# Discussion 4

Hanying Jiang

## If you haven't download the Discussion 4 handout, please download it now.

## Information

- Collaboration on assignments is encouraged. However, each student must do their own work, perform their own analyses, and write up their own submission, all in their own words.
- Make sure you only submit one file in Canvas.

## Warmup

A certain vending machine offers 20-ounce bottles of soda for \$1.50. Let the random variable $X$ reflect the number of bottles bought from the machine on any day. RV $X$ has a mean of 50 and a standard deviation of 15.

Now let the random variable $Y$ equal the total revenue from this machine on a given day, so . $Y = 1.50(X)$. Assume that the machine works properly and that no sodas are stolen from the machine.

Select the option below that reflects the mean and standard deviation of Y.

- $\mu y = 1.50; \sigma y = 22.50$
- $\mu y = 1.50; \sigma y = 33.75$
- $\mu y = 75; \sigma y = 18.37$
- $\mu y = 75; \sigma y = 22.50$
- $\mu y = 75; \sigma y = 33.75$

We have $Y = 1.50(X)$.

To calculate its expectation, we use the Rule b learned in Tuesday's lecture: $\mathrm{E}(c \times X) =$ ___.

To calculate its standard deviation, we use the Rule g(ii) in Tuesday's lecture: $\mathrm{SD}(c \times X) =$ ___.

## Exercise 1: Normal Random Variables

### R: pnorm()

`pnorm()`: "p" stands for probability and "norm" stands for normal distribution.

If $X \sim N(a, b^2)$, to calculate the probability of $X$ being smaller than $v$, we can use `pnorm(v, a, b)`. If $X \sim N(0.1)$, we can omit the mean $a$ and the sd $b$. For example, if $X \sim N(7.5, 0.8^2)$, $P(X < 7)$ is

```r
pnorm(7, 7.5, 0.8)
```

```
## [1] 0.2659855
```

To compute $P(X > 7)$

```r
1 - pnorm(7, 7.5, 0.8)
```

```
## [1] 0.7340145
```

For $X \sim N(0, 1)$, $P(X < 1)$ is

```r
pnorm(1, 0, 1)
```

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

You can omit the mean and sd:

```r
pnorm(1)
```

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

#### Practice

$X \sim N(3.1, 0.5^2)$, calculate $P(2.9 < X < 3.7)$.

Hint: the answer is 0.5403521.

## Question

Weights of watermelons of a certain variety (A) are well approximated by a normal distribution with mean 7.5 kg and standard deviation of 0.8 kg $X \ \sim N(7.5, 0.8^2)$.

### 1(a)

What proportion of watermelons of that variety (A) have weights between 7.2 and 8 kg?

$$P(7.2 < X < 8) = P(X < 8) - P(X < 7.2),$$

where $X \ \sim N(7.5, 0.8^2)$.

```r
# the way to use pnorm: pnorm(value, mean, sd)
pnorm(8, 7.5, .8)-pnorm(7.2, 7.5, .8)
```

```
## [1] 0.3801842
```

### 1(b)

A watermelon of that variety (A) has a weight that is 0.5 standard deviations above the mean. What proportion of watermelons of that variety (A) are heavier than this one?

The weight of this watermelon is $mean + 0.5 * SD = 7.5 + 0.5 * 0.8 = 7.9$. Then the proportion of watermelons of that variety (A) are heavier than this one is $P(X > 7.9) = 1 - P(X < 7.9)$.

```r
1 - pnorm(7.9, 7.5, 0.8)
```

```
## [1] 0.3085375
```

Can we solve question 1(b) without knowing the mean and standard deviation of the watermelon weights?

The Z-score of this watermelon is 0.5. Watermelons having a larger weight is the same as having a larger Z-score. Thus, $P(X > \text{weight of this watermelon}) = P(Z > 0.5) = 1 - P(Z < 0.5)$.

```r
1 - pnorm(0.5) # same as 1 - pnorm(0.5, 0, 1)
```

```
## [1] 0.3085375
```

### 1(c)

Ten watermelons are chosen at random from that variety (A). What is the probability that exactly 2 of them weigh more than 8.5 kg? Assume that the weights of the 10 watermelons are identical to X and independent since the population is very large. (Sampling such a small number of watermelons without replacement will not functionally change the probability of the outcome of subsequent watermelons.)

View choosing each watermelon as an event, the outcome that the watermelon has a weight more than 8.5 Kg as successful. Then what is the distribution of the number of success?

What is the probability of success?

$P(success) = P(X > 8.5)$

```r
1 - pnorm(8.5, 7.5, 0.8)
```

```
## [1] 0.1056498
```

Then the number of watermelons $> 8.5$ Kg $Y \sim Bin(10, 0.1056498)$.

$P(Y = 2) = \frac{10!}{2!(10-2)!}(0.1056498)^2(1 - 0.1056498)^{10-2} = 0.205$.

Alternatively, you can use `dbinom()` to compute the probability.

```r
dbinom(2, 10, 0.1056498)
```

```
## [1] 0.2055941
```

### 1(d)

Watermelons from another variety (B) have weights well approximated by a normal distribution with mean 8.6 kg and standard deviation of 0.5 kg $X \sim N(8.6, 0.5^2)$. What weight watermelon from variety B would be at the same percentile as a watermelon from variety A that weighs 8.3 kg relative to their respective populations?

If two watermelons from different varieties are at the same percentile relative to their respective populations, they should have the same z-score.

Z-score for the watermelon from variety A that weighs 8.3 kg:

$$z = \frac{8.3 - 7.5}{0.8} = 1.$$

The watermelon from variety B at the same percentile that weighs ? should have the same z-score:

$$z = \frac{? - 8.6}{0.5} = 1.$$

Thus, it weighs 9.1 Kg.

# Exercise 2

## R: pbinom()

Similar to `pnorm`, "P" stands for probability and "binom" stands for binomial distribution.

If $X \sim Bin(n, q)$, then $P(X \leq m)$ for $0 \leq m \leq n$ can be calculated by `pbinom(m, n, q)`.

For example, if $X \sim Bin(4, 0.45)$, what is the probability that $X$ is not larger than 1?

$P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{4!}{0!4!}(0.45)^0(1-0.45)^4 + \frac{4!}{1!3!}(0.45)^1(1-0.45)^3 = 0.3909813.$

Check with R:

```r
pbinom(1, 4, 0.45)
```

```
## [1] 0.3909813
```

#### Practice

$X \sim Bin(10, 0.3)$, use R to compute the probability that $X$ is at least 5.

Hint: the answer is 0.1502683.

## Question

Work on Exercise 2 with your neighbors.

For a part time job, Darion is helping his neighbor shovel, rake, and weed throughout the year. As payment for each hour of work, Darion selects a single bill from a bag of bills. The neighbor restocks the bag each time he selects so it always has 20 bills:11 \$10s, 8 \$20s and 1 \$50.

Let X correspond to the amount Darion will be paid for an hour of work. The distribution of X is summarized in the table below. It can be confirmed that $\mu_X = 16$ and $\sigma_X^2 = 84$

| x | 10 | 20 | 50 |
|---|---|---|---|
| P(X = x) | 11/20 | 8/20 | 1/20 |

#### 2(a)

What is the probability that if Darion helps for 5 hours (and draws from the bag 5 times), he selects at least 3 bills that are more than 10 dollars? Interpret the probability value in terms of percent of samples that could be observed.

Is every selection independent? Why?

Success = selecting more than \$10 = \$20 or \$50.

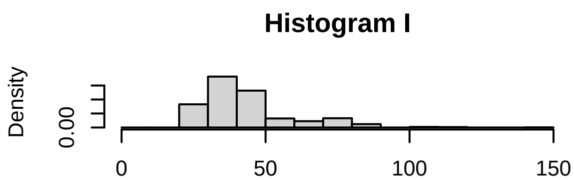
$P(success) = \frac{9}{20}$.

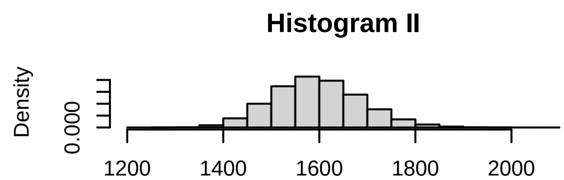
```r
1 - pbinom(2, 5, 0.45)
```

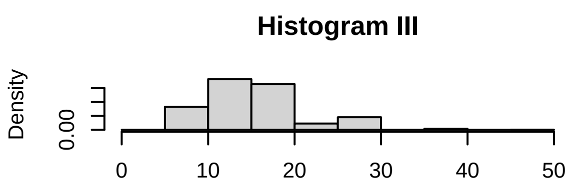
```
## [1] 0.4068731
```

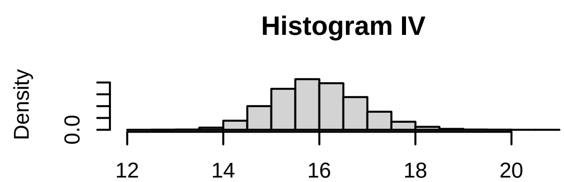
### 2(b)

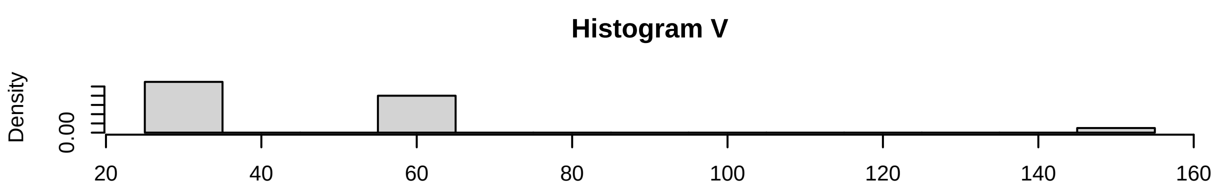
Match the formulaic and verbal description of the random variable with the histogram that most closely matches the probability histogram of the RV's sampling distribution. That is, what probability histogram displays all values the statistic could take and their probability under repeated sampling?




RV formulas:

- RV1: $Y = X_1 + X_2 + X_3$
- RV2: $Y = \frac{1}{3}(X_1 + X_2 + X_3)$
- RV3: $Y = X_1 + X_2 + X_3 + .... + X_{100}$
- RV4: $Y = \frac{1}{100}(X_1 + X_2 + X_3 + .... + X_{100})$
- RV5: $Y = 3X$

RV descriptions:

- A: Let Y be the total amount of money Darion earns from working 100 hours.
- B: Let Y be the amount of money Darion earns if he works for 1 hour but the employer triples his draw.
- C: Let Y be the total amount of money Darion earns from working 3 hours.
- D: Let Y be the average amount of money Darion earns from working 3 hours.
- E: Let Y be the average amount of money Darion earns from working 100 hours.

Expectation of X:

$$E(X) = 10 \times \frac{11}{20} + 20 * \frac{8}{20} + 50 \times \frac{1}{20} = 16$$

$$Var(X) = (10-16)^2 \times \frac{11}{20} + (20-16)^2 * \frac{8}{20} + (50-16)^2 \times \frac{1}{20} = 84.$$