大數據統計分析與預測 第七章作業

Created by 黃彥鈞 (Weber Huang) m946108006

---

13. According to the Massachusetts Department of Health, 224 women who gave birth in the state of Massachusetts in 1988 tested positive for the HIV antibody. Assume that, in time, 25% of the babies born to such mothers will also become HIV-positive.

(a) If samples of size 224 were repeatedly selected from the population of children born to mother with the HIV antibody, what should be the mean number of the infected children per sample ?

The mean number of the infected children is kind of **binomial distribution**, so the answer is : $np = 224 \times (0.25) = 56$

(b) What would be the standard deviation?

The standard deviation is : $\sigma = \sqrt{npq} = \sqrt{np(1-p)} = 6.4807$

(c) Use Chebychev's inequality to describe this distribution.

Chebychev's inequality allows us to say that for any number k that is greater than or equal to 1, at least $[1-(\frac{1}{k})^2]$ of the measurements in the set of the data lie with in k standard deviations of their mean.

For example, we can say that there is 75% of measurements lie within $56 \pm 2 \times (6.5)$, that is $(43, 69)$.

15. In a particular county, the average number of suicides reported each month is 2.75. Assume that the number of suicides follows a Poisson distribution.

(a) What is the probability that "no suicides" will be reported during a given month?

The algorithm of Poisson distribution is $P(X) = \frac{e^(-\lambda)\lambda^x}{x!}$

$x = 0$ (no suicide)

$\lambda = 2.75$

$P(nosuicides) = \frac{e^(-2.47)2.75^0}{0!} = e^(- 2.75) = 0.063927$

(b) What is the probability that at most four suicides will be reported?

$P(AtMostFourSuicides) = \frac{e^(-2.47)2.75^0}{0!} + \frac{e^(-2.47)2.75^1}{1!} + \frac{e^(-2.47)2.75^2}{2!} + \frac{e^(-2.47)2.75^3}{3!} + \frac{e^(-2.47)2.75^4}{4!} = 0.8553$

(c) What is the probability that six or more suicides will be reported?

The probability that six or more suicides will be (1 - the probability less than 6) :

$$P(6orMore) = 1 - P(<6) = 1 - P(AtMostFourSuicides) - \frac{e^(-2.47)2.75^5}{5!} = 0.06083531$$

19. The distribution of weights for population of males in the United States is approximately normal with mean $\mu = 172.2$ pounds and standard deviation $\sigma = 29.8$ pounds.

(a) What is the probability that a randomly selected man weighs less than 130 pounds?

In R : Cumulative Distribution Function

$P(X < 130) =$`pnorm(130, 172.2, 29.8) = 0.07837203`

(b) What is the probability that he weighs more than 210 pounds?

$P(X > 210) = 1 - P(x <= 210)$ = `pnorm(210, 172.2, 29.8,lower.tail=FALSE) = 0.1023175`

(c) What is the probability that among five males selected at random from the population, at least one will have a weight outside the range 130 to 210 pounds?

At least one outside the range = 1 - all not in the range :

$$P(AtLeastOneInTheRange) = 1 - (1 - 0.07837203 - 0.1023175)^5 = 0.6308163$$