Problem Set 8 - Solutions

Statistics - NYU, Summer 2016
Ercan Karadas

[1] Random variable: X: Weight of a cereal box, and we know that $X \sim N(20, (0.6)^2)$.

a) $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n} = \frac{(0.6)^2}{4} \Rightarrow$ the standard error of the sample is $\sigma_{\bar{X}} = \frac{(0.6)}{2} = 0.3$.
And the distribution of the sample (of size 4) means: $\bar{X} \sim N(20, (0.3)^2)$.

b) $P(\bar{X} < 19.7) = P(\frac{\bar{X}-20}{0.3} < \frac{19.7-20}{0.3}) = P(Z < \frac{19.7-20}{0.3}) = P(Z < -1) = 0.1587$

c) $P(\bar{X} > 20.6) = P(Z > \frac{20.6-20}{0.3}) = P(Z > 2) = 0.0228$

d) $P(19.7 < \bar{X} < 20.5) = P(\frac{19.7-20}{0.3} < Z < \frac{20.5-20}{0.3}) = P(-1 < Z < 1.67) = 0.7938$

e) First four boxes are chosen and then two of the four boxes, but since at both steps selections are random we can think of as if we are choosing samples of size two, instead of four. Then,
the standard error of the sample is $\sigma_{\bar{X}} = \frac{(0.6)}{\sqrt{2}} = \frac{0.6}{\sqrt{2}}$.
And the distribution of the sample (of size 2) means: $\bar{X} \sim N(20, (\frac{0.6}{\sqrt{2}})^2)$

Finally, $P(19.5 < \bar{X} < 20.5) = P(\frac{19.5-20}{\frac{0.6}{\sqrt{2}}} < Z < \frac{20.5-20}{\frac{0.6}{\sqrt{2}}}) = P(-1.18 < Z < 1.18) = 0.762$

[2] Random variable: X: time spent studying, and we know that $X \sim N(?, (8)^2)$, $n = 4$. And the distribution of the sample (of size 4) means: $\bar{X} \sim N(?, (8)^2/4)$, and note that $\sigma_{\bar{X}} = 4$

a) $P(\bar{X} - \mu > 2) =?$
$P(\frac{\bar{X}-\mu}{4} > \frac{2}{4}) = P(Z > \frac{1}{2}) = 0.3085$

b) $P(\bar{X} - \mu < -3) =?$
$P(\frac{\bar{X}-\mu}{4} < \frac{-3}{4}) = P(Z < \frac{-3}{4}) = 0.2266$

c) $P(|\bar{X} - \mu| > 4) =?$
$P(\bar{X} - \mu < -4 \quad \text{or} \quad \bar{X} - \mu > 4) = P(\bar{X} - \mu < -4) + P(\bar{X} - \mu > 4) = P(Z < -1) + P(Z > 1) = 0.3174$

d) Here we don't know the desired sample size yet, so the best we can do is to leave the standard error of the sample mean as a function of sample size: $\sigma_{\bar{X}} = \frac{8}{\sqrt{n}}$.
The sample size n should satisfy the following:
$P(|\bar{X} - \mu| > 1.0) < 0.10$
$P(\bar{X} - \mu < -1 \quad \text{or} \quad \bar{X} - \mu > 1) = P(\frac{\bar{X}-\mu}{8/\sqrt{n}} < \frac{-1}{8/\sqrt{n}}) + P(\frac{\bar{X}-\mu}{8/\sqrt{n}} > \frac{1}{8/\sqrt{n}}) = 0.1$

Here we find the desired sample size n for the equality case, and for the inequality case it will be enough to say "sample size should be at least n".
Since standard normal distribution is symmetric, each probability in the last equation should be equal to 0.05. Therefore the problem reduces down to finding an n satisfying:
$P(\frac{\bar{X}-\mu}{8/\sqrt{n}} > \frac{1}{8/\sqrt{n}}) = 0.05$. In turn this is equivalent to write:
$P(Z > \frac{1}{8/\sqrt{n}}) = 0.05$. Then, from Z-table:
$\frac{1}{8/\sqrt{n}} = 1.645$ which gives $n = 174$ (after rounding up).

Therefore, if the sample size is at least 174, then the probability that the sample mean differs from the population mean by more than 1.00 will be less than 0.10

[3] Let X denote the nicotine content of the cigarettes. Then $X \backsim N(20, 5^2)$.

a) $P(X > 23) = P(Z > \frac{23-20}{5}) = P(Z > 0.6) = 0.2743$.
b) $P(\overline{X} > 23) = P(Z > \frac{23-20}{5/\sqrt{25}}) = P(Z > 3) = 0.00135$.

[4] a) The Uniform distribution on [1,10] best fits the data. In addition, we would need to assume uniformity over the year.

b) $\mu = 5.51$ and $\sigma = 2.87$.

c) Let X denote the failing time of the device (in years). Then using the uniform distribution, $P(X > \frac{80}{12}) = (\frac{1}{10})(\frac{40}{12}) = \frac{1}{3}$.

d) Since $n = 36 > 25$, by the CLT, $\overline{X} \backsim N(5.51, \frac{2.87^2}{36})$. Then $P(\overline{X} < \frac{80}{12}) = P(Z < \frac{80/12-5.51}{2.87/\sqrt{36}}) = P(Z < 2.418) = 0.9922$.

[5] Let X denote the daily balances in the savings accounts. Then $X \backsim N(108, 15^2)$.

a) $B \backsim N(108, (\frac{15}{\sqrt{4}})^2)$. Then $P(B > 116) = P(Z > \frac{116-110}{15/\sqrt{4}}) = P(Z > 1.067) = 0.1423$.

b) Now $B \backsim N(108, (\frac{15}{\sqrt{16}})^2)$. Then $P(B > 116) = P(Z > \frac{116-110}{15/\sqrt{16}}) = P(Z > 2.133) = 0.01659$.

c) The standard deviation of B fell as the sample size n increased. With more data, the distribution of B becomes more precise, as given by the formula $\sigma_B = \frac{\sigma}{\sqrt{n}}$.

d) The probability of observing $B = \$116$ fell as the sample size n increased. This follows directly from the fall in the standard deviation of B - tail probabilities then become smaller.

[6] Skip this problem.

[7] a) $P(s > a) = 0.01 \Rightarrow a =?$
Equivalently $P(s^2 > a^2) = 0.01 \Rightarrow a =?$
Note that it is better to have s^2 rather than s because in the χ^2 distribution that we are going to use below s^2 will appear not s.
$P(\frac{s^2(n-1)}{\sigma^2} > \frac{a^2(n-1)}{\sigma^2}) = 0.01$
$P(\chi^2_{14} > \frac{a^2(n-1)}{\sigma^2}) = 0.01$
From χ^2-table, we can write $29.14 = \frac{a^2(n-1)}{\sigma^2} = \frac{a(14)}{(1.8)^2}$ so $a^2 = 6.74$, and finally $a = 2.596$.

b) $P(s < a) = 0.025 \Rightarrow a =?$
Equivalently $P(s^2 < a^2) = 0.025 \Rightarrow a =?$

$P(\frac{s^2(n-1)}{\sigma^2} < \frac{a^2(n-1)}{\sigma^2}) = 0.025$
$P(\chi^2_{14} < \frac{a^2(n-1)}{\sigma^2}) = 0.01$
From χ^2-table, we can write $5.63 = \frac{a^2(n-1)}{\sigma^2} = \frac{a^2(14)}{(1.8)^2}$, so $a^2 = 1.303$, and finally $a = 1.141$.

c) $P(a < s < b) = 0.9 \Rightarrow a, b =?$
Equivalently $P(a^2 < s^2 < b^2) = 0.9 \Rightarrow a, b =?$

Assuming that we leave equal areas in the upper and the lower tails, we can write:
$P(a^2 < s^2 < b^2) = 0.9$ if and only if $P(s^2 < a^2) = 0.05$ and $P(s^2 > b^2) = 0.05$
Then following the same steps in parts (a) and (b) we can obtain the following two equations:
$6.57 = \frac{a^2(14)}{(1.8)^2}$ and $23.68 = \frac{b^2(14)}{(1.8)^2}$
Solving these equations yield: $a = 1.233$ and $b = 2.341$

[8] X : Undergraduate GPA; $X \sim N(?, \quad .45^2)$; $n = 25$; $\bar{X} = 2.90$

a) $(1 - \alpha)100 = 95 \Rightarrow \alpha = 0.05$
95% confidence interval for the population mean (μ):
$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ and $z_{\alpha/2} = z_{0.025} = 1.96$
$2.9 \pm 1.96\frac{0.45}{\sqrt{25}} \Rightarrow [2.7236, 3.0764]$

Interpretation: If we select samples of size 25, and calculate 95% confidence interval for the population mean (μ) repeatedly, as we did above for one such sample, 95% of these intervals would contain the true population mean (μ).

Remark: This does NOT mean that $P(2.7236 \leq \mu \leq 3.0764) = 0.95$. Because true population mean (μ) is a constant number (though unknown at that point), and therefore this specific interval contains the true population mean or not.

b) If everything else is the same, a higher confidence level results in a wider interval than as found in part (a).

c) Higher standard deviation means dispersion of the random variable is higher, which in turn result in a wider confidence interval than found in part (a). You could see this just by looking at how changing the standard deviation effects the confidence interval in the formula of C.I. without doing any calculations.

d) Increasing the sample size reduces the variance of the sample means ($\sigma_{\bar{X}^2} = \frac{\sigma^2}{n}$), which in turn means $\bar{X}$ becomes a better estimator of the true population mean (μ), and therefore the confidence interval will be narrower than the one found in part (a). You can see this just looking at how changing n effects the confidence interval in the formula of C.I. without doing any calculations.

e) $2.99 - 2.90 = ME = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
$0.09 = z_{\alpha/2}\frac{0.45}{\sqrt{25}}$
Solving this equation yields $z_{\alpha/2} = 1$
Then $\alpha = 2[1 - F(1)] = 0.3174$, and therefore confidence level is $100(1 - .3174)\% = 68.26\%$.

[9] a) We should have $z_{0.025}\frac{\sigma}{\sqrt{n}} = \frac{\sigma}{4}$ or substituting $z_{0.025} = 1.96$ we need to solve the equation $1.96\frac{\sigma}{\sqrt{n}} = \frac{\sigma}{4}$. Solving for n yields $n = 62$.

b) Similarly solving for n in $1.96\frac{\sigma}{\sqrt{n}} = k\sigma$ gives $n = (\frac{1.96}{k})^2$

[10] a) $X \backsim N(1000, 150^2)$.
Then $P(X > 1150) = P(Z > \frac{1150-1000}{150}) = P(Z > 1) = 0.1587$.

b) $P(925 < X < 1150) = P(\frac{925-1000}{150} < Z < \frac{1150-1000}{150}) = P(-0.5 < Z < 1) = 0.84134 - 0.30854 = 0.5328$.

c) We want to solve for a such that $P(X < a) = 0.05$:

$$\begin{aligned} P(Z < \frac{a-1000}{150}) &= 0.05 \\ \frac{a-1000}{150} &= z_{0.05} \\ \frac{a-1000}{150} &= -1.64 \\ a &= 754 \end{aligned}$$

Thus the warranty should cover bulbs with a life of 754 hours or less.

[11] a)

$$\begin{aligned} P\big(18 < \overline{X} < 26) &= P(\frac{18-24.7}{19.3/\sqrt{64}} < Z < \frac{26-24.7}{19.3/\sqrt{64}}) \\ &= P(-2.79 < Z < 0.54) \\ &= .4974 + .2054 = .7028 \end{aligned}$$

b) No. Since $n = 64 > 25$, the Central Limit Theorem would allow to say $\overline{X} \backsim N(24.7, \frac{(19.3)^2}{64})$, nothing would change.

[12] $X \backsim N(500, 75^2)$. Since $n = 50$, $\overline{X} \backsim N(500, \frac{75^2}{50})$.

Then $P(475 < \overline{X} < 525) = P(\frac{475-500}{75/\sqrt{50}} < Z < \frac{525-500}{75/\sqrt{50}}) = P(-2.357 < Z < 2.357) = 0.98172$.

[13] Let X denote the number of miles per gallon of the Honda. Then $X \backsim N(40, 5^2)$.

a) $P(X > 45) = P(Z > \frac{45-40}{5}) = P(Z > 1) = 0.1587$.

b) $P(\overline{X} > 45) = P(Z > \frac{45-40}{5/\sqrt{100}}) = P(Z > 10) = 0$.

c) Let Y be the number of cars, out of 100, that get more than 45 miles per gallon. Then $Y \backsim Bin(100, 0.1587)$. Since $Var(Y) = 100(0.1587)(0.8413) = 13.35 > 5$, we can use the normal approximation to the binomial and write $Y \backsim N(15.87, 13.35)$.

Then $P(Y > 35) = P(Z > \frac{35-15.87}{\sqrt{13.35}}) = P(Z > 5.236) = 0$. We would expect 15.87 cars, out of the 100, to get more than 45 miles per gallon.