Econ 250 Winter 2009
Assignment 2 - Solutions

1. For a restaurant, the time it takes to deliver pizza (in minutes) is uniform over the interval (25, 37). Determine the proportion of deliveries that are made in less than half an hour.

   Solution: Let X be the time it takes to deliver pizza. Then,

   $$X \sim U[25, 37].$$

   The proportion of deliveries that are made in less than half an hour is simply

   $$\begin{aligned} P(X < 30) &= \int_{27}^{35} \frac{dx}{37-25} \\ &= 5 \times \frac{1}{12} \\ &= 0.4167 \end{aligned}$$

2. The length of an aluminum-coated steel sheet manufactured by a certain factory is approximately normal with mean 75 centimeters and standard deviation 1 centimeter. Find the probability that a randomly selected sheet manufactured by this factory is between 74.5 and 75.8 centimeters.

   Solution: Let L denote the length of an aluminum-coated steel sheet. Then,

   $$L \sim N[75, 1].$$

   The probability that L is between 74.5 and 75.8 cm is

   $$\begin{aligned} P(74.5 \leq L \leq 75.8) &= P(\frac{74.5-75}{\sqrt{1}} \leq \frac{L-75}{\sqrt{1}} \leq \frac{75.8-75}{\sqrt{1}}) \\ &= P(-0.5 \leq L \leq 0.5) \\ &= F_z(0.5) - F_z(-0.5) \\ &= F_z(0.5) - (1 - F_z(0.5)) \\ &= 2F_z(0.5) - 1 = 2 \times 0.6915 - 1 = 0.383 \end{aligned}$$

3. Let X, the grade of a randomly selected student in the first midterm of ECON250, be a normal random variable. A professor is said to grade the test *on the curve* if he finds the average μ and the standard deviation σ of the grades and then assigns letter grades according to the following table.

Range of the grade	Letter grade
$X \geq \mu + \sigma$	A
$\mu + \sigma > X \geq \mu$	B
$\mu > X \geq \mu - \sigma$	C
$\mu - \sigma > X \geq \mu - 2\sigma$	D
$X > \mu - 2\sigma$	F

Suppose that the professor grades the test on the curve. Determine the percentage of the students who will get A, B, C, D, and F, respectively. Note: this kind of grading scheme is actually not permitted at Queen's.

Solution: We can interpret the probability that one student will get a particular grade as the *fraction* of all students that will get that grade. So the percentage of students that will get an A is simply

$$\begin{aligned} P(X \geq \mu + \sigma) &= P(\frac{X - \mu}{\sigma} \geq \frac{\mu + \sigma - \mu}{\sigma}) \\ &= P(Z \geq 1) \\ &= 1 - P(Z \leq 1) \\ &= 1 - F_z(1) = 0.1587 \end{aligned}$$

The percentage of students that will get a B is

$$\begin{aligned} P(\mu + \sigma > X \geq \mu) &= P(\frac{\mu + \sigma - \mu}{\sigma} > \frac{X - \mu}{\sigma} \geq \frac{\mu - \mu}{\sigma}) \\ &= P(1 > Z \geq 0) \\ &= F_z(1) - F_z(0) \\ &= 0.8413 - 0.5000 = 0.3413 \end{aligned}$$

The percentage of students that will get a C is

$$\begin{aligned} P(\mu > X \geq \mu - \sigma) &= P(\frac{\mu - \mu}{\sigma} > \frac{X - \mu}{\sigma} \geq \frac{\mu - \sigma - \mu}{\sigma}) \\ &= P(0 > Z \geq -1) \\ &= F_z(0) - F_z(-1) \\ &= 0.5000 - (1 - 0.8413) = 0.3413 \end{aligned}$$

The percentage of students that will get a D is

$$\begin{aligned} P(\mu - \sigma > X \geq \mu - 2\sigma) &= P(\frac{\mu - \sigma - \mu}{\sigma} > \frac{X - \mu}{\sigma} \geq \frac{\mu - 2\sigma - \mu}{\sigma}) \\ &= P(-1 > Z \geq -2) \\ &= F_z(-1) - F_z(-2) \\ &= (1 - 0.8413) - (1 - 0.9772) = 0.1559 \end{aligned}$$

The percentage of students that will fail is

$$\begin{aligned}P(X > \mu - 2\sigma) &= P(\frac{X-\mu}{\sigma} > \frac{\mu - 2\sigma - \mu}{\sigma}) \\ &= P(Z \geq -2) \\ &= 1 - P(Z \leq 2) \\ &= 1 - F_z(2) = 0.0228\end{aligned}$$

4. Suppose that 90% of the patients with a certain disease can be cured with a certain drug.

    a. What is the approximate probability that, of 50 such patients, at least 45 can be cured with the drug?

    b. Approximate the same probability but using the continuity correction.

    Solution: This is a binomial question. A success occurs is a patient can be cured with a certain drug. The probability of a success is 0.9. Let X be the number of successes in 50 trials. Then, the probability that there are at least 45 successes in 50 trials is

$$P(X \geq 45) = \sum_{x=45}^{50} \binom{50}{45} 0.9^x 0.1^{50-x}$$

This is difficult to compute directly, so we can approximate it using the normal distribution. We know that when $n\pi(1-\pi) > 9$ the approximation is good. In this case, $n\pi(1-\pi) = 50 \times 0.9 \times 0.1 = 4.5 < 9$ so the approximation may not be very good. Proceeding anyway, we know that approximately $X \sim N(50 \cdot 0.9, 0.9 \cdot 0.1 \cdot 50)$ or $X \sim N(45, 4.5)$.

a. Then,

$$\begin{aligned}P(X \geq 45) &\approx P(\frac{X-45}{\sqrt{4.5}} \geq \frac{45-45}{\sqrt{4.5}}) \\ &= P(Z \geq 0) \\ &= 1 - F_z(0) \\ &= 1 - 0.5 = 0.5\end{aligned}$$

b. Applying the continuity correction we have

$$\begin{aligned} P(Xle45) &\approx P(X \geq 45.5) \\ &= P(\frac{X-45}{\sqrt{4.5}} \geq \frac{44.5-45}{\sqrt{4.5}}) \\ &= P(Z \geq -0.24) \\ &= 1 - F_z(-0.24) \\ &= 1 - (1 - F_z(0.24)) = 0.5948 \end{aligned}$$

The exact value for the binomial probability is 0.616123. So we see that when the normal approximation is not very good, the continuity correction can be very large and provide a large improvement in accuracy.

5. The weight of a single 9-ounce bag of potato chips has an $N(9.12, 0.15)$ distribution. Consider taking a sample of bags of chips and calculating the average weight of the sample.

   a. What is the probability of a 3-bag average falling between 9.08 and 9.16 ounces?

   b. What is the probability of a 30-bag average falling between 9.08 and 9.16 ounces?

   c. What is the probability of a 150-bag average falling between 9.08 and 9.16 ounces?

   d. If the weights were known to have a mean of 9.12 ounces and a standard deviation of 0.15 ounces but the distribution of the weights was unknown, what would you be able to say about the probabilities you calculated in parts (a), (b), and (c)? Would any of them still be reasonable accurate?

   Solution: Let W be the weight of a single 9-ounce bag of potato chips. Then,

$$W \sim N(9.12, 0.15)$$

   a. Since the data is normally distributed and independent, the sampling distribution of the sample mean, $\bar{X}$ is

$$N(9.12, \frac{0.15}{3})$$

for a 3-bag sample. Then,

$$\begin{aligned}P(9.08 \leq \bar{X} \leq 9.16) &= P(\frac{9.08 - 9.12}{\sqrt{0.15/3}} \leq Z \leq \frac{9.16 - 9.12}{\sqrt{0.15/3}}) \\ &= P(-0.18 \leq Z \leq 0.18) \\ &= F_z(0.18) - F_z(-0.18) \\ &= F_z(0.18) + F_z(0.18) - 1 = 0.1428\end{aligned}$$

b. Again, since the data is normally distributed and independent, the sampling distribution of the sample mean, $\bar{X}$ is

$$N(9.12, \frac{0.15}{30})$$

for a 30-bag sample. Then,

$$\begin{aligned}P(9.08 \leq \bar{X} \leq 9.16) &= P(\frac{9.08 - 9.12}{\sqrt{0.15/30}} \leq Z \leq \frac{9.16 - 9.12}{\sqrt{0.15/30}}) \\ &= P(-0.57 \leq Z \leq 0.57) \\ &= F_z(0.57) - F_z(-0.57) \\ &= F_z(0.57) + F_z(0.57) - 1 = 0.4314\end{aligned}$$

c. With a 150-bag sample, the sampling distribution of the sample mean, $\bar{X}$ is

$$N(9.12, \frac{0.15}{150})$$

. Then,

$$\begin{aligned}P(9.08 \leq \bar{X} \leq 9.16) &= P(\frac{9.08 - 9.12}{\sqrt{0.15/150}} \leq Z \leq \frac{9.16 - 9.12}{\sqrt{0.15/150}}) \\ &= P(-0.18 \leq Z \leq 0.18) \\ &= F_z(1.26) - F_z(-1.26) \\ &= F_z(1.26) + F_z(1.26) - 1 = 0.7924\end{aligned}$$

d. As the above calculations make clear, the larger the sample size, the higher the chances that the sample mean will be close to the population mean. In fact, for sample size of 30, the probability that the sample mean is within 0.08 of the population is about 43% which is not terribly high. It shoots to about 80% for a sample size of 150. If the distribution shape is unknown, then the sampling distribution is approximately normal for large n due to the CLT. So we would expect that if we didn't know the shape of the distribution that the result in c) would still be valid while the result in a) would be very suspicious. The result in b) would be a very rough answer but not completely invalid.

6. Plastic bags used for packaging produce are manufactured so that the breaking strengths of the bags are normally distributed with a standard deviation of 1.8 pounds per square inch. A random sample of 16 bags is selected.

   a. The probability is 0.01 that the sample standard deviation of breaking strengths exceeds what number?

   b. The probability is 0.15 that the sample mean exceeds the population mean by how much?

   c. The probability is 0.05 that the sample mean differs from the population mean by how much?

   Solution: Let B be the breaking strength of a plastic bag. Then,

   $$B \sim N(\mu, 1.8^2)$$

   where the population mean is denoted by μ as it is not given. Let $\bar{X}$ be the sample mean of the 16 plastic bags. Then, the sampling distribution of $\bar{X}$ is $N(\mu, \frac{1.8^2}{16})$.

   a. Please ignore, we didn't cover this.

   b. Let the sample mean exceed the population mean by at least x so that $\bar{X} > \mu + x$. Then, we are given that

   $$P(\bar{X} > \mu + x) = 0.15$$

   After standardization, we have

   $$\begin{aligned} P(\bar{X} > \mu + x) &= P(\frac{\bar{X} - \mu}{1.8/4} > \frac{\mu + x - \mu}{1.8/4}) \\ &= P(Z > x/0.45) = 0.15 \end{aligned}$$

   or

   $$1 - P(Z \leq x/0.45) = 0.15$$

   implying

   $$P(Z \leq x/0.45) = 0.85$$

Now, doing a reverse lookup on the tables we find that $F_z(1.04) = 0.8508$. Therefore, $x/0.45 = 1.04$ implying that $x = 0.468$. Hence, the probability that the sample mean exceeds the population means by 0.47 is 0.15.

c. We essentially need to apply the same logic as in b) except that we need to allow the sample mean to be either larger or smaller than the population mean. Let the difference be d so that

$$\mu - d \leq X \leq \mu + d$$

Now, we know that $P(\mu - d \leq X \leq \mu + d) = 0.05$. Then, proceeding as in b) we have

$$\begin{aligned} P(\mu - d \leq X \leq \mu + d) &= P(\frac{\mu - d - \mu}{0.45} \leq \frac{X - \mu}{0.45} \leq \frac{\mu + d - \mu}{0.45}) \\ &= P(\frac{-d}{0.45} \leq Z \leq \frac{d}{0.45}) \\ &= F_z(d/0.45) - F_z(-d/0.45) \\ &= F_z(d/0.45) - (1 - F_z(d/0.45)) \\ &= 2F_z(d/0.45) - 1 = 0.05 \end{aligned}$$

Hence,

$$F_z(d/0.45) = 0.525$$

Moreover, doing a reverse lookup on the tables we have $F_z(0.06) = 0.524$ which is close enough. Hence, $d/0.45 = 0.06$ or $d = 0.027$. In other words, the probability that the sample mean differs from the population mean by 0.06 is 0.05.

7. Independent random samples of accounting professors and information systems (IS) professors where asked to provide the number of hours they spend in preparation for each class. The sample of 321 IS professors had a mean time of 3.01 preparation hours, and the sample of 94 accounting professors had a mean time of 2.88 hours. From similar past studies that population standard deviation for the IS professors is assumed to be 1.09 and similarly the population standard deviation for the accounting professors is 1.01 and the population means are the same. What is the probability that the IS professors mean preparation time is at least 30 minutes more than accounting professors? Repeat the calculations for at least 15 minutes and at least 5 minutes and comment.

Solution: This is a difference in means question. Let $X_{1,i}$ denote the preparation time of the i^{th} accounting professor and let $X_{2,i}$ denote the preparation time of the i^{th} IS professor. Then, let $\bar{X}_1 = \sum_{i=1}^{n_1} \frac{X_{1,i}}{n_1}$ and $\bar{X}_2 = \sum_{i=1}^{n_2} \frac{X_{2,i}}{n_2}$ be the sample means of the preparation times of the accounting and IS professors respectively. Also, let μ_1, μ_2 and σ_1, σ_2 be the population mean preparation times and standard deviations for the two types of professors.

From the question, we know that $\bar{X}_1 = 3.01$, $n_1 = 94$ and $\bar{X}_2 = 2.88$, $n_2 = 321$. The populations standard deviations are 1.01 and 1.09 for accounting and IS professors respectively. Then, we want to know what is the probability that $\bar{X}_2$ will exceed $\bar{X}_1$ by at least 30 minutes. Noting that both n_1 and n_2 are large and appealing to the CLT, we know that the sampling distribution of the difference in means is $N(\mu_2 - \mu_1, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$. Then,

$$\begin{aligned}
P(\bar{X}_2 - \bar{X}_1 \geq 30) &= 1 - P(\bar{X}_2 - \bar{X}_1 \leq 0.5) \\
&= 1 - P(\frac{\bar{X}_2 - \bar{X}_1 - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq \frac{0.5 - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}) \\
&= 1 - P(Z \leq \frac{0.5}{\sqrt{\frac{1.01^2}{94} + \frac{1.09^2}{321}}}) \\
&= 1 - F_z(4.14) = 0
\end{aligned}$$

That is the probability of such a large difference in the sample means is practically zero. Repeating the calculation for 15 minutes, we have

$$\begin{aligned}
P(\bar{X}_2 - \bar{X}_1 \geq 30) &= 1 - P(\bar{X}_2 - \bar{X}_1 \leq 0.25) \\
&= 1 - P(\frac{\bar{X}_2 - \bar{X}_1 - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq \frac{0.25 - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}) \\
&= 1 - P(Z \leq \frac{0.25}{\sqrt{\frac{1.01^2}{94} + \frac{1.09^2}{321}}}) \\
&= 1 - F_z(2.07) = 0.02
\end{aligned}$$

The probability of a 15 minute difference is about 2%. In other words, if we were to sample repeatedly using these sample sizes about 2%

of the time, we would get a difference exceeding 15 minutes. Lastly, repeating the calculation for 5 minutes, we have

$$
\begin{aligned}
P(\bar{X}_2 - \bar{X}_1 \geq 30) &= 1 - P(\bar{X}_2 - \bar{X}_1 \leq 0.0833) \\
&= 1 - P(\frac{\bar{X}_2 - \bar{X}_1 - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq \frac{0.0833 - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}) \\
&= 1 - P(Z \leq \frac{0.0833}{\sqrt{\frac{1.01^2}{94} + \frac{1.09^2}{321}}}) \\
&= 1 - F_z(0.69) = 0.25
\end{aligned}
$$

Here we see the probability has jumped to about a quarter. So, a 5 minute difference in the sample mean preparation times occurs in about 1 out 4 samples.