140.615.HW.02.Jin.Vincent

Vincent Jin

2023-02-17

Homework 02

Vincent Jin

1. Suppose that 1/100,000 bacterial cells contain a mutation providing resistance to substance A. Suppose I create a bunch of plates with 200,000 cells per plate. Let X be the number of cells on a plate that are resistant to substance A.

(a) What is the mean number of cells in a plate that are resistant to A?

Answer

```
res_p <- 1 / 100000
total_c <- 200000
mean_x <- res_p * total_c
print(mean_x)
```

[1] 2

The mean number of cells in a plate that are resistant to A is 2.

(b) What is SD(X)?

Answer

As the number of cells in a plate that are resistant to A follows a poisson distribution, the variance will be equal to the mean, which means the variance is equal to 2.

```
sd_x <- sqrt(mean_x)
print(round(sd_x, 2))
```

[1] 1.41

Therefore, SD(X) equals to the square root of mean of x, which is 1.41.

(c) Calculate $Pr(X = 0)$.

Answer

```
Pr0 <- dpois(0, mean_x)
print(round(Pr0, 2))
```

[1] 0.14

The probability of X = 0 is 0.14.

(d) Calculate Pr(X = 5).

Answer

```
Pr5 <- dpois(5, mean_x)
print(round(Pr5, 2))
```

[1] 0.04

The probability of X = 5 is 0.04.

(e) Calculate Pr(X > 2).

Answer

```
Pr2 <- 1 - ppois(2, mean_x)
print(round(Pr2, 2))
```

[1] 0.32

The probability of X > 2 is 0.32.

2. Suppose Y is a random variable with E(Y) = 30 and SD(Y) = 5.

(a) Let Z = (Y − 30)/5. Calculate E(Z) and SD(Z).

Answer

$$Z = (Y - 30)/5 = Y/5 - 6$$

```
E_Y <- 30
SD_Y <- 5
E_Z <- (E_Y - 30) / 5
SD_Z <- 1 / 5 * SD_Y
print(E_Z)
```

[1] 0

```
print(SD_Z)
```

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

Therefore, we can see that: $E(Z) = 0$ $SD(Z) = 1$

  (b) Let $X = -Y$. Calculate $E(X)$ and $SD(X)$.

Answer

```
E_X <- -1 * E_Y
SD_X <- abs(-1) * SD_Y
print(E_X)
```

```
## [1] -30
```

```
print(SD_X)
```

```
## [1] 5
```

Therefore, we can see that: $E(X) = -30$ $SD(X) = 5$

  (c) Let $R = 5 + Y/3$. Calculate $E(R)$ and $SD(R)$.

Answer

```
E_R <- 5 + 1/3 * E_Y
SD_R <- 1/3 * SD_Y
print(E_R)
```

```
## [1] 15
```

```
print(SD_R)
```

```
## [1] 1.666667
```

Therefore, we can see that: $E(R) = 15$ $SD(R) = 1.67$

3. Below is the joint probability distribution of viral infection and strain in a population of rats.

Strain A B C Infected 0.25 0.19 0.03 Not infected 0.19 0.17 0.17

  (a) If we randomly sample a rat, what is the probability it is infected with the virus?

Answer

```
rows <- c("Infected", "Not Infected")
cols <- c("A", "B", "C")
joint_p <- matrix(c(0.25, 0.19, 0.03, 0.19, 0.17, 0.17), nrow = 2, byrow = TRUE, dimnames = list(rows,
P_inf <- as.numeric(joint_p[1, 1]) + as.numeric(joint_p[1, 2]) + as.numeric(joint_p[1, 3])
print(P_inf)
```

[1] 0.47

Therefore, the probability of randomly sample an infected rat is 0.47.

(b) If we randomly sample a rat, what is the probability it is of strain A?

Answer

```
P_A <- joint_p[1, 1] + joint_p[2, 1]
print(P_A)
```

[1] 0.44

Therefore, the probability of randomly sample an strain A rat is 0.44.

(c) What is the strain distribution among infected rats?

Answer

```
P_Ainf <- round(joint_p[1, 1] / P_inf, 2)
P_Binf <- round(joint_p[1, 2] / P_inf, 2)
P_Cinf <- round(joint_p[1, 3] / P_inf, 2)
inf <- matrix(c(P_Ainf, P_Binf, P_Cinf), nrow = 1, byrow = TRUE, dimnames = list("Infected", cols))
print(inf)
```

```
##               A    B    C
## Infected 0.53 0.4 0.06
```

Therefore, among infected rats, the probability of being strain A, B, and C is 0.53, 0.40, 0.06, respectively.

(d) Are the random variables strain and infection independent?

Answer

```
Ind_TF <- (P_inf * P_A) == joint_p[1, 1]
print(Ind_TF)
```

[1] FALSE

Since the probability of being infected and is strain A does not equal to the production of being infected
and probability of being strain A, the random variables strain and infection are not independent.

4. Assume blood-glucose levels in a population of adult women are normally distributed with mean 90 mg/dL and standard deviation 13 mg/dL. What percentage of individuals would be have 'abnormal'

glucose levels, and what is the normal range of glucose levels (in units of mg/dL), if 'abnormal' was defined as glucose levels outside of 2 standard deviations of the mean?

Answer

```
bg_mean <- 90
bg_sd <- 13
normal_upper <- bg_mean + 2 * bg_sd
normal_lower <- bg_mean - 2 * bg_sd
print(c(normal_lower, normal_upper))
```

```
## [1]  64 116
```

Since there will be 95% of population within 2 standard deviation from the mean, if abnormal was defined as gluciose levels outside of 2 standard deviations of the mean, so that there will be 1 - 0.95 = 0.05 = 5% of adult women would be having abnormal glucose levels. The normal range of glucose level is 64 mg/dL - 116 mg/dL.

5. Suppose we have 100 independent draws from a normal distribution with population mean 20 and population standard deviation 5.

  (a) What is the chance that the sample mean is within 0.1 units of the population mean?

Answer

Since we sampled from a population, we can know that in this sample, the sample mean will be equal to the population mean, and the sample mean standard deviation will be the standard error of $s/\sqrt{n}$.

```
se_mean <- 5 / sqrt(100)
P_0.1 <- pnorm(20 + 0.1, 20, se_mean) - pnorm(20 - 0.1, 20, se_mean)
print(round(P_0.1, 2))
```

```
## [1] 0.16
```

The chance that the sample mean is within 0,1 units of the population mean 16%.

  (b) What is the chance that the sample mean exceeds the population mean by at least 0.25 units?

Answer

```
P_ex0.25 <- 1 - pnorm(20 + 0.25, 20, se_mean)
print(round(P_ex0.25, 2))
```

```
## [1] 0.31
```

The chance that the sample mean exceeds the population mean by at least 0.25 units is 31%.