Homework 2

Yuki Joyama

2023-10-10

Problem 1

p, the probability of having at least one dental checkup during a two-year period, is 0.73.

(a) Let X be the the number of people (out of 56 random individuals) that have at least one dental checkup.
X~Bin(56, 0.73)

$P(X = 40) = f(x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{56}{40} (0.73)^{40} (1 - 0.73)^{56-40} = 0.11$

Therefore, the probability that exactly 40 of these individuals will have at least one dental check up is **11.33%**

(b) The probability that at least 40 of these individuals will have at least one dental checkup can be denoted as: $P(X \geq 40) = 1 - P(X < 40) = 1 - P(X \leq 39)$

$P(X \leq 39) = P(X = 0) + P(X = 1) + ... + P(X = 39) = 0.33$
Therefore, $P(X \geq 40) = 1 - P(X \leq 39) = 0.67$
$\rightarrow$ **66.79%**

(c)

$np = 56 \times 0.73 = 40.88 > 10$ and $nq = 56 \times 0.73 = 15.12 > 10$
Thus we can approximate X to the normal distribution X~N(40.88, 3.32) where $P(X = 40)$ becomes $P(X = 40 - \frac{1}{2})$ and $P(X \geq 40)$ becomes $P(X \geq 40 - \frac{1}{2})$

When X~N(40.88, 3.32),
$P(X = 39.5) =$ **0.11**
$P(X \geq 39.5) =$ **0.66**

These are similar to the results of (a) and (b).

(d) The expected value of X~Bin(56, 0.73) is:
$\mu = E(X) = np =$ **40.88**

(e) The variance of X~Bin(56, 0.73) is:
$\sigma^2 = var(X) = np(1-p) = 11.04$
Therefore, the standard deviation is **3.32**

Problem 2

Suppose the number of tornadoes in the U.S. follows a Poisson distribution with parameter $\lambda = 6$ tornadoes per year.
Let X denote the number of tornadoes in the U.S. per year. X~Poi(6)

(a)

The probability of having fewer than 3 tornadoes in the U.S. next year is
$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$
$P(X = x) = f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$ so,

$P(X = 0) = \frac{6^0 e^{-6}}{0!} = 0$
$P(X = 1) = \frac{6^1 e^{-6}}{1!} = 0.01$
$P(X = 2) = \frac{6^2 e^{-6}}{2!} = 0.04$
Therefore, $P(X < 3) = \mathbf{0.06}$

(b)

The probability of having exactly 3 tornadoes in the U.S. next year is
$P(X = 3) = \frac{6^3 e^{-6}}{3!} = \mathbf{0.09}$

(c)

The probability of having more than 3 tornadoes in the U.S. next year is
$P(X > 3) = 1 - P(X \leq 3) = 1 - (P(X < 3) + P(X = 3)) = \mathbf{0.85}$

Problem 3

Assume the systolic blood pressure (SBP) of 20-29 year old American males is normally distributed with population mean 128.0 and population standard deviation 10.2.

Let X denote the systolic blood pressure of 20-29 year old American males.
Now,
X~N(128.0, 10.2), $\mu = 128.0$, $\sigma = 10.2$

(a) The probability that a randomly selected American male between 20 and 29 years old has a SBP above 137.0 is $P(X > 137.0) = 1 - P(X \leq 137.0)$

$P(X < x) = \int_{-\infty}^{x} f(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$
So, $P(X \leq 137.0) = \int_{-\infty}^{137} \frac{1}{10.2\sqrt{2\pi}} e^{\frac{-(137-128)^2}{2\times 10.2^2}} = 0.81$
Therefore, $P(X > 137.0) = 1 - P(X \leq 137.0) = \mathbf{0.19}$

(b) Let $\bar{X}$ denote the sampling distribution (n=50). The underlying population distribution is normal, so $\bar{X}$~$N(\mu, \frac{\sigma^2}{n}) = \bar{X}$~$N(128.0, \frac{10.2}{\sqrt{50}})$

The probability of sample mean of blood pressure of 50 males between 20 and 29 yaers old to be less than 125.0 is $P(\bar{X} < 125.0) = \mathbf{0.02}$

(c) When sample size is 40, the sampling distribution $\bar{X}$~$N(128.0, \frac{10.2}{\sqrt{40}})$
The 90th percentile of this distribution is **130.07**

Problem 4

Let $\bar{X}$ denote the sample distribution (n=40): $\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{(n-1)}$

(a) The population standard deviation is unknown, so 95% confidence interval for the population mean pulse rate of young females with fibromyalgia is

$\bar{X} - t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}$
$t_{39,.975} = 2.02$

Therefore, 95% confidence interval is
$[80-\ 2.02\ (10)/\sqrt{40}, 80+\ 2.02\ (10)/\sqrt{40}]$ = **(76.8, 83.2)**

(b) When we repeat drawing samples from the same population, among 95% confidence intervals calculated from each sample, 95% of them has the population mean μ. And (76.8, 83.2) is one of the ranges that was calculated from this particular experiment.

(c) Let $\mu_0 = 70$
H_0 : the mean pulse = 70, H_1 : the mean pulse $\neq$ 70 ($\alpha = 0.01$)

The test statistic is $t = \frac{\bar{X}-\mu_o}{s/\sqrt{n}} = \frac{80-70}{10/\sqrt{40}} = 6.32$

$t_{n-1,1-\alpha/2} = 2.71$
$|t| > t_{n-1,1-\alpha/2}$ so we reject the null hypothesis at a significance level of .01.

Thus, the mean pulse of young women suffering from fibromyalgia does differ significantly from 70.