Soda can volumes vary with a distribution that has mean 360 milliliters (ml) and standard deviation 8 ml. Assume (for now) that the variation follows a Normal distribution.

1. What is the mean volume in fluid ounces?  
   ___________________ [nearest 0.001 fl oz]

2. What is the probability a can has volume between 340 and 380 ml?  
   __________

Consider random samples of 10 cans.

3. What is the distribution of the sample mean volume for the 10 cans?
   Shape = ____________  Mean = ________ml  SD = ________ ml [nearest 0.01]

4. Determine the probability the sample mean volume for 10 cans is between 358 and 362 ml.  
   ____________ [nearest 0.0001]

5. Interpret your result from #4:
   ____________ [← answer to 4] of __________________________ [← units]  
   have ___________________________ [← variable] between 358 and 362 ml.

Instead consider random samples of 50 cans.

6. What is the distribution of the sample mean volume for the 50 cans?
   Shape = ____________  Mean = ________ml  SD = ________ ml [nearest 0.01]

7. Determine the probability that the sample mean volume for 50 cans is between 359.6 and 360.4 ml.  
   ____________ [nearest 0.0001]

8. How do answers to 2 through 7 change if the variation in volumes of (single) cans does not follow a Normal distribution?
30% of all incoming email messages for a college’s computer system are spam.

1. The value 30% is a ________ [circle the correct answer] parameter statistic

2. What symbol is used to represent this value? __________

Consider random samples of 234 incoming email messages, calculating for each the proportion of these messages that are spam.

3. The sample proportion is a ________ parameter statistic

4. What symbol is used to represent the sample proportion? __________

5. Determine the mean and standard deviation of the distribution of sample proportions.
   Mean = __________ SD = __________ [nearest 0.001 at least]

6. Compute np and n(1 – p): np = __________ n(1 – p) = __________

   These are both at least 10. Consequently, the sample proportion has a Normal distribution.
   (Technically the distribution is only approximately Normal. But it’s close enough to Normal that treating it as Normal gives very good results.)

7. What is the probability that the percent of spam in a randomly selected sample exceeds 35%? __________ [nearest 0.0001]

8. Interpret your result from #7:
   __________ [↩ answer to 7 goes here] of ____________________________ [↩ units]
   have ____________________________ [↩ variable] exceeding 35%.

Now consider a sample that is 4 times larger: a random sample of 936 incoming email messages.

9. Determine the mean and standard deviation of the distribution of the sample proportion.
   Mean = __________ SD = __________ [nearest 0.001 at least]

10. What is the probability that the percent of spam in a sample (of size 936) exceeds 35%? __________

Now consider a sample of size 26 (9 times smaller than 234).

11. Now what is the standard deviation of the sample proportion? __________ [nearest 0.001]

12. Using the Normal distribution, can one determine the probability that the percent of spam in a sample (of size 26) exceeds 35%? Explain.
Solutions

1. 12.173 fl oz. (You can just type “convert 360 milliliters to fluid ounces” into a search box to learn this.)

2. Z scores are ± 2.4; the probability is 0.9876.

3. Normal; Mean 360; SD = 8/√10 = 2.53.

4. Z scores are ± 0.79; the probability is 0.5708.

5. 57.08% of all samples of 10 cans have a sample mean volume between 358 and 362.

6. Normal; Mean 360; SD = 8/√50 = 1.13.

7. Z scores are ± 0.35. The probability is 0.2763.

8. 2 can no longer be answered. In 3, the shape is no longer Normal, consequently 4 and 5 can no longer be answered. However, results to 6 and 7 stay the same, because with this large a sample the sample mean will have (approximately) a Normal distribution.

Questions 2, 4 and 7 all address the question: What is the probability the total weight of the sampled cans is within 20 of the mean value.

1. Parameter
2. p
3. Statistic
4. p̂
5. The mean is 0.30 with standard deviation \(\sqrt{\frac{0.3 \times 0.7}{234}} = 0.0300\)

6. np = 70.2 \(n(1 - p) = 163.8\)

7. The Z score is \((0.35 - 0.3)/0.03 = 1.67\). The probability is 0.0478.

8. 4.78% of all samples of 234 emails have a sample percent of spam over 35%.

9. The mean is still 0.30; the standard deviation is \(\frac{1}{2}\) as big: 0.015.

10. The Z-score is now 3.33 and the probability is 0.0004.

11. The standard deviation is 0.0899 – 3 times larger.

12. The probability can’t be calculated here. Since np = 26(0.3) = 7.8 < 10, a Normal distribution does not apply.

Notice how the standard deviation “goes with” the square root of the sample size. This is as close to a hard and fast rule as one can get in statistics; with exceptions (which usually come close to obeying this rule of thumb) only when there are “factors other than sample size” playing a role.