Homework 2

P8130 Biostatistics Method I - Fall 2025

Yongyan Liu (yl6107)

December 16, 2025

Due: September 30, 2025 at 11:59pm

Problem 1 (10 points)

Suppose the probability that a randomly selected adult receives at least one flu shot during a single flu season is **68%**. A sample of **30 adults** is observed.

- For part (a), work **by hand** and show all intermediate steps.
- For parts (b)–(e), you may use tables or software.

a) What is the probability that exactly 18 of these 30 adults receive at least one flu shot?

$$P(X = 18) = \binom{30}{18} * 68\%^{18} * (1 - 68\%)^{30-18}$$

$$= \frac{30!}{12!18!} * 68\%^{18} * 32\%^{12}$$

And the result is

```
p = factorial(30) / factorial(18) / factorial(12) * (0.68 ** 18) * (0.32 ** 12)
p # show the probablity
```

```
## [1] 0.09637009
```

```
p - dbinom(18, 30, 0.68) < 1e-8 # check whether the result is correct
```

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

b) What is the probability that at least 18 of these adults receive at least one flu shot (exact calculation)?

```
1 - pbinom(17, 30, 0.68)   #  1 - (at most 17 people receive at least one flu shot)
```

[1] 0.8708277

c) For part (b), is it appropriate to use an approximation method? If yes, compute the probabilities using an approximation (with continuity correction) and compare to the exact values.

```
mu = 30 * 0.68                     # mean
sigma = sqrt(30 * 0.68 * 0.32)     # standard deviation
1 - pnorm(17.5, mu, sigma)         # show approximation results using normal distribution
```

[1] 0.8718189

```
# show the differences
pbinom(17, 30, 0.68) - pnorm(17.5, mu, sigma)
```

[1] 0.0009912017

d) What is the expected number of adults who receive at least one flu shot?

```
mu        # as calcualted in (c)
```

[1] 20.4

e) What is the standard deviation of the number who receive at least one flu shot?

```
sigma    # as calcualted in (c)
```

[1] 2.554995

Problem 2 (10 points)

During peak flu season, a city health department records an average of **15 new laboratory-confirmed influenza cases per day**. Assume the daily number of new cases follows a **Poisson(λ = 15)** distribution.

- For part (a), work **by hand** and show all intermediate steps.
- For parts (b)–(d), you may use tables or software.

a) What is the probability that the city records exactly 15 new cases tomorrow?

$$P(X = 15) = \frac{15^{15} * e^{-15}}{15!}$$

The result is

```
p = 15 ** 15 * exp(-15) / factorial(15) # calculate 'by hand'
p                                       # show result
```

```
## [1] 0.1024359
```

```
p - dpois(15, 15) < 1e-8                         # Verify the result is correct
```

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

b) What is the probability that the city records fewer than 3 new influenza cases tomorrow?

```
ppois(2, 15)
```

```
## [1] 3.930845e-05
```

c) What is the probability that the city records more than 20 new cases tomorrow?

```
1 - ppois(20, 15)
```

```
## [1] 0.08297091
```

d) The Poisson distribution can sometimes be approximated by a normal distribution. Calculate approximations based on a $\mathbf{N}(\lambda, \lambda)$ distribution for (a)–(c) (with continuity correction). Does the approximation seem good?

```
mu = 15             # mean
sigma = sqrt(15)    # standard deviation

# (a) approx
pnorm(15.5, mu, sigma) - pnorm(14.5, mu, sigma)
```

```
## [1] 0.102721
```

```
# (b) approx
pnorm(2.5, mu, sigma)
```

```
## [1] 0.0006244155
```

```
# (c) approx
1 - pnorm(20.5, mu, sigma)
```

```
## [1] 0.07779017
```

The approximation looks better when X is close to λ, but not so good when X is a few standard deviations away from λ.

Problem 3 (10 points)

Assume the **systolic blood pressure (SBP)** of 20–29 year-old American males is normally distributed with: - Population mean: $\mu = 128$ - Population standard deviation: $\sigma = 10.2$

a) What is the probability that a randomly selected American male between 20 and 29 years old has a systolic blood pressure above 137?

```
1 - pnorm(137, 128, 10.2)
```

```
## [1] 0.188793
```

b) What is the probability that the sample mean blood pressure of 50 males between 20 and 29 years old is less than 125?

```
sigma = 10.2 / sqrt(50)     # calculate the standard deviation for the sample mean
pnorm(125, 128, sigma)      # show the result
```

```
## [1] 0.01877534
```

c) What is the 90th percentile of the sampling distribution of the sample mean with a sample size of 40?

```
sigma = 10.2 / sqrt(40)     # calculate the standard deviation for the sample mean
qnorm(0.9, 128, sigma)      # show the result
```

```
## [1] 130.0668
```

Problem 4 (10 points)

Researchers studied the **mean pulse rate** of young women suffering from **fibromyalgia**. A random sample of **40 women** gave: - Sample mean $= 80$ - Sample standard deviation $= 10$

a) Compute the 95% confidence interval for the population mean pulse rate of young females with fibromyalgia and provide the correct interpretation.

```
# calculate the standard deviation for sample mean
sigma = 10 / sqrt(40)

# 95% CI is roughly 2 times standard deviation around population mean
c(80 - 2 * sigma, 80 + 2 * sigma)
```

```
## [1] 76.83772 83.16228
```

b) Test the null hypothesis that the mean pulse of young women with fibromyalgia equals 70 (two-sided), at $\alpha = 0.01$. Conduct this hypothesis test and interpret the result.

```
se = 10 / sqrt(40)
p_value = 2 * (1 - pt((80 - 70) / se, df = 39))
p_value      # show the p value
```

```
## [1] 1.835006e-07
```

From the very small p-value, we can see the null hypothesis H_0 is rejected.