Think and Answer requires you to answer two questions. For each question, do not simply provide an answer; make sure you explain how you arrived at that answer. Even if your reasoning is wrong, you will still be credited for participation.

Answer the following question:

1. We define probabilities for various normal distributions. However, we use one single table (Standard Normal Distribution Table) for all of them. Explain why that is possible.

All scores from normal distribution families with different means and standard deviations can be "standardized" calculating the Z score. In the Z-score calculation, the distance from the mean is expressed in terms of standard deviations, which effectively transforms the standard deviation to one and the center of the distribution to zero, which are the parameters of the standard normal distribution). As a result of this standardization, only one normal distribution table is required to compute the probability of observing values in certain ranges provided that the underlying distribution is normal even if its parameters differ from those of the standard normal distribution.

True or False?

2. The more lottery tickets you buy, the higher is your chance to win a prize.

True, the probability of winning a prize from purchasing one lottery ticket is infinitesimally small. However, the probability of winning increases (although by a small magnitude) with an increase in the number of tickets bought as the probabilities are added. If you buy all lottery tickets printed, you are guaranteed to win a prize.

3. Health insurance policy is designed in such a way that a healthy person has a negative expected value, and an ailing one has a positive expected value.

False, the expected value to insurance companies for insurance companies is positive as multiplying the probability of no illness (which is high and close to one) by the premium payments yields a positive expected value while the expected value for ailing persons is negative as the product of the probability of illness (which is high) and insurance payouts exceeds the

product of the probability of no illnesses (which is low) and premium payments.

4. Because the tails of the normal distribution curve are infinitely long, the total area under the curve is also infinite.

False. The total area under the curve represents the probability of occurrence of the range of values and is equal to one as probability cannot be more than one.

HW 3

Ch. 5

Sec 5-1:

1. **Random Variable** The accompanying table lists probabilities for the corresponding numbers of girls in four births. What is the random variable, what are its possible values, and are its values numerical?

Number of Girls in Four Births

The random variable is number of girls and its possible values are 0,1, 2, 3, and 4. Yes, the values are numerical

Number of Girls x	1. P(x)
1. 0	1. 0.063
1. 1	1. 0.250
1. 2	1. 0.375
1. 3	1. 0.250
1. 4	1. 0.063

1.

1. **Discrete or Continuous?** Is the random variable given in the accompanying table discrete or continuous? Explain.

The random variable is discrete as it can take a finite number of values.

Identifying Probability Distributions In **Exercises 7–14**, determine whether a probability distribution is given. If a probability distribution is given, find its mean and standard deviation. If a probability distribution is not given, identify the requirements that are not satisfied.

9. **Pickup Line** Ted is not particularly creative. He uses the pickup line "If I could rearrange the alphabet, I'd put U and I together." The random variable *x* is the number of women Ted approaches before encountering one who reacts positively.

x	***P*(*x*)**
1	0.001
2	0.009
3	0.030
4	0.060

This is not a probability distribution as the cumulative probability is 0.1, which is less than 1.

10. **Fun Ways to Flirt** In a Microsoft Instant Messaging survey, respondents were asked to choose the most fun way to flirt, and the accompanying table is based on the results.

	***P*(*x*)**
E-mail	0.06
In person	0.55

Instant message	0.24
Text message	0.15

This is not a probability distribution since the random variable, most fun way to flirt, is non-numeric.

13 (a,c)

Cell Phone Use In a survey, cell phone users were asked which ear they use to hear their cell phone, and the table is based on their responses (based on data from "Hemispheric Dominance and Cell Phone Use," by Seidman et al., *JAMA Otolaryngology—Head & Neck Surgery,* Vol. 139, No. 5).

	***P*(*x*)**
Left	0.636
Right	0.304
No preference	0.060

This is not a probability distribution since the random variable, ear used to hear cell phone, is non-numeric.

Genetics. In **Exercises 15–20**, refer to the accompanying table, which describes results from groups of 8 births from 8 different sets of parents. The random variable *x* represents the number of girls among 8 children.

Number of Girls *x*	***P*(*x*)**
0	0.004
1	0.031

2	0.109
3	0.219
4	0.273
5	0.219
6	0.109
7	0.031
8	0.004

15. **Mean and Standard Deviation** Find the mean and standard deviation for the numbers of girls in 8 births.
Mean = 0 * 0.004 + 1 * 0.031 + 2 * 0.109 + 3 * 0.219 + 4*0.273 + 5 * 0.219 + 6 *0.109 + 7*0.031 + 8 * 0.004
Mean = 3.996 girls
Variance = (3.996 – 0)^2 * 0.004 + (3.996 – 1)^2 * 0.031 + (3.996 – 2)^2 * 0.109 + (3.996 – 3)^2 * 0.219 + (3.996 – 4)^2 * 0.273 + (3.996 – 5)^2 * 0.219 + (3.996 – 6)^2 * 0.109 + (3.996 – 7)^2 * 0.031 + (3.996 – 8)^2 * 0.004
Variance = 1.996
Standard deviation = √1.996 = 1.4128

16. **Range Rule of Thumb for Significant Events** Use the range rule of thumb to determine whether 1 girl in 8 births is a significantly low number of girls.
Significantly low cut-off = 3.996 – 2*1.4128
Significantly low, cut-off = 1.17
Since 1 girl is less than the 1.17 girls cut-off, 1 girl in 8 births is a significantly low number of girls.

18.Using Probabilities for Significant Events

- Find the probability of getting exactly 7 girls in 8 births.
  P of exactly 7 girls = 0.031
- Find the probability of getting 7 or more girls in 8 births.
  P of 7 or more girls = P (X = 7) + P (X = 8)
  P of 7 or more girls = 0.031 + 0.004 = 0.035

- Which probability is relevant for determining whether 7 is a significantly high number of girls in 10 births: the result from part (a) or part (b)?
  The result from part b
- Is 7 a significantly high number of girls in 8 births? Why or why not?
  7 is significantly high number of girls in 8 births as the probability of getting at least 7 girls in 8 birth is significantly low at 0.035.

28. **Expected Value in Roulette** When playing roulette at the Venetian casino in Las Vegas, a gambler is trying to decide whether to bet $5 on the number 27 or to bet $5 that the outcome is any one of these five possibilities: 0, 00, 1, 2, 3. From **Example 6**, we know that the expected value of the $5 bet for a single number is –26
c (slash through it). For the $5 bet that the outcome is 0, 00, 1, 2, or 3, there is a probability of 5/38 of making a net profit of $30 and a 33/38 probability of losing $5.

    a. Find the expected value for the $5 bet that the outcome is 0, 00, 1, 2, or 3.
    Expected value = 5/38 * 5 + 33/38 * -5
    Expected value = -3.68
    b. Which bet is better: a $5 bet on the number 27 or a $5 bet that the outcome is any one of the numbers 0, 00, 1, 2, or 3? Why?
    The bet that the outcome is one of five numbers is better as it has a higher expected value. In the long-run betting on five numbers results in lower losses than betting on a single number

29. **Expected Value for Life Insurance** There is a 0.9986 probability that a randomly selected 30-year-old male lives through the year (based on data from the U.S. Department of Health and Human Services). A Fidelity life insurance company charges $161 for insuring that the male will live through the year. If the male does not survive the year, the policy pays out $100,000 as a death benefit.

    a. From the perspective of the 30-year-old male, what are the monetary values corresponding to the two events of surviving the year and not surviving?
    The monetary value from surviving and not surviving is -$161 and $100,000, respectively.
    b. If a 30-year-old male purchases the policy, what is his expected value?
    Expected value = 0.9986 * -161 + (1 – 0.9986) * 100,000
    Expected value = -$20.77
    c. Can the insurance company expect to make a profit from many such policies? Why?
    Yes, the expected value to the insurance company is $20.77.

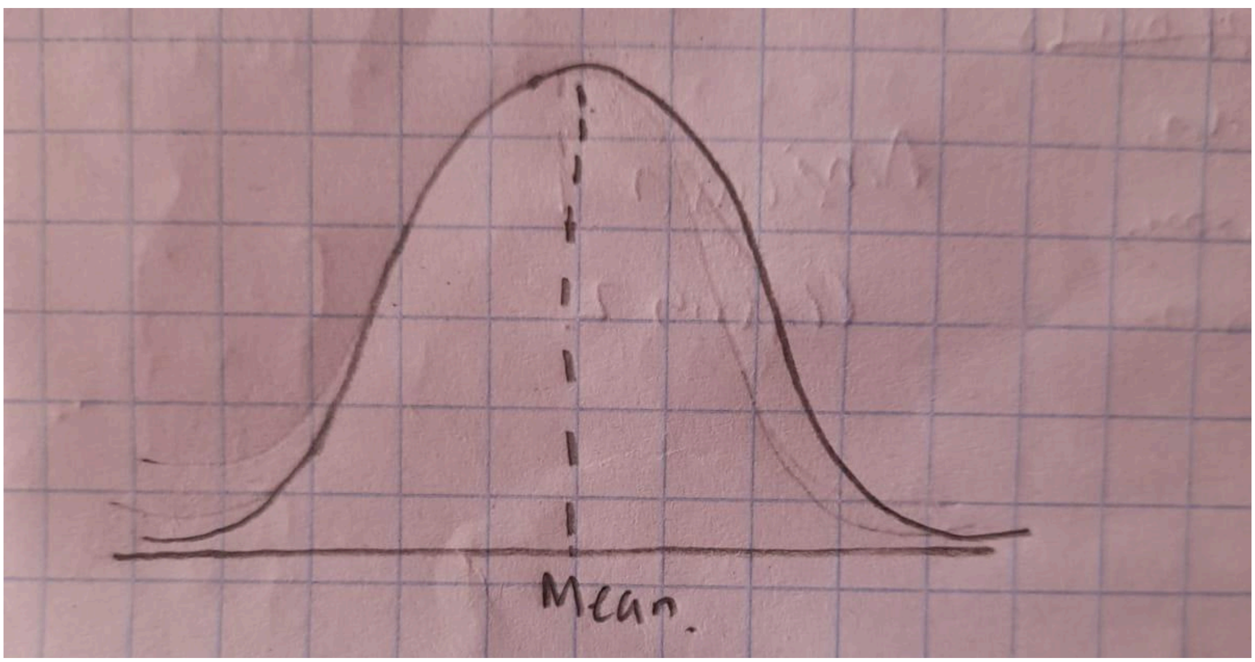
Ch. 6

Sec.6-1:

1. **Normal Distribution** What's wrong with the following statement? "Because the digits 0, 1, 2, . . ., 9 are the normal results from lottery drawings, such randomly selected numbers have a normal distribution."

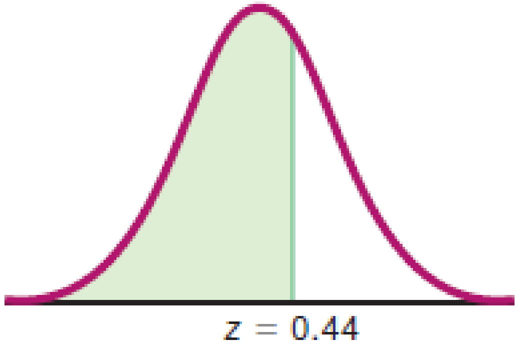
   A normal distribution does not imply arising from normal results. It refers to a bell-shaped distribution where extreme values to the left and right of the mean are unlikely with the likelihood of the occurrence of values increasing as you move closer to the mean. The selected numbers from lottery drawings should have a uniform distribution if selected randomly as random selection implies equal chance of occurrence, which is a characteristic of uniform distributions.

1. **Normal Distribution** A normal distribution is informally described as a probability distribution that is "bell-shaped" when graphed. Draw a rough sketch of a curve having the bell shape that is characteristic of a normal distribution.



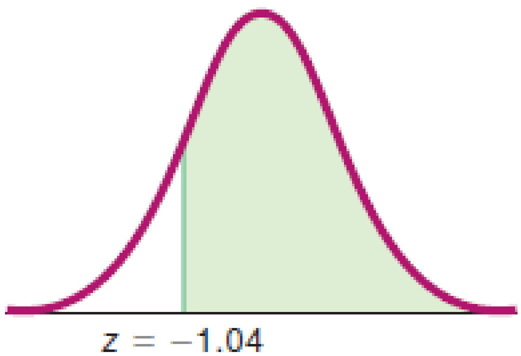
Standard Normal Distribution. In **Exercises 9–12**, find the area of the shaded region. The graph depicts the standard normal distribution of bone density scores with mean 0 and standard deviation 1.

9.



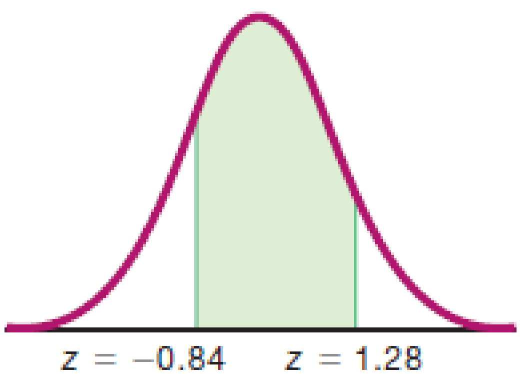
P (Z < 0.44) = 0.67

10.



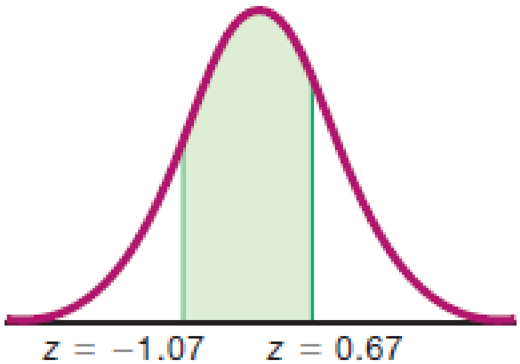
P (Z > -1.04) = 1 – P (Z < -1.04)
P (Z > -1.04) = 0.8508

11.



P (-0.84 < Z < 1.28) = P (Z < 1.28) – P(Z < 0.84)
P(-0.84 < Z < 1.28) = 0.6993

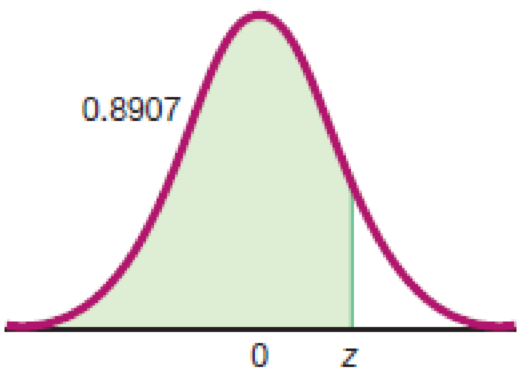
12.



P (-1.07 < Z < 0.67) = P (Z < 0.67) – P(Z < 1.07)
P (-1.07 < Z < 0.67) = 0.6063

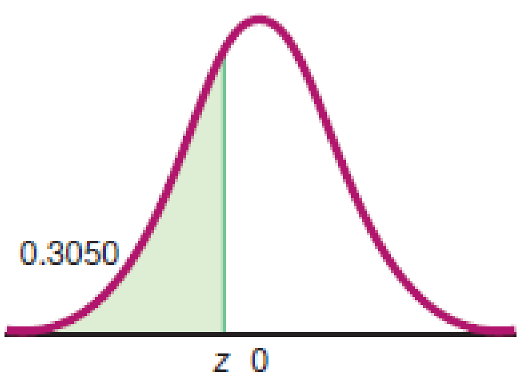
Standard Normal Distribution. In **Exercises 13–16**, find the indicated z score. The graph depicts the standard normal distribution of bone density scores with mean 0 and standard deviation 1.

13.



P (Z <?) = 0.8907
Z = 1.23

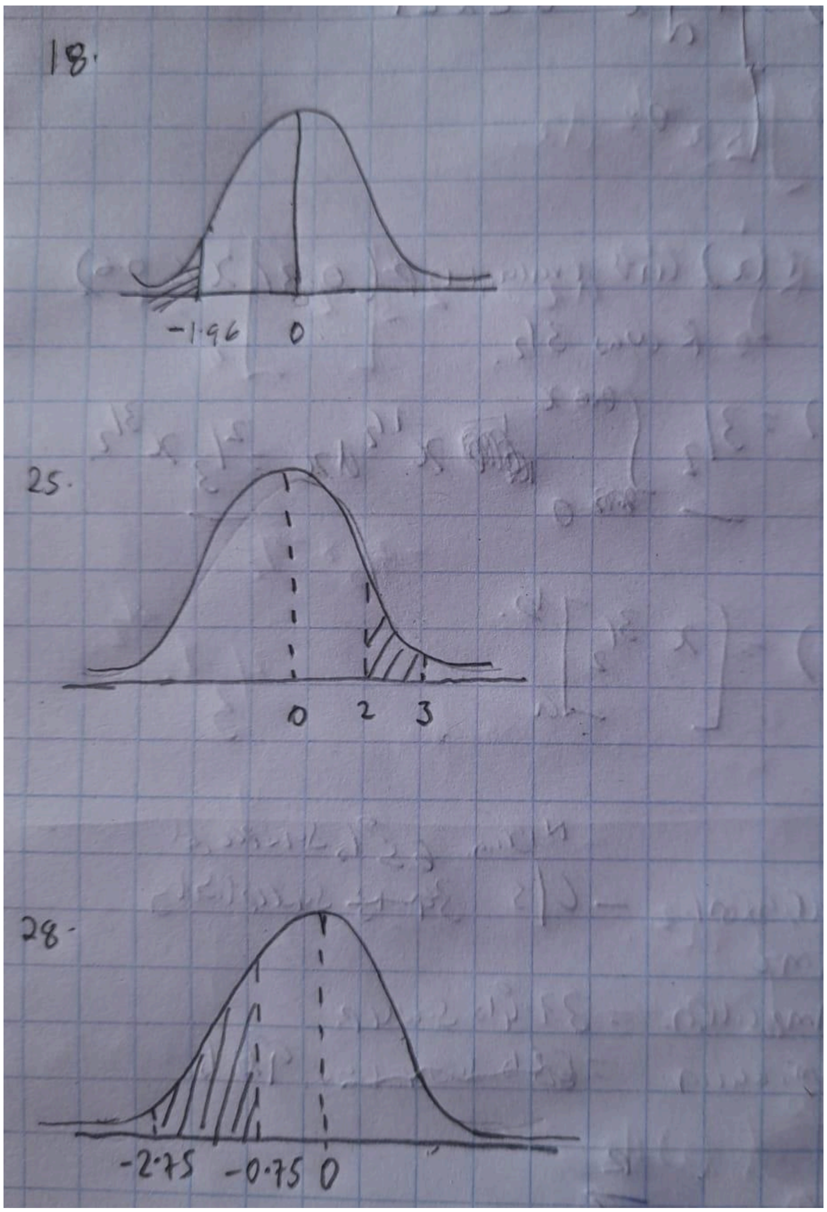
13.



P (Z <?) = 0.305

$Z = -0.51$

Standard Normal Distribution. In **Exercises 17–36**, assume that a randomly selected subject is given a bone density test. Those test scores are normally distributed with a mean of 0 and a standard deviation of 1. In each case, draw a graph, then find the probability of the given bone density test scores. If using technology instead of **Table A-2**, round answers to four decimal places.



18. Less than –1.96

$P(Z < -1.96) = 0.025$

25. Between 2.00 and 3.00

$P(2 < Z < 3) = P(Z < 3) - P(Z < 2)$

P (2< Z < 3) = 0.0214
28. Between −2.75 and −0.75
P (-2.75 < Z < -0.75) = P (Z < -0.75) – P (Z < -2.75)
P (-2.75 < Z < -0.75) = 0.2236

Critical Values. In **Exercises 41–44**, find the indicated critical value. Round results to two decimal places.

41, $z_{0.10}$
Area below the normal distribution = 0.10
Z = -1.28
Sec. 6-3:

1. **Births** There are about 11,000 births each day in the United States, and the proportion of boys born in the United States is 0.512. Assume that each day, 100 births are randomly selected and the proportion of boys is recorded.
   A. What do you know about the mean of the sample proportions?
   The mean of the sample proportions is equal to the population proportion of 0.512
   B. What do you know about the shape of the distribution of the sample proportions?
   Since each sample comprises of 100 births and is adequately large (N > 25), the sample proportions are normally distributed.

2. **Sampling with Replacement** The Orangetown Medical Research Center randomly selects 100 births in the United States each day, and the proportion of boys is recorded for each sample.
   A. Do you think the births are randomly selected with replacement or without replacement?
   I think the births are selected without replacement as it is more convenient to select all 100 births in one step rather than repeatedly after replacement
   C. Give two reasons why statistical methods tend to be based on the assumption that sampling is conducted *with* replacement, instead of without replacement.

Sampling with replacement yields a true random sample as the probability of selection is equal for all the items in the population unlike in sampling without replacement where the probability of selection increases as items are selected. Sampling with replacement also ensures the independence condition of statistical testing is met as the observations are independent of each other unlike in sampling without replacement the probability of selection is dependent on the previous items selected.

4. **Sampling Distribution** Data Set 4 "Births" in **Appendix B** includes a sample of birth weights. If we explore this sample of 400 birth weights by constructing a

histogram and finding the mean and standard deviation, do those results describe the sampling distribution of the mean? Why or why not?

No, the histogram and statistics describe the center, dispersion, and shape of the sample of 400 birth weights, not the sampling distribution of the mean. In order to describe the sampling distribution of the mean, we must take many samples of 400 birth weights and calculate the mean of their means and the standard deviation of the means. The histogram of the sampling distribution of the mean depicts the distribution of the means of the samples.

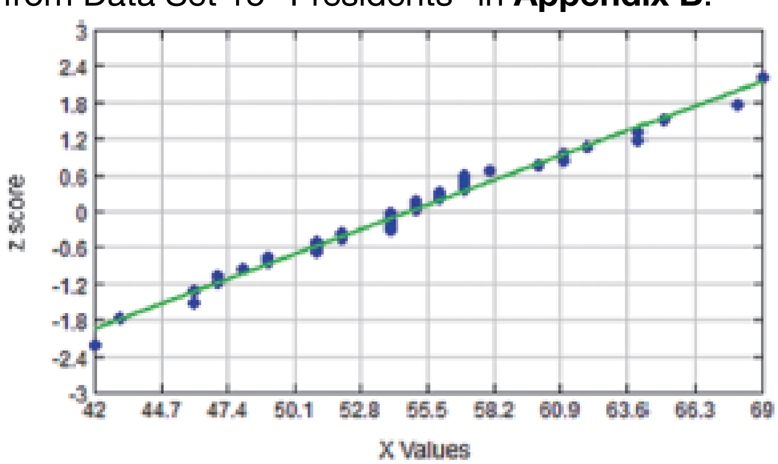
6. **College Presidents** There are about 4200 college presidents in the United States, and they have annual incomes with a distribution that is skewed instead of being normal. Many different samples of 40 college presidents are randomly selected, and the mean annual income is computed for each sample.

A. What is the approximate shape of the distribution of the sample means (uniform, normal, skewed, other)?
   The shape of the sampling distribution is approximately normal as all sampling distributions for adequately large samples (N > 25) are approximately normally distribution regardless of the shape of the distribution from which the samples were taken.
B. What value do the sample means target? That is, what is the mean of all such sample means?
   The mean of all sample means is equal to the population mean

Sec 6-4:

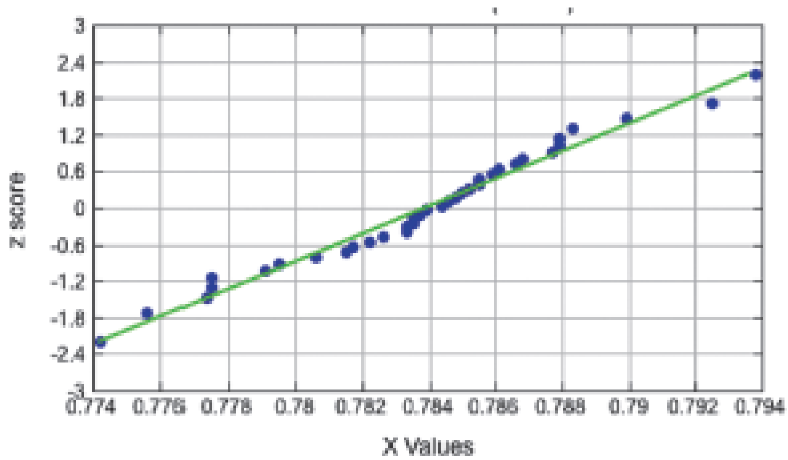
Interpreting Normal Quantile Plots. In **Exercises 5–8**, examine the normal quantile plot and determine whether the sample data appear to be from a population with a normal distribution.

5. **Ages of Presidents** The normal quantile plot represents the ages of presidents of the United States at the times of their inaugurations. The data are from Data Set 15 “Presidents” in **Appendix B**.



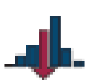
The sample data appears to be from a normally distributed population as the observations follow a linear trend and there is minimal deviation from the theoretical quantiles line

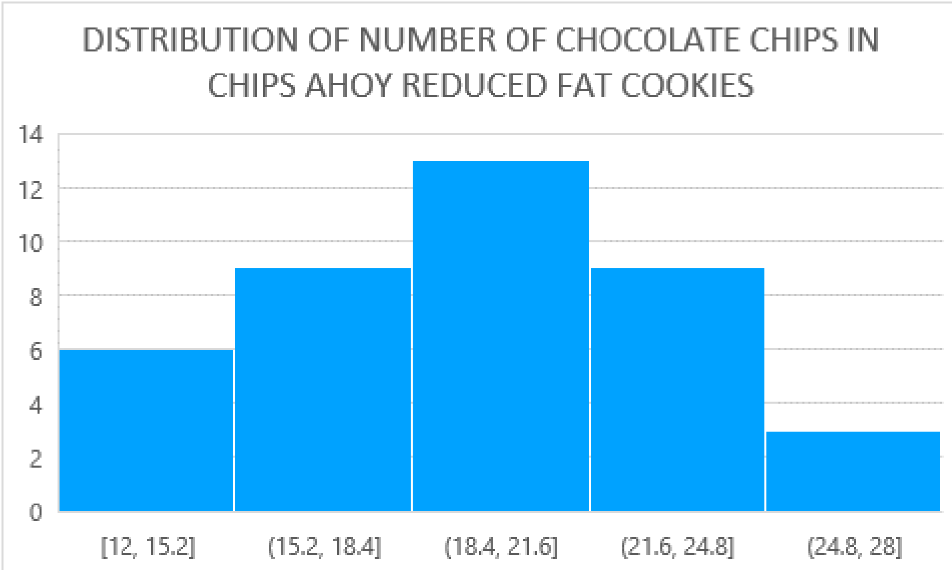
*6. **Diet Pepsi** The normal quantile plot represents weights (pounds) of the contents of cans of Diet Pepsi from Data Set 26 "Cola Weights and Volumes" in **Appendix B**.*



There is a slight curvature in the middle of the normal quantiles plot. However, the curvature is muted and there is minimal deviation from the mean. The population weight of contents of cans of diet Pepsi appears to be approximately normally distributed.

Determining Normality. *In **Exercises 9–12**, refer to the indicated sample data and determine whether they appear to be from a population with a normal distribution. Assume that this requirement is loose in the sense that the population distribution need not be exactly normal, but it must be a distribution that is roughly bell-shaped.*

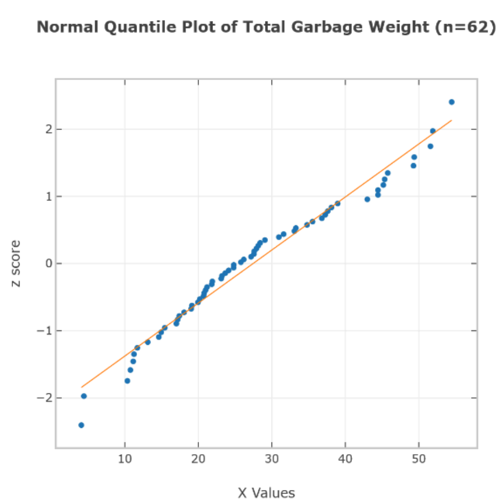
9.  ***cookies** The numbers of chocolate chips in Chips Ahoy (reduced fat) cookies, as listed in Data Set 28 "Chocolate Chip Cookies" in **Appendix B**.*



The distribution of number of chocolate chips in a Chips Ahoy Reduced fat cookie is approximately normal as indicated by the bell-shaped curve with few values at the tails, modal category at the center and the rough symmetricity of the distribution

Using Technology to Generate Normal Quantile Plots. *In* ***Exercises 13–16****, use the data from the indicated exercise in this section. Use software (such as Statdisk, Minitab, Excel, or StatCrunch) or a TI-83/84 Plus calculator to generate a normal quantile plot. Then determine whether the data come from a normally distributed population.*

14, *Exercise 11 "Garbage"*



There is a distinct curvature in the data with deviation from the theoretical quantiles line in the top and bottom corner of the distribution. The data came from a skewed population.