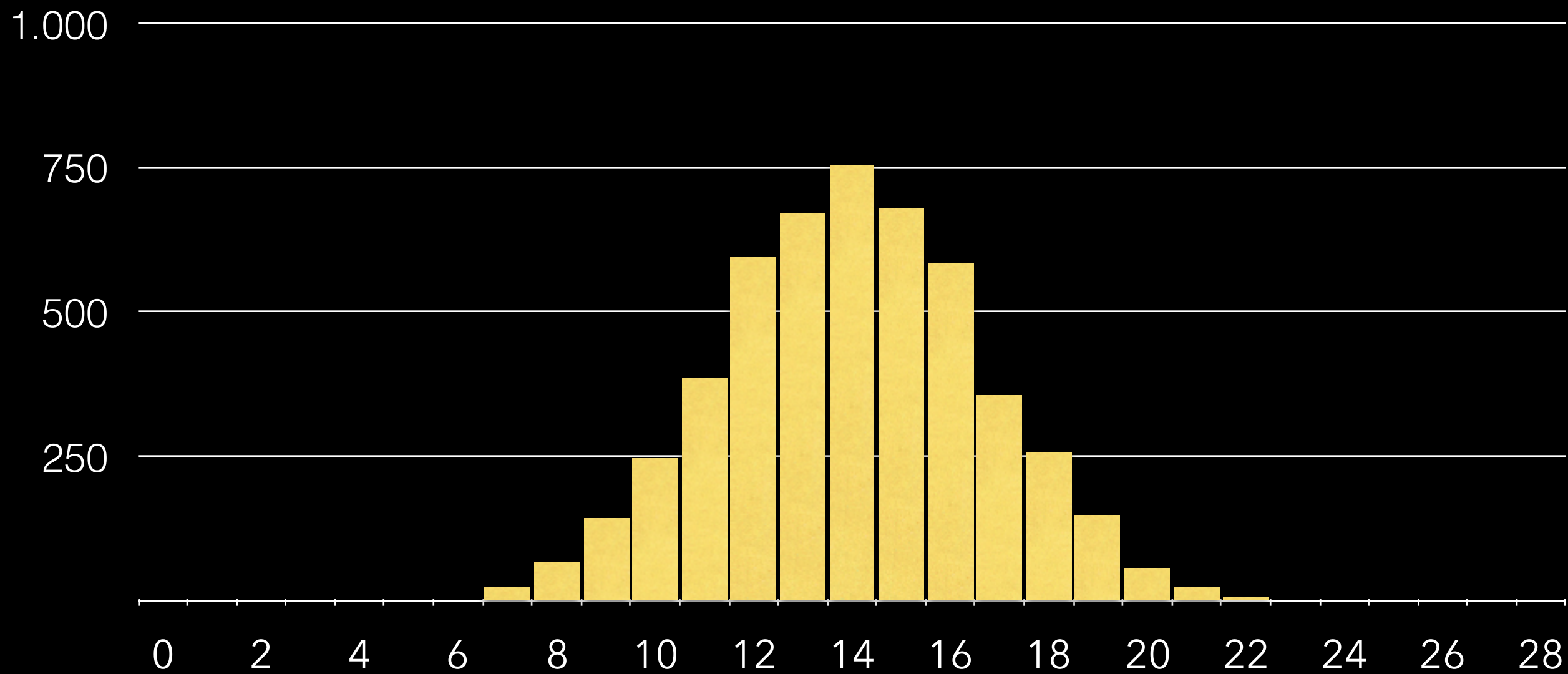




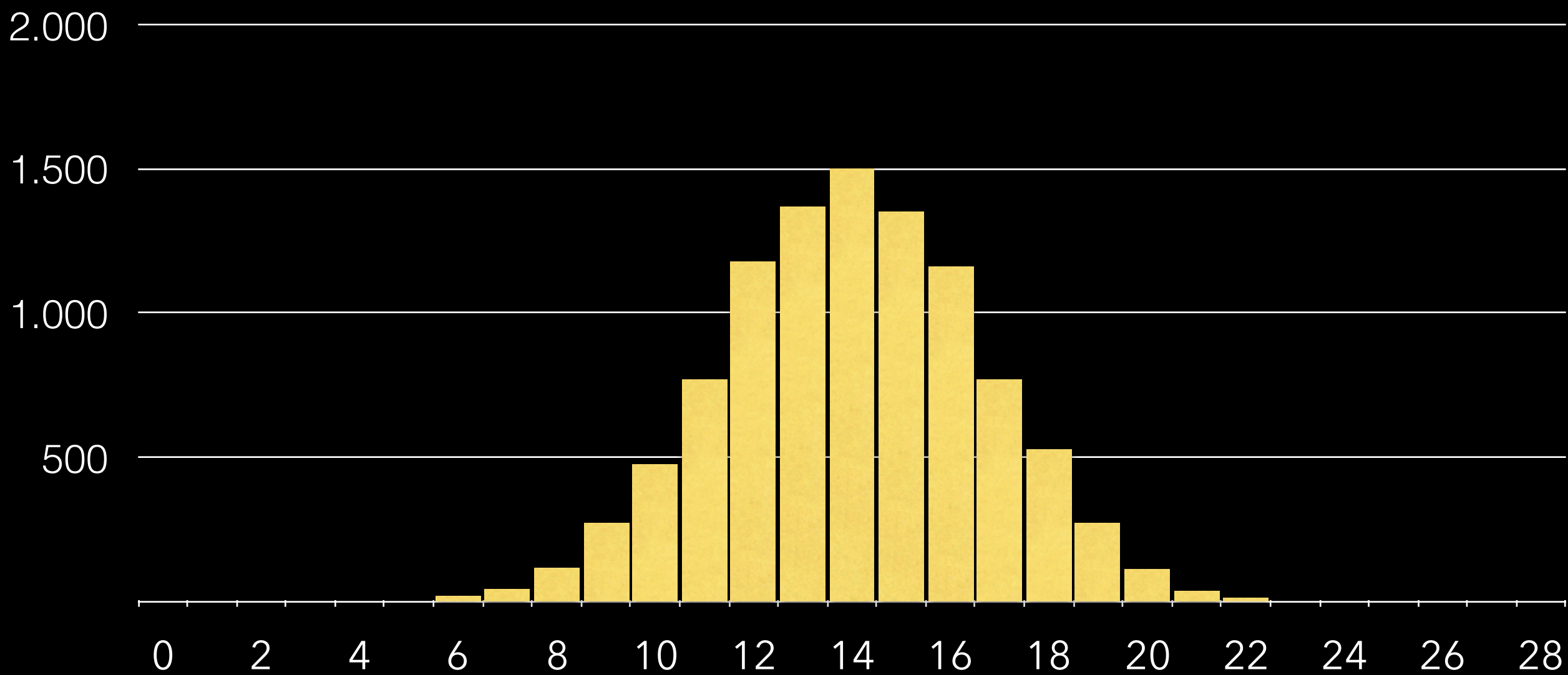




$n = 2000$



$n = 5000$



$n = 10000$

# Hekersko preverjanje hipotez

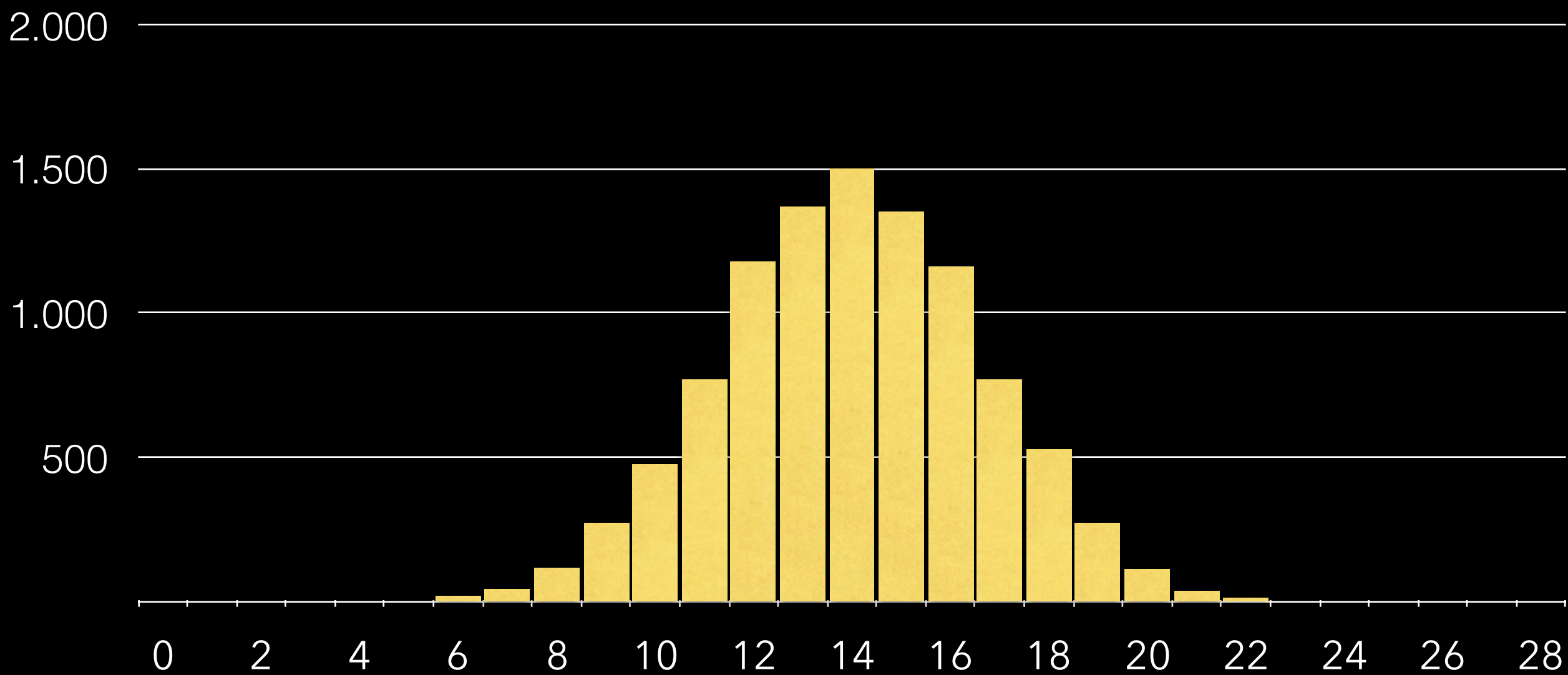
1. postavimo ničelno hipotezo



2. generiramo veliko dogodkov

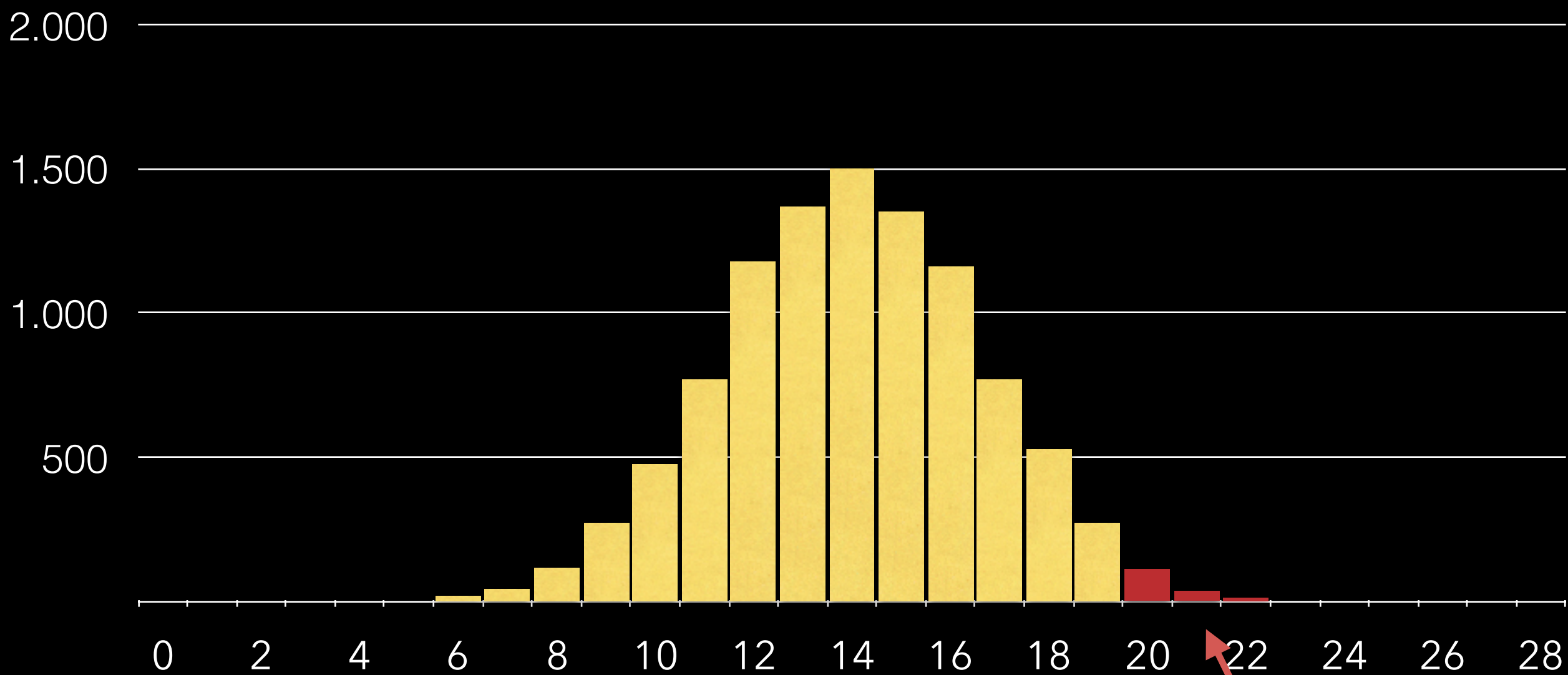


3. ocenimo  $p$ -vrednost



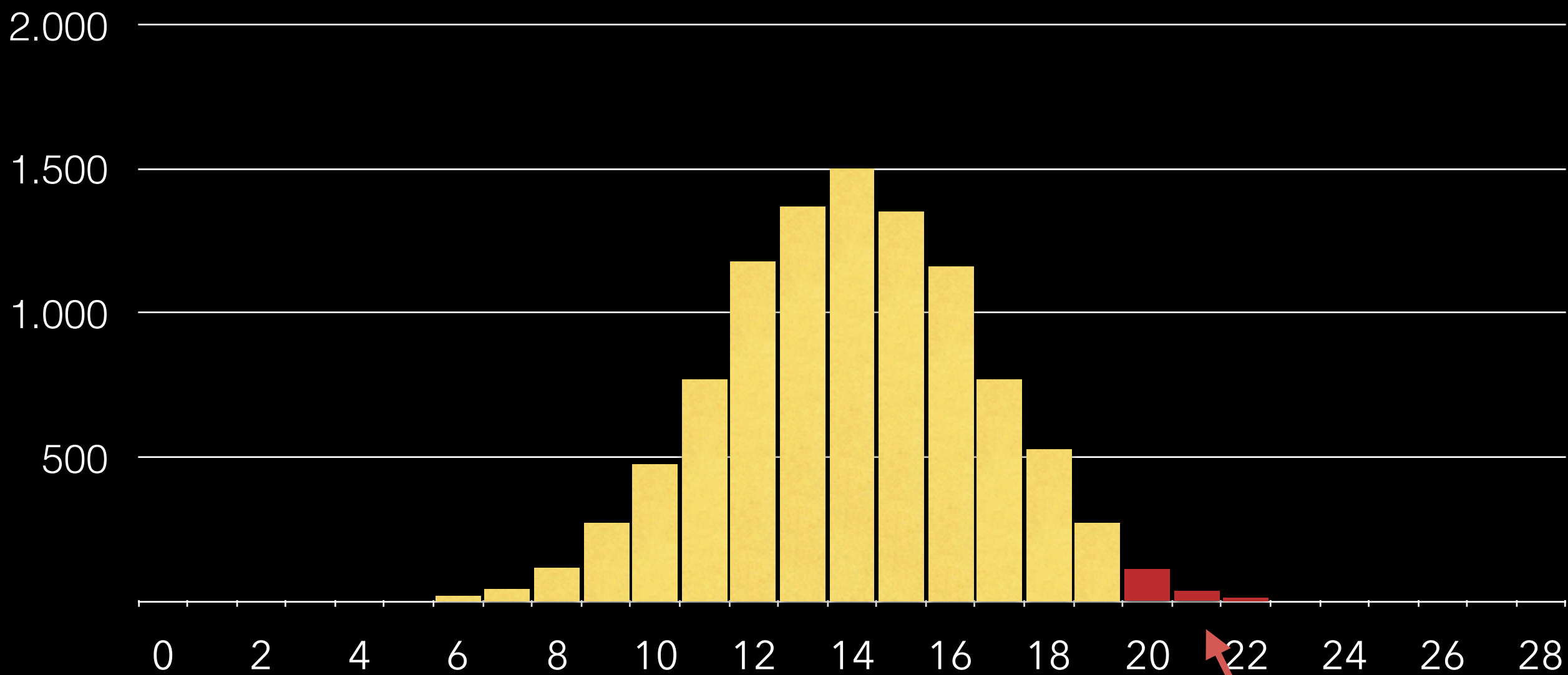
$n = 10000$





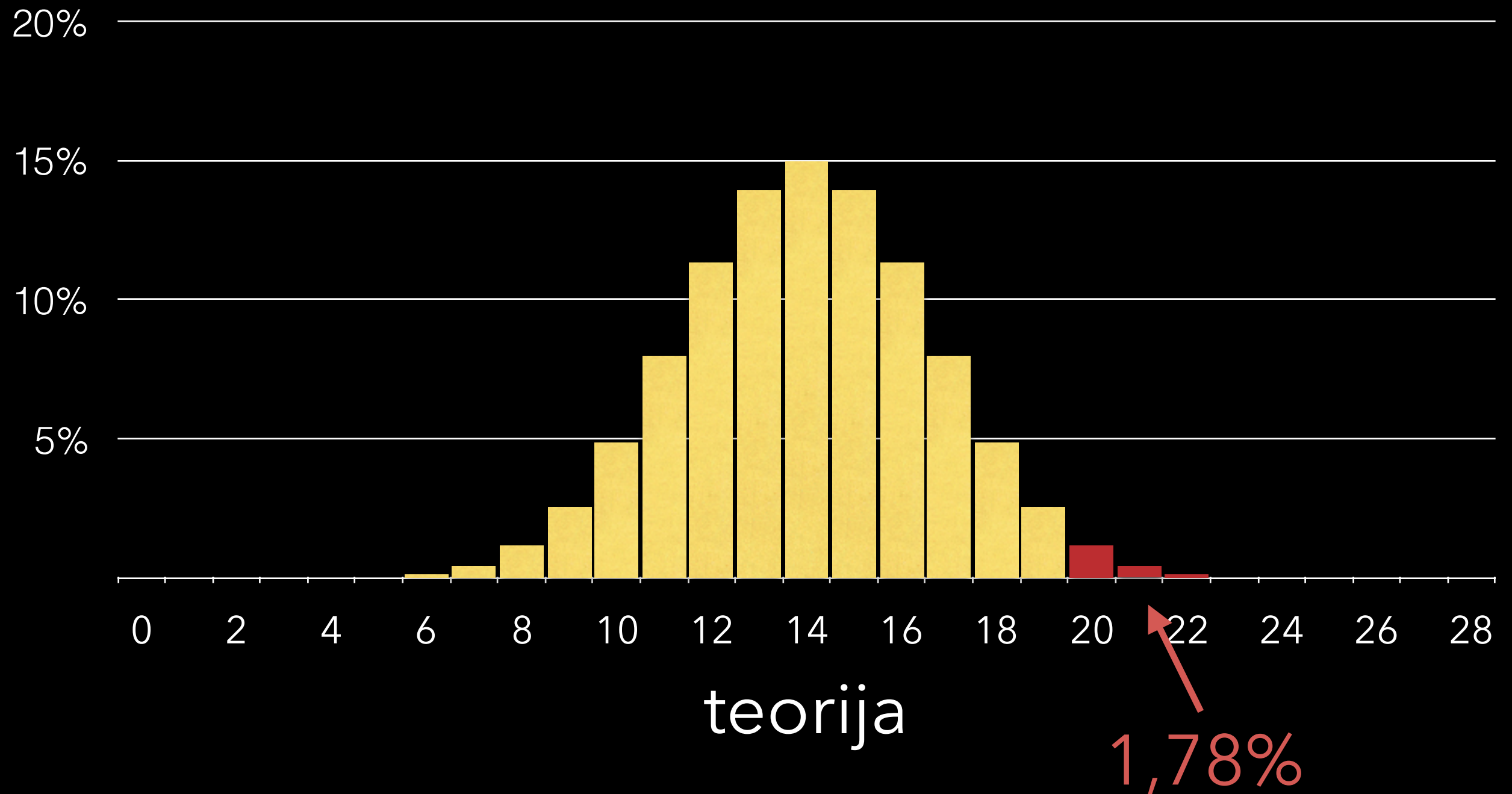
$n = 10000$

171



$n = 10000$

1,71%



# Hekersko preverjanje hipotez

1. postavimo ničelno hipotezo
2. generiramo veliko dogodkov
3. ocenimo  $p$ -vrednost





Kako izračunamo število padlih cifer v  $n$  metih?

Kako izračunamo število padlih cifer v  $n$  metih?

Števec cifer nastavimo na 0.

Kako izračunamo število padlih cifer v  $n$  metih?

Števec cifer nastavimo na 0.

Naslednje korake ponovimo  $n$ -krat:



Kako izračunamo število padlih cifer v  $n$  metih?

Števec cifer nastavimo na 0.

Naslednje korake ponovimo  $n$ -krat:

Vržemo kovanec.

Kako izračunamo število padlih cifer v  $n$  metih?

Števec cifer nastavimo na 0.

Naslednje korake ponovimo  $n$ -krat:

Vržemo kovanec.

Če je padla cifra:

Kako izračunamo število padlih cifer v  $n$  metih?

Števec cifer nastavimo na 0.

Naslednje korake ponovimo  $n$ -krat:

Vržemo kovanec.

Če je padla cifra:

števec cifer povečamo za 1.

Kako izračunamo število padlih cifer v  $n$  metih?

Števec cifer nastavimo na 0.

Naslednje korake ponovimo  $n$ -krat:

Vržemo kovanec.

Če je padla cifra:

števec cifer povečamo za 1.

Število cifer je ravno vrednost števca cifer.

Kako izračunamo število padlih cifer v **n** metih?

**Števec cifer** nastavimo na **0**.

Naslednje korake ponovimo **n**-krat:

Vržemo kovanec.

Če je padla cifra:

**števec cifer** povečamo za **1**.

Število cifer je ravno vrednost **števca cifer**.

```
def stevilo_padlih_cifer(stevilo_metov):
```

Kako izračunamo število padlih cifer v **n** metih?

**Števec cifer** nastavimo na **0**.

Naslednje korake ponovimo **n**-krat:

Vržemo kovanec.

Če je padla cifra:

**števec cifer** povečamo za **1**.

Število cifer je ravno vrednost **števca cifer**.

```
def stevilo_padlih_cifer(stevilo_metov):  
    stevec_cifer = 0
```

Kako izračunamo število padlih cifer v **n** metih?

**Števec cifer** nastavimo na **0**.

Naslednje korake ponovimo **n**-krat:

Vržemo kovanec.

Če je padla cifra:

**števec cifer** povečamo za **1**.

Število cifer je ravno vrednost **števca cifer**.

```
def stevilo_padlih_cifer(stevilo_metov):  
    stevec_cifer = 0  
    for _ in range(stevilo_metov):
```

Kako izračunamo število padlih cifer v **n** metih?

**Števec cifer** nastavimo na **0**.

Naslednje korake ponovimo **n**-krat:

Vržemo kovanec.

Če je padla cifra:

**števec cifer** povečamo za **1**.

Število cifer je ravno vrednost **števca cifer**.

```
def stevilo_padlih_cifer(stevilo_metov):  
    stevec_cifer = 0  
    for _ in range(stevilo_metov):  
        padlo_je = random.choice(['cifra', 'grb'])
```



Kako izračunamo število padlih cifer v  $n$  metih?

Števec cifer nastavimo na 0.

Naslednje korake ponovimo  $n$ -krat:

Vržemo kovanec.

Če je padla cifra:

števec cifer povečamo za 1.

Število cifer je ravno vrednost števca cifer.

```
def stevilo_padlih_cifer(stevilo_metov):  
    stevec_cifer = 0  
    for _ in range(stevilo_metov):  
        padlo_je = random.choice(['cifra', 'grb'])  
        if padlo_je == 'cifra':
```

Kako izračunamo število padlih cifer v  $n$  metih?

Števec cifer nastavimo na 0.

Naslednje korake ponovimo  $n$ -krat:

Vržemo kovanec.

Če je padla cifra:

števec cifer povečamo za 1.

Število cifer je ravno vrednost števca cifer.

```
def stevilo_padlih_cifer(stevilo_metov):  
    stevec_cifer = 0  
    for _ in range(stevilo_metov):  
        padlo_je = random.choice(['cifra', 'grb'])  
        if padlo_je == 'cifra':  
            stevec_cifer += 1
```

Kako izračunamo število padlih cifer v  $n$  metih?

Števec cifer nastavimo na 0.

Naslednje korake ponovimo  $n$ -krat:

Vržemo kovanec.

Če je padla cifra:

števec cifer povečamo za 1.

Število cifer je ravno vrednost števca cifer.

```
def stevilo_padlih_cifer(stevilo_metov):  
    stevec_cifer = 0  
    for _ in range(stevilo_metov):  
        padlo_je = random.choice(['cifra', 'grb'])  
        if padlo_je == 'cifra':  
            stevec_cifer += 1  
    return stevec_cifer
```

*praktični  
primer*

57	71	74	35	60	83	74	62
43	65	63	30	51	62	57	72
64	51	43	55	85	54	77	45
69	46	97	59	73	65	57	73

**fantje: 59,13 dekleeta: 63,82**

**razlika: 4,69**

# Hekersko preverjanje hipotez

1. postavimo ničelno hipotezo
2. generiramo veliko dogodkov
3. ocenimo  $p$ -vrednost

**ničelna hipoteza:**

ocene so

neodvisne od spola

# Hekersko preverjanje hipotez

1. postavimo ničelno hipotezo ✓
2. generiramo veliko dogodkov
3. ocenimo  $p$ -vrednost



57	71	74	35	60	83	74	62
43	65	63	30	51	62	57	72
64	51	43	55	85	54	77	45
69	46	97	59	73	65	57	73

**fantje: 59,13 dekleta: 63,82**

**razlika: 4,69**

46	85	73	77	51	43	63	45
57	55	60	59	35	57	71	74
54	62	83	57	73	72	43	65
65	51	62	69	74	64	97	30

**fantje: 58,47** **dekleta: 64,41**

**razlika: 5,95**

45	74	54	72	62	57	57	85
64	69	71	55	30	51	77	51
74	57	35	46	65	63	73	62
83	60	97	59	73	43	65	43

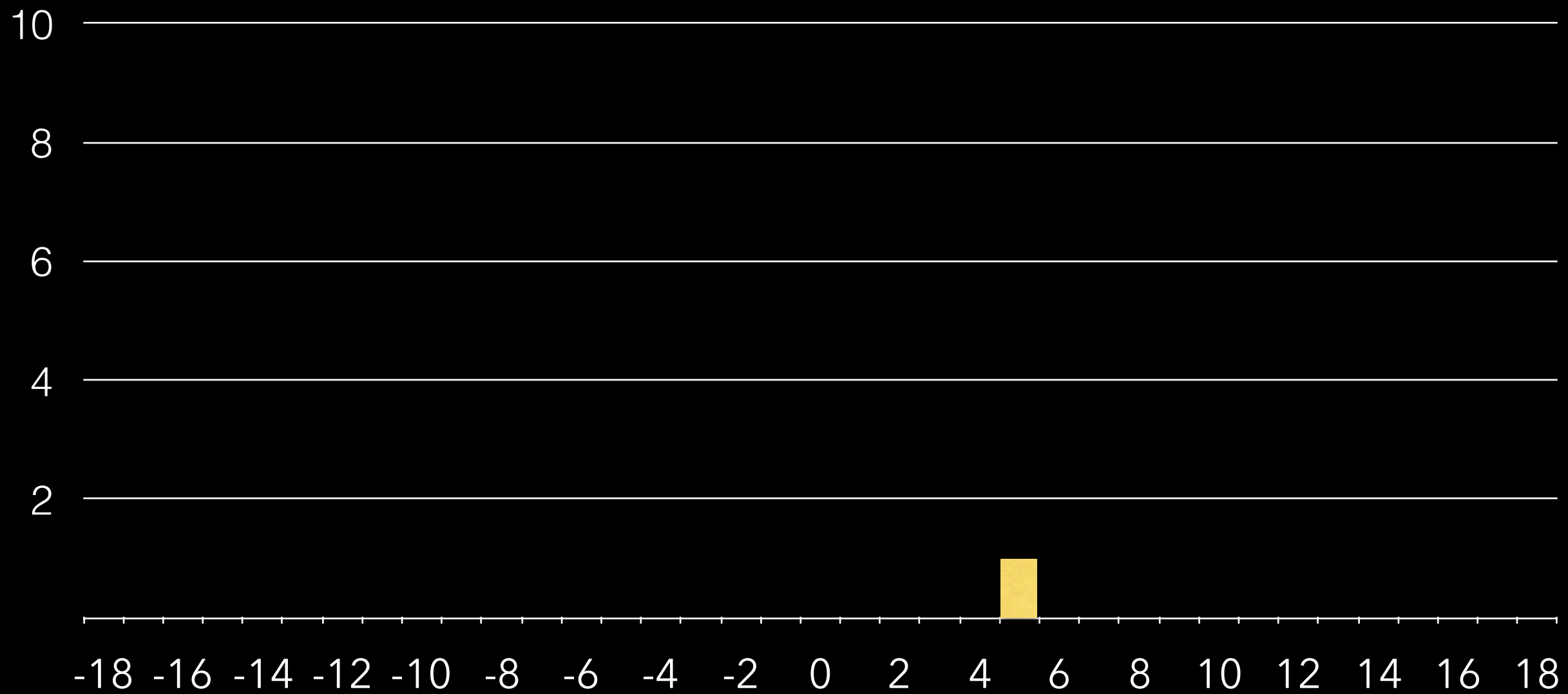
**fantje: 61,53 dekleeta: 61,71**

**razlika: 0,18**

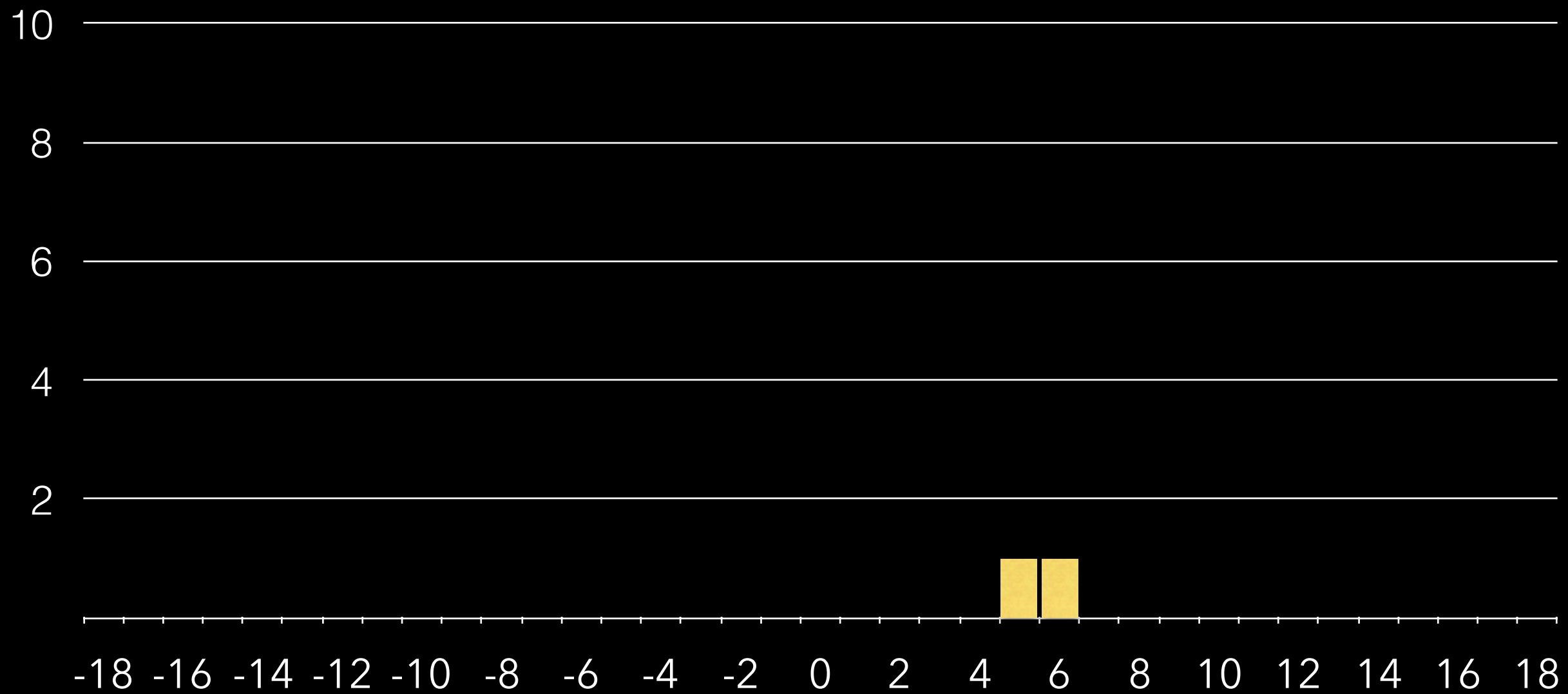
62	73	83	46	57	72	55	63
77	64	30	43	59	97	71	45
51	69	73	54	57	85	62	74
65	35	57	43	60	51	74	65

**fantje: 63,64 dekleta: 60,00**

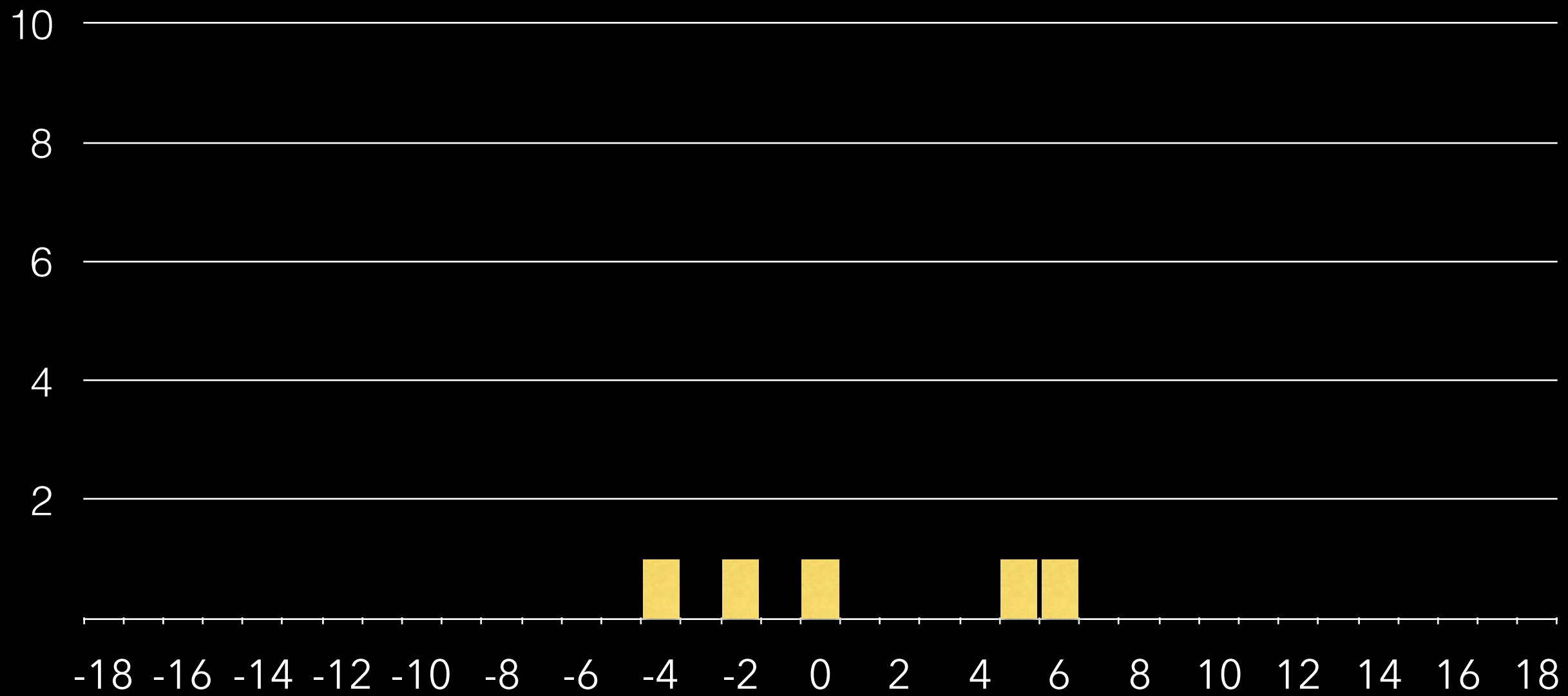
**razlika: -3,64**



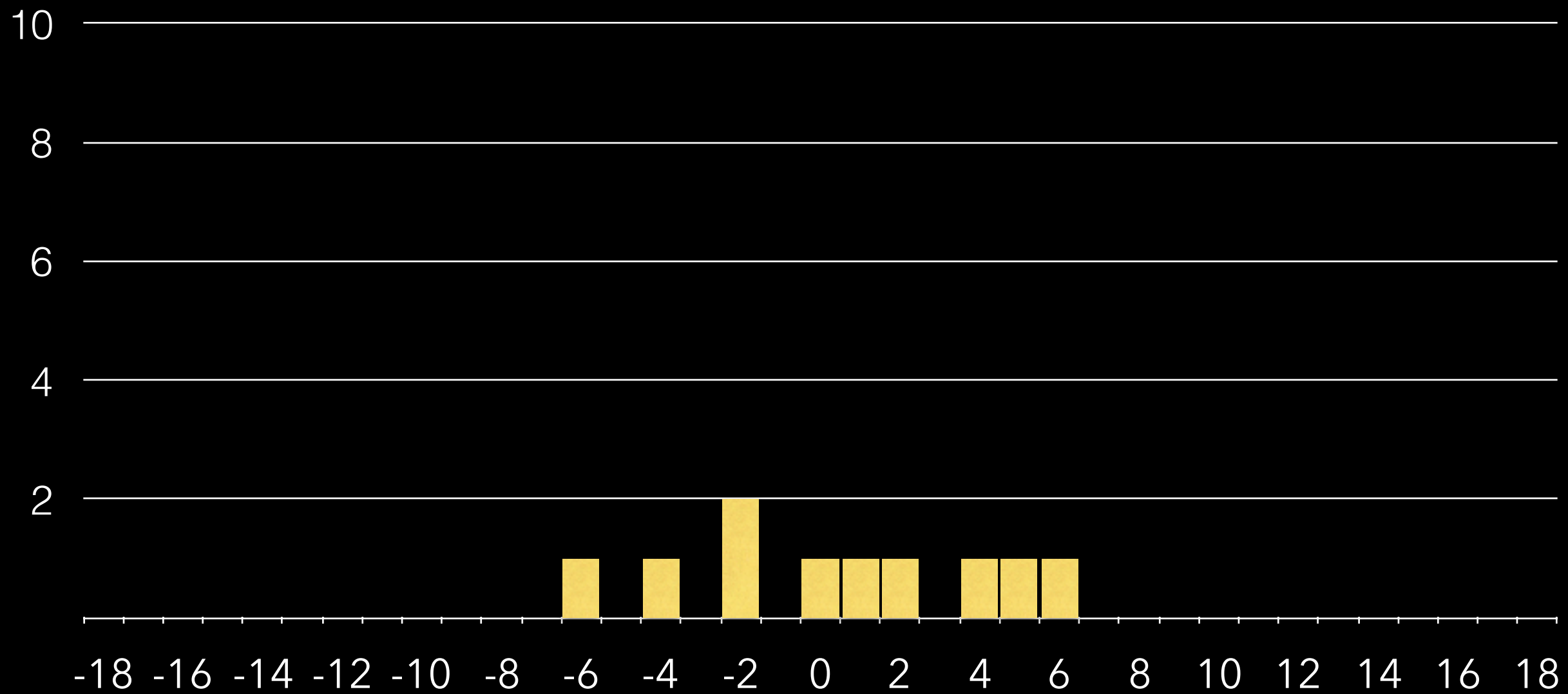
$n = 1$



$n = 2$

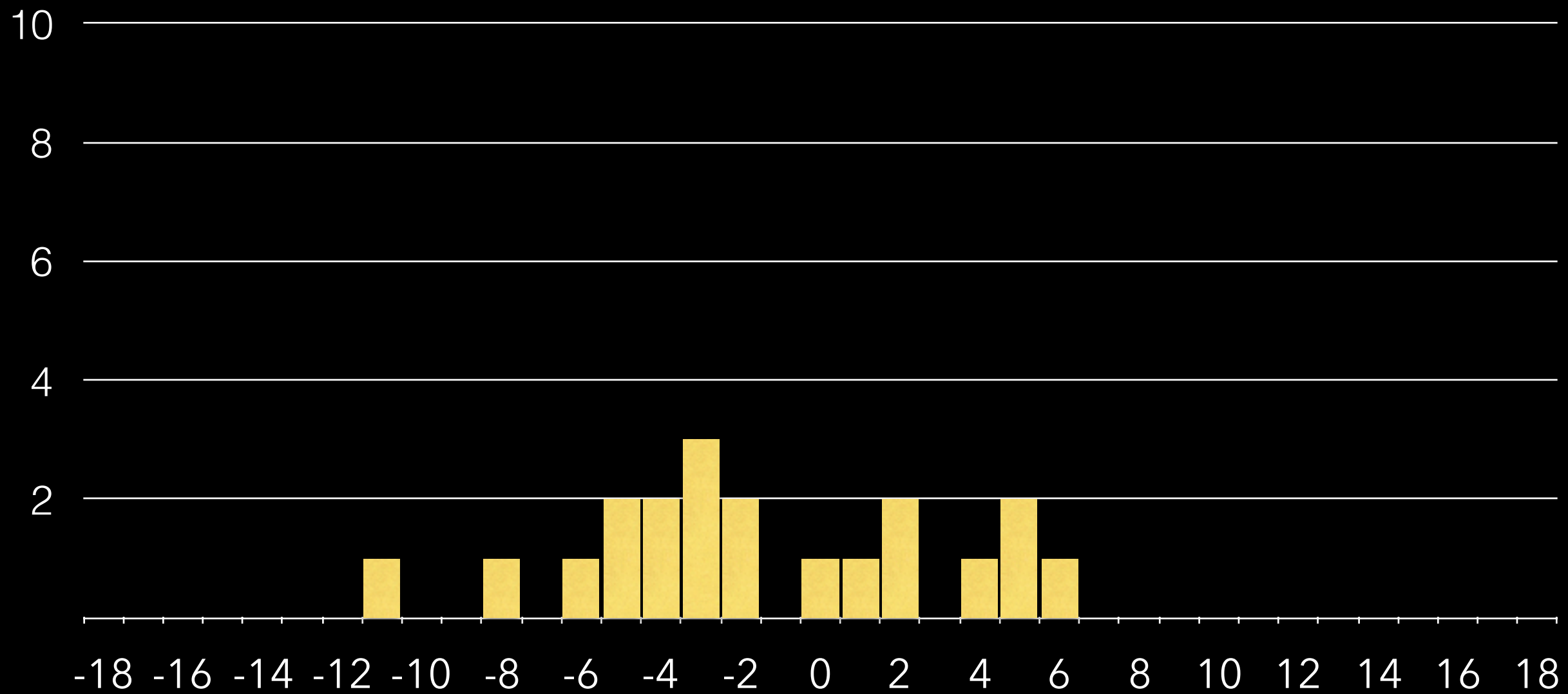


$n = 5$

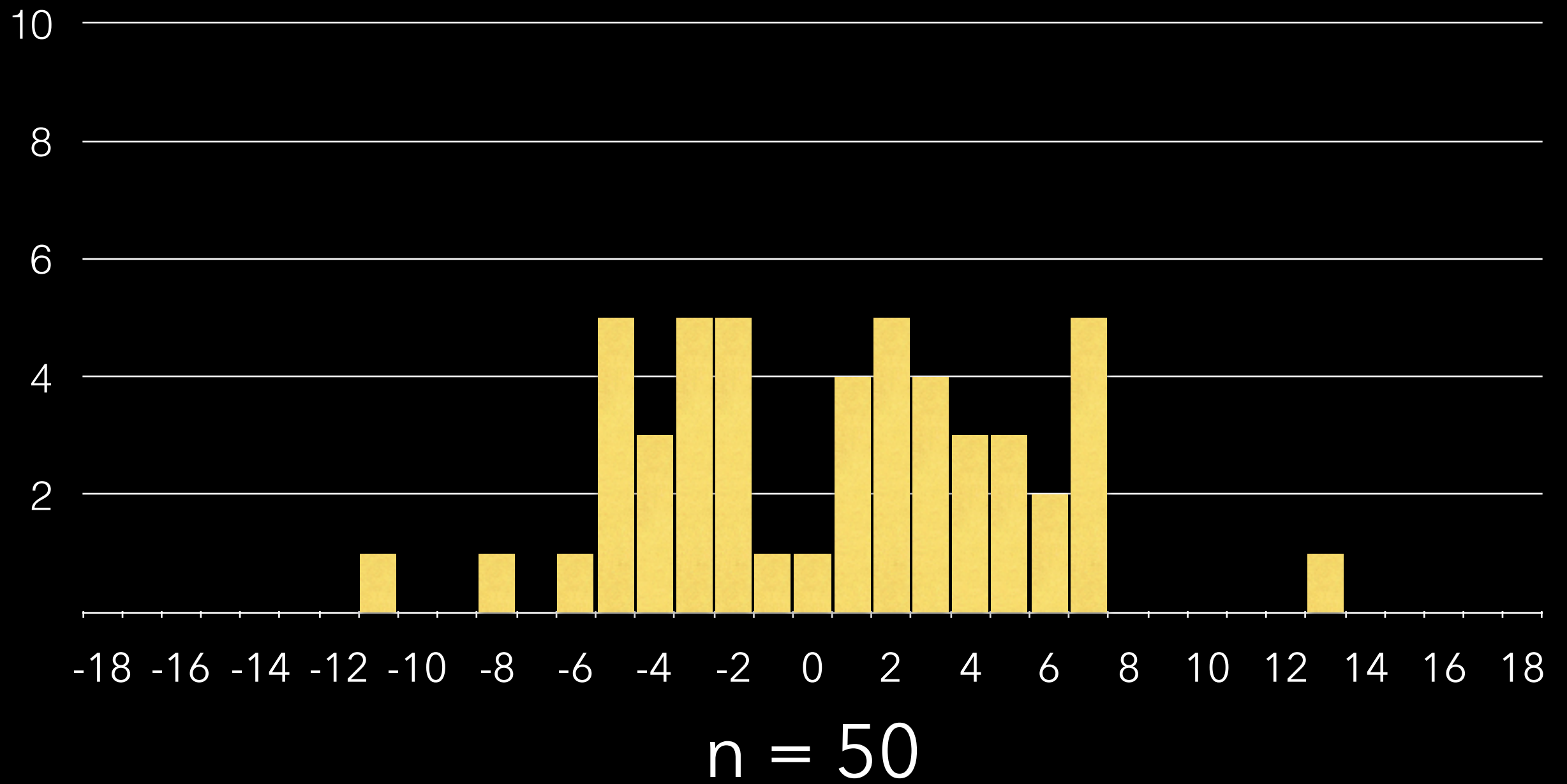


$n = 10$

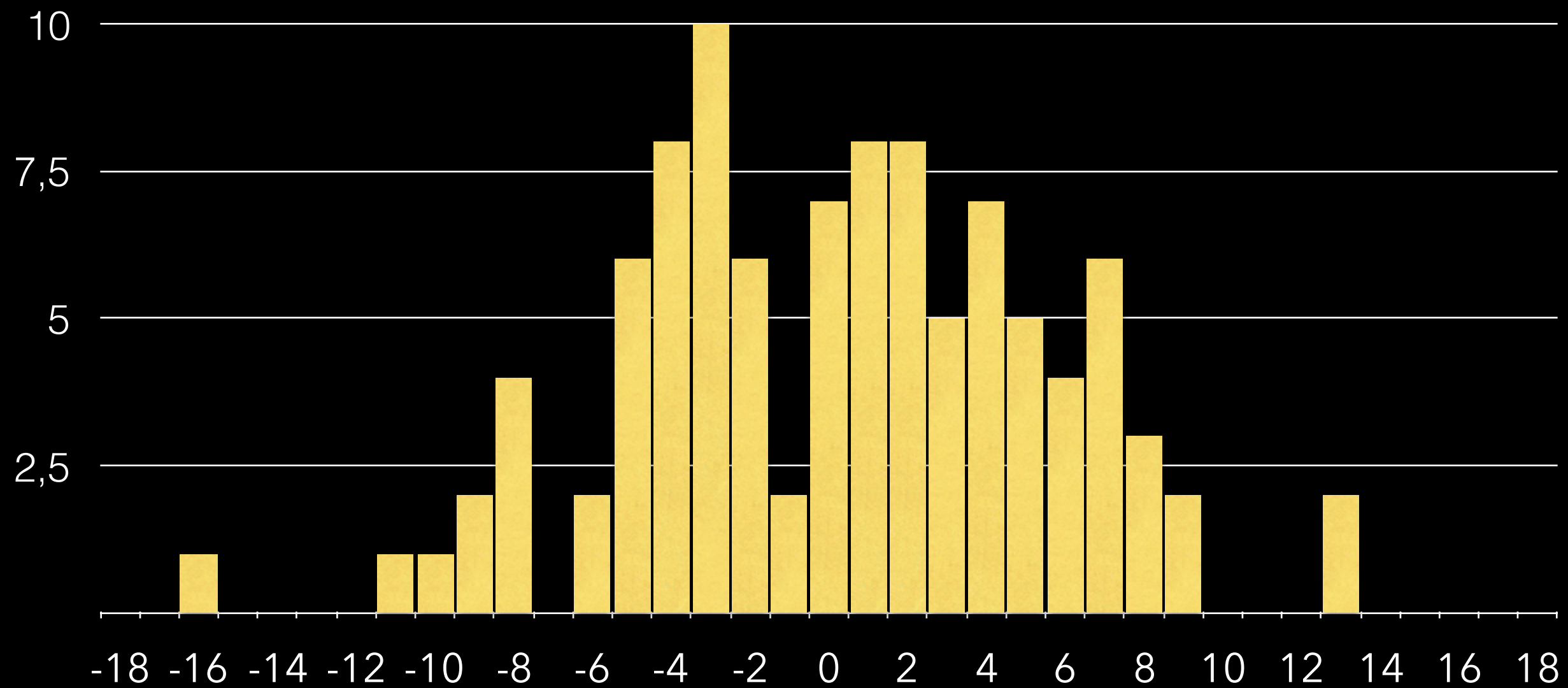




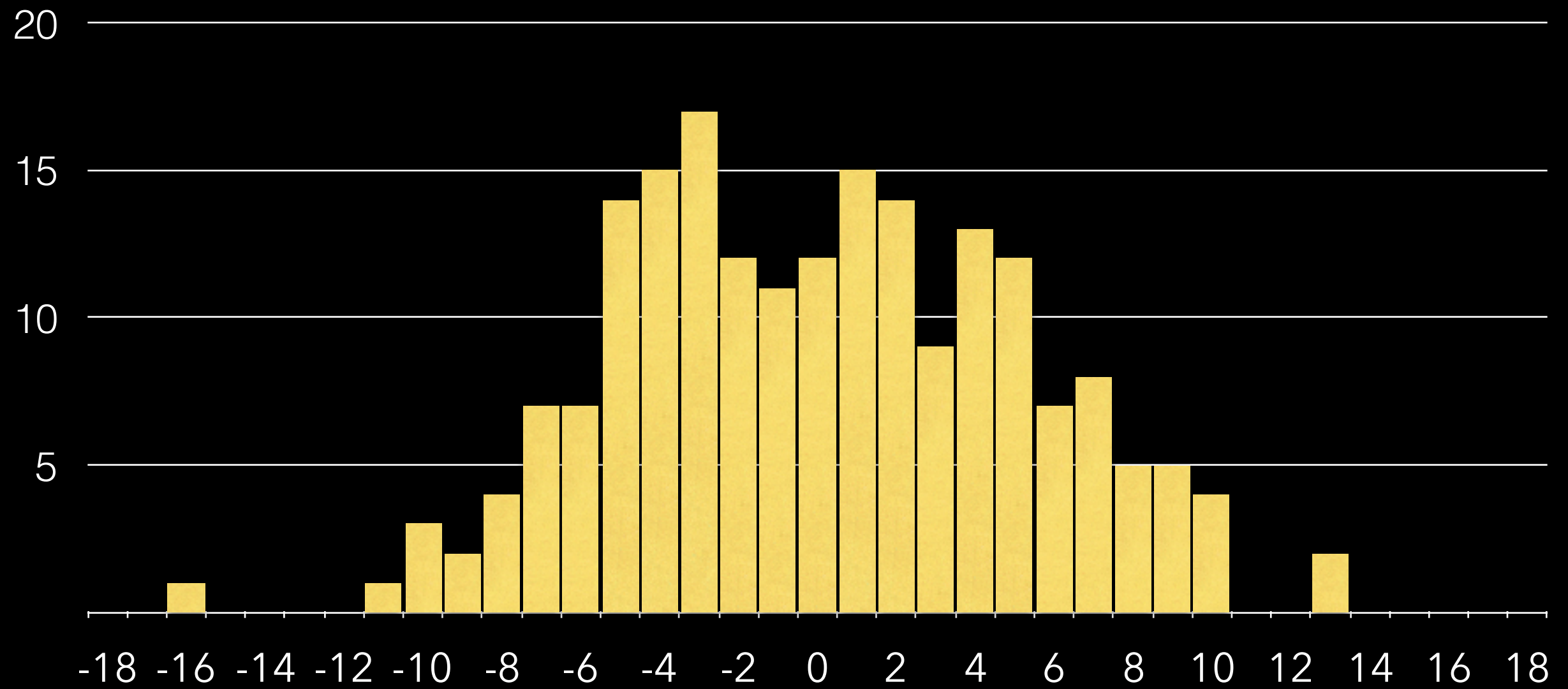
$n = 20$



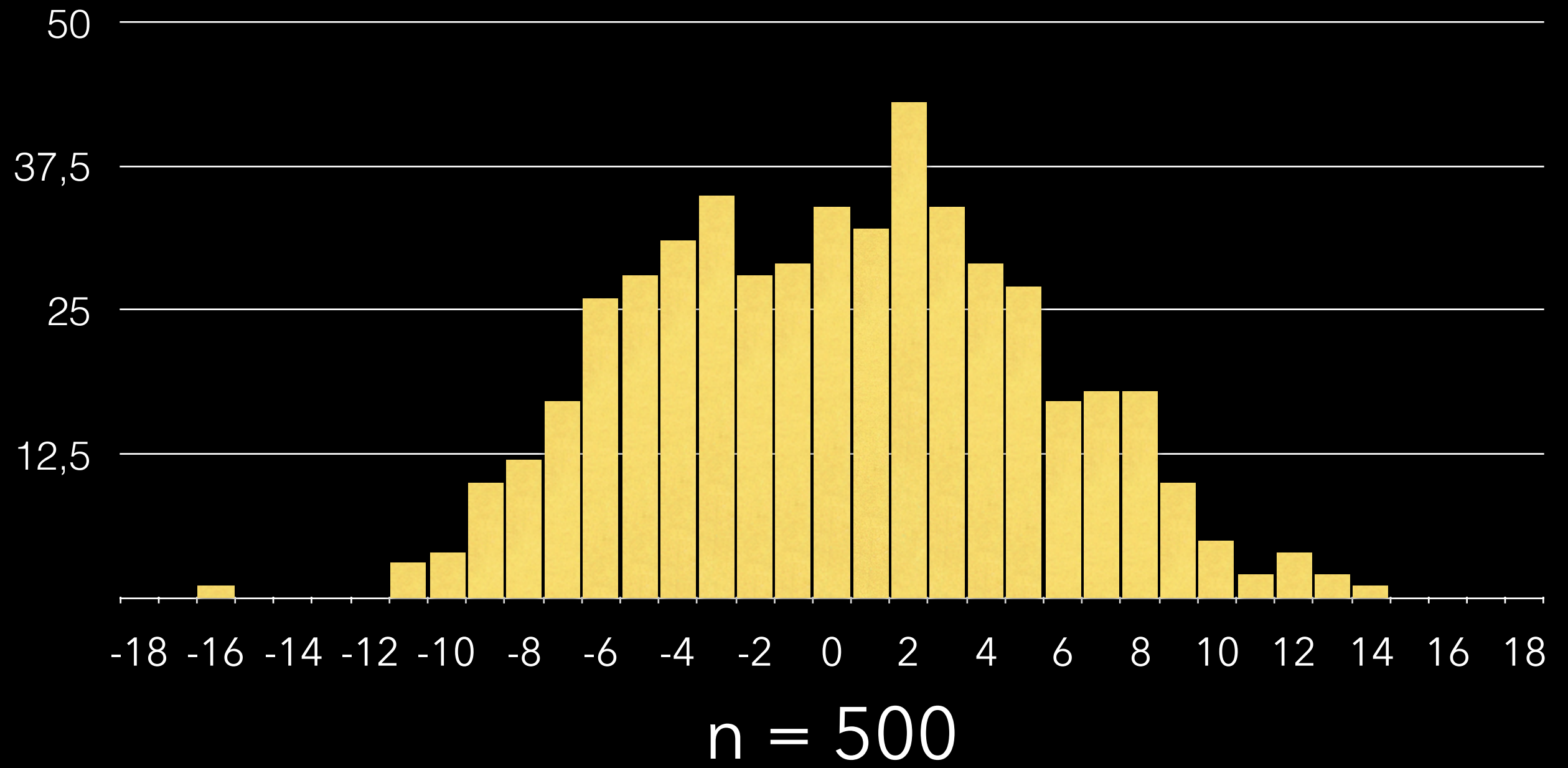


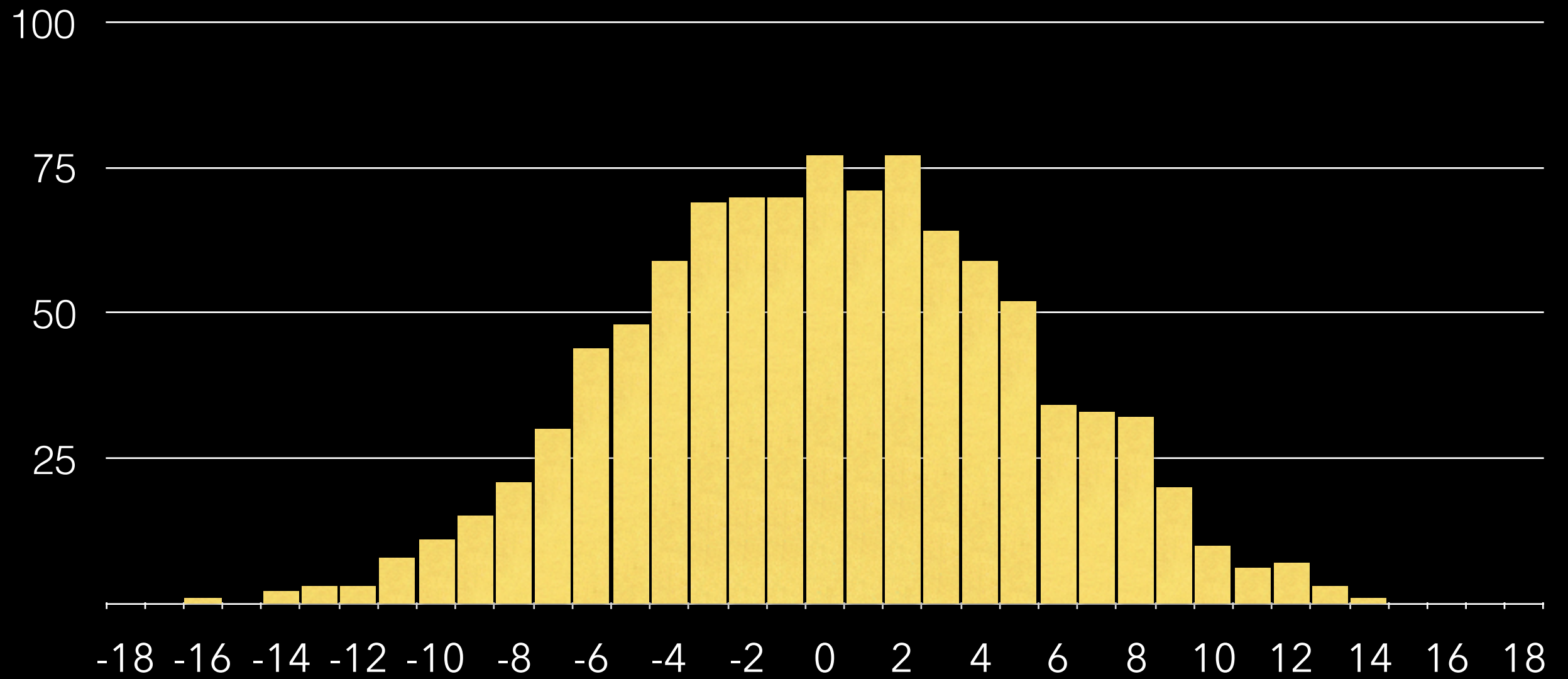


$n = 100$

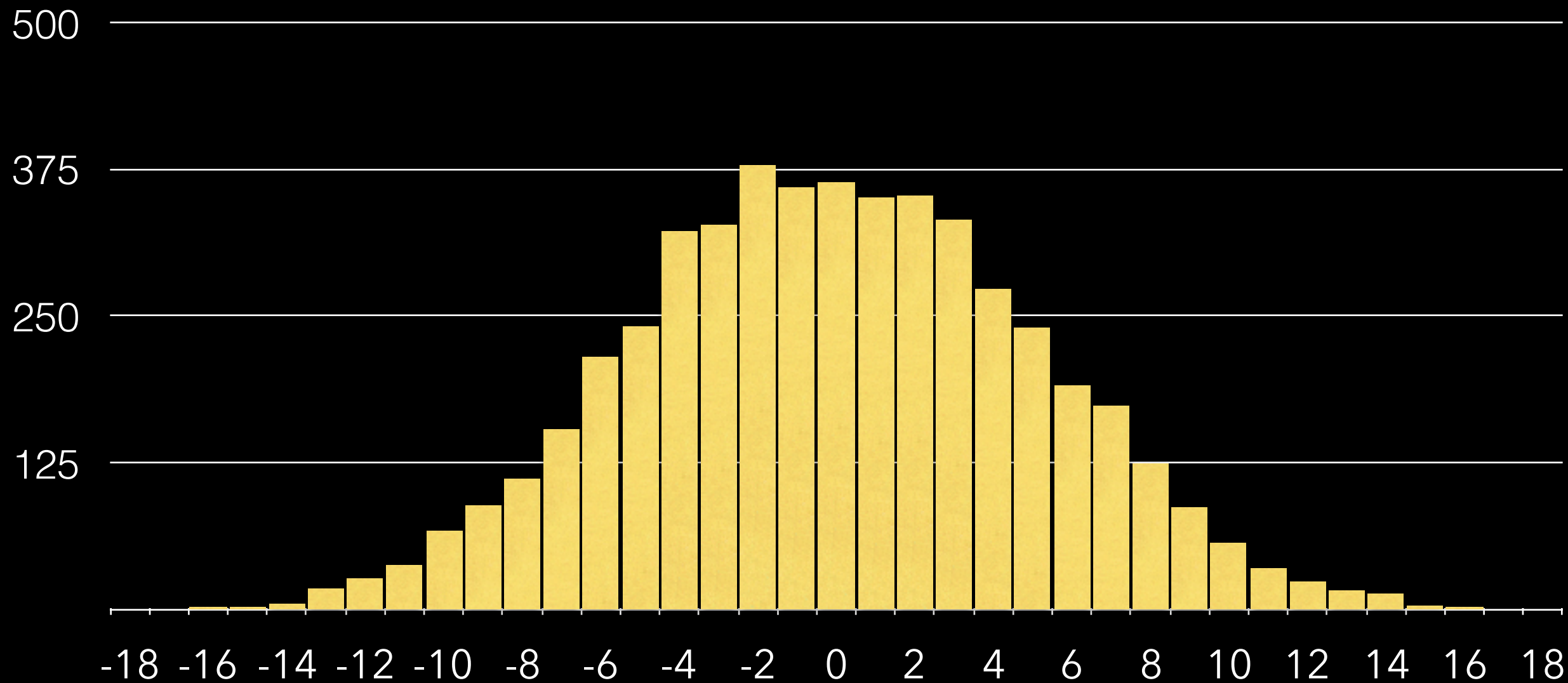


$n = 200$



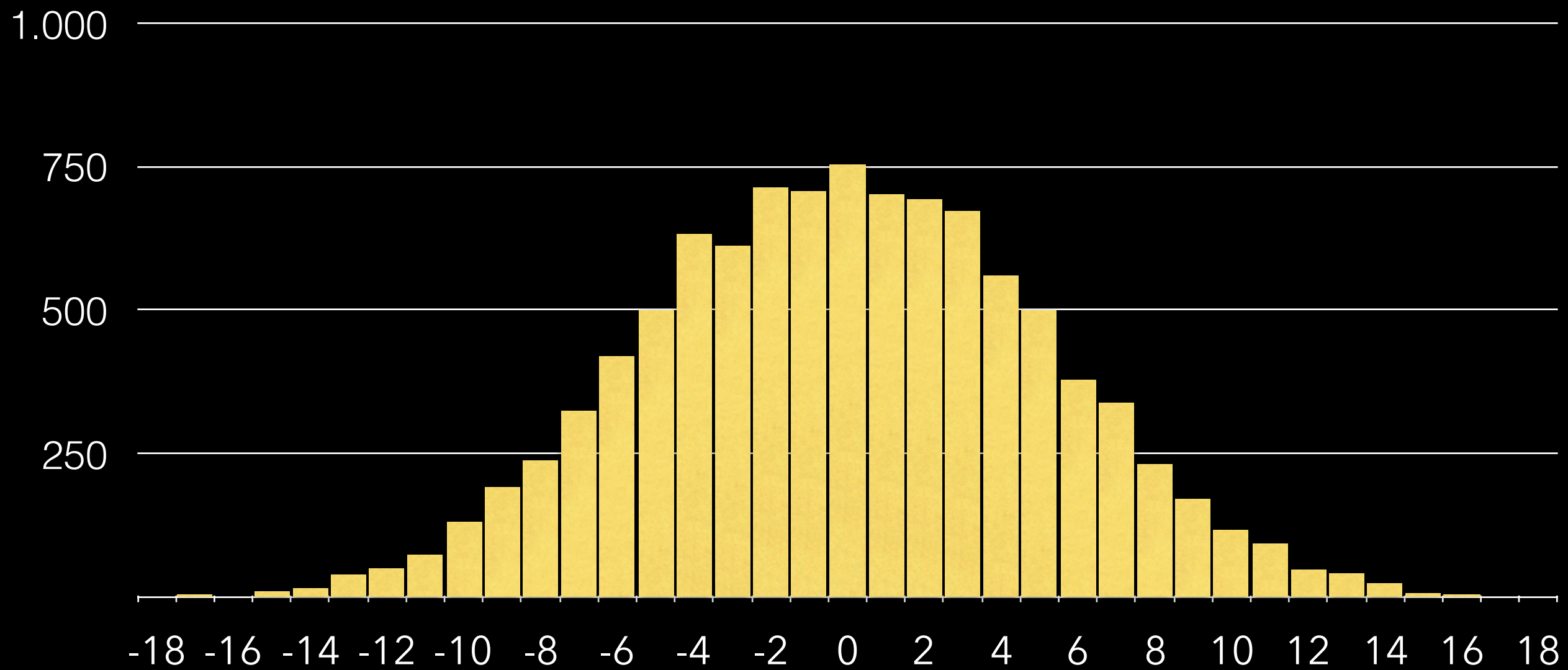


$n = 1000$



$n = 5000$





$n = 10000$

# Hekersko preverjanje hipotez

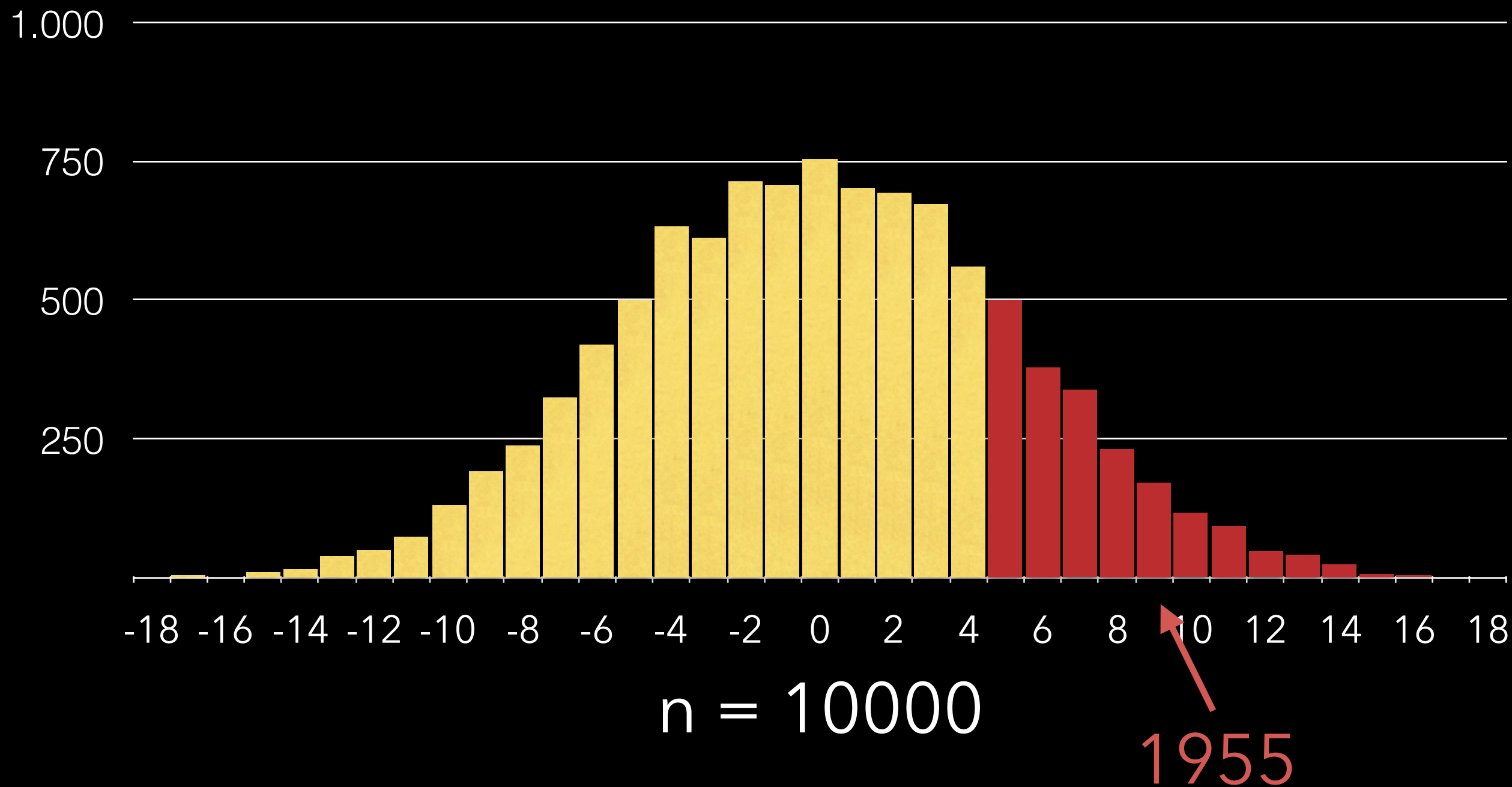
1. postavimo ničelno hipotezo

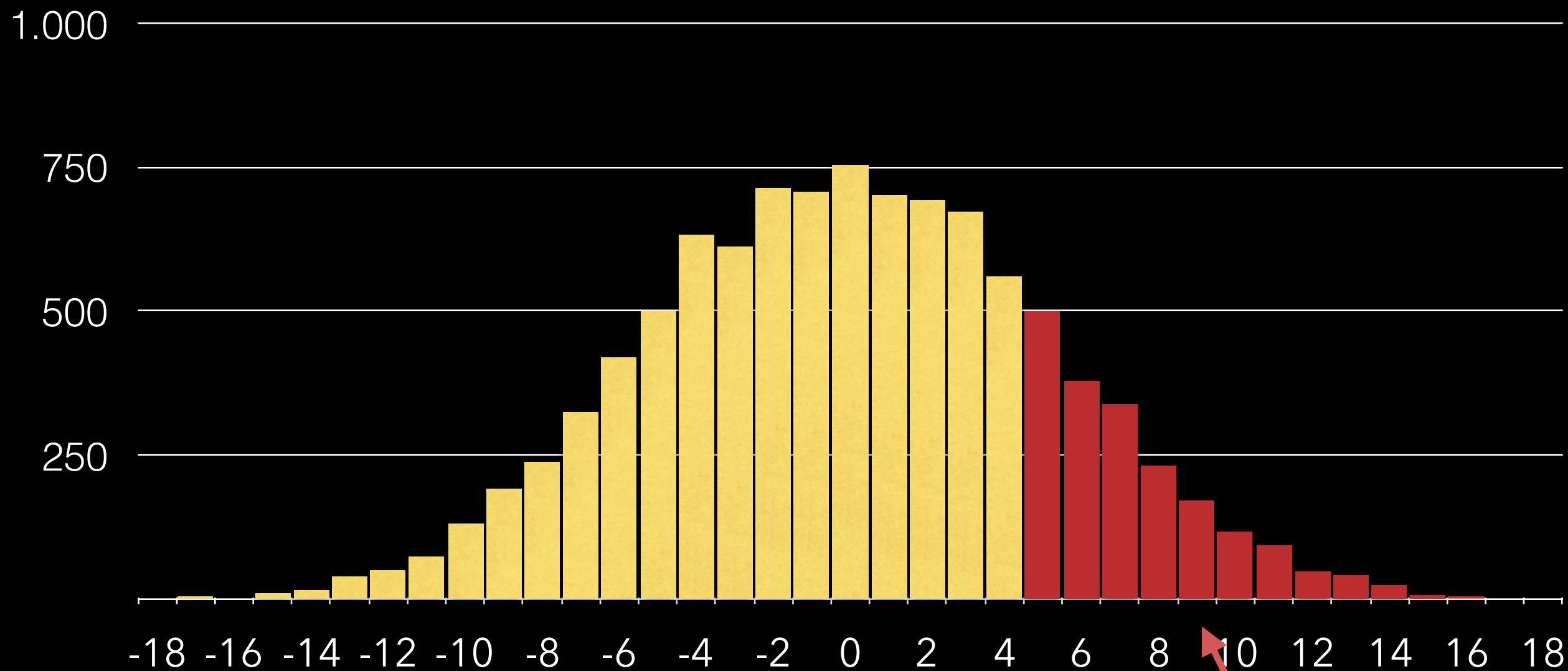


2. generiramo veliko dogodkov



3. ocenimo  $p$ -vrednost





$n = 10000$

19,55%

# Hekersko preverjanje hipotez

1. postavimo ničelno hipotezo
2. generiramo veliko dogodkov
3. ocenimo  $p$ -vrednost



*praktični  
primer*

*intermezzo*

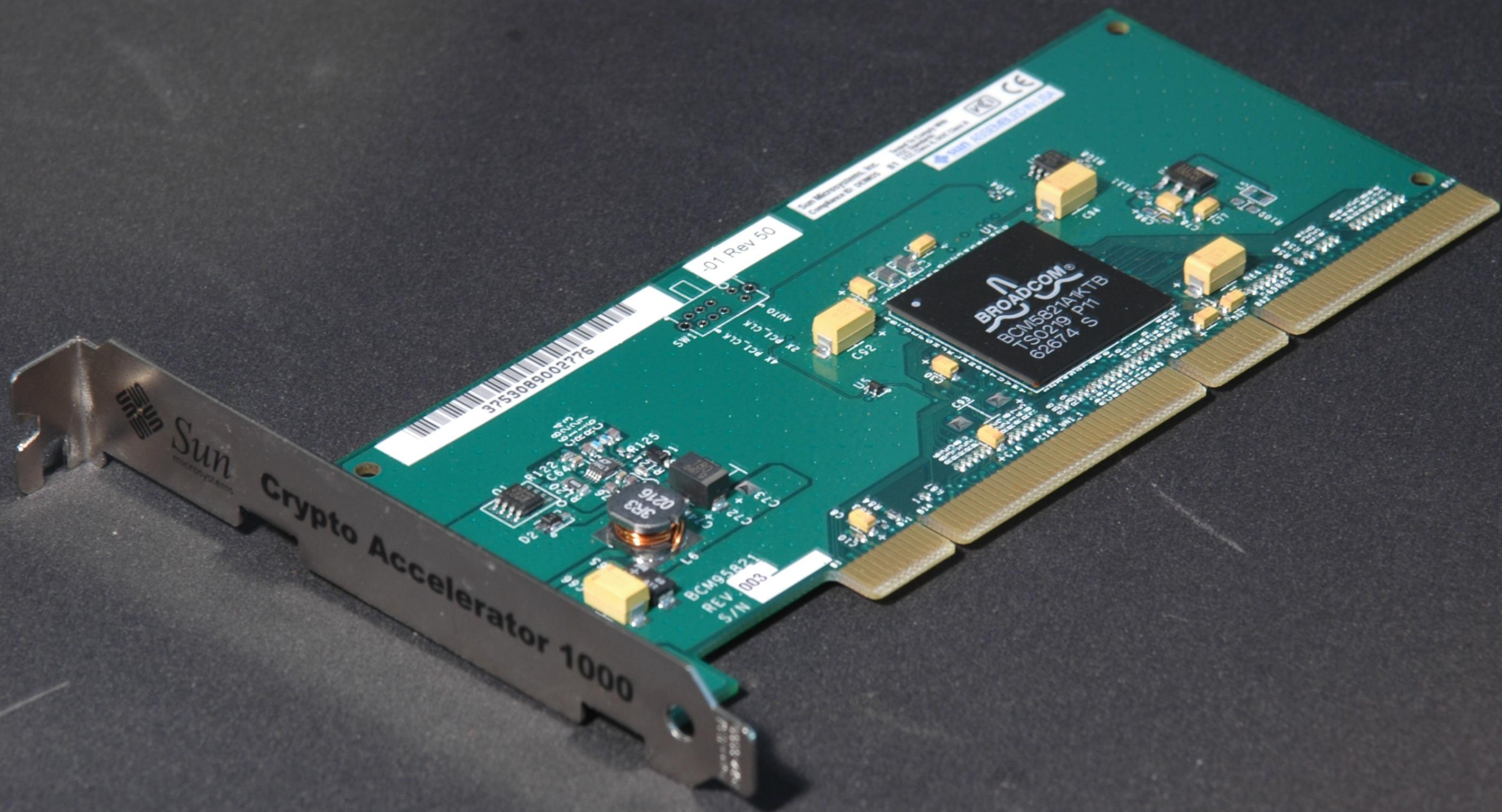
Kako dobimo  
naključne  
vrednosti?







```
int getRandomNumber()  
{  
    return 4; // chosen by fair dice roll.  
              // guaranteed to be random.  
}
```



Kako dobimo  
psevdonaključne  
vrednosti?

# Metoda srednjega kvadrata

# Metoda srednjega kvadrata

seme

# Metoda srednjega kvadrata

seme

$$1234^2 = 01522756$$

# Metoda srednjega kvadrata

$$1234^2 = 01522756$$

$$5227^2 = 27321529$$



# Metoda srednjega kvadrata

$$1234^2 = 01522756$$

$$5227^2 = 27321529$$

$$3215^2 = 10336225$$



# Metoda srednjega kvadrata

$$1234^2 = 01522756$$

$$5227^2 = 27321529$$

$$3215^2 = 10336225$$

$$3362^2 = 11303044$$

# Metoda srednjega kvadrata

$$1234^2 = 01522756$$

$$5227^2 = 27321529$$

$$3215^2 = 10336225$$

$$3362^2 = 11303044$$

$$3030^2 = 09180900$$

# Metoda srednjega kvadrata

# Metoda srednjega kvadrata

$$2916^2 = 08503056$$

# Metoda srednjega kvadrata

$$2916^2 = 08503056$$

$$5030^2 = 25300900$$

# Metoda srednjega kvadrata

$$2916^2 = 08503056$$

$$5030^2 = 25300900$$

$$3009^2 = 09054081$$

# Metoda srednjega kvadrata

$$2916^2 = 08503056$$

$$5030^2 = 25300900$$

$$3009^2 = 09054081$$

$$0540^2 = 00291600$$

# Metoda srednjega kvadrata

$$2916^2 = 08503056$$

$$5030^2 = 25300900$$

$$3009^2 = 09054081$$

$$0540^2 = 00291600$$

$$2916^2 = 08503056$$



# Metoda srednjega kvadrata

# Metoda srednjega kvadrata

$$3792^2 = 14379264$$

$$3792^2 = 14379264$$

$$3792^2 = 14379264$$

$$3792^2 = 14379264$$

$$3792^2 = 14379264$$

$$3792^2 = 14379264$$

$$3792^2 = 14379264$$

# Linearni kongruenčni generator

# Linearni kongruenčni generator

$$123 \cdot 1234 + 567 = 2349 \pmod{10000}$$

# Linearni kongruenčni generator

$$123 \cdot 1234 + 567 = 2349 \pmod{10000}$$

$$123 \cdot 2349 + 567 = 9494 \pmod{10000}$$

# Linearni kongruenčni generator

$$123 \cdot 1234 + 567 = 2349 \pmod{10000}$$

$$123 \cdot 2349 + 567 = 9494 \pmod{10000}$$

$$123 \cdot 9494 + 567 = 8329 \pmod{10000}$$

# Linearni kongruenčni generator

$$123 \cdot 1234 + 567 = 2349 \pmod{10000}$$

$$123 \cdot 2349 + 567 = 9494 \pmod{10000}$$

$$123 \cdot 9494 + 567 = 8329 \pmod{10000}$$

$$123 \cdot 8329 + 567 = 5034 \pmod{10000}$$

# Linearni kongruenčni generator

$$123 \cdot 1234 + 567 = 2349 \pmod{10000}$$

$$123 \cdot 2349 + 567 = 9494 \pmod{10000}$$

$$123 \cdot 9494 + 567 = 8329 \pmod{10000}$$

$$123 \cdot 8329 + 567 = 5034 \pmod{10000}$$

$$123 \cdot 5034 + 567 = 9749 \pmod{10000}$$



# Linearni kongruenčni generator

$$123 \cdot 1234 + 567 = 2349 \pmod{10000}$$

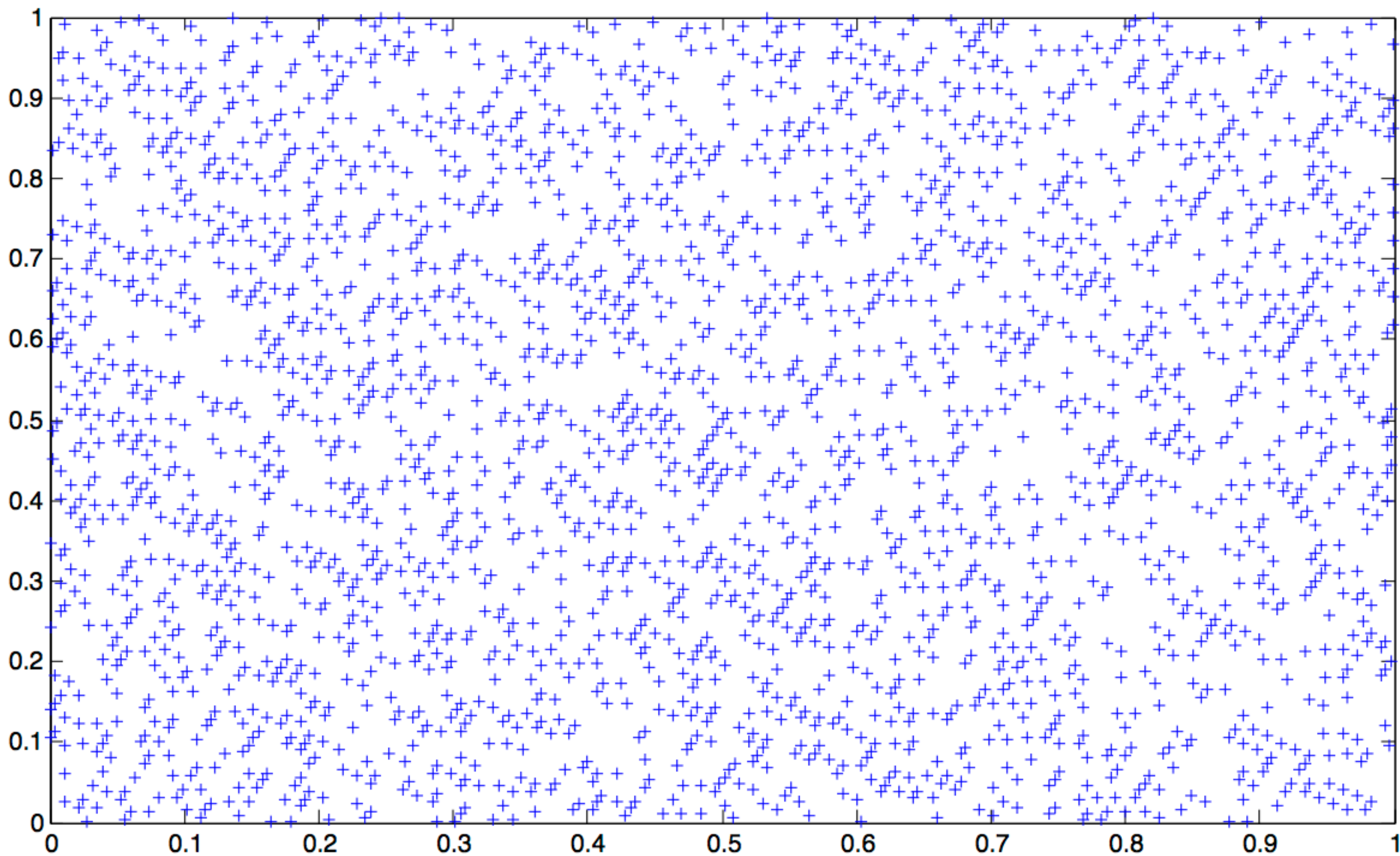
$$123 \cdot 2349 + 567 = 9494 \pmod{10000}$$

$$123 \cdot 9494 + 567 = 8329 \pmod{10000}$$

$$123 \cdot 8329 + 567 = 5034 \pmod{10000}$$

$$123 \cdot 5034 + 567 = 9749 \pmod{10000}$$

$$x_{n+1} = 1103515245 \cdot x_n + 12345 \pmod{2^{31}}$$



generator psevdonaključnih števil



naključno naravno število med 0 in  $2^{31}$

(naključno naravno število med 0 in  $2^{31}$ ) mod  $(n + 1)$



naključno naravno število med 0 in  $n$

$m + (\text{naključno naravno število med } 0 \text{ in } n - m)$



naključno naravno število med  $m$  in  $n$

naključno naravno število med 1 in  $n$



naključna izbira iz seznama dolžine  $n$

$(\text{naključno naravno število med } 0 \text{ in } 2^{31}) / 2^{31}$



naključno realno število med 0 in 1

$(b - a)(\text{naključno realno število med } 0 \text{ in } 1) + a$



naključno realno število med  $a$  in  $b$



(naključno realno število med 0 in 1)  $< p$



izid naključnega dogodka z verjetnostjo  $p$

nepoštén kovanec



?

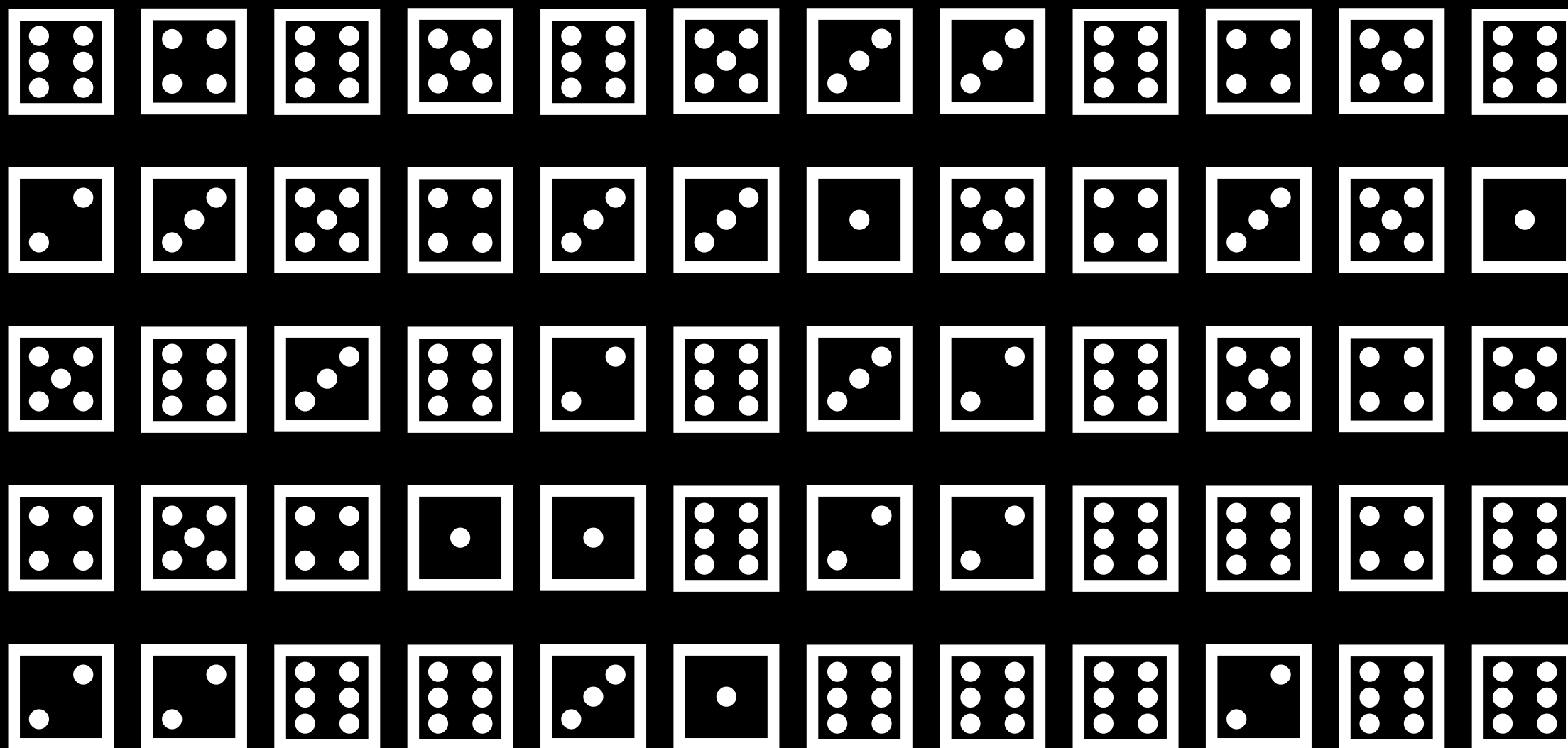


poštena izbira

?



naključno število po Gaussovi razporeditvi



5× 8× 9×

8× 10× 20×

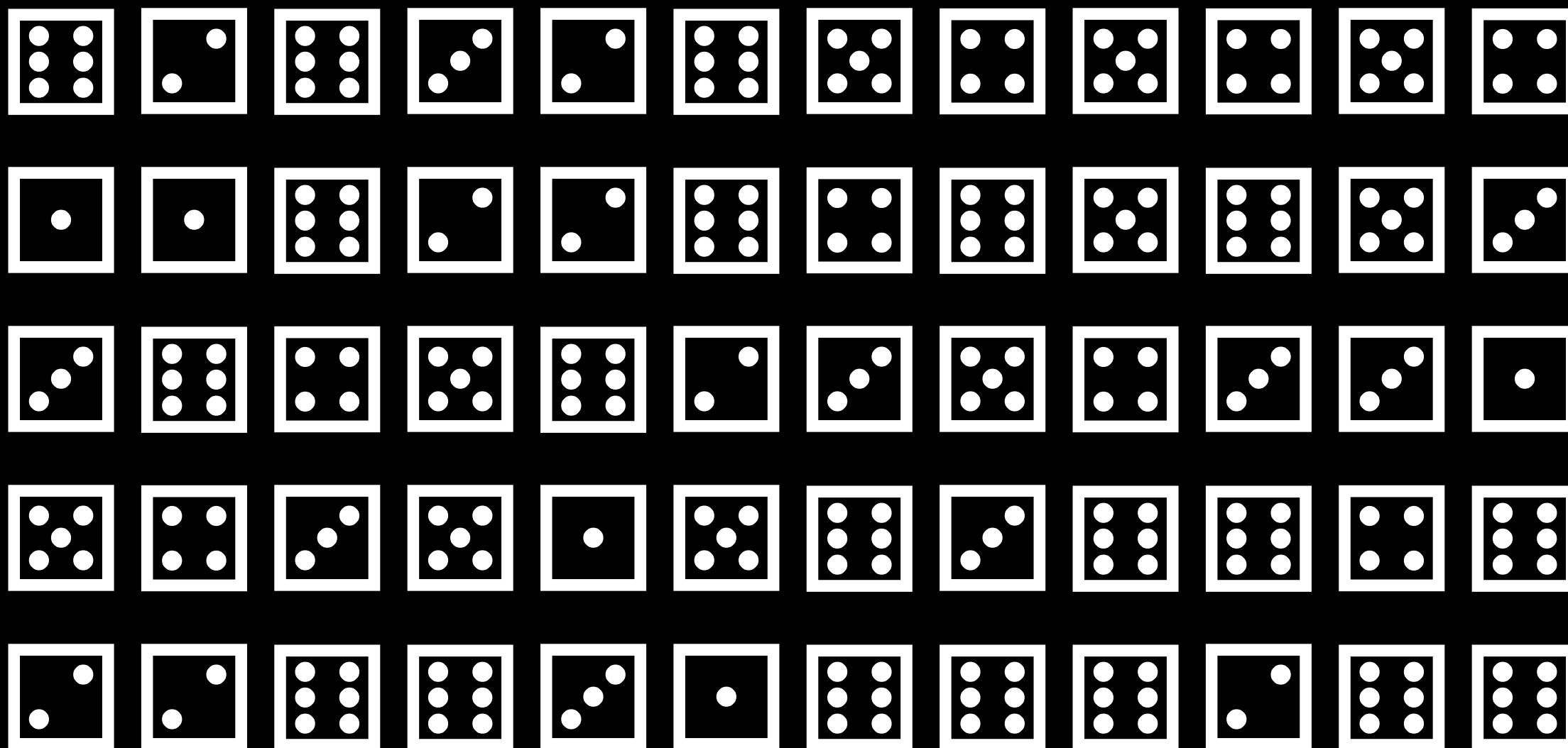
# Hekersko preverjanje hipotez

1. postavimo ničelno hipotezo
2. generiramo veliko dogodkov
3. ocenimo  $p$ -vrednost

**ničelna hipoteza:**  
kocka je poštena

# Hekersko preverjanje hipotez

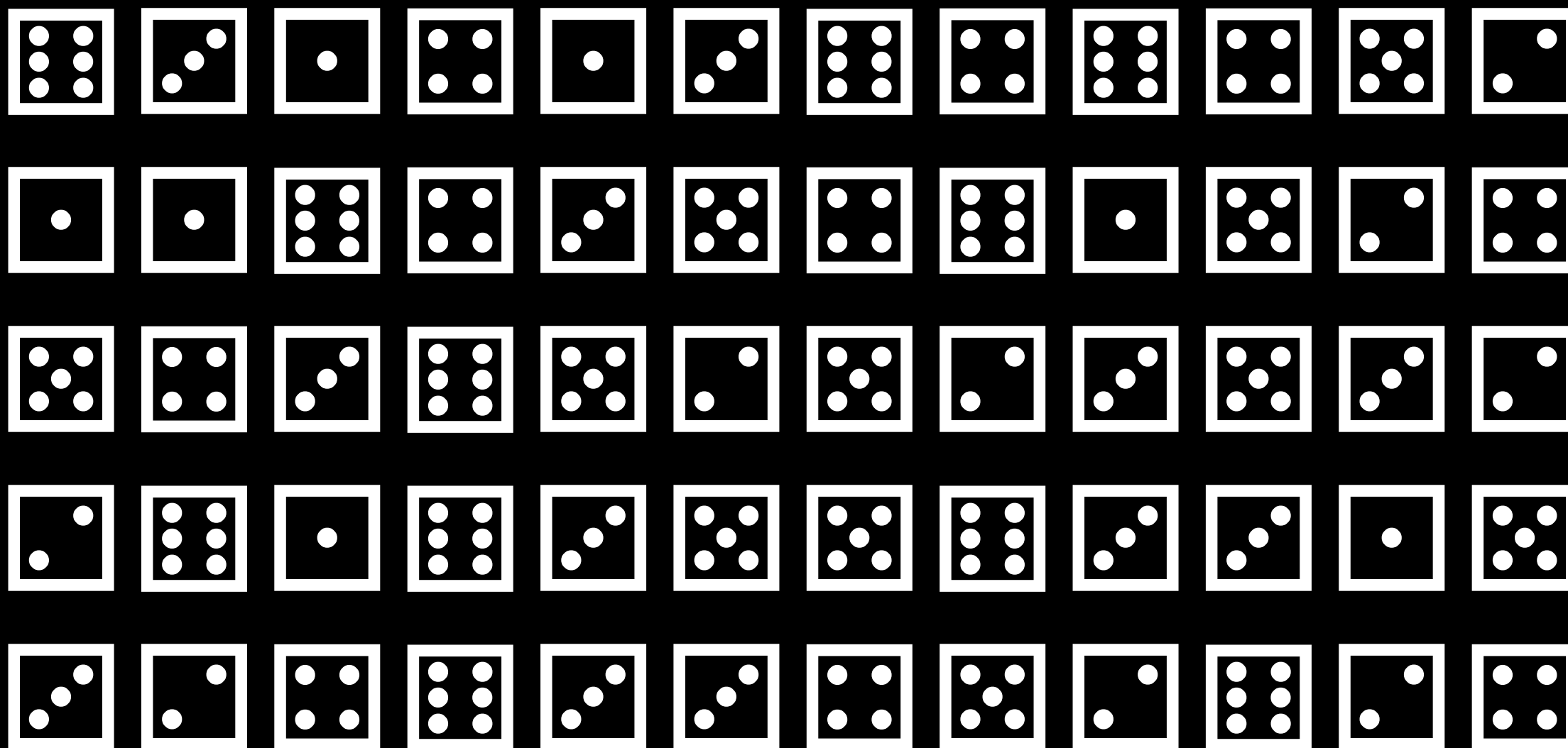
1. postavimo ničelno hipotezo ✓
2. generiramo veliko dogodkov
3. ocenimo  $p$ -vrednost

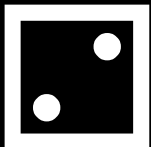


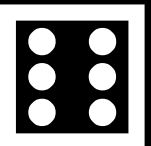
5× 8× 9×

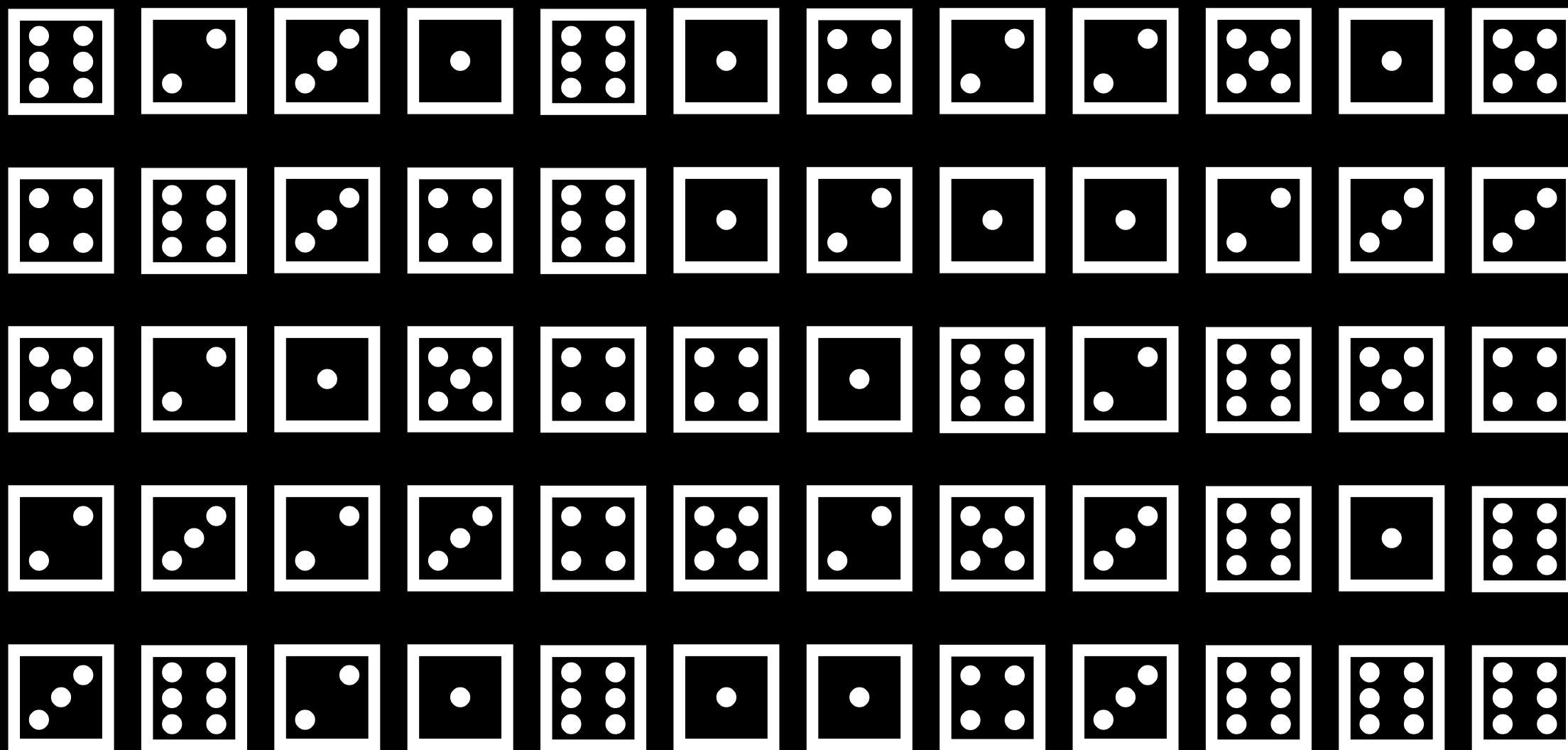
8× 10× 20×



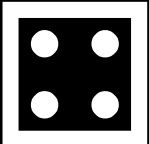


7× 9× 12×

10× 11× 11×



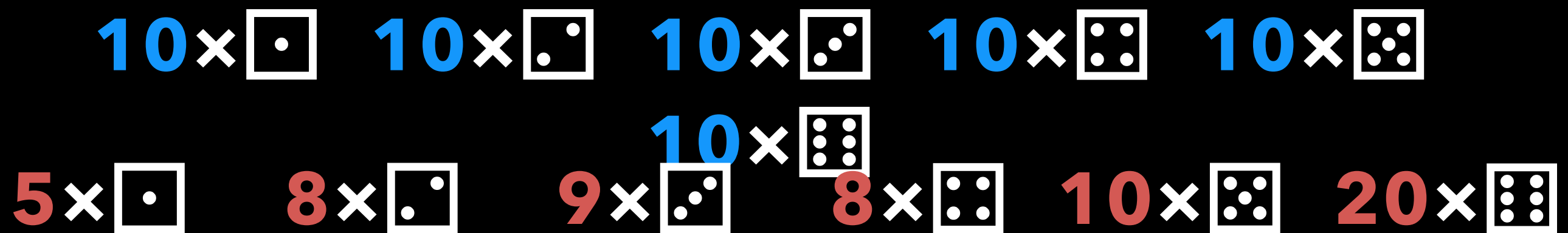
12× 11× 9×

8× 7× 13×

# Hekersko preverjanje hipotez

1. postavimo ničelno hipotezo ✓
2. generiramo veliko dogodkov ✓
3. ocenimo  $p$ -vrednost

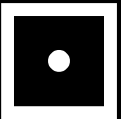
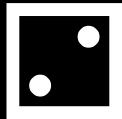
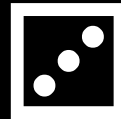
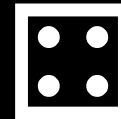
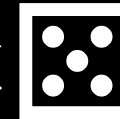

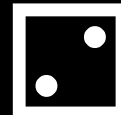


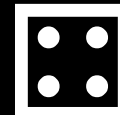

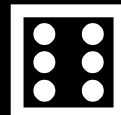
$$\chi^2 = \sum_{i=1}^n \frac{(P_i - D_i)^2}{P_i}$$



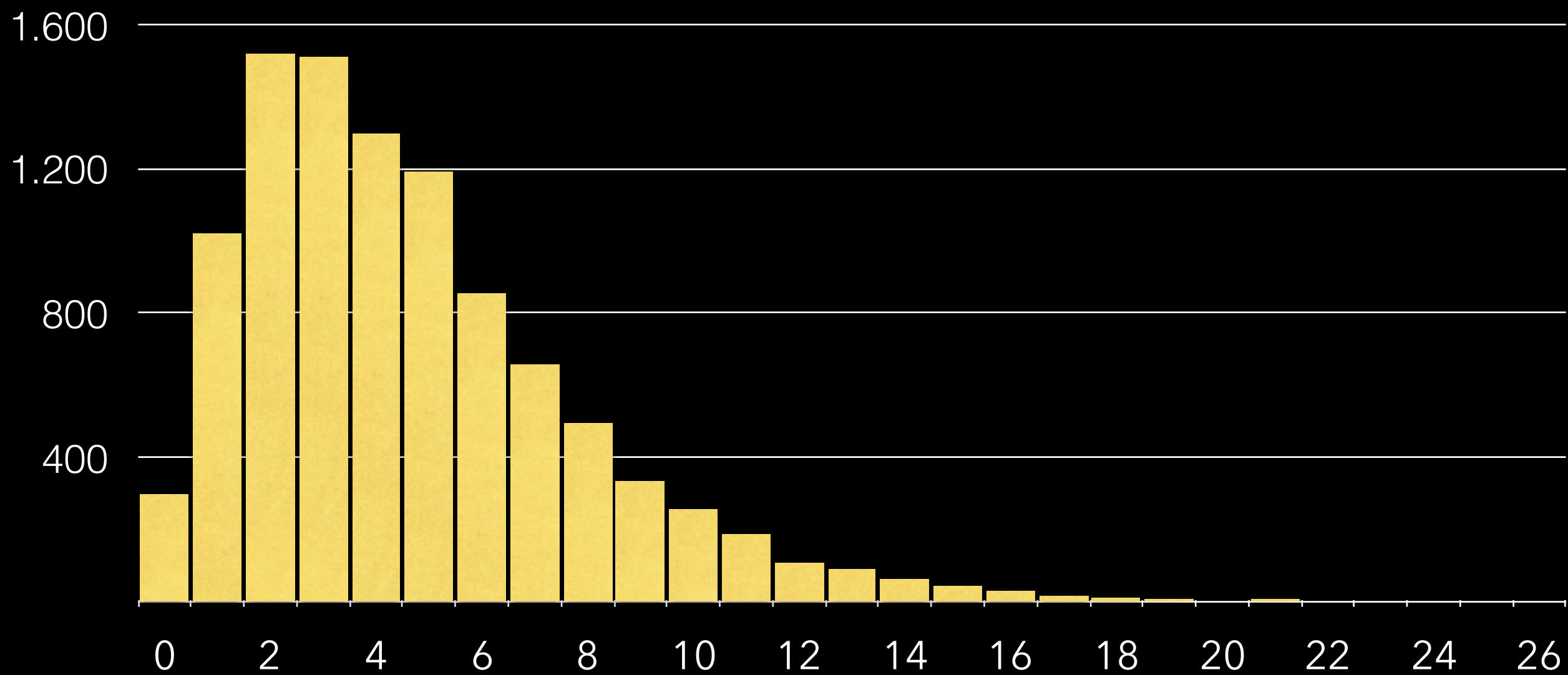
$$\begin{aligned}
 & (5-10)^2/10 + (8-10)^2/10 + (9-10)^2/10 + \\
 & (8-10)^2/10 + (10-10)^2/10 + (20-10)^2/10 = \\
 & 2,5 + 0,4 + 0,1 + 0,4 + 0 + 10 = \\
 & \mathbf{13,4}
 \end{aligned}$$

$10 \times \square_{\cdot}$     $10 \times \square_{\cdot\cdot}$     $10 \times \square_{\cdot\cdot\cdot}$     $10 \times \square_{\cdot\cdot\cdot\cdot}$     $10 \times \square_{\cdot\cdot\cdot\cdot\cdot}$   
 $7 \times \square_{\cdot}$     $9 \times \square_{\cdot\cdot}$     $12 \times \square_{\cdot\cdot\cdot}$     $10 \times \square_{\cdot\cdot\cdot\cdot}$     $11 \times \square_{\cdot\cdot\cdot\cdot\cdot}$     $11 \times \square_{\cdot\cdot\cdot\cdot\cdot\cdot}$

$$\begin{aligned}
 & (7-10)^2/10 + (9-10)^2/10 + (12-10)^2/10 + \\
 & (10-10)^2/10 + (11-10)^2/10 + (11-10)^2/10 = \\
 & 0,9 + 0,1 + 0,4 + 0 + 0,1 + 0,1 = \\
 & \mathbf{1,6}
 \end{aligned}$$

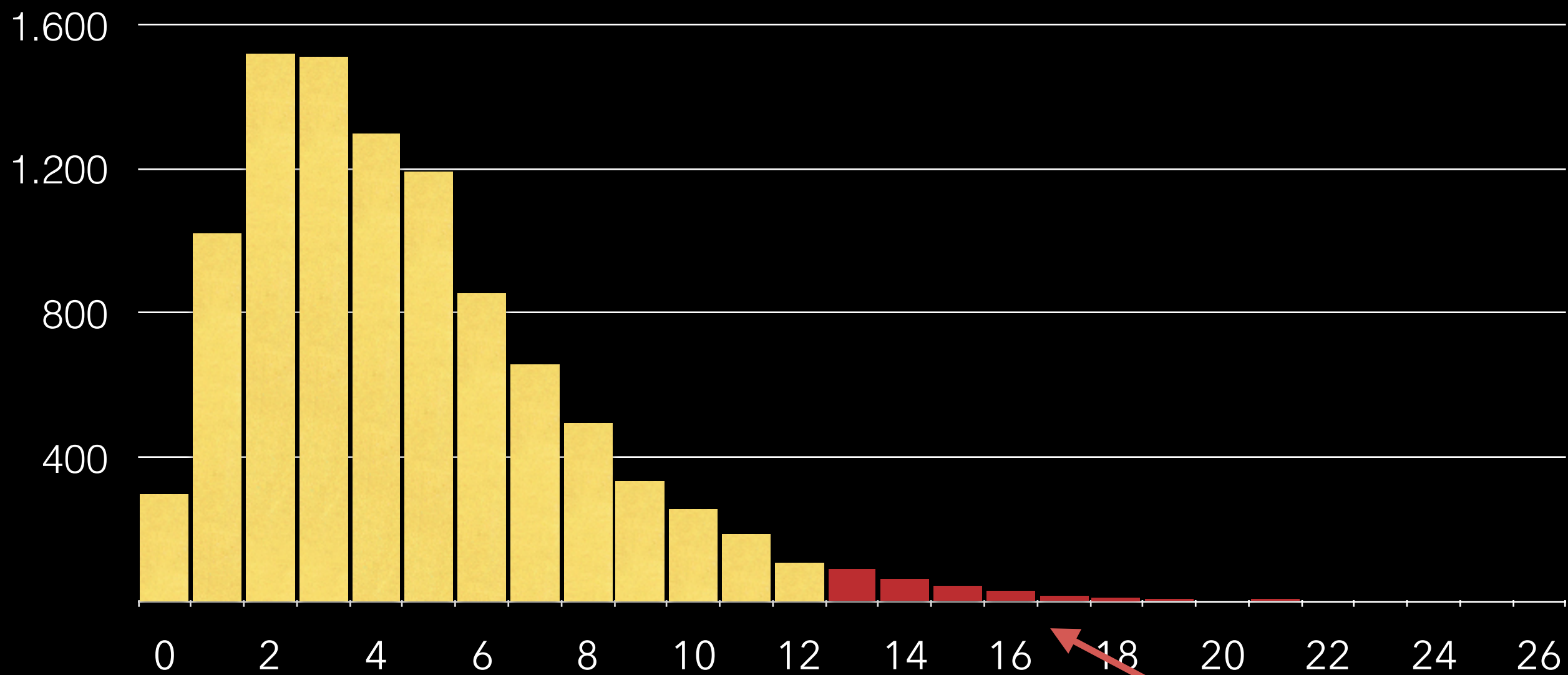
10× 10× 10× 10× 10×  
 12× 11× 9× 10× 8× 7× 13×

$$\begin{aligned}
 & (12-10)^2/10 + (11-10)^2/10 + (9-10)^2/10 + \\
 & (8-10)^2/10 + (7-10)^2/10 + (13-10)^2/10 = \\
 & 0,4 + 0,1 + 0,1 + 0,4 + 0,9 + 0,9 = \\
 & \mathbf{2,8}
 \end{aligned}$$



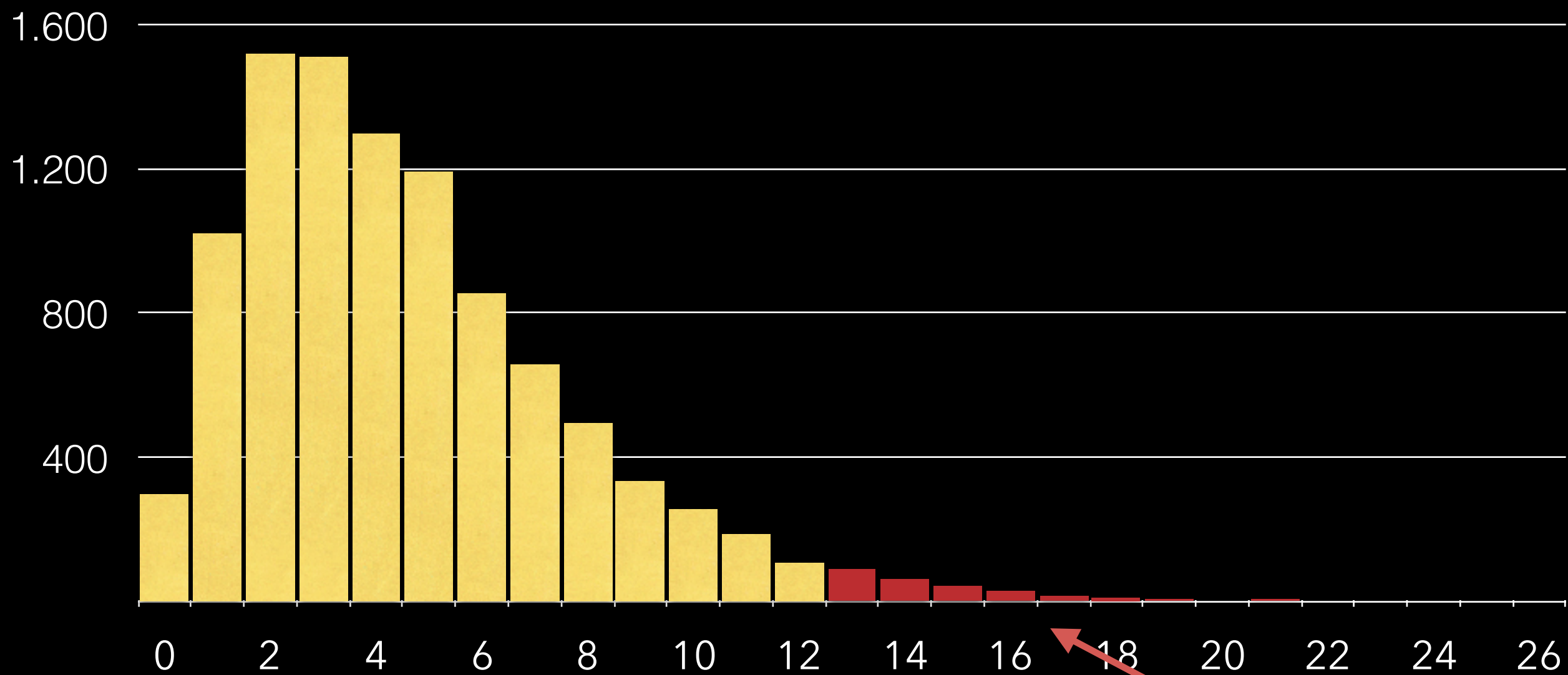
$n = 10000$





n = 10000

2,73%



n = 10000

2,73%

# Hekersko preverjanje hipotez

1. postavimo ničelno hipotezo


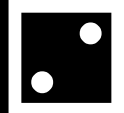
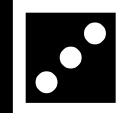
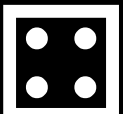
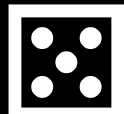
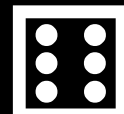


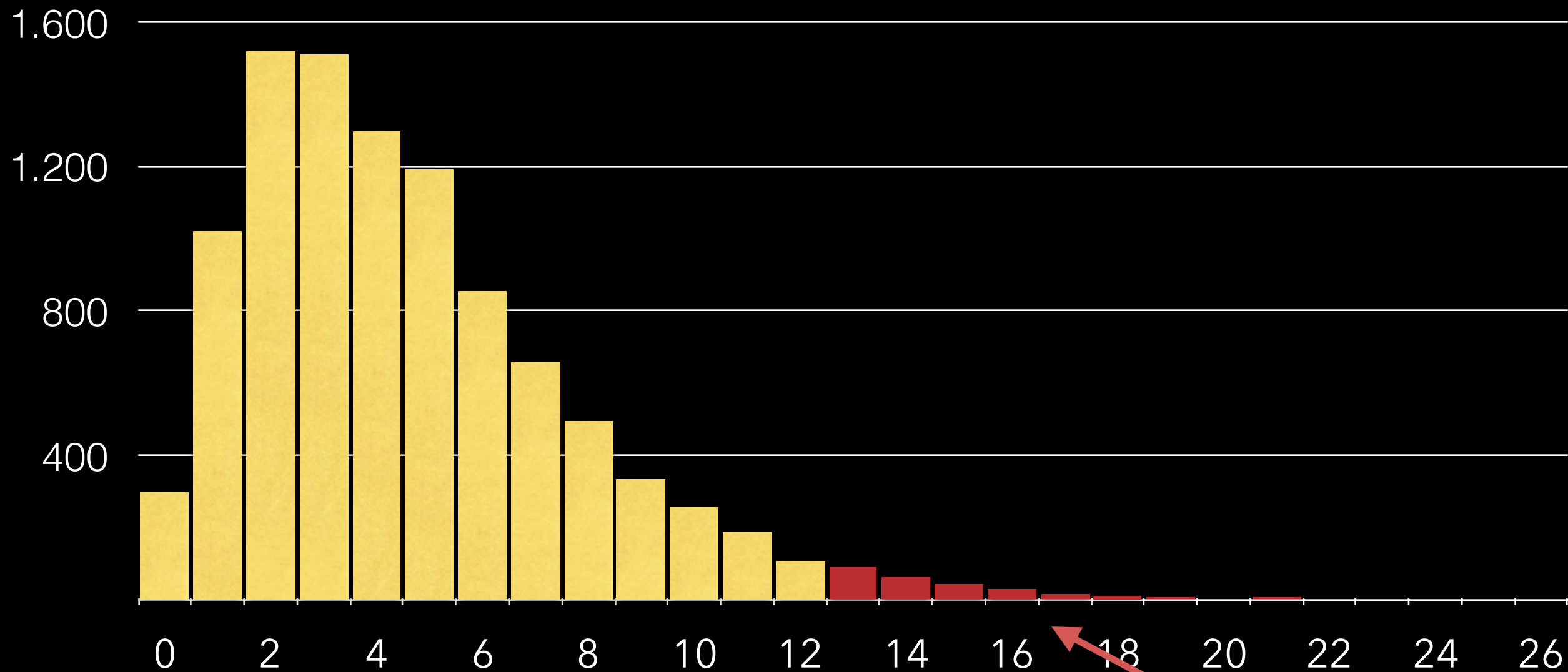
2. generiramo veliko dogodkov



3. ocenimo  $p$ -vrednost



**10×**  **10×**  **10×**   
**10×**  **10×**  **10×** 

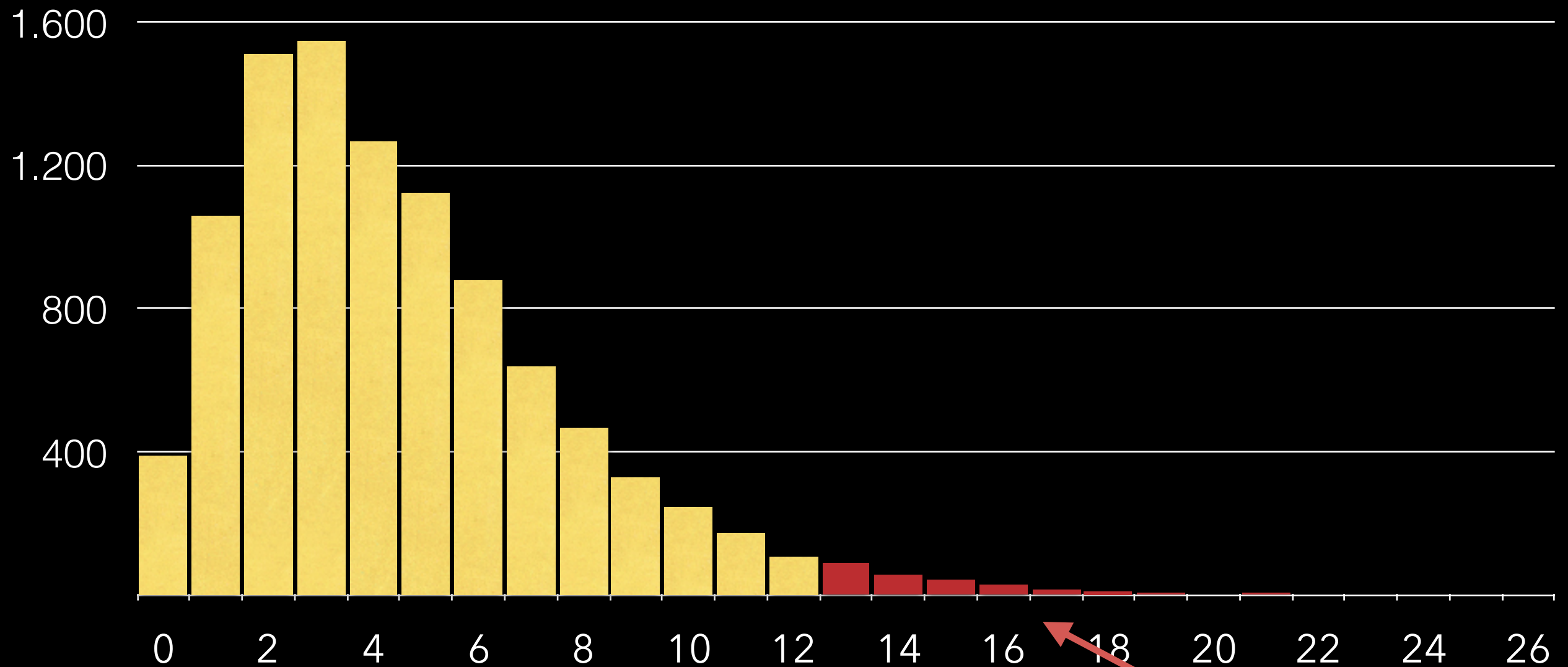


$n = 10000$

2,73%

9.99×☐<sup>•</sup> 10.00×☐<sup>•</sup> 9.97×☐<sup>••</sup>

10.01×☐<sup>••</sup> 9.97×☐<sup>•••</sup> 10.04×☐<sup>•••</sup>



$n = 10000$

2,64%

*praktični  
primer*



13

12

5



9

5

9


# Hekersko preverjanje hipotez

1. postavimo ničelno hipotezo
2. generiramo veliko dogodkov
3. ocenimo  $p$ -vrednost



**ničelna hipoteza:**  
izbiri sta neodvisni

# Hekersko preverjanje hipotez

1. postavimo ničelno hipotezo 
2. generiramo veliko dogodkov
3. ocenimo  $p$ -vrednost



13

12

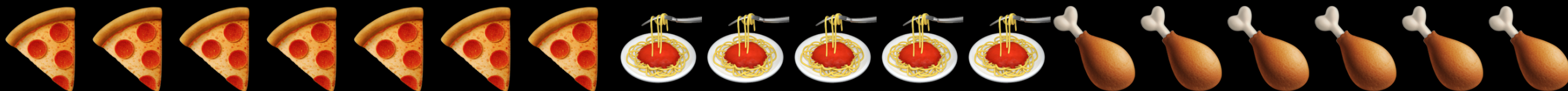
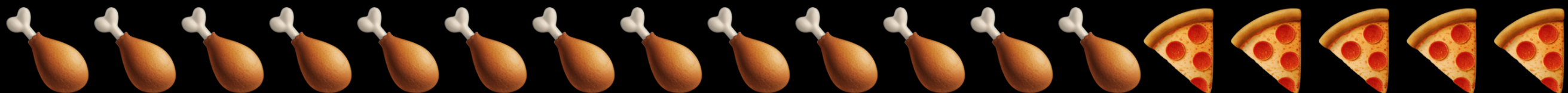
5

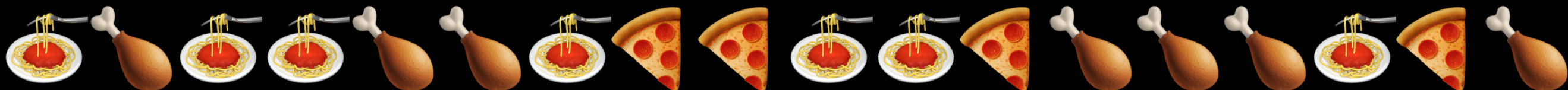
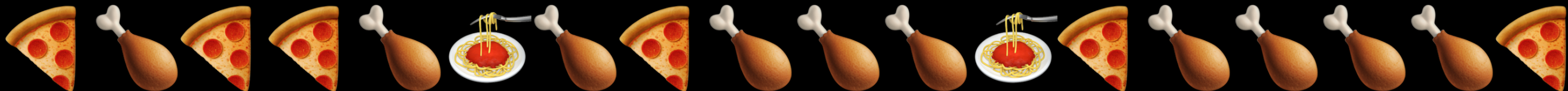


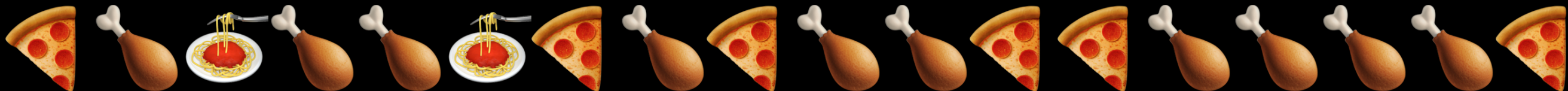
9

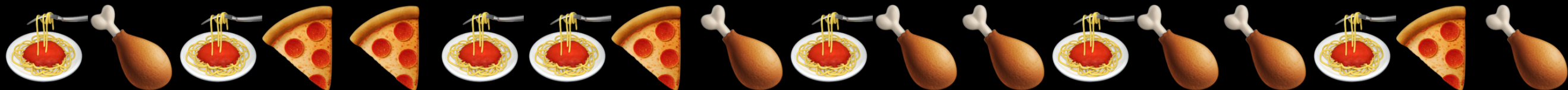
5

9









# Hekersko preverjanje hipotez

1. postavimo ničelno hipotezo



2. generiramo veliko dogodkov



3. ocenimo  $p$ -vrednost





13

12

5

30



9

5

9

23

22

17

14

**53**



13

12

5

57%



9

5

9

43%

41%

32%

27%

**53**



23%

18%

16%

57%



18%

14%

11%

43%

41%

32%

27%

**53**



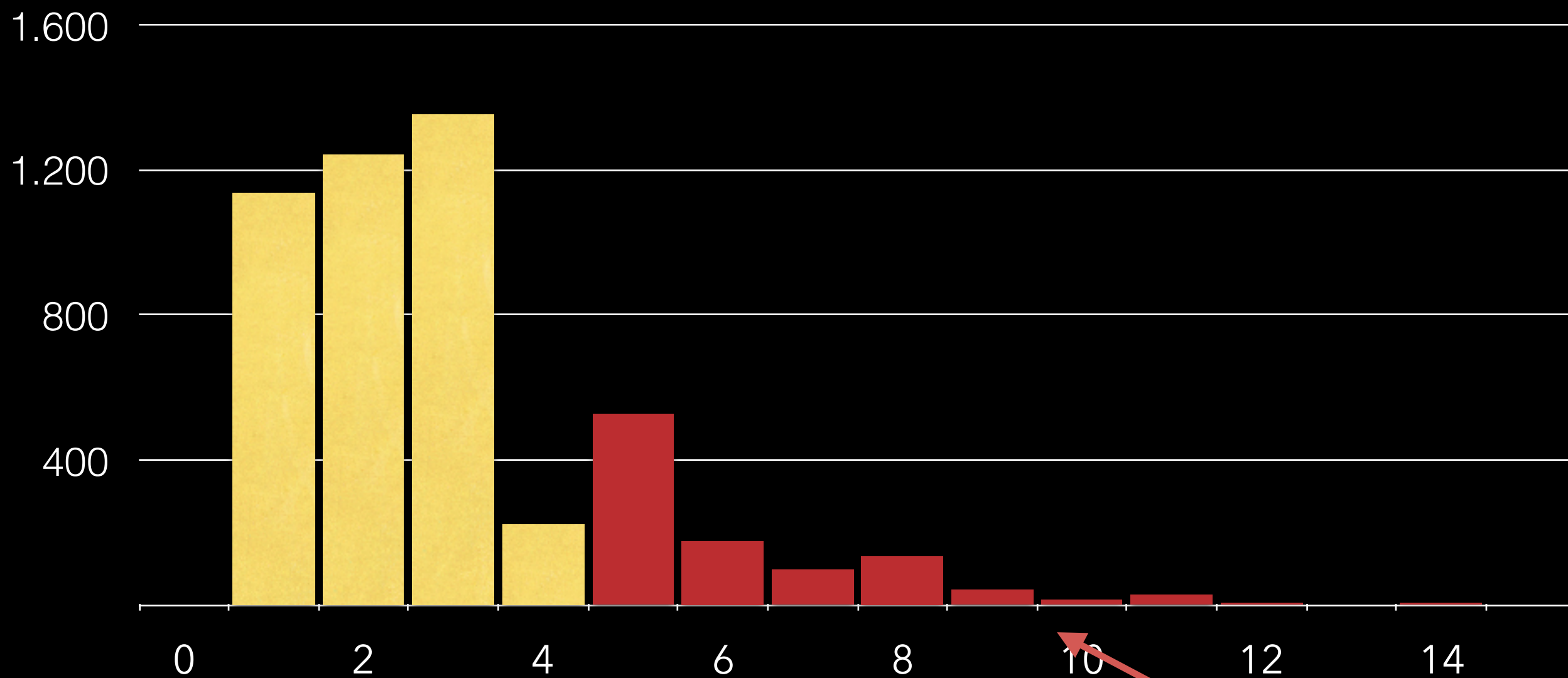
12,83 9,67 8,16 57%

9,34 7,29 6,15 43%

41% 32% 27% **53**

								
	13	12	5			12,8	9,7	8,2
	10	3	10			9,3	7,3	6,2

$$\begin{aligned}
 & (13 - 12,8)^2 / 12,8 + (12 - 9,7)^2 / 9,7 + \\
 & (5 - 8,2)^2 / 8,2 + (10 - 9,3)^2 / 9,3 + \\
 & (3 - 7,3)^2 / 7,3 + (10 - 6,2)^2 / 6,2 \\
 & = \\
 & \mathbf{4,91}
 \end{aligned}$$



n = 5000

20,8%

# Hekersko preverjanje hipotez

1. postavimo ničelno hipotezo



2. generiramo veliko dogodkov



3. ocenimo  $p$ -vrednost



*praktični  
primer*



*primer*

*v živo*

**zaključek**



grajenje intuicije

grajenje intuicije



grajenje intuicije



hitre ocene

grajenje intuicije



hitre ocene



grajenje intuicije



hitre ocene



brez znanja statistike



grajenje intuicije



hitre ocene



brez znanja statistike



grajenje intuicije



hitre ocene



brez znanja statistike



vedno enostavno

grajenje intuicije



hitre ocene



brez znanja statistike



vedno enostavno

