Exercise 9: Statistics 213 (L05) - Fall 2007

1. The random variable Y defined over the interval, $1 \leq Y \leq 6$, is uniformly distributed with probability density function $f(Y) = \frac{1}{5}$, $\mu = 3.5$ and variance $\sigma^2 = 2.083$.
   (a) If a random sample of size 50 is taken from this distribution, find $P(\overline{Y} < 3.91)$.
   (b) If a random sample of size 10 is taken from this distribution, could you use the central limit theorem to find $P(\overline{Y} < 3.91)$? Why or Why not ?

2. The number of customers per week at each store of a very large supermarket chain has a population mean of 5000 and a standard deviation of 500.
   (a) If a random sample of 64 stores is selected what is the probability that the sample mean will be below 5075 customers per week?
   (b) If a random sample of 16 stores is selected,
      i. Give the assumptions that must be made in order to solve parts (ii) and (iii) below. Next do the questions based on these assumptions holding
      ii. What is the probability that the sample mean will be below 5075 customers per week?
      iii. There is an 85% probability that the sample mean will be above this number of customers per week. How many customers per week?

3. Scores on an aptitude test used for determining admission to graduate study in business are known to be normally distributed with a mean of 500 and a population standard deviation of 100. How large sample would have to be taken to be 98% confident of estimating the true mean score of Stephen College applicants to within 30 points?

4. A sample survey was designed to estimate the proportion of sports utility vehicles being driven in the state of California. Suppose that a sample of 500 registrations selected from a Department of Motor Vehicles database contained 68 that were classified as sports utility vehicles.
   (a) Determine a 95% confidence interval for the proportion of sports utility vehicles in California. Interpret the confidence interval.
   (b) Using the data provided in the problem statement as a pilot study, how many registration would have to be sampled to estimate the proportion of sports utility vehicles to within 0.025 with 95%

5. An internet server conducted a survey of 250 of its customers and found that the average amount of time spent online was 10.5 hours per week with a standard deviation of 5.2 hours.
   (a) Construct a 95% confidence interval for the mean online time for all users of this Internet server.
   (b) If the online times for the population of all individuals using the Internet server were not normally distributed, would that invalidate any conclusions based on the confidence interval in (a) ? Explain.
   (c) If the Internet server claimed that its users averaged 12 hours of use per week, would you agree or disagree ? Explain.

6. The effectiveness of various drugs used to treat horses is discussed in the paper. One characteristic of interest is the *half-life* of a drug (the length of time until the concentration of the drug in the blood is one-half of the initial value). Among the drugs studied was sulfadimethoxine. The researchers treated 100 hoses with this drug, and measured the half-life of the drug on each horses. They determined a sample mean half-life over the 100 horses of 10.62 hours, and a sample standard deviation of 2.56 hours. Determine a 99% confidence interval for the mean half-life of sulfadimethoxine.

7. In a study to determine whether decisions to purchase were based on price or quality, a sample of 137 consumers was surveyed. Exactly 100 of these consumers indicated that they base their buying decisions on price.

   (a) Find 95% confidence interval for the proportion of consumers who base their buying decisions on price.

   (b) Using the $\hat{p}$, how many consumers must be sampled to estimate the proportion of all consumers who base their buying decisions on price to within 0.04 with 95% confidence ?

Solutions

---

1. (a). CLT applies, $\overline{X} \sim \mathcal{N}(3.5, 2.083/50)$. Want $P(\overline{Y} < 3.91) \approx P(Z < 2.01) \approx 0.9778$ (b). No, not with a sample size as small as 10 unless the distribution of population is mounded-shape and symmetric. The uniform distribution is symmetric, but not mounded-shape.

2. Let Y be the number of customers per week at store $E(Y) = 5000$, $\sigma_Y = 500$, $E(\overline{Y}) = 5000$, $\sigma_{\overline{Y}} = \frac{500}{\sqrt{n}}$. (a). Since n=64 is large enough that CLT applies, i.e. $\overline{Y} \sim$ Normal. Want $P(\overline{Y} < 5075) = 0.8849$. (b). (i) Since $n = 16$ is small, the distribution of Y must be such that it can be approximated by a normal distribution. Assuming normality of Y. (ii) $P(\overline{Y} < 5075) = 0.7257$. (iii). Want y_0 such that $P(\overline{Y} > y_0) = 0.85$, $y_0 = 4870$. i.e. an 85% chance $\overline{Y}$ will be above 4870 customers per week.

3. $z_{0.99} = 2.33, se(\overline{X}) = \frac{\sigma}{\sqrt{n}} = \frac{100}{\sqrt{n}}, K = 30,\ n = 60.32$, need a sample size of 61

4. (a). $\hat{p} = 0.136$, $z_{1-\alpha/2} = z_{0.975} = 1.96$, $se(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.0153$, 95% confidence interval for p: $(0.136 \pm 0.03) = (0.106, 0.166)$. The level of confidence, 95%, gives the proportion of intervals created by repeated sampling that would cover p, the proportion of sports utility vehicles in California. (b).$\hat{p} = 0.136, z_{1-\alpha/2} = z_{0.975} = 1.96, K = 0.025, n = 722.2 \approx 723$

5. $n = 250$, $\overline{X} = 10.5$, $\hat{\sigma} = 5.2$, $z_{0.975} = 1.96$, $se(\overline{X}) = 0.3289$, the 95% confidence interval for μ: $(10.5 \pm 0.6446) = (9.85554, 11.1446)$. (b). No, it is not necessary to assume that the online times are normally distributed as $n > 30$, and the CLT holds. (c). We would disagree, as the confidence interval in (a) does not cover 12.

6. $n = 100, \bar{x} = 10.62, \hat{\sigma} = 2.56, se(\overline{X}) = \frac{2.56}{\sqrt{100}} = 0.256, z_{0.995} = 2.576$; 99% confidence interval for μ: $(10.62 \pm 2.56 \times 0.256) = (9.961, 11.279)$.

7. (a). $z_{0.975} = 1.96, se(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.0379$, 95% confidence interval of p: $(0.7299 \pm 0.0743) = (0.6556, 0.8043)$ (b). $z_{0.975} = 1.96, K = 0.04, \hat{p} = 0.7299, n = 473.3 \approx 474$