

This simulation follows on from CP575.

### Sampling from binomial distribution

A machine makes three times as many black lollies as red and randomly packs them into packets of 20.

90% confidence intervals are constructed based on the mean number of red lollies per packet from random samples of 25 packets.

If  $X$  is the number of red lollies per packet so that  $X \sim B(n,p)$  then  $\bar{x} = np = 5$  and  $sd = \sqrt{np(1-p)} = 1.936$ .

What fraction of 90% confidence intervals constructed in this way contain the known population mean of 5?

The z-score for a 90% CI is 1.645.  
Calculate the associated standard error.

Simulate taking a random sample of 25 packets and use the mean number of reds per packet to determine the sample mean.

Calculate the lower and upper bounds of the 90% confidence interval.

Judge whether the interval contains the known population mean of 5.

Tap back on the top line and press EXE to repeat the simulation...

In Main, take a copy of the line to calculate the sample mean and then start a new spreadsheet, pasting the formula into cell A1.

A1 =mean(randBin(20,0.25,25))

randBin(20, 0.25) 6

randBin(20, 0.25, 25)  
{8, 3, 5, 4, 4, 3, 5, 10, 7, 5, 9, ...}

mean(randBin(20, 0.25, 25)) 5.36

invNormCDF("C", 0.9, 1, 0)  
-1.644853627

$\sqrt{\frac{20 \times 0.25(1-0.25)}{25}} \rightarrow se$   
0.3872983346

mean(randBin(20, 0.25, 25))  $\rightarrow p$   
5.88

invNormCDF("C", 0.9, 1, 0)  
-1.644853627

$\sqrt{\frac{20 \times 0.25(1-0.25)}{25}} \rightarrow se$   
0.3872983346

mean(randBin(20, 0.25, 25))  $\rightarrow p$   
5.88

$p - 1.645 \times se \rightarrow lo$   
5.24289424

$p + 1.645 \times se \rightarrow hi$   
6.51710576

judge(lo < 5 < hi)  
FALSE

	A	B	C
1	5.04		
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			

=mean(randBin(20, 0.25, 25))

A1 5.04

In cells B1 and C1 calculate the lower and upper bounds of the 90% confidence interval.

In cell D1 judge if the interval contains the population proportion.

B1 =A1-1.645\*0.387

C1 =A1+1.645\*0.387

D1 =judge(B1<5<C1)

Adjust the column widths - select

columns A to E, tap Edit, Format, Column Width and enter a width of 62.

Make ten copies of the first row.

We'd expect 9 out of 10 TRUE's for a 90% confidence interval.

Use File, Recalculate to repeatedly examine how many of the ten intervals contain the population proportion.

You may want to File, Save your spreadsheet.

	B	C	D
1	4.40339	5.67662	TRUE
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			

Formula bar: =A1+1.645\*0.387

Status bar: C1 5.676615

	A	B	C	D
1	5.04	4.403	5.677	TRUE
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				

Formula bar: =mean(randBin(20, 0.25, 2

Status bar: A1:D1

	A	B	C	D
1	5.04	4.403	5.677	TRUE
2	4.88	4.243	5.517	TRUE
3	4.76	4.123	5.397	TRUE
4	4.76	4.123	5.397	TRUE
5	5.56	4.923	6.197	TRUE
6	4.36	3.723	4.997	FALSE
7	4.88	4.243	5.517	TRUE
8	4.92	4.283	5.557	TRUE
9	5.32	4.683	5.957	TRUE
10	4.72	4.083	5.357	TRUE
11				
12				
13				
14				
15				
16				

Formula bar: =mean(randBin(20, 0.25, 2

Status bar: A10:D10

	C	D		
1	4.797	FALSE		
2	5.637	TRUE		
3	5.117	TRUE		
4	5.717	TRUE		
5	5.437	TRUE		
6	5.76	5.123	6.397	FALSE
7	5.08	4.443	5.717	TRUE
8	5.16	4.523	5.797	TRUE
9	5.56	4.923	6.197	TRUE
10	4.92	4.283	5.557	TRUE
11				
12				
13				
14				
15				
16				

Formula bar: =mean(randBin(20, 0.25, 2

Status bar: A10:D10