Math 1350 - Exam 1 Study Guide

Daniel Havens
dwhavens@unm.edu

September 11, 2026

Contents

Summary and Disclaimer

This is a study guide for the first exam for math 1350 at the University of New Mexico (Intro to Statistics). The exam covers chapters 1-3 of *The Basic Principles of Statistics*, by Moore, Notz, and Flinger. As such, this study guide is focused on that material. I assume that the student reading this study guide is familiar with the material from chapter 0 of the textbook, as well as basic algebra. If a you feel that you need to review this material, you can send me an email, or reach out to on-campus tutoring resources provided by the Center for Teaching and Learning:

https://ctl.unm.edu/undergraduate-students/stem-tutoring.html

If you are not in my class, I cannot guarantee how much these notes will help you. With that said, if your TA or instructor has shared these with you, then you will most likely get some use out of them.

Here, I do not focus on examples like the textbook does. These notes assume that you are already familiar with the basic concepts, and just need a quick refresher before an exam. If, as you are going through this review, you find that haven't seen one of these topics, go ahead and review the relevant section in the textbook and try a few related problems, and then come back.

Methods and Techniques

This exam is concerned with making sure you understand the basics of exploratory data analysis, the general idea of a distribution, as well as how normal distributions work. This is because normal distributions are the most common statistical distributions you will run across in day-to-day life.

Our first concern is about how to categorize variables. In particular, it is good to have a formal idea of when we are treating a variable as a number, and when we are treating a variable as a "category" that things can fall into.

Categorical and Quantitative Variables

There are two types of variables: quantitative variables and categorical variables. In general, we label anything that is a number as a quantitative variable, and anything that isn't a number as a categorical variable.

Different kinds of variables have different kinds of graphs. For instance, categorical variables lend themselves well to graphs that highlight differences between categories. Whereas, in contrast, quantitative variables usually lend themselves more cleanly to variables that highlight numerical differences and show off trends.

Plots for Categorical Variables

In general, our options for visualizing categorical variables are fairly limited. We tend to use the following plots/graphs for categorical variables:

- Bar graphs
- Pie charts

In contrast, quantitative variables are much easier to work with, since they're numbers (and numbers lend themselves quite well to analysis):

Plots for Quantitative Variables

In general, we tend to use the following plots/graphs for quantitative variables:

- Stem and Leaf plots (split or unsplit).
- Histograms.
- Box (box-and-whisker) plots.
- Dot plots.

Sometimes we want to see roughly how data is spread out. For this, we tend to use split (or sometimes unsplit) stem and leaf plots. The difference is highlighted below:

Split and Unsplit Stem and Leaf Plots

The idea of a stem and leaf plot is to list out the ten's place (or higher) in one column, and the one's place in the second column. This allows you to see how your data are roughly spread.
For a split stem plot, each "ten's place" column will have two entries: one for one's places between 0 and 4, and another for entries between 5 and 9.

Next up are bar graphs and histograms. They also provide a nice way of seeing how much data is in certain categories. They are super similar visually, but have a few key differences.

Bar Graphs and Histograms

Bar graphs and histograms look almost identacle at first glance. There are a number of differences. In order of importance, the main two are:

- Bar graphs are for categorical variables, histograms are for quantitative variables.
- Bar graphs (usually) have a gap between the bars, histograms don't.

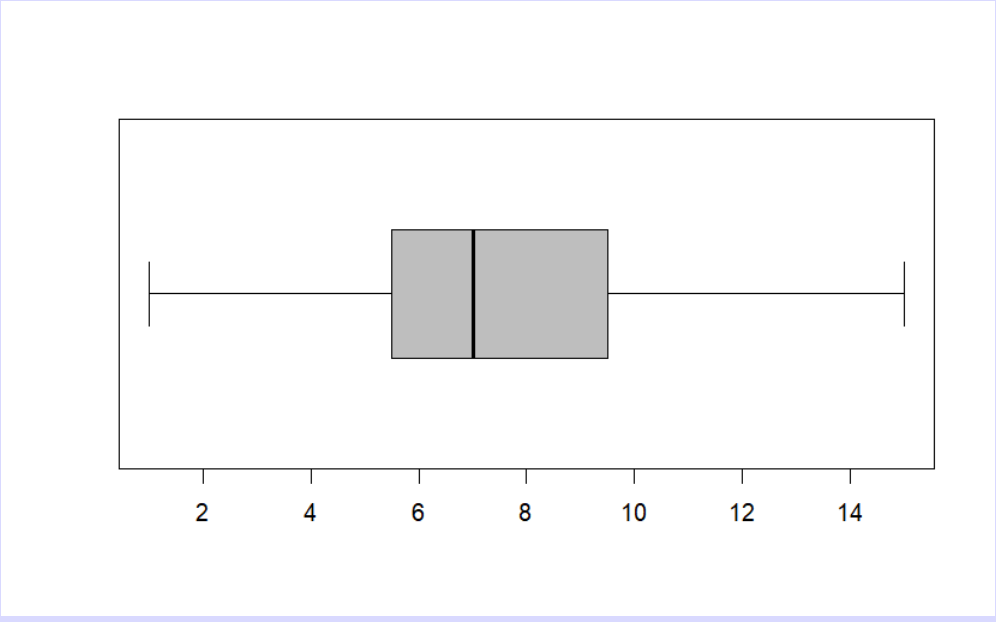
Median and Quartiles: 5-Number Summaries

A five-number summary consists of (as you guessed) five numbers: the minimum, the maximum, the median, and the first and third quartiles.
To find the median, we want to find the datum which is in the exact middle of the data set. So write your data set in order from least to greatest, and remove items from each side until you find the one in the middle. If there are two in the middle, take their average, and that's your median.
The median splits the data set in half, and each of those halves has their own "median", called the quartiles. The first quartile is the one that splits the first half in half, and the third quartile is the one that splits the second half in half. You find them the same way you found the median, only removing items from each half instead of from the full data set.

We often pair 5-number summaries with a plot that highlights them. We call these "box and whisker plots", or sometimes just "box plots".

Box, or Box-And-Whisker, Plots

Box plots provide a nice way of laying out a five-number summary visually. Above a number line, make a bar where each of the five numbers of your five-number summary are. Then, connect them such as below to get a box plot:



We're also concerned with other ways to measure the average, as well as how to measure the spread of data. We usually use mean and standard deviation to measure these things. Mean is a measure of average, often denoted μ, while standard deviation is a measure of spread, often denoted σ.

Mean and Standard Deviation

The mean of a set of n datum, $x_1, x_2, \ldots, x_n$, is

$$\mu = \bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

For standard deviation, we compute σ^2 with a definition, and you take the square root of it to get σ.

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n - 1}.$$

That $n - 1$ isn't a typo! It has to do with what's called the *degrees of freedom* of the data set. But that's a topic that we cover in more advanced courses. So you don't need to worry about the why, so much as the fact that it's there.

Next up, it's good to have a nice rule of thumb for detecting outliers. In general, as the subject expert, you will know when a piece of data is an outlier. But when you aren't an expert on the subject, this can help you determine if a piece of data is an outlier.

Outliers

Begin by recalling the definition of the inter-quartile range (IQR):

$$IQR = Q3 - Q1.$$

If a datum is more than $1.5 \cdot IQR$ away from the nearest quartile, then we can mark it as an outlier.

Next up, we want to address situations where the data you may be working with could be skewed one way or another, and how this changes how the mean and median look compared to each other.

Skew Distributions

Sometimes, the data has a "tail" that you can see on a histogram. If the tail is to the left, we call the distribution a left-skewed distribution.
If the tail is to the right, we call the distribution a right-skewed distribution.
The mean tends to go towards the tail, while the median stays closer to the center "bump" of the distribution. This means that on right-skewed distributions, the median is bigger than the mean, and on left-skewed distributions, the median is smaller than the mean.

We can now switch our focus to normal distributions. They occur in nature all over the place, so you're likely to run across them in real world applications of stats. The reason we like normal distributions so much is that we have a nice way of understanding how much data is close to the center. We measure this in terms of the standard deviation, giving us the following rule:

68-95-99.7 Rule

This is a rule of thumb, telling us that on a normal distribution, 68% of the data is within one standard deviation of the mean (that is, between $\mu - \sigma$ and $\mu + \sigma$), 95% is within two standard deviations of the mean (between $\mu - 2\sigma$ and $\mu + 2\sigma$) and 99.7% is within three standard deviations of the mean (between $\mu - 3\sigma$ and $\mu + 3\sigma$).

The last thing we want you to know how to do is how to interpret data from a normal distribution when the situation isn't quite as nice. That is, when you can't use the 68-95-99.7 rule, what do you do? The answer: Use a Z-table.

Z-Score and Z-Tables

To use a Z-table, we first need to compute the Z-score. In the formula, X is the cutoff for where we are trying to find out how much data is below that value, μ is the mean, and σ is the standard deviation. Our Z-score is then:

$$Z = \frac{X - \mu}{\sigma}.$$

To find the corresponding percentage of data below X, you then look up this Z value on a Z-table.

With this, you should now know everything that could show up on the exam! So we now move on to some worked examples. We will then move on to practice problems, and finally some unsolved problems.

Worked Examples

We will now work through some examples, beginning charts, and working our way towards normal distributions. The goal of this section is to show you the work right next to the problem, so that you can see how to do it right away.

We'll start with a classic: computing the mean and standard deviation of a data set. This can be tedious, and a little long, but just stick with it! In the end, to us as instructors, seeing that you understand the process is often more important than you actually getting the correct final answer. After all, sign errors happen to everyone. We care more about you knowing how to attack the problem.

Example: Find the mean and standard deviation of the following data set:

$$1, 4, 5, 6, 7, 10$$

We begin by noting that there are 6 numbers, so $n = 6$ in our formulae. Computing the mean, we get that it is:

$$\frac{1+4+5+6+7+10}{} = \frac{33}{6} = \frac{11}{2} = 5.5.$$

On the other hand, the square of the standard deviation is:

$$\frac{(1-5.5)^2 + (4-5.5)^2 + (5-5.5)^2 + (6-5.5)^2 + (7-5.5)^2 + (10-5.5)^2}{6-1}$$

Which is equal to:

$$\frac{(-4.5)^2 + (-1.5)^2 + (0.5)^2 + (0.5)^2 + (1.5)^2 + (4.5)^2}{5}$$

So,

$$\sigma^2 = \frac{45.5}{5}$$

Thus, we have that,

$$\sigma = \sqrt{\frac{45.5}{5}}$$

Next up, we will do an example dealing with skew distributions. These questions tend to be more of the trivia type, asking about how the mean and median relate to each other. Here's an example of my personal favorite kind of question:

Example: On a right-skew distribution, is the mean minus the median: positive, negative, or zero? Why?
With a right-skew distribution, we have a tail on the right-hand side. This drags the mean further to the right than the median. This means that the mean is larger than the median, since it is to the right of it on the number line. Thus, the mean minus the median must be positive.

Another common answer deals with five number summaries. Here, I won't include the follow-up question of "make a box plot for the data", since my graph would be made by a computer, and not representative of how you'd actually do it. But if you're following along, I recommend you give that a try while you read through this one.

Example: Give the five number summary of the following data

$$1, 1, 1, 3, 3, 4, 5, 6, 6, 7, 8, 9$$

We know the minimum is 1 and the maximum is 9. There are 12 numbers, so half way down the middle are the sixth and seventh numbers: 4 and 5. So the median is half-way between them, 4.5. Finally, the first quartile is half way between 4 and 1, which means it's half way between the final 1 and the first 3, making $Q1 = 2$. For the third quartile, we similarly have that it is half way between 6 and 7, meaning $Q3 = 6.5$. So, our five number summary is:

Min: 1, Max: 9
Median: 4.5
Q1: 2, Q3: 6.5

Next up, normal distributions. We will go through a couple of common problems, starting with the 68-95-99.7 rule. The goal of these problems is to get you to realize that the numbers we give you are some number of standard-distributions away from the mean, and then to use the rule to get the answer.

Example: Using the 68-95-99.7 rule, if you have a normal distribution with $\mu = 5$ and $\sigma = 3$, determine how much data is between -1 and 11.
We begin by noting that $11 = 5 + 6 = \mu + 2\sigma$. So, we suspect that our region is two standard deviations in width. We confirm by checking the other side. Seeing that $-1 = 5 - 6 = \mu - 2\sigma$, our suspicion has been confirmed. Thus, by the 68-95-99.7 rule, we know that 95% of the data is between -1 and 11 on our distribution.

Next, up, Z-tables. These are the other kind of normal distribution question we can ask you. And it's worth knowing how to do them. Try drawing a picture of the normal distribution to go with it, since that can often help you visualize what's going on.

Example: Use a Z-table to find how much data of a normal distribution with $\mu = 10$ and $\sigma = 10$ lies below 12.4.
We begin by computing the Z-score. Here, since we are trying to find what lies below the cutoff 12.4, we take $x = 12.4$. This gives us

$$Z = \frac{x - \mu}{\sigma} = \frac{12.4 - 10}{10} = \frac{2.4}{10} = 0.24.$$

Looking this up on a Z-table, the corresponding entry for our Z-score is 0.59483. So, 59.483% of the data lies below our x value, which is 12.4.

Often times, I find that a more complicated example can be useful too. This one asks for the data above the given number. And the given number is a weird one that you can't just simplify away. You have to actually use a calculator for it (or have a numerical estimate memorized).

Example: Using a Z-table, find how much data of a normal distribution with $\mu = 2$ and $\sigma = 1$ lies above π.
We once again start by finding our Z-score. Here, $x = \pi$, so we have that

$$Z = \frac{x - \mu}{\sigma} = \frac{\pi - 2}{1} = \pi - 2 \approx 1.14.$$

So, looking up our score on the Z-table, we see that 0.87286 is the corresponding entry. This means that 87.286% of the data is below π.
Since we want to find out how much data is above π, we just subtract this number from 100%, giving us that $100\% - 87.286\% = 12.714\%$ of the data is above π on our distribution.

Here's another type of Z-table question we could possibly ask. This kind has to deal with finding what's between two values. This is done by taking the large Z-score's table value, and subtracting the table value of the smaller Z-score table value. It can help to think of this as "The amount of stuff between a and b is the same as the amount of stuff that's below b, but not below a".

Example: Using a Z-table, find how much of the data of a normal distribution with $\mu = 3$ and $\sigma = \frac{2}{3}$ lies between 2 and 4.
Here, since 4 is the larger x value, we look it up first. We have that the Z-score is

$$Z = \frac{x - \mu}{\sigma} = \frac{4 - 3}{\frac{2}{3}} = 1.5.$$

The corresponding table entry tells us that 0.93319, or 93.319% of the data lies below 4.
On the other hand, for 2 as the smaller value, we know that for $x = 2$, we have that our Z-score is:

$$Z = \frac{2 - 3}{\frac{2}{3}} = -1.5.$$

The corresponding table entry gives us that 0.06681, or 6.681%, of the data lies below 2.
Subtracting these two to find out how much lies between 4 but not below 2 (and thus, between 4 and 2), we find that $93.319\% - 6.681\% = 86.638\%$ of the data is between 2 and 4.

Next up, we will cover the most tedious kind of question: determining what kind of variable each item in a particular study is.

Example: Suppose you are running a study on grain yield in a certain region's fields. You also gather data on soil type, type of fertilizer used, quantity of fertilizer used, and average water usage per day per square foot.
Determine which variables are response variables and which are explanatory. Explain which are categorical and which are quantitative.
We know there is only one response variable: the grain yield itself. Every other variable is a response variable: soil type, type of fertilizer, quantity of fertilizer, and water usage.
The following variables are numbers, and thus quantitative: grain yield, quantity of fertilizer, and water usage. The remaining variables are categorical, since they aren't numbers: soil type and type of fertilizer.

Practice Problems

We now move on to some practice problems. Solutions are in the next section, so don't worry! The goal here is to give you the space to try problems without seeing the solutions right away.

1. Find the mean and standard deviation of the data set:

$$1, 4, 5, 5, 5$$

2. If you have a normal distribution with $\mu = 2$ and $\sigma = 1$, use the 68-95-99.7 rule to find the probability that the observed datum is between 0 and 4.

3. If you have a normal distribution with $\mu = 4$ and $\sigma = 3$, use the 68-95-99.7 rule to find the probability that the observed datum is less than 7.

4. Using a Z-table, if you have a normal distribution with $\mu = 1$ and $\sigma = 0.5$, find the probability that the observed datum is less than 1.6.

5. On a normal distribution (say, with $\mu = 0$ and $\sigma = 1$), is the mean minus the median positive, negative, or zero? Why?

6. Determine the explanatory and response variables in the following situation: You are studying cat ages. In addition to the cat's actual ages, you measure hours of daily exercise, cat lifestyle (indoors or outdoors), coloration, and daily lasagna intake in grams.

7. In the previous problem, which of the variables are categorical, and which are quantitative? Explain why.

8. Determine which of the variables in the following study are categorical and which are quantitative variables, and which are response and which are explanatory variables. Consider a study where you are determining product sales. You also consider packing type (box, bubble mailer, etc.), product sell value, and marketing budget.

Practice Problem Solutions

1.

Solution: We first note that there five data points, so $n = 5$. To find the mean, we add them up and divide by n:

$$\bar{x} = \frac{1+4+5+5+5}{5} = \frac{20}{5} = 4.$$

To find the standard deviation, we first find the variance, σ^2, and take the square root of it:

$$\sigma^2 = \frac{(1-4)^2 + (4-4)^2 + (5-4)^2 + (5-4)^2 + (5-4)^2}{4}$$

So, simplifying we have that

$$\sigma^2 = \frac{(-3)^2 + 0^2 + 1^2 + 1^2 + 1^2}{4} = \frac{9+1+1+1}{4} = \frac{12}{4} = 3.$$

Thus, $\sigma = \sqrt{3}$.

2.

Solution: If we have a normal distribution with $\mu = 2$ and $\sigma = 1$, then $0 = 2-2\cdot 1$, or, written in terms of our variables, $0 = \mu - 2\sigma$. Similarly, $4 = 2+2\cdot 1 = \mu + 2\sigma$. Thus, we want to know how much data is within two standard deviations of the mean. The 68-95-99.7 rule tells us that this is 95% of the data.

3.

Solution: We have a normal distribution with $\mu = 4$ and $\sigma = 3$. Since we know that $7 = 4 + 3 = \mu + \sigma$, we know we are working one standard deviation away from the mean, which contains 68% of the data. Put differently, 68% of the data is between 1 and 7 (since this is one standard deviation away from the mean).
This means that we haven't accounted for 32% of the data (which is sitting in the tails of the distribution). Of that, half is below 1 and half is above 7. To rephrase, $\frac{32}{2}\% = 16\%$ of the data is below 1.
Combining this with our data that is between 1 and 7, we know that $16\%+68\% = 84\%$ of the data is less than either less than 1, or between 1 and 7. Put together, this means that 84% of the data is less than 7.

4.

Solution: We begin by noting that our cutoff, our "x" is 1.6. Combining this with the known values of $\mu = 1$ and $\sigma = 0.5$, we have a Z-score of

$$Z = \frac{x - \mu}{\sigma} = \frac{1.6 - 1}{0.5} = \frac{1.1}{0.5} = 2.2.$$

So, using a Z-table, we have that 0.98610 of the data, or 98.61% of the data are less than, 1.6.

5.

Solution: We know that the mean is at the exact center of a normal distribution. We also know that normal distributions are symmetric, so the median is in the exact center as well. Thus, they are the exact same number, and their difference is zero.

6.

Solution: As we are looking to study it, the cat's ages are our response variable. This means that everything else is an explanatory variable: daily exercise, lifestyle, coloration, and even daily lasagna intake.

7.

Solution: Since they are numbers, the quantitative variables are: cat age, daily exercise, and lasagna intake.
The remaining variables are all categorical: namely lifestyle and coloration.

8.

Solution: The response variable is product sales, and the remaining are all explanatory: packaging type, retail price, and marketing budget.
The quantitative variables are: product sales, retail price, and marketing budget.
While the only categorical variable is packaging type.

Unsolved Problems

For more problems than the practice test(s) given in class and on canvas, time yourself doing the following problems. I recommend picking (about) ten of them at random to make your own personalized "practice exam". It's important that you be prepared for whatever can show up, so make sure you don't take it easy on yourself, and give yourself a few hard ones.

1. Make an unsplit stem and leaf plot for the following data set:

$$0, 1, 5, 6, 10, 11, 12, 14, 21, 22, 31, 35.$$

2. Make an unsplit stem and leaf plot for the following data set:

$$31, 37, 51, 52, 53, 55, 57, 61, 72, 102.$$

3. Make a split stem and leaf plot for the following data set:

$$10, 11, 11, 13, 15, 16, 16, 18, 21, 24, 26.$$

4. Make a split stem and leaf plot for the following data set:

$$44, 57, 59, 61, 67, 100, 101, 107.$$

5. Make a histogram for the following data set:

$$10, 14, 14, 15, 17, 17, 17, 20, 21.$$

6. Make a histogram for the following data set:

$$-1, -1, 0, 4, 7, 10.$$

7. Determine the explanatory and response variables in the following situation:

   In trying to study the deer population, you track the population of deer, the number of young deer, the season, how much forest cover there is in the region, the most common predator type, and the total population of predators in the region.

8. Determine the explanatory and response variables in the following situation:

   To study how long it takes gravity to start affecting a coyote (in seconds), you also track the number of ACME product sales, as well as the number of roadrunners spotted.

9. Compute the mean of the following data set:

$$4, 15, 21, 22, 25, 28, 31, 40, 49.$$

10. Compute the mean of the following data set:

$$10, 15, 16, 16, 18, 19, 20.$$

11. Compute the standard deviation of the following data set:

$$8, 9, 10, 10, 11, 12.$$

12. Compute the standard deviation of the following data set:

$$1, 4, 5, 5, 6.$$

13. If you have a left-skew distribution, is the mean minus the median: positive, negative, or zero? Why?

14. If you have a normal distribution, is the mean left of the median, right of the median, or the same as the median? Why?

15. For the following data set, write the five number summary and draw a box plot:

$$1, 1, 2, 4, 5, 7, 8, 9, 9, 10, 12$$

16. For the following data set, write the five number summary and draw a box plot:

$$1, 1, 1, 1, 5, 7, 8$$

17. In the following situation, determine which variables are categorical, and which are quantitative:

    Suppose you are studying cat age. You collect data for the following variables: cat age, daily cat food in grams, cat coloration, and living situation (indoor vs outdoor).

18. In the following situation, determine which variables are categorical, and which are quantitative:

    You are studying frog population in a local wetland. You gather the following variables: number of frogs in a pond, pond pH level, most common frog species in that pond, and month in which the observation was made.

19. Using the $68-95-99.7$ rule, if you have a normal distribution with $\mu = 10$ and $\sigma = 3$, how much of the data is between 7 and 13?

20. Using the $68-95-99.7$ rule, if you have a normal distribution with $\mu = 10$ and $\sigma = 3$, how much of the data is under 10?

21. Using the $68-95-99.7$ rule, if you have a normal distribution with $\mu = 10$ and $\sigma = 3$, how much of the data is below 13?

22. Using the $68-95-99.7$ rule, if you have a normal distribution with $\mu = 10$ and $\sigma = 3$, how much of the data is above 16?

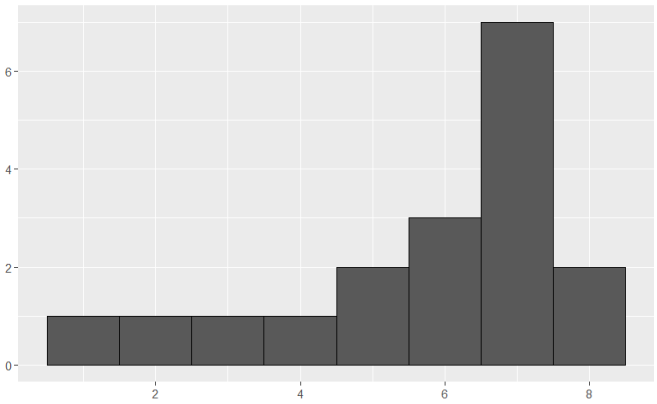
23. Using a Z-table, if you have a normal distribution with $\mu = 5$ and $\sigma = 3$, how much of the data is below 7?

24. Using a Z-table, if you have a normal distribution with $\mu = 5$ and $\sigma = 3$, how much of the data is below 1?

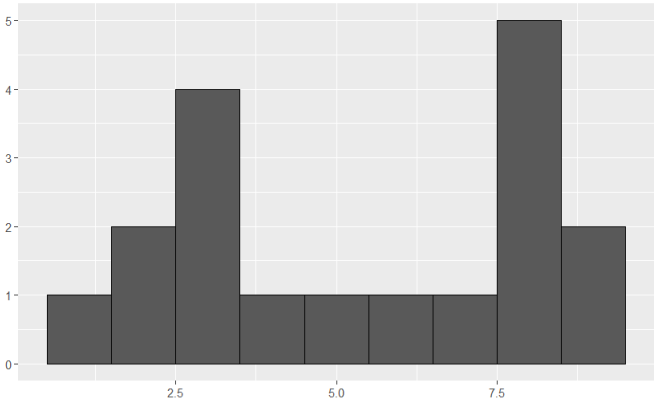
25. Using a Z-table, if you have a normal distribution with $\mu = 5$ and $\sigma = 3$, how much of the data is above 4?

26. Using a Z-table, if you have a normal distribution with $\mu = 5$ and $\sigma = 3$, how much of the data is between 3 and 6?

27. Does the following histogram show data from a left-skew distribution, right-skew distribution, or neither?



28. Does the following histogram show data from a left-skew distribution, right-skew distribution, or neither?



29. Are there any outliers in the following dataset that you can identify with methods from this course? If so, what are they, and why are they outliers? If not, explain why.

$$2, 4, 8, 8, 8, 9, 9, 9, 10, 10, 10, 10, 11$$

30. Are there any outliers in the following dataset that you can identify with methods from this course? If so, what are they, and why are they outliers? If not, explain why.

$$1, 1, 1, 2, 2, 2, 3, 4, 4, 5, 5, 6$$