SOLUTIONS

(1) Which of the following would be an example of panel (longitudinal) data? **(c)**

(2) Which of the following histograms is positively skewed? **(e)**

(3) Consider a sample taken from a normal population. With numbers recorded to the second decimal place, which of the following statements about the percent of observations within 1 standard deviation (s.d.) of the mean is true? **(b)**

(4) Based on the tabulation of the variable X below, what is the sample median of X? **(c)**

(5) What can you reasonably conclude about the sample based on this statistic? **(d)**

(6) Suppose X is a normally distributed variable with mean 10 and variance 20: X ~ N(10, 20). What is the probability that X is negative (rounded to nearest hundredth)? **(b)**

(7) What can you reasonably conclude about the sample based only on the above information? **(b)**

(8) What can you reasonably infer about the population based only on the above information? **(e)**

(9) Which of the following statements about the sample range and population range is true? **(a)**

(10) Which is closest to the sample standard deviation? **(c)**

(11) Which is closest to the sample 90th percentile? **(e)**

(12) Suppose hours per year employees spend surfing the internet at work is normally distributed. If 13.4% of employees spend more than 160 hours, how many standard deviations above the mean is 160? **(a)**

(13) Which is the expected number of complaints for each type of sales representative (round answer to nearest tenth)? **(e)**

(14) If the manager estimates that each complaint results in $250 in lost sales and there are 100 sales reps in total, how much money in total does the manager expect to lose due to complaints (round answer to nearest dollar)? **(a)**

(15) Suppose a random variable X is binomially distributed. In which of the following cases can the probabilities associated with the binomial distribution be reasonably approximated by the probabilities associated with the normal distribution? **(d)**

(16) Consider the population density function given below. If 1 observation is drawn you can be 90% confident that the sample mean will fall in what interval (round to the nearest tenth)? **(c)**

(17) Given the sampling distribution of the difference between these means shown below, what is the probability that due to sampling noise we obtain two samples where the average profit margins of the duopolists are actually higher than the monopolists? (Round to nearest hundredth.) **(b)**

(18) Using the simulation results, compute the interval that should contain the sample standard deviation with 90% confidence. Which show the correct values rounded to nearest tenth? **(a)**

(19) What is the point estimate of μ? **(e)**

(20) Which of the following would increase the width of the interval estimate? **(d)**

SOLUTIONS

(1)

(a) p = 0.02

$$E[\hat{P}] = 0.02 \qquad \sigma_{\hat{P}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.02(1-0.02)}{500}} = 0.0063$$

Rule of thumb reveals that the normal approximation for the sampling distribution of the sample proportion (p-hat) is appropriate:

0.02 – 3*0.0063 = 0.0011, which is greater than 0
0.02 + 3*0.0063 = 0.0389, which is less than 1

$$P(\hat{P} > 0.03) = P\left(Z > \frac{0.03-0.02}{0.0063}\right) = P(Z > 1.60) = 0.5 - 0.4452 = 0.0548$$

(b) There is a 5.48% chance that simply because of sampling noise more than 3% of the 500 sampled products will need a service call IF what the manufacturer said is true.

(c) At conventional significance levels (5% level) we would not conclude from this sample that the manufacturer is lying. However, there is SOME evidence that the claim is not true because there is only a 5.48% chance of getting such a high level due to sampling noise.

Another way to say this is that there is more than a 1 in 20 chance that such a high sample proportion is obtained NOT because the manufacturer is lying but simply because of sampling noise. Had the answer come out to 0.00548 we could be much more confident in calling the manufacturer a liar.

(d)

$$p = 0.02$$

$$E[\hat{P}] = 0.02$$

$$\sigma_{\hat{P}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.02(1-0.02)}{100}} = 0.014$$

Rule of thumb reveals that the normal approximation for the sampling distribution of the sample proportion (p-hat) is **NOT** appropriate:

0.02 – 3*0.014 = **-0.022, which is negative.** Impossible for sample proportion to be negative.
0.02 + 3*0.014 = 0.062, which is less than 1

Cannot use the normal approximation.

*** Students who showed that the normal approximation is not appropriate earn substantial partial marks even if they did not compute the correct probabilities below. Students who use the normal approximation even though it is clearly wrong get no partial marks (this bad approximation results in a large error: obtain probability of 24% instead of 14%) ***

Can use the Poisson approximation to Binomial by checking the rule of thumb: n*p = 100*0.02 = 2 ≤ 7 and n > 20. With a sample size of 100, if 0, 1, 2, or 3 customers need a service call then 3% or less would need a service call (0%, 1%, 2%, or 3%).

λ = n*p = 100*0.02 = 2

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad for\ x = 0,1,2,....$$

$$p(0) = \frac{e^{-2}2^0}{0!} = 0.1353$$

$$p(1) = \frac{e^{-2}2^1}{1!} = 0.2706$$

$$p(2) = \frac{e^{-2}2^2}{2!} = 0.2706$$

$$p(3) = \frac{e^{-2}2^3}{3!} = 0.1804$$

$$P(X > 3) = 1 - (0.1353 + 0.2706 + 0.2706 + 0.1804) = 0.1431$$

(2)

(a) You could draw a tree diagram if you have trouble identifying the possible samples and probabilities.

Possible samples	Sample median	Probability
111	1	$(0.4)^3$ = 0.064
115	1	$(0.4)^2(0.6)^1$ = 0.096
151	1	$(0.4)^2(0.6)^1$ = 0.096
511	1	$(0.4)^2(0.6)^1$ = 0.096
155	5	$(0.4)^1(0.6)^2$ = 0.144
515	5	$(0.4)^1(0.6)^2$ = 0.144
551	5	$(0.4)^1(0.6)^2$ = 0.144
555	5	$(0.6)^3$ = 0.216

Sample median	Probability
1	0.352
5	0.648

(b) $E[M] = 1 * 0.352 + 5 * 0.648 = 3.592$

The population median is 5: if sort balls in box obtain 1 1 5 5 5, which shows the median is 5. Hence the sample median is a biased estimate of the population median: 3.592 < 5. It is biased downward.

(3)

(a)

$$\overline{X} = \frac{\sum_{i=1}^{6} X_i}{6} = \frac{29 + 34 + 33 + 38 + 15 + 19}{6} = 28$$

$$s = \sqrt{\frac{\sum_{i=1}^{6} (X_i - 28)^2}{6 - 1}} = \sqrt{\frac{(1)^2 + (6)^2 + (5)^2 + (10)^2 + (-13)^2 + (-9)^2}{5}} = 9.08$$

$t_{\alpha/2} = t_{0.025} = 2.571$ with $\nu = 5$

$$\overline{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

$$28 \pm 2.571 \frac{9.08}{\sqrt{6}}$$

$$LCL = 18.47 \approx 18$$

$$UCL = 37.53 \approx 38$$

(b) We are 95% confident that the interval from 18 – 38 hours includes the mean number of hours spent on campus per week for all University of Toronto students that do not live on campus.

(c) We can no longer be 95% confident that the interval from 18 – 38 hours includes the mean number of hours spent on campus per week for all University of Toronto students that do not live on campus. In fact we have evidence that our point estimate of 28 is biased downward, which would shift the whole interval such that it would be much less than 95% likely that the interval contains the true population mean.