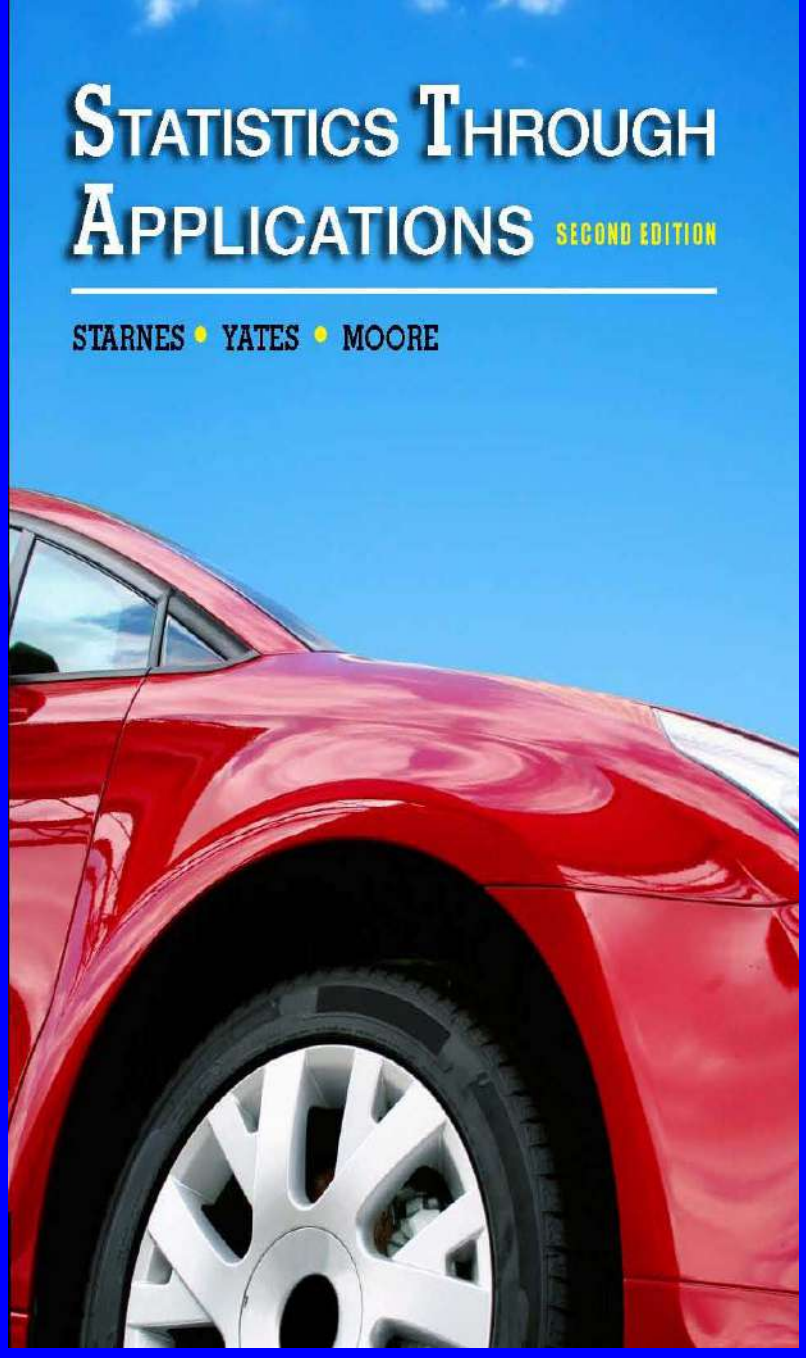


Chapter 5

Sampling and Surveys

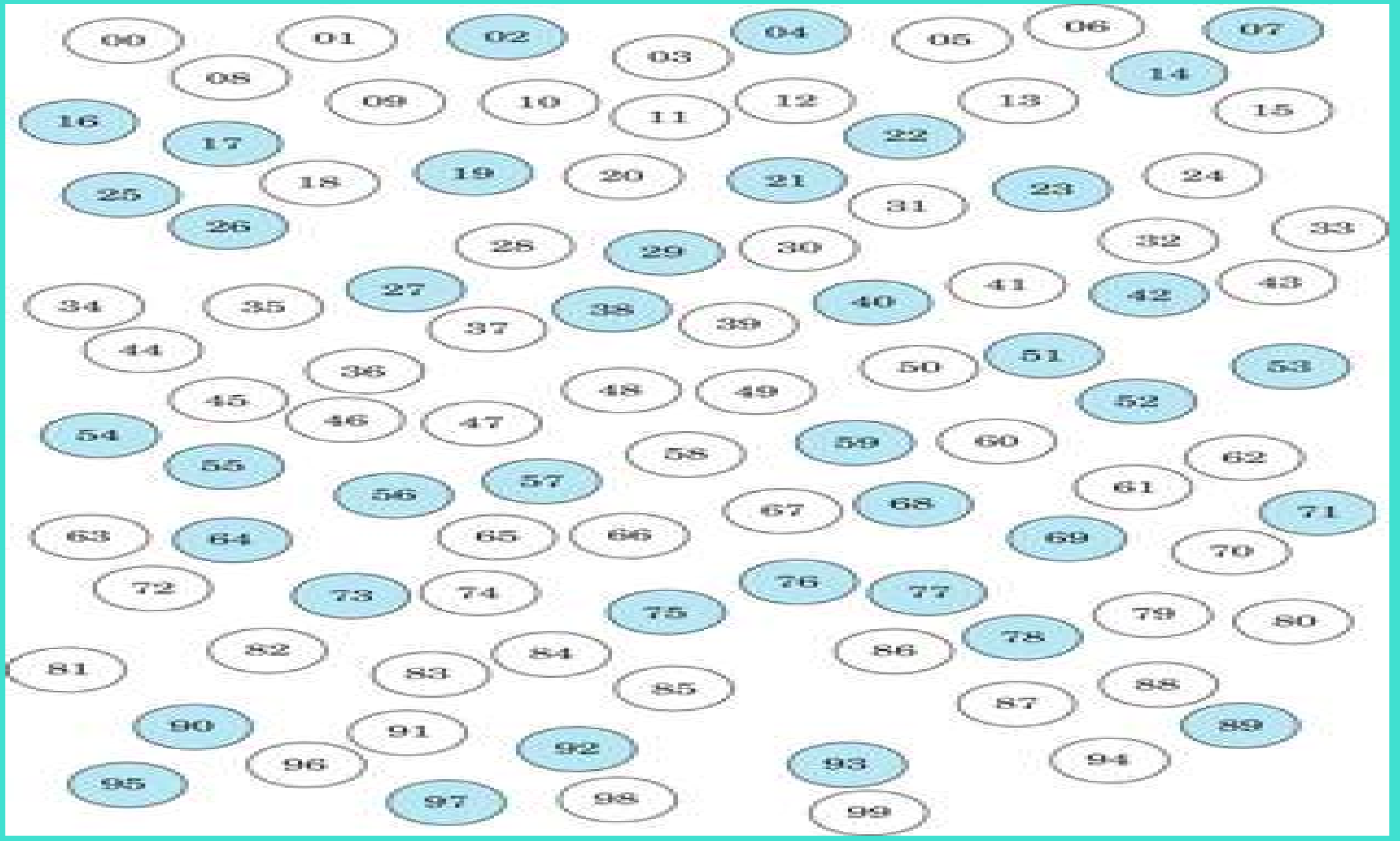
Section 5.2
What do samples tell us?

How well did our village sample data line up with the population data?

Village Sample Data charts.docx

Should gambling be legal?

The next screen shows a small population. Each circle represents an adult. Colored circles disapprove of legal gambling, and white circles are people who approve.
60 out of 100 approve of gambling = 0.6
Use a random generator to select a sample of ten. Compare results. Repeat for 10 trials. Calculate the average of your results.


00
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

Let's make a dotplot of our results...

From sample to population

Parameter

A number that describes the population

A fixed number, but in practice we don't know its value

Statistic

A number that describes a sample

Value is known when we have taken a sample, but can change from sample to sample

Used to estimate an unknown parameter

This is how Gallup polls come up with numbers

Sampling variability

If we take a SRS, then a second one on the same population, we may not get exactly the same results

If this variability is too great from repeated samples, we can't trust the results

Advantages of random samples

Eliminates favoritism by reducing bias

If we take lots of samples of the same size from the same population, the variation from sample to sample would follow a predictable pattern

This tells us that results of ***bigger samples*** are ***less variable*** than the results of smaller samples

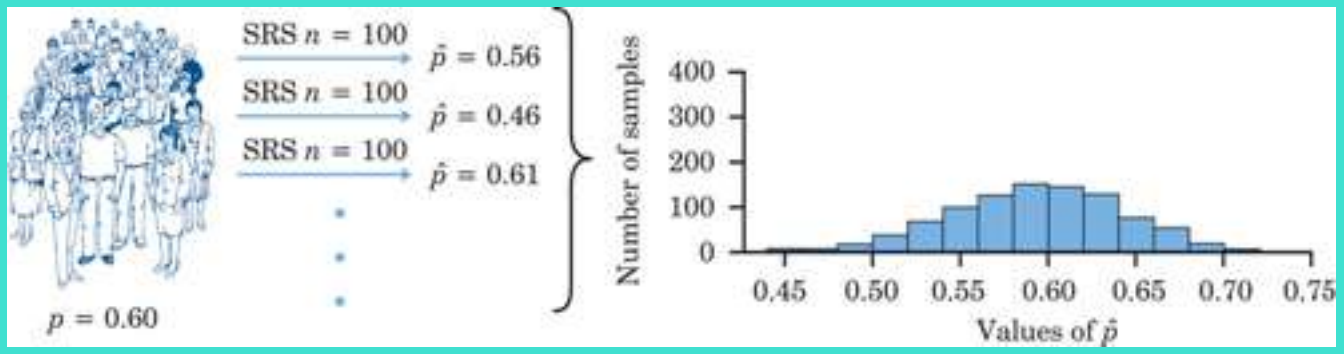


Figure 5.5
The results of many SRSs have a regular pattern. Here we draw 1000 SRSs of size 100 from the same population. The population proportion is $p = 0.6$. The sample proportions vary from sample to sample, but their values center at the truth about the population.

Here p has no bias whether the samples were large or small, however the variability was less when the samples were large

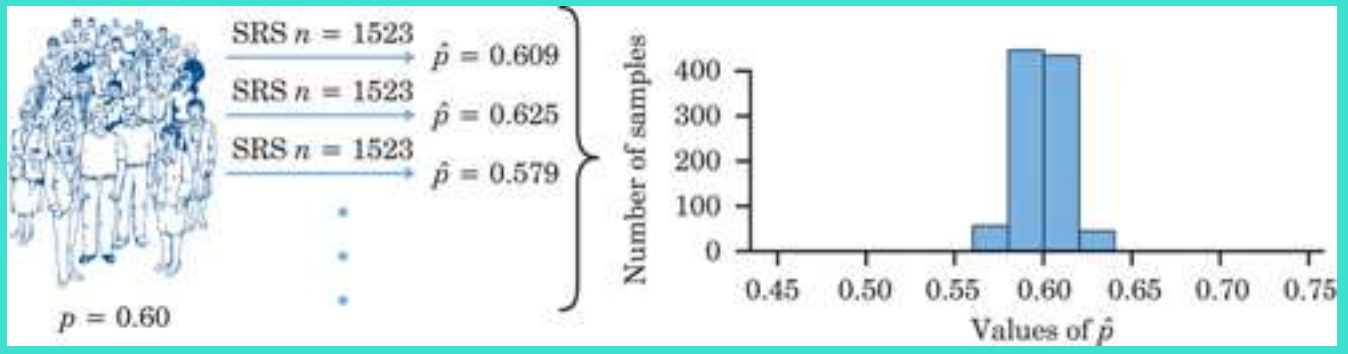


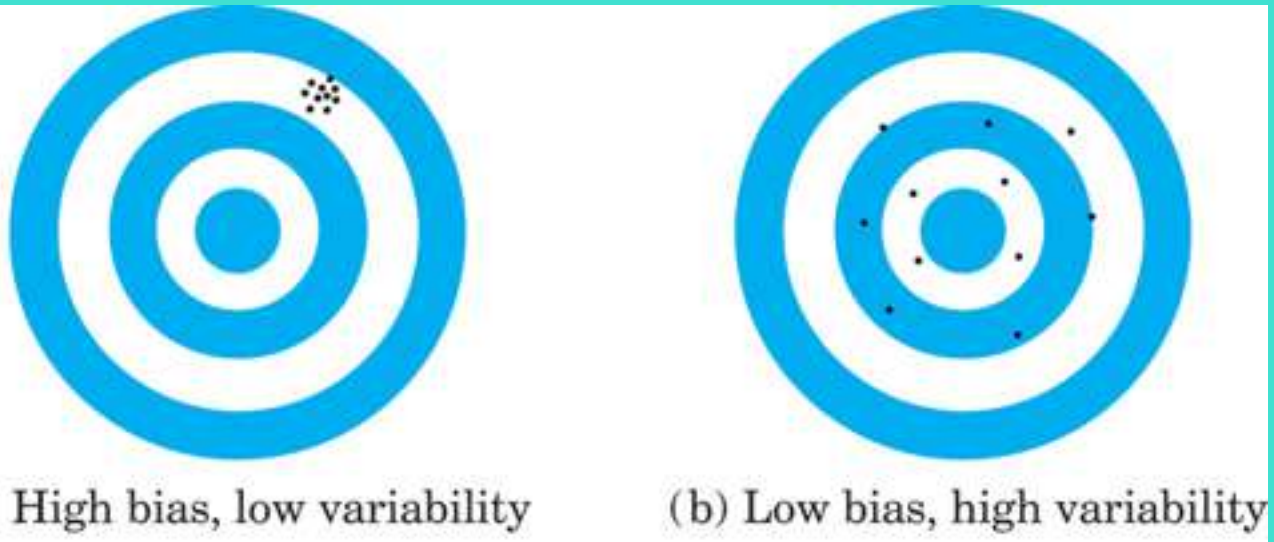
Figure 5.6
Draw 1000 SRSs of size 1523 from the same population as in Figure 5.5. The 1000 values of the sample proportion are much less spread out than was the case for the smaller samples.

Types of error in estimation

Bias Consistent, repeated deviation of the sample statistic from the population parameter in the same direction when we take many samples.

Variability How spread out the values of the sample statistic are when we take many samples. Large variability means that the result of sampling is not repeatable.

Good sampling methods have both small bias and small variability



(a) High bias, low variability

(b) Low bias, high variability

(c) High bias, high variability

(d) The ideal: low bias, low variability

If a parameter is the bull's eye,
bias means the aim is off consistently in the same direction,
variability means the results are scattered.

Managing bias and variability

To reduce bias,
use random sampling

To reduce the variability of an SRS,
use a larger sample

Margin of error and all that

"Margin of error plus or minus 3 percentage points" is shorthand for saying:

If we took many samples using the same method we used to get this one sample, 95% of the samples would give a result within plus or minus 3 percentage points of the truth about the population.

That means 5% of the samples miss by more than the margin of error!

A quick method for the margin of error

If you use the sample proportion p from a simple random sample of size n to estimate an unknown population proportion p, then the margin of error for 95% confidence is roughly equal to $1/\sqrt{n}$.

In a sample of size 1523, the margin of error for 95% confidence will be about

$$\frac{1}{\sqrt{1523}} = \frac{1}{39.03} = 0.026$$

(that is, 2.6%)

Confidence statements

margin of error – how close the sample statistic lies to the population parameter
level of confidence – what percent of all possible samples satisfy the margin of error

Interpreting confidence statements...

Conclusion applies to the population
Never completely certain
Can choose confidence level higher than 95%
95% level is usual
Smaller margin of error with same confidence? Take a larger sample!

Calculating the confidence interval

If the population parameter is .6 and we say that we have 95% confidence that our sample results are within .2 points of the truth we calculate the confidence interval this way:

95% of samples will show the statistic between .4 and .8

$(.6 - .2) = .4$ $(.6 + .2) = .8$

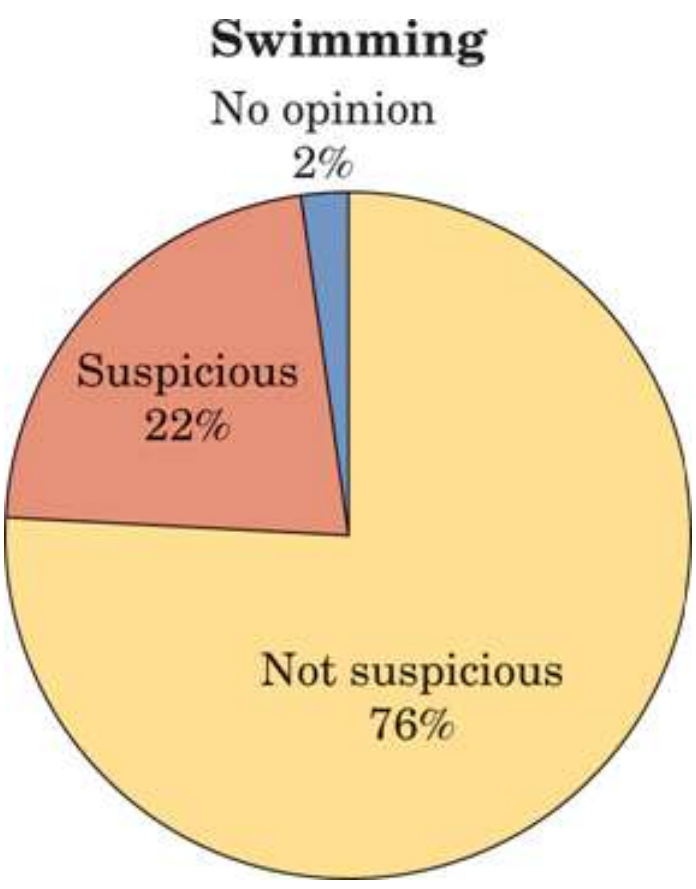
Sampling from large populations

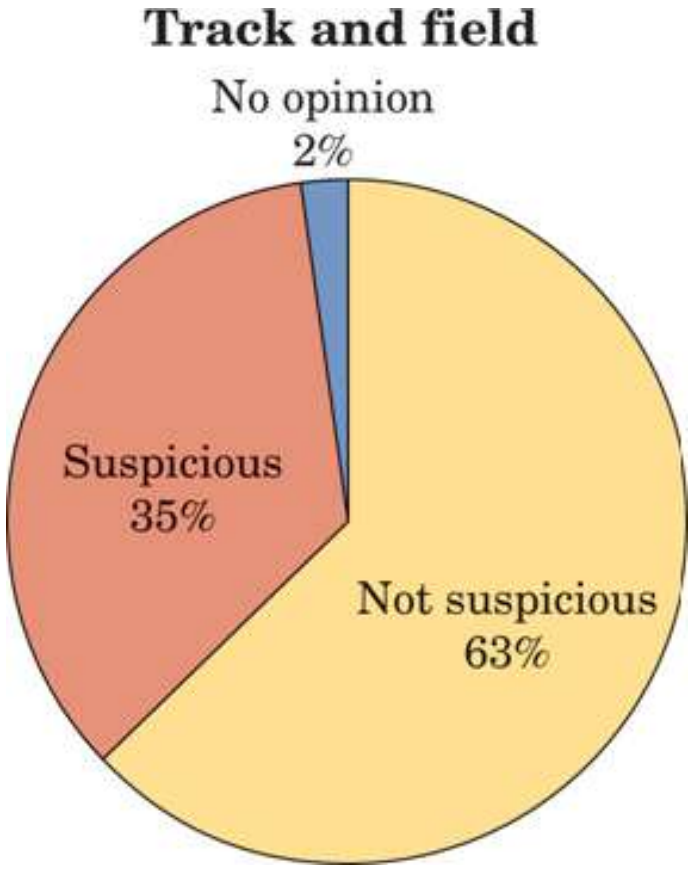
Population size doesn't matter in margin of error...

The variability of a statistic from a random sample does not depend on the size of the population, as long as the population is at least 10 times larger than the sample

It only depends on the size of the sample!

Remember margin of error = $\frac{1}{\sqrt{n}}$





Poll sampled 626 fans
"When you see or hear about an athlete breaking a world record in swimming or track and field, are you suspicious or not suspicious that the athlete used performance-enhancing drugs?"

What does a margin of ±4 percentage points mean?

Draw a bar graph that could be used for easy comparison.

Attachments

Village Sample Data charts.docx