IE 7300 Homework 2 – Xinan Wang

Question 1: While looking at a person's chickenpox in 2021, it turns out that 1% of people in the world are affected. Consider a group of 20 people. (20 points)

(a) Write down the random variable in this problem. (2 points)

Ans: The random variable is the number of people in this 20 people group that having chickenpox.

(b) How is this a binomial experiment? (3 points)

Ans: This is a binomial distribution because: 1) We have 2 results for each experiment. The success case is that person is having the chickenpox, and the failure case is that the person doesn't have the chickenpox. No exceptions. 2) Each observation is independent with each other, which means the probability for each person having chickenpox is independent of each other. 3) The probability for each person having chickenpox is fixed (p = 0.01) 4) We have fixed number observations. (N = 20)

Therefore, we can develop a binomial distribution that:

$$X \sim \text{Binomial}\ (\Theta = 0.01, N = 20)$$

$$P\ (X \mid \Theta, N) = \binom{N}{X} \Theta^X (1-\Theta)^{N-X} = \binom{20}{X} (0.01)^X (0.99)^{20-X}$$

(c) Find none of them have chickenpox. (3 points)

Ans: In this case, X = 0, N = 20, Θ = 0.01. So, we could apply the binomial distribution formula that:

$$P\ (X = 0 \mid \Theta, N) = \binom{N}{X} \Theta^X (1-\Theta)^{N-X} = \binom{20}{0} (0.01)^0 (0.99)^{20} = 0.8179$$

Therefore, in theory, we have 81.79% probability that none of these 20 people groups have chickenpox. This is a large possibility.

(d) Find 10 of them have chickenpox. (3 points)

Ans: In this case, X = 10, N = 20, Θ = 0.01. So, we could apply the binomial distribution formula that:

$$P\ (X = 10 \mid \Theta, N) = \binom{N}{X} \Theta^X (1-\Theta)^{N-X} = \binom{20}{10} (0.01)^{10} (0.99)^{10} = 1.6709 * 10^{-15}$$

Therefore, in theory, we have $1.6709 * 10^{-15}$ probability that 10 of these 20 people groups have chickenpox. This is an extremely small probability. In practice, if we randomly choose 20 people from the world and we do not influence by other features such as region, it's very unlikely to happen that 10 of these 20 people have chickenpox.

(e) At most 4 of them have chickenpox. (3 points)

Ans: In this case, we will add up the cases that X = 0, X = 1, X = 2, X = 3 and X = 4. And our N = 20, Θ = 0.01.

P (X ≤ 4 | Θ, N) = P (X = 0 | Θ, N) + P (X = 1 | Θ, N) + P (X = 2 | Θ, N) + P (X = 3 | Θ, N) + P(X = 4 | Θ, N) = $\binom{20}{0}(0.01)^0(0.99)^{20}$ + $\binom{20}{1}(0.01)^1(0.99)^{19}$ + $\binom{20}{2}(0.01)^2(0.99)^{18}$ + $\binom{20}{3}(0.01)^3(0.99)^{17}$ + $\binom{20}{4}(0.01)^4(0.99)^{16}$ = 0.99999863

Therefore, in theory, we have 99.999863% probability that at most 4 of these 20 people groups have chickenpox. This is a very large probability. In practice, if we randomly choose 20 people from the world and we do not influence by other features such as region, it's very likely to happen that at most 4 of them have chickenpox.

(f) At least 5 of them have chickenpox. (3 points)

Ans: In this case, because that the total probability is 1, we will use 1 to minus the possibility that at most 4 of them have chickenpox. By doing this, we will get the possibility that at least 5 of them have chickenpox.

P (X ≥ 5 | Θ, N) = 1 - P (X ≤ 4 | Θ, N) = 1 - 0.99999863 = 0.00000137

Therefore, in theory, we have 0.000137% probability that at least 5 of them have chickenpox. This is a very small probability. In practice, if we randomly choose 20 people from the world and we do not influence by other features such as region, it's very unlikely to happen that at least 5 of them have chickenpox.

(g) In Africa, four people out of twenty have chickenpox. Is this expected? Did you see any statistical evidence? (3 points)

Ans: Expected number of people in 5 of them have chickenpox is:

$$E(X) = P(X) * X = 0.01 * 20 = 0.2$$

We can find that for 20 people group, there is expected 0.2 person of 20 people group that will have chickenpox, which is greatly less than 4 people out of 20 in Africa.

Also, the probability for Africa people having the chickenpox is $\frac{4}{20}$ = 0.2, which is greatly smaller than the world mean value 0.01.

Therefore, it is not expected.

The statistical evidence I found is that there are multiple features that could influence the probability result. In this question, Africa has greatly higher possibility than other region for people to get chickenpox. So, we should consider "region" as a feature that could influence the chickenpox probability. When we do chickenpox analysis for Africa region, we should consider 20% as a more accurate probability instead of 1%.

Question 2: In the Snohomish County, Family Income ~ N ($25000, $1000^2). If the low-income level is $15,000, what percentage of the population lives under the low-income level? (Hint: Normal distribution) (5 points)

Ans: In this question, $\mu = 25000$, $\sigma = 1000$, and therefore, N ~ ($\mu = 25000$, $\sigma = 1000$)

P (X < 15000) = P ($\frac{X - \mu}{\sigma} < \frac{15000-25000}{1000}$) = P (Z < -10) = 0

Therefore, in theory, we have 0% of the population lives under the low-income level. That means there are no population in Snohomish County that live in the low-income level. And they are all above the low-income level.

Question 3: The Darigold uses a filling machine to fill milk bottles with 2% milk. The milk bottles should contain 350ml of milk. But the container will vary with a mean of 348 ml and a standard deviation of 3 ml. (10 points)

(a) What is the probability that a bottle contains less than 3ml of the original size? (5 points)

350ml – 3ml = 347ml

P (X < 347) = P ($\frac{X - \mu}{\sigma} < \frac{347-348}{1000}$) = P (Z < -$\frac{1}{3}$) = 0.36944

Therefore, in theory, we have 36.944% probability that a bottle contains less than 3ml of the original size. In practice, we may have around 1 out of 3 bottles that the bottle contains less than 3ml of the original size, which is a little common for us to see that a bottle contains less than 3ml of the original size.

(b) What is the probability that the mean of the six-pack bottles is less than 345 ml? (5 points)

Ans: Sample size = 6, therefore, $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} = \frac{3}{\sqrt{6}}$ = 1.2247

P ($\bar{X}$ < 345) = P ($\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{345-348}{1.225}$) = 0.00716

Therefore, in theory, we have 0.716% probability that the mean of the six-pack bottles is less than 345ml. This is a very small possibility. In practice, it will be very unlikely to happen that the mean of the six-pack bottles is less than 345ml if we choose the bottles randomly and we do not consider other features.