P8130_hw2_ym2813

Anna Ma

10/6/2021

Problem 1:

Suppose that probability of having at least one dental check up during a two-year period is 65%. 40 random individuals are being studied for their health care utilization practices over a two-year period.

a) What is the probability that exactly 24 of these individuals will have at least one dental checkup?

1-1.jpg

Let x be the random variable denoting the number of People come to dental check up.

$X \sim Bin(40, 0.65)$ where $n=40$, $p=0.65$ (Probability of coming to the check-ups)

$$P(X=24) = f(24) = \binom{40}{24} \cdot 0.65^{24} (1-0.65)^{40-24}$$

$$= \frac{40!}{24!\ 16!} 0.65^{24} \cdot 0.35^{16}$$

$$\boxed{= 0.103}$$

Figure 1: Hand written answer for part a

b) What is the probability that at least 24 of these individuals will have at least one dental checkup?

$P(X \geqslant 24) = 1 - P(X < 24) = 1 - P(X \leqslant 23)$

```
p_at_least_24 = 1 - pbinom(23, 40, 0.65)
p_at_least_24
```

```
## [1] 0.7977604
```

The probability that at least 24 of the individuals has at least one dental checkup is $P(X \geqslant 24) = 0.7977604$

c) Could you use an approximation method to calculate the probabilities above? If yes, calculate the probabilities using approximations and compare to the exact values; otherwise, explain why approximations methods are not appropriate.

Case 1: Use Poisson distribution to approximate Binomial distribution. This approximation has two condition: 1). n > 100; 2). p < 0.01. In our case, n = 40 and p = 0.65, both of which fails to satisfy the condition for approximating binomial with poisson. Therefore, we cannot use this approximation method.

Case 2: Use Normal distribution to approximate Binomial distribution This approximation has two conditions: 1). np $\geqslant$ 10; and 2). n(1-p) $\geqslant$ 10. In our case, np = 40 $\times$ 0.65 = 26, which satisfies the first condition; n(1-p) = 40 $\times$ (1-0.65) = 14 which satisfies the second condition. Therefore, we can use this approximation.

$\mu = np = 40 \times 0.65 = 26$, $\sigma = \sqrt{np(1-p)} = \sqrt{40 \times 0.65 \times (1-0.65)} = 3.012$

In a), $P(X = 24)$, rewrite with the continuity correction factor for normal distribution and we get $P(24 - \frac{1}{2} \leqslant X \leqslant 24 + \frac{1}{2})$.

Applying Z transformation:

$$\begin{aligned} P(24 - \frac{1}{2} \leqslant X \leqslant 24 + \frac{1}{2}) &= P(\frac{23.5 - 26}{3.02} \leqslant Z \leqslant \frac{24.5 - 26}{3.02}) \\ &= P(-0.83 \leqslant Z \leqslant -0.497) \\ &= P(0.497) - P(0.83) \\ &= 0.1058829 \end{aligned} \tag{1}$$

The normal approximation calculates that the probability that exactly 24 individuals came to the check up is $P(23.5 \leqslant X \leqslant 24.5) = 0.10588$. This is very close to the exact probability, which is 0.103, meaning that the approximation is appropriate.

In b), $P(X \geqslant 24)$, rewrite with the continuity correction factor for normal distribution and we get $P(X > 24 - \frac{1}{2}) = P(X > 23.5)$

Applying Z transformation:

$$\begin{aligned} P(X > 23.5) &= P(Z > \frac{23.5 - 26}{3.02} \\ &= P(Z > -0.83) \\ &= 1 - P(Z \leqslant -0.83) \\ &= 0.7963748 \end{aligned} \tag{2}$$

The normal approximation calculates that the probability that at least 24 individuals came to the check up is 0.7963.This is very close to the exact probability, which is 0.7977, meaning that the approximation is appropriate.

d) How many individuals do you expect to have at least one dental checkup?

$E(x) = np = 40 \times 0.65 = 26$

Therefore, we expect 26 individuals to have at lease one dental checkup.

e) What is the SD of the number of individuals who will have at least one dental checkup?

sd = $\sqrt{Var} = \sqrt{np(1-p)} = \sqrt{40 \times 0.65 \times (1-0.65)} = 3.0166206$

The standard deviation is 3.0166206

Problem 2

Suppose the number of tornados in the United States follows a Poisson distribution with parameter is 7 tornados per year. Compute using tables or R.Show the formula for "part a" (it can be handwritten and embedded in the pdf file).

a) What is the probability of having fewer than 3 tornados in the United States next year?

Let X be the random variable denoting the number of tornados in the United States. X ~ $Poi(\lambda)$ where $\lambda = 7$

$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

$P(x < 3) = P(x \leqslant 2) = \sum_{x=0}^{2} \frac{-7^x e^{-7}}{x!} = \frac{-7^0 e^{-7}}{0!} + \frac{-7^1 e^{-7}}{1!} + \frac{-7^2 e^{-7}}{2!}$=0.0296362

The probability of having fewer than 3 tornados in the United States next year is 0.0296362

b) What is the probability of having exactly 3 tornados in the United States next year?

$P(x = 3) = P(3) = \frac{-7^3 e^{-7}}{3!}$= 0.0521293

The probability of having exactly 3 tornados in the United States next year is 0.0521293

c) What is the probability of having more than 3 tornados in the United States next year?

$$
\begin{aligned}
P(X > 3) &= 1 - P(x \leqslant 3) \\
&= 1 - \sum_{x=0}^{3} \frac{-7^x e^{-7}}{x!} \\
&= 1 - (\frac{-7^0 e^{-7}}{0!} + \frac{-7^1 e^{-7}}{1!} + \frac{-7^2 e^{-7}}{2!} + \frac{-7^3 e^{-3}}{3!}) \\
&= 0.9182346
\end{aligned}
\tag{3}
$$

The probability of having more than 3 tornados in the United States next year is 0.9478707

Problem 3

Assume the systolic blood pressure of 20 -29 years old American males is normally distributed with population mean 132.0 and population standard deviation 9.8.

a) What is the probability that a randomly selected American male between ages of 20 and 29 to have a systolic blood pressure above 137.0?

Let X be the random variable denoting the number of 20-29 years old American males who have a systolic blood pressure. X ~ N(132,9.8)

```
1 - pnorm(137, mean = 132, sd = 9.8)
```

```
## [1] 0.3049542
```

$P(x > 137) = 1 - P(x \leqslant 137)$= 0.3049542

The probability that a randomly selected American male between ages of 20 and 29 to have a systolic blood pressure above 137.0 is 0.3049542

b) What is the probability that the sample mean for blood pressure of 40 males between ages of 20 and 29 will be less than 129?

Since $n = 40 \geqslant 30$, we can use the central limit theory, which states that the distribution of the sample mean is approximately normal with $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$, it follows that $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$P(\bar{X} < 129) = P(Z < \frac{129-132}{9.8/\sqrt{40}}) = P(Z < -1.94)$= 0.0264284

The probability that the sample mean for blood pressure of 40 males between ages of 20 and 29 less than 129 is 0.0264284

c) What is the 90 quantile of the sampling distribution of the sample mean X for a sample size of 40?

The distribution of the sample mean is approximately normal with $\bar{X} \sim N(132, \frac{9.8}{\sqrt{40}})$ The formula for a normal distribution quantile is $X = \mu + Z\sigma$. Using Z=1.282 from the z-score table for the 90th quantile, $X = 132 + 1.282(\frac{9.8}{\sqrt{40}}) = 133.98$

```
qnorm(0.9,mean = 132,sd = 9.8/sqrt(40))
```

```
## [1] 133.9858
```

Thus, the 90 quantile of the sample mean is 133.9857847

Problem 4

The researchers are interested in the mean pulse of young women suffering from fibromyalgia. They selected a random sample of 30 young females suffering from fibromyalgia. Thesample mean of their pulses was 75 and the sample standard deviation was 8.

a) Compute the 95% confidence interval for the population mean pulse rate of young females suffering from fibromyalgia.

The formula for the 95% confidence interval for the population mean is

$\bar{X} - t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}} \leqslant \mu \leqslant \bar{X} + t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}$

Given n = 30, $\bar{X}$ = 75,s = 8, α=0.05

$75 - t_{29,0.975}\frac{8}{\sqrt{30}} \leqslant \mu \leqslant 75 + t_{29,0.975}\frac{8}{\sqrt{30}}$

```
lower = 75 - qt(0.975, df = 29) * 8/sqrt(30)
upper = 75 + qt(0.975, df = 29) * 8/sqrt(30)
c(lower, upper)
```

```
## [1] 72.01275 77.98725
```

$72.01275 \leqslant \mu \leqslant 77.98725$

The 95% confidence interval for the population mean is (72.0127509, 77.9872491)

b) Interpret the calculated confidence interval

We are 95% confident that the true population mean pulse rate of young females suffering from fibromyalgia lies between the lower limit 72.01275 and the upper limit 77.98725.

c) Suppose the researchers now want to test the null hypothesis that the mean pulse of young women suffering from fibromyalgia is equal to 70, against the alternative that mean pulse is not equal to 70, at the significance level of is 0.05. Would it be possible to answer this question without doing any additional calculations? If yes, state the conclusion of their test, and explain how you arrived at the answer.

Yes, we can answer the question without further calculation. This is because we found in part b) that the 95% confidence interval for the mean pulse of young women suffering from fibromyalgia is (72.0127509, 77.9872491), and 70 does not fall into this interval. That is, we are 95% confident that 70 is not the population mean.