STOCKHOLM UNIVERSITY
Department of Statistics

Emma Pettersson
2018-03-20

Solutions to Plenary Exercises: Plenary Exercise 3

Basic Statistics for Economists, 15 ECTS, STE101

EXERCISE 1

Before we can answer the questions, we need to determine the distribution of a package of paper.

Let X_k be the weight of sheet number $k = 1, \dots, 5000$. Assume that the weights of the sheets are independent from one another and that they are identically distributed.

$$X_k \sim N(80, 4^2)$$

In other words, the weight of a paper sheet is normally distributed with mean $\mu_X = 80$ g and variance $\sigma_X^2 = 4^2 = 16\ \text{g}^2$ or standard deviation $\sigma_X = 4$ g.

Let Y be the weight of a whole package:

$$Y = \sum_{k=1}^{5000} X_k = X_1 + X_2 + \cdots + X_{5000}$$

<u>Expected value of Y</u>:

$$\mu_Y = E(Y) = E(X_1 + X_2 + \cdots + X_{5000}) = E(X_1) + E(X_2) + \cdots + E(X_{5000})$$

$$= \underbrace{\mu_X + \mu_X + \cdots + \mu_X}_{5000} = 5000\mu_X = 5000 \cdot 80 = \mathbf{400\ 000g} = \mathbf{400\ kg}$$

<u>Variance of Y</u> (remember that the sheets of paper are independent of one another):

$$\sigma_Y^2 = Var(Y) = Var(X_1 + X_2 + \cdots + X_{5000}) = Var(X_1) + Var(X_2) + \cdots + Var(X_{5000})$$

$$= \underbrace{\sigma_X^2 + \sigma_X^2 + \cdots + \sigma_X^2}_{5000} = 5000\sigma_X^2 = 5000 \cdot 16 = \mathbf{80\ 000\ g^2}$$

<u>Standard deviation of Y</u>: $\sigma_Y = \sqrt{80\ 000} = \mathbf{282.843\ g} = \mathbf{0.282843\ kg}$

Variance of Y in kg^2: $\sigma_Y^2 = \mathbf{0.282843^2} = \mathbf{0.08\ kg^2}$

The distribution of Y:

If all $X_k \sim N \implies Y \sim N$ i.e. if the weights of all the sheets of paper are normally distributed, then the package of 5000 sheets of paper will also be normally distributed.

In this case $\boldsymbol{Y \sim N(\mu_Y = 400\text{kg},\ \sigma_Y = 0.282843\text{kg})}$

a. Since the customer will return the package if it weighs less than 399.5 kg, we want to find the probability that this happens:

$$P(Y < 399.5) = P\left(Z < \frac{399.5 - \mu_Y}{\sigma_Y}\right) = P\left(Z < \frac{399.5 - 400}{0.282843}\right) \approx P(Z < -1.77)$$

$$= [\text{normal distribution is symmetrical}] = P(Z > 1.77)$$

$$= [\text{complement rule}] = 1 - P(Z < 1.77)$$

$$= [\text{from Table 1}] = 1 - 0.96164 = \mathbf{0.03836}$$

b. Assume that all $n = 1000$ packages are independent of one another. If the weights of different sheets of paper are independent, then it is reasonable to assume that the weights of the packages are also independent.

Define the random variable W as the number of packages (out of 1000) that are returned.

Distribution of W: $W \sim Bin(n = 1000, P = 0.03836)$

Expected value of W i.e. the number of packages that are expected to be returned:

$$\mu_W = E(W) = \boldsymbol{np} = 1000(0.03836) = \mathbf{38.36}$$

c. Let the profit be denoted and defined by $V = price - cost$.

Profit per package if package is not returned:

$$V = 5000 \cdot 0.4 - 5000 \cdot 0.3 = 2000 - 1500 = 500$$

Profit per package if package is returned: $V = 1200 - 1500 = -300$

Sample space of V: $S_V = \{-300, 500\}$

Probability distribution of V: $P(V = -300) = 0.03836$ $P(V = 500) = 0.96164$

Expected profit per package:

v	-300	500	Sum
$P(V = v)$	0.03836	0.96164	1
$v \cdot P(V = v)$	-11.508	480.82	$\mathbf{469.312}$

The expected profit (per package): $\mu_V = \sum_{v \in S_V} v \cdot P(V = v)$

$$(-300) \cdot 0.03836 + 500 \cdot 0.96164 = \mathbf{469.312}$$

d. The average weight of each sheet of paper is reduced to 79.8 g; the standard deviation is unchanged but the production cost drops to 0.28 SEK.

The expected weight of a package:

$$\mu_Y = 5000 \cdot 79.8 = 399\,000 \text{ g} = \mathbf{399\ kg}$$

The probability that a package is returned:

$$P(Y < 399.5) = P\left(Z < \frac{399.5 - \mathbf{399}}{0.282843}\right) = P(Z < 1.77) = 0.96164$$

Profit per package if package is not returned:

$$V = 5000 \cdot 0.4 - 5000 \cdot 0.28 = 2000 - 1400 = 600$$

Profit per package if package is returned: $V = 1200 - 1400 = -200$

Expected profit per package:

v	-200	600	Sum
$P(V = v)$	0.96164	0.03836	1
$v \cdot P(V = v)$	-192.328	23.016	$\boldsymbol{\mu_V = -169.312}$

Conclusion: On average we'll be losing 169.31 SEK per package. The idea wasn't that good after all, so don't reduce the average weight!

EXERCISE 2

Let X be the weight of a random tablet. $X \sim N(\mu_X = 0.65 \text{ g}, \sigma_X = 0.02 \text{ g})$

a. $P(0.64 < X < 0.66) = P(X < 0.66) - P(X < 0.64)$

$$= [\text{standardize}] = P\left(Z < \frac{0.66 - \mu_X}{\sigma_X}\right) - P\left(Z < \frac{0.64 - \mu_X}{\sigma_X}\right)$$

$$= P\left(Z < \frac{0.66 - 0.65}{0.02}\right) - P\left(Z < \frac{0.64 - 0.65}{0.02}\right) = P(Z < 0.5) - P(Z < -0.5)$$

$$= [\text{by symmetry}] = P(Z < 0.5) - P(> 0.5)$$

$$= [\text{complement rule}] = P(Z < 0.5) - [1 - P(Z < 0.5)] = 2 \cdot P(Z < 0.5) - 1$$

$$= [\text{from Table 1}] = 2 \cdot 0.69146 - 1 = \mathbf{0.38292}$$

b. Assume that the weights X in a sample of size $n = 30$ are independent of each other and that $X \sim N(\mu_X, \sigma_X^2)$. Then the sample mean will have a distribution of $\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$.

The probability that the mean weight of the tablets lies in the given interval is:

$$P(0.64 < \bar{X} < 0.66) = P(\bar{X} < 0.66) - P(\bar{X} < 0.64)$$

$$= [\text{standardize}] = P\left(Z < \frac{0.66 - \mu_X}{\sigma_X/\sqrt{n}}\right) - P\left(Z < \frac{0.64 - \mu_X}{\sigma_X/\sqrt{n}}\right)$$

$$= P\left(Z < \frac{0.66 - 0.65}{0.02/\sqrt{30}}\right) - P\left(Z < \frac{0.64 - 0.65}{0.02/\sqrt{30}}\right) \approx P(Z < 2.74) - P(Z < -2.74)$$

$$= [\text{by symmetry and the complement rule}] = 2 \cdot P(Z < 2.74) - 1$$

$$= [\text{from Table 1}] = 2 \cdot 0.99693 - 1 = \mathbf{0.99386}$$

The probability that the mean weight of the tablets does <u>not</u> lie in the interval is:

$$P(\bar{X} < 0.64 \cup \bar{X} > 0.66) = 1 - P(0.64 < \bar{X} < 0.66) = 1 - 0.99386 = \mathbf{0.00614}$$

c. The probability that the mean weight lies <u>outside</u> the interval can be at most 0.05.

$$1 - P(0.64 < \bar{X} < 0.66) < 0.05 \quad \Leftrightarrow \quad P(0.64 < \bar{X} < 0.66) > 0.95$$

$$P(0.64 < \bar{X} < 0.66) = [\text{standardize}] = P\left(Z < \frac{0.66 - 0.65}{0.02/\sqrt{n}}\right) - P\left(Z < \frac{0.64 - 0.65}{0.02/\sqrt{n}}\right)$$

$$P\left(Z < \frac{0.01}{0.02/\sqrt{n}}\right) - P\left(Z < \frac{-0.01}{0.02/\sqrt{n}}\right) = P\left(Z < \sqrt{n}/2\right) - P\left(Z < -\sqrt{n}/2\right)$$

$$= [\text{by symmetry and the complement rule}]$$

$$= 2 \cdot P\left(Z < \sqrt{n}/2\right) - 1 > 0.95 \quad \Leftrightarrow \quad \boxed{P\left(Z < \sqrt{n}/2\right) > \frac{(1 + 0.95)}{2} = 0.975}$$

Look in Table 1 to find the value of z that gives $P(Z < z) = 0.975$ or look in Table 2 to find the value of z that gives the complement probability $P(Z > z) = 0.025$.

From the tables we get that $\boldsymbol{z} = \mathbf{1.96}$ so:

$$P\left(Z < \sqrt{n}/2\right) > P(Z < 1.96) \quad \Leftrightarrow \quad \sqrt{n}/2 > 1.96 \quad \Leftrightarrow \quad n > 15.3664$$

The answer is that we need at least $\boldsymbol{n} = \mathbf{16}$ in the sample i.e. $n \geq 16$.

EXERCISE 3

The size of the target group is unknown, interpret that as $N \to \infty$. The sample size is $n = 1000$.

Select persons at random and let the random variable $X_k = 0$ if person k doesn't know about the product, and $X_k = 1$ if person k does know about the product.

Let $Y = \sum_{k=1}^{1000} X_k$ be the number of people in the sample that have an awareness of the product.

Distribution of Y: Assume independence between individuals, and that the probability that a random person knows about the product is constant $p = 0.2$. Then Y is binomially distributed:

$$Y \sim Bin(1000, 0.2)$$

a. Mean, variance and standard deviation:

$$\boldsymbol{\mu_Y} = E(Y) = \boldsymbol{np} = 1000(0.2) = \mathbf{200}$$

$$\boldsymbol{\sigma_Y^2} = V(Y) = \boldsymbol{np(1-p)} = 1000(0.2)(0.8) = \mathbf{160}$$

$$\boldsymbol{\sigma_Y} = \sqrt{160} = \mathbf{12.6491}$$

b. We want the probability that 150 individuals or less have awareness of the product, i.e. $P(Y \leq 150)$. Since the tables do not provide probabilities for $n = 1000$ we can consider using the **normal distribution as an approximation** of the binomial distribution. Checking the Rule of thumb:

$$np(1-p) = \sigma_Y^2 = 160 > \mathbf{5} \Rightarrow \text{ok to approximate with a normal distribution!}$$

Use $\mu_Y = 200$ and $\sigma_Y = 12.6491$:

$$P(Y \leq 150) = [\text{standardize}] = P\left(Z \leq \frac{150 - \mu_Y}{\sigma_Y}\right) = P\left(Z \leq \frac{150 - np}{np(1-p)}\right)$$

$$= P\left(Z \leq \frac{150 - 200}{12.6491}\right) = P(Z \leq -3.95) = [\text{by symmetry}] = P(Z \geq 3.95)$$

$$= [\text{complement rule}] = 1 - P(Z < 3.95) \approx [\text{Table 1}] = 1 - 0.99996 = \mathbf{0.000040}$$

NOTE: Remember that the probability is an approximation, it's not the true probability. We could calculate the true probability by using the probability function of the binomial:

$$P(Y \leq 150) = P(Y = 0) + \cdots + P(Y = 150) = \sum_{y=0}^{150} \binom{1000}{y} 0.2^y 0.8^{150-y}$$

but this is obviously too cumbersome to do by hand. We could use computer software, Excel even, to do it for us in which case we arrive at $P(Y \leq 150) = \mathbf{0.000026}$.

c. No. If we assume that 20% of the population are aware of the product, the probability of collecting a sample where less than 151 people knew about the product was only 0.004%. The results from the sample indicate that it's very unlikely that the true population proportion is 20%, we should expect a larger proportion than 15% in the sample.