生物統計 **HW6** 解答

#1:

Measurements in the lab. Juan makes a measurement in a chemistry laboratory and records the result in his lab report. The standard deviation of students' lab measurements is $\sigma = 10$ milligrams. Juan repeats the measurement 3 times and records the mean $\overline{x}$ of his 3 measurements.

(a)

What is the standard deviation of Juan's mean result? (That is, if Juan kept on making 3 measurements and averaging them, what would be the standard deviation of all his $\overline{x}$'s?)

Sol):

$$Var(\overline{X}) = \frac{\sigma^2}{n} \to \sigma(\overline{\mathrm{X}}) = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{3}} \approx 5.7735$$

(b)

How many times must Juan repeat the measurement to reduce the standard deviation of $\overline{x}$ to 5? Explain to someone who knows no statistics the advantage of reporting the average of several measurements rather than the result of a single measurement.

Sol):

$$\frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{n}} = 5 \to n = 4$$

若僅做一次試驗，結果可能高估或低估，做多次試驗取平均可消弭此狀況，並降低變異數，故做多次試驗後平均會比做單次試驗好。

#2:

National math scores. The scores of 12th-grade students on the National Assessment of Educational Progress year 2000 mathematics test have a distribution that is approximately Normal with mean $\mu = 300$ and standard deviation $\sigma = 35$.

(a)

(a) Choose one 12th-grader at random. What is the probability that his or her score is higher than 300? Higher than 335?

Sol):

$X \sim N(300, 35^2)$

$$\Pr(X > 300) = \Pr\left(\frac{X - 300}{35} > \frac{300 - 300}{35}\right) = \Pr(Z > 0) = 0.5$$

$$\Pr(X > 335) = \Pr\left(\frac{X - 300}{35} > \frac{335 - 300}{35}\right) = \Pr(Z > 1) \approx 0.1587$$

(b)

Now choose an SRS of four 12th-graders. What is the probability that their mean score is higher than 300? Higher than 335?

Sol):

$X_i \sim N(300, 35^2), \mathrm{i} = 1, 2, 3, 4$ <i.i.d.>

$$\rightarrow \overline{X} = \frac{1}{4}\sum_{i=1}^{4} X_i \sim N(300, \frac{35^2}{4})$$

$$\Pr(\overline{X} > 300) = \Pr\left(\frac{\overline{X} - 300}{35/2} > \frac{300 - 300}{35/2}\right) = \Pr(Z > 0) = 0.5$$

$$\Pr(\overline{X} > 335) = \Pr\left(\frac{\overline{X} - 300}{35/2} > \frac{335 - 300}{35/2}\right) = \Pr(Z > 2) \approx 0.0228$$

#3:

Flaws in carpets. The number of flaws per square yard in a type of carpet material varies with mean 1.6 flaws per square yard and standard deviation 1.2 flaws per square yard. The population distribution cannot be Normal, because a count takes only whole-number values. An inspector samples 200 square yards of the material, records the number of flaws found in each square yard, and calculates $\bar{x}$, the mean number of flaws per square yard inspected. Use the central limit theorem to find the approximate probability that the mean number of flaws exceeds 2 per square yard.

Sol):

$E(X) = 1.6, Var(X) = 1.2$

By CLT, $\dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

$$\rightarrow \Pr(\overline{\mathrm{X}} > 2) = \Pr\left(\frac{\overline{\mathrm{X}} - \mu}{\sigma / \sqrt{n}} > \frac{2 - \mu}{\sigma / \sqrt{n}}\right) = \Pr\left(\mathrm{Z} > \frac{2 - 1.6}{1.2 / \sqrt{200}}\right) \approx 0$$

#4:

Mark McGwire's home runs. In 1998, Mark McGwire of the St. Louis Cardinals hit 70 home runs, a new major league record. Was this feat as surprising as most of us thought? In the three seasons before 1998, McGwire hit a home run in 11.6% of his times at bat. He went to bat 509 times in 1998. McGwire's home run count in 509 times at bat has approximately the binomial distribution with $n = 509$ and $p = 0.116$. What is the mean number of home runs he will hit in 509 times at bat? What is the probability of 70 or more home runs? (Use the Normal approximation.)

Sol):

$X \sim binomial(n, p)$

(a) $E(X) = np = 509 \times 0.116 = 59.044$

(b) Let $X = \sum_{i=1}^{n} Y_i$,where $Y_i \sim Bernoulli(p)$ <i.i.d.>

By CLT, $\dfrac{\sum_{i=1}^{n} Y_i - np}{\sqrt{np(1-p)}} = \dfrac{\sum_{i=1}^{n} Y_i - np}{\sqrt{np(1-p)}} \sim N(0,1)$

$$\Pr(X \geq 70) = \Pr\left(\frac{X - np}{\sqrt{np(1-p)}} \geq \frac{70 - np}{\sqrt{np(1-p)}}\right) = \Pr\left(\mathrm{Z} \geq \frac{70 - 509 \times 0.116}{\sqrt{509 \times 0.116 \times (1 - 0.116)}}\right) \approx 0.0647$$

#5:

Checking for survey errors. One way of checking the effect of undercoverage, nonresponse, and other sources of error in a sample survey is to compare the sample with known facts about the population. About 12% of American adults are black. The number X of blacks in a random sample of 1500 adults should therefore vary with the binomial ($n = 1500$, $p = 0.12$) distribution.

(a)

What are the mean and standard deviation of X?

Sol):

$X \sim binomial(n, p)$

$E(X) = np = 1500 \times 0.12 = 180$

$\sigma(X) = \sqrt{np(1-\mathrm{p})} = \sqrt{1500 \times 0.12 \times (1-0.12)} \approx 12.5857$

(b)

Use the Normal approximation to find the probability that the sample will contain 170 or fewer blacks. Be sure to check that you can safely use the approximation.

Sol):

Let $X = \sum_{i=1}^{n} Y_i$,where $Y_i \sim Bernoulli(p)$ <i.i.d.>

By CLT, $\dfrac{\sum_{i=1}^{n} Y_i - np}{\sqrt{np(1-p)}} = \dfrac{\sum_{i=1}^{n} Y_i - np}{\sqrt{np(1-p)}} \sim N(0,1)$

$$\Pr(X \le 170) = \Pr\left(\frac{X-np}{\sqrt{np(1-p)}} \le \frac{170-np}{\sqrt{np(1-p)}}\right) = \Pr\left(\mathrm{Z} \le \frac{170-1500\times 0.12}{\sqrt{1500\times 0.12\times(1-0.12)}}\right) \approx 0.2134$$

#6:

Chebyshev's inequality. This inequality points out another useful property of the standard deviation. In particular, it states that "The probability that any random variable X falls within k standard deviations of its mean is at least $1 - 1/k^2$." For example, if we know that X has mean 3 and standard deviation 1, then we can conclude that the probability that X lies between 1 and 5 ($k = 2$ standard deviations from the mean) is at least $1 - 1/2^2 = .75$.

(a)

Let X denote the amount of rainfall received per week in a region. Assume that $\mu = 1.00$ inch and $\sigma = .25$ inch. Would it be unusual for this region to receive more than 2 inches of rain in a given week? Explain on the basis of Chebyshev's inequality.

Sol):

$X \equiv$ amount of rainfall

ChebyShev's inequality : $\Pr(|X-\mu| \le k\sigma) \ge 1 - \frac{1}{k^2} \Leftrightarrow \Pr(|X-\mu| > k\sigma) \le \frac{1}{k^2}$

$\rightarrow \Pr(|X-1| > 4\times 0.25) \le \frac{1}{4^2}$

$\rightarrow \Pr(X > 2) + \Pr(X < 0) \le \frac{1}{16}$

$\rightarrow \Pr(X > 2) \le \frac{1}{16} - \Pr(X < 0) \le \frac{1}{16} = 0.0625$

So, it's unusual for this region to receive more than 2 inches of rain in a given week.

(b)

Let X denote the number of cases of rabies reported in a given state per week. Assume that $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{1}{25}$. Would it be unusual to observe two cases in a given week? Explain on the basis of Chebyshev's inequality.

Sol):

$X \equiv$ the number of cases of rabies

$$\Pr(X > 2) + \Pr(X < -1) = \Pr(|X - \frac{1}{2}| > \frac{3}{2}) = \Pr(|X - \frac{1}{2}| > \frac{15}{2}\times\frac{1}{5}) \le \frac{1}{(15/2)^2} = \frac{4}{225}$$

$$\Pr(X > 2) \le \frac{4}{225} - \Pr(X < -1) \le \frac{4}{225} \approx 0.01\overline{7}$$

So, it's unusual to observe more than two cases in a given week.