1

A

Since the values of Y are identical distributed and independent across the population with mean 10 and variance 4.

Given that the sample size is 500, then it is large enough ($500 \gg 30$) to use the central limit theorem. So we have enough information to solve that problem.

$$\overline{Y} = \frac{\sum_{i=1}^{500} Y_i}{500} \asymp N(10, \frac{4}{500})$$

$$\frac{\overline{Y} - 10}{\sqrt{\frac{4}{500}}} \asymp N(0, 1)$$

$$P(9.8 \leq \overline{Y} \leq 10.1) = P(-\sqrt{5} \leq \frac{\overline{Y} - 10}{\sqrt{\frac{4}{500}}} \leq \frac{\sqrt{5}}{2}) = \Phi(\frac{\sqrt{5}}{2}) - \Phi(-\sqrt{5}) = 0.8555501$$

B

If the sample size is 10, then it is not large enough (30) to use the central limit theorem. So we don't have enough information to solve that problem.

Assume that $Y \sim N(10, 4)$:

$$\overline{Y} = \frac{\sum_{i=1}^{10} Y_i}{10} \sim N(10, \frac{4}{10})$$

$$\frac{\overline{Y} - 10}{\sqrt{\frac{4}{10}}} \sim N(0, 1)$$

$$P(9.8 \leq \overline{Y} \leq 10.1) = P(-\frac{\sqrt{10}}{10} \leq \frac{\overline{Y} - 10}{\sqrt{\frac{4}{10}}} \leq \frac{\sqrt{10}}{20}) = \Phi(\frac{\sqrt{10}}{20}) - \Phi(-\frac{\sqrt{10}}{10}) = 0.1869017$$

C

Because the sample sizes in A and B are different, so the variances of sample mean in A and B are also different. The difference of variance lead to difference of the answers in parts A and B.(Assume $\overline{Y}$ is normally distributed both in A and B.)

2

A

$X_i = 1$: the i^{th} people in the sample have been victim of violence.
$X_i = 0$: the i^{th} people in the sample have not been victim of violence.
The first way:
$\sum_{i=1}^{500} X_i \sim B(500, 0.83)$
then calculate $P(\sum_{i=1}^{500} X_i > 415) = P(\sum_{i=1}^{500} X_i = 416) + P(\sum_{i=1}^{500} X_i = 417) \cdots + P(\sum_{i=1}^{500} X_i = 500)$

The second way:
since the sample size is greater than 30 and X_i is identical and independent distributed, we can use the central limit theorem:
$\sum_{i=1}^{500} X_i \asymp N(415, 70.55)$
then calculate $P(\sum_{i=1}^{500} X_i > 415)$

B

Use the second way:
$\sum_{i=1}^{500} X_i \asymp N(415, 70.55)$
$P(\sum_{i=1}^{500} X_i > 415) = \Phi(\frac{415-415}{\sqrt{70.55}}) = \Phi(0) = 0.5$

C

$\sum_{i=1}^{8} X_i \sim B(8, 0.83)$
$P(\sum_{i=1}^{8} X_i > 6) = P(\sum_{i=1}^{8} X_i = 7) + P(\sum_{i=1}^{8} X_i = 8) = 0.5942795$

3

A

Suppose that X is a random variable with a binomial distribution:
$X \sim B(n, p)$
then the variance of X is :$\sigma^2 = np(1-p)$
Mathematically, with the n increases, the value of σ^2 increase.
Intuitively, with the n increase, the range of X (0 to n) get larger, so the variance of X will also get larger.

B

A: True
B: False. To use the Central Limit Theorem, the population being sampled doesn't need to be normally distributed.

C

X: the number of odd tosses

If I choose two times:

$X \sim B(2, 0.5)$

The probability that the number of odd tosses equal to the number of even tosses is: $P(X = 1) = C_2^1 \times 0.5 \times 0.5 = 0.5$

If I choose 100 times:

$X \sim B(100, 0.5)$

The probability that the number of odd tosses equal to the number of even tosses is: $P(X = 50) = C_{100}^{50} \times 0.5^{50} \times 0.5^{50} = 0.07958924$

0.5>0.07958924, so it is better to choose two times.

4

If $X \sim B(n, p)$, this equal to do n times Bernoulli experiments. Each experiments has the same distribution $B(1, p)$ and they are independent to each other. So when n is large enough, the normal distribution can be used as an approximation to the binormal distribution.

For example, when $X \sim B(n, p)$ and n is large enough, we can use $N(np, np(1-p)$ as the approximation to X. The approximation becomes better with increasing n.

5

A

X_i= the weight of the i^{th} people.

According to central limit theorem, when the weights of these guests are independent and identically distributed random variables, we can use $N(5664, 28800)$ as the approximation of the distribution of total weight.

$P(\sum_{i=1}^{32} X_i \leq 6000) = \Phi(\frac{6000-5664}{\sqrt{28800}}) = 0.9761426$

B

The central limit theorem enabled me to answer this question and I use R language to calculate it.

6

A

The finite sample correction factor make the standard error of the sample mean smaller.

When the sample size is infinite, the value of correction factor equals to 1.

Because $n > 1$, when the sample size is finite, the value of correction factor less than 1.

B

When N=200 and n=75:

$\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{200-75}{200-1}} = 0.7925533$

C

When N=1000000 and n=400:

$\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{1000000-400}{1000000-1}} = 0.9998005$

D

The situation in Part(C) is likely to be most relevant for much of the type of social science research I am likely to do.

7

A

X: the number of returns that qualify for a refund in 250 samples.

According to the central limit theorem:

$X \asymp N(200, 40), \quad 0.81 \times 250 = 202.5$

$P(X \leq 202) = \Phi(\frac{202-200}{\sqrt{40}}) = 0.6240852$

B

Assume that the proportion of refunds in the sample that I drew is p.

$X \sim B(250, p)$

According to the central limit theorem:

$X \asymp N(250p, 250p(1-p))$

$250 \times 0.8 = 200$

$P(X < 200) = 0.95 \Rightarrow \frac{200-250p}{\sqrt{250p(1-p)}} = 1.644854 \Rightarrow p = 0.755275$

8

A

Because $\overline{X}$ is the function of X_i(the i^{th} sample), given that every X_i is a random variable, the $\overline{X}$ is also a random variable(whose value is uncertain and depend on outcomes of a random phenomenon).

Similarly, $\overline{Y}$ is also a random variable.

B

Because of central limit theorem, when the sizes of X and Y are all larger than 30 and they are all random samples, we can use normal distribution as the approximation for them:

$\overline{X} \sim N(7, \frac{15}{80})$
$\overline{Y} \sim N(9, \frac{12}{59})$

C

Because X and Y are independent random samples, the distribution of W is approximately normal distribution.

Mean of W:$E(\overline{X} + 2\overline{Y} = 7 + 2 \times 9 = 25$
Variance of W:$\frac{15}{80} + 4 \times \frac{12}{59} = \frac{945}{944}$
$W \sim N(25, \frac{945}{944})$

9

A

$E(X) = 0 \times 0.3 + 1 \times 0.2 + 2 \times 0.2 + 5 \times 0.3 = 2.1$
$E(X^2) = 0 \times 0.3 + 1 \times 0.2 + 4 \times 0.2 + 25 \times 0.3 = 8.5$
$Var(X) = E(X^2) - E^2(X) = 4.09$

B

$\overline{X}$	0	0.5	1	1.5	2	2.5	3	3.5	5
P	0.09	0.12	0.16	0.08	0.04	0.18	0.12	0.12	0.09

C

$E(B) = 0.09 * 0 + 0.12 * 0.5 + 0.16 * 1 + 0.08 * 1.5 + 0.04 * 2 + 0.18 * 2.5 + 0.12 * 3 + 0.12 * 3.5 + 0.09 * 5 = 2.1$

D

$S^2 = \frac{(X_1-\overline{X})^2+(X_2-\overline{X})^2}{2-1} = (X_1-\overline{X})^2+(X_2-\overline{X})^2 = \frac{(X_1-X_2)^2}{2}$

S^2	0	0.5	2	4.5	8	12.5
P	0.26	0.2	0.12	0.12	0.12	0.18

E

$E(D) = 0.26*0+0.2*0.5+0.12*2+0.12*4.5+0.12*8+0.18*12.5 = 4.09$

Because sample variance S^2 is an unbiased estimator of the population variance.

10

A

$X_1 \sim N(0,(0.2h)^2)$
$X_2 \sim N(0,(0.4h)^2)$
$X_3 \sim N(0,(0.6h)^2)$
Since the three measurements are independent random variables:
$X_1+X_2+X_3 \sim N(0,(0.2h)^2+(0.4h)^2+(0.6h)^2)$
$\overline{X} = \frac{X_1+X_2+X_3}{3} \sim N(0, \frac{(0.2h)^2+(0.4h)^2+(0.6h)^2)}{9}) \Rightarrow \frac{X_1+X_2+X_3}{3} \sim N(0, \frac{14}{225}h^2)$
$P(0.7h < \overline{X} < 1.3h) = \Phi(\frac{1.3}{\sqrt{\frac{14}{225}}}) - \Phi(-\frac{0.7}{\sqrt{\frac{14}{225}}}) = 0.9974938$

B

If we didn't assume that the measurements were independent random variables, we can't derived the distribution of $X_1+X_2+X_3$ then we can't make the calculation in A.

I didn't use the central limit theorem in A to solve the problem, because given that the measurements are independent random variables, I can derive the distribution of $X_1+X_2+X_3$ directly.