Patel’s 02 normal poisson handout

Harsh Patel

2024-09-26

Normal Distribution

```r
#probability X is less than (or equal to) 105
pnorm(105, 100, 5)
```

```
## [1] 0.8413447
```

```r
#probability X is greater than 105
pnorm(105, 100, 5, lower.tail = FALSE)
```

```
## [1] 0.1586553
```

```r
#probability Z is less than (or equal to) 1
pnorm(1)
```

```
## [1] 0.8413447
```

```r
#probability Z is greater than 1
pnorm(1, lower.tail = FALSE)
```

```
## [1] 0.1586553
```

```r
#identify X value
qnorm(0.841, 100, 5)
```

```
## [1] 104.9929
```

```r
qnorm(0.159, 100, 5, lower.tail = FALSE)
```

```
## [1] 104.9929
```

```r
#identify Z value
qnorm(0.841)
```

```
## [1] 0.9985763
```

```r
qnorm(0.159, lower.tail = FALSE)
```

```
## [1] 0.9985763
```

1.

a) A woman aged around 38 (37.8) years is in the top 2.5% of the age distribution.

```
qnorm(0.975, mean = 26.4, sd = 5.8)

## [1] 37.76779
```

b) P(X≥21)=0.824; this means that 82.4% of women giving birth are at least 21 years old.

```
pnorm(21, mean = 26.4, sd = 5.8, lower.tail = FALSE)

## [1] 0.8240821
```

2. To find P(90≤X≤100), calculate P(X≤100)−P(X≤90). The chance that a randomly chosen male in this population is mildly hypertensive is 0.154.

```
pnorm(100, mean = 80, sd = 12)- pnorm(90, mean = 80, sd = 12)

## [1] 0.154538
```

3.

a) P(X≥115.0)=0.0228; 2.28% of children between the ages of 1 and 14 are hypertensive.

```
pnorm(115, mean = 105, sd = 5, lower.tail = FALSE)

## [1] 0.02275013
```

b) P(Y≥140.0)=0.0668; 6.68% of individuals aged 15 to 44 are hypertensive.

```
pnorm(140, mean = 125, sd = 10, lower.tail = FALSE)

## [1] 0.0668072
```

c) Let C represent the number of hypertensive children in a family (modeled as a binomial random variable), and A represent the number of hypertensive adults in a family. C follows Bin(2,0.0228), and A follows Bin(2,0.0668). A family is considered hypertensive if at least one adult and one child are hypertensive. Let H be the event that a family is hypertensive. Assuming C and A are independent:

#P(H)=P(C≥1)×P(A≥1)=0.0058. #So, the probability that a family is hypertensive is 0.0058.

```
pbinom(0, 2, 0.0228, lower.tail = FALSE) * pbinom(0, 2, 0.0668, lower.tail =
FALSE)

## [1] 0.005821551
```

d) Let K represent the number of hypertensive families in the community. The probability that there are between 1 and 5 hypertensive families is P(1≤K≤5)=P(K≤5)−P(K=0)=0.475.

```
pbinom(5, 1000, 0.0058)- dbinom(0, 1000, 0.0058)

## [1] 0.4749525
```

Poisson distribution

```
#probability X equals 5
dpois(5, 3)

## [1] 0.1008188

#probability X is less than or equal to 5
ppois(5, 3)

## [1] 0.9160821

#probability X is greater than 5
ppois(5, 3, lower.tail = FALSE)

## [1] 0.08391794
```

4.

a) The expected number of children born with Down syndrome in Boston in 2017, assuming the number of live births is unchanged, is (1/800)×8,011=10.01.

```
lambda = 1/800; n = 8011
n*lambda

## [1] 10.01375
```

b) Since the mean of a Poisson distribution equals the rate, the expected number of trisomy 21 births is also 10.01. The chance of 12 or more Down syndrome births in 2017 is P(X≥12), where X~Pois(10.01). The probability is 0.304.

```
ppois(11, lambda = 10.01, lower.tail = FALSE)

## [1] 0.3043618
```

c) No, this information is insufficient. Simply adjusting for fewer births from women over 35 isn't enough since women in that age range have a higher likelihood of having a Down syndrome birth. Assuming the rate for trisomy 21 in this group is 1 in 800 live births isn't accurate.

d) The statement could be valid even if Black and Latino women have a higher birth rate than the overall average. This would imply that the birth rates for other groups, like Caucasian or Asian women, could be lower than the overall average of 44.5 births per 1,000 women aged 15-44.

5.

a) If there are equal numbers of male and female births annually, there are around 2,000,000 male births in the U.S. each year. For this population, the hemophilia rate is $(15/1000) \times 2{,}000{,}000 = 400$. Since the mean of a Poisson distribution equals the rate, the expected number of male births with hemophilia is 400. The chance of 380 or fewer newborn males having hemophilia is $P(X \leq 380)$, where $X \sim Pois(400)$, and this probability is 0.165.

```
rate = (1/5000)*(4000000/2)
ppois(380, lambda = rate)

## [1] 0.164859
```

b) The probability that 450 or more newborn males have hemophilia in a year is 0.007.

```
ppois(449, lambda = rate, lower.tail = FALSE)

## [1] 0.007454327
```

c) In a hypothetical country with 750,000 male births annually (12×1,500,000), and if the hemophilia rate is 1 in 5,000 male births, the expected number of hemophilia cases per year is (1/5000)×750,000=150. Over five years, the rate becomes 150×5=750. With a Poisson distribution, both the mean and the variance are 750, making the standard deviation 750=27.39.

```
new.yearly.rate = (1/5000)*(1500000/2); new.yearly.rate

## [1] 150

years = 5
five.year.rate = new.yearly.rate*years; five.year.rate

## [1] 750

sqrt(five.year.rate)

## [1] 27.38613
```

6.

a) The probability of having 146 or more overdose deaths in a given year is 2.62×10^-40.

```
non.hispanic.whites = 769000 * 0.73
lambda.essex = (6.8/100000) * non.hispanic.whites
ppois(145, lambda.essex, lower.tail = FALSE)

## [1] 2.621073e-40
```

b) The observed rate is about 26 deaths per 100,000 non-Hispanic whites annually, which is much higher than the national average reported by the CDC.

```
non.hispanic.whites = 769000 * 0.73
obs.rate = (146/non.hispanic.whites) * 100000
obs.rate

## [1] 26.0078
```

c) Using the rate from part b), find P(X≥165)=P(X>164). Since Essex County's population didn't change from 2014 to 2015, the rate remains 146, which is already the death rate per year among non-Hispanic whites in Essex County. The likelihood of 165 or more opioid-related deaths is approximately 6.5% (0.065).

```
#show that the observed rate lambda is 146
obs.lambda.essex = (146/non.hispanic.whites) * non.hispanic.whites
obs.lambda.essex

## [1] 146

ppois(164, obs.lambda.essex, lower.tail = FALSE)

## [1] 0.0650307
```

d) Even after adjusting for the overdose deaths in 2014, the probability of 165 or more opioid-related deaths in Essex County in 2015 is still low. The opioid death rate in 2014 was already significantly above the national average, and the situation worsened in 2015, indicating a rapid escalation in opioid abuse within Essex County compared to the national trend.