CHAPTER 6

Theoretical Distributions Including The Normal Distribution

Summary

Empirical distributions are frequency distributions of observed scores. *Theoretical distributions* are distributions based on logic or mathematical formulas. The current chapter focuses on finding probabilities by using theoretical distributions and then using those probabilities to answer questions about empirical distributions.

The concept of probability is central to understanding inferential statistics. To find the empirical probability of an event, form a ratio with successes in the numerator and the total of successes and failures in the denominator. To find the theoretical probability of an event, determine the proportion of a theoretical curve that corresponds to the events in question.

Major portions of this chapter are about using theoretical curves to find probabilities of events. With the theoretical curves in this chapter, the probability of an event is equal to the proportion of the curve that corresponds to that event.

Three different theoretical distributions are described – *rectangle distributions* (for example, playing cards), *binomial distributions* (for example, flipping a coin) and the *normal distribution* (lots of measures in the social sciences are normally distributed). Each distribution allows us to know the probability of events (or sets of events).

The normal curve is a bell-shaped distribution (sometimes called the "bell curve"). The unit of measurement on the X axis is the *z score*, which extends from approximately –3 on the left to 0 in the middle to approximately +3 on the right side. The points marked on the curve in between are at *z* scores of –1 and +1.

Chapter 6

Any empirical distribution that is not normally distributed can be converted to a normal distribution using z scores $\left(\frac{X-\mu}{\sigma}\right)$. Using z scores and the table of proportions of the normal curve, you can find:

1. the proportion of the empirical distribution that is beyond a particular score
2. the proportion of the empirical distribution that is between two particular scores
3. a particular score that separates out a proportion of the curve
4. two particular scores that encompass a given proportion of the curve
5. the number of cases with particular scores
6. the number of cases within a particular proportion of the curve

Drawing a picture of a theoretical distribution is highly recommended (even ifthe drawing doesn't quite look like the one in your book). Drawing a curve and labeling the known and unknown will help you understand the problem more completely. Once you understand it, finding an answer involves only a little algebra. Of course, to write an interpretation, it is necessary that you understand the problem.

Multiple-Choice Questions _____

1. The difference between an empirical distribution and a theoretical distribution is that a theoretical distribution
   (1) is based on many more observations;
   (2) is theory and cannot be used;
   (3) is based on mathematics and logic;
   (4) is based solely on observations.

2. Which of the following is an empirical distribution?
   (1) the given names and their frequencies of all high school graduates in the United States for the year 2005;
   (2) the scores expected from an infinite number of throws of one die;
   (3) the normal distribution in Table C in your text;
   (4) all of the above.

3. To use the theoretical normal curve, which of the following things must be known about the population?
   (1) mean;
   (2) standard deviation;
   (3) the form of the distribution;
   (4) all of the above.

4. The term *normal distribution* was adopted because
   (1) the results were found only with normal, healthy individuals;
   (2) Sir Francis Normal was the first to write the equation for the curve;
   (3) results were first applied by teachers who had been trained in teachers' colleges, which, in those days, were called Normal Schools;
   (4) none of the above.

5. The area under the curve of a standard normal distribution is
   (1) dependent on the number of frequencies;
   (2) dependent on the size of the mean;
   (3) 1.00;
   (4) none of the above.

6. The theoretical normal curve has a mean equal to _____ and a standard deviation equal to _____.
   (1) 1.00, 0.00;
   (2) 0.00, 1.00;
   (3) 1.00, 1.00;
   (4) 1.00, the standard deviation of the population.

7. If you were given one z score from a population of measurements and nothing else, you could determine
   (1) the mean of the population;
   (2) the standard deviation of the population;
   (3) both (1) and (2)
   (4) neither (1) nor (2)

8. .4332 of the normal curve lies between μ and 1.5σ. The proportion between μ and $.75\sigma$ is
   (1) .8664
   (2) .2166
   (3) .0668
   (4) not determinable from the information given.

9. If a normal distribution of empirical scores is converted to a distribution of *z* scores,
   (1) the new mean will be zero;
   (2) the new standard deviation will be 1;
   (3) both (1) and (2)
   (4) neither (1) or (2)

10. Which of the following is a theoretical distribution?
   (1) Your statistics professor stayed in the residence hall one Friday night and flipped a coin 10,000 times. The number of heads and tails was recorded.
   (2) The price of every house sold in the last five years in Hampden County was obtained from courthouse records.
   (3) The number of persons who arrived late was recorded every time a statistics course met during the semester.
   (4) None of the above.

11. Bud and Lou were arguing about scores on the Ace Slap-Stick Comedy Test. The scores are distributed normally with a mean of 50. They agreed that 10% of the population had scores of 60 or better (and they were correct on this). Bud also claimed that 10% of the population had scores of 40 or below.
   (1) Bud is correct;
   (2) Bud is correct, but only because each score point is worth one percentage point;
   (3) Bud is mistaken;
   (4) More information is necessary before a decision can be made.

12. Continuing the example of the Ace Slap-Stick Comedy Test, Bud claimed that, because 10% of the population had scores of 60 or better, which is 10 points from the mean, 5% must have had scores of 70 or better, because doubling the score distance always halves the percentage.
    (1) Bud is correct;
    (2) Bud is mistaken;
    (3) More information is necessary before a decision can be made.

13. Suppose the mean of a particular normal distribution is 3.95. The median of this distribution will be
    (1) larger than 3.95;
    (2) smaller than 3.95;
    (3) 3.95;
    (4) not determinable from the information given.

14. Suppose that if K should occur, it will be called a success. If j should occur, it will be called a failure. The ratio $\frac{k}{j}$ is
    (1) the empirical probability of k;
    (2) the empirical probability of j;
    (3) the theoretical probability of k;
    (4) none of the above.

15. Suppose that if k should occur, it will be called a success. If j should occur it will be called a failure. The ratio $\frac{k}{k+j}$ is
    (1) the empirical probability of k;
    (2) the empirical probability of j;
    (3) the theoretical probability of k;
    (4) the theoretical probability of j.

16. Asymptotic means
    (1) the highest point;
    (2) the lowest point;
    (3) approaching a line;
    (4) not possible.

17. The advantage of calculating *z scores* is that
    (1) they allow you to know the sample mean;
    (2) they allow you to know the sample standard deviation;
    (3) they allow you to compare distributions that come from different populations;
    (4) none of the above.

18. If you flip a coin a thousand times and plot the results, the distribution will be
    (1) a normal distribution;
    (2) a rectangular distribution;
    (3) a relative distribution;
    (4) a binomial distribution.

19. Probability is important in statistics because
    (1) it allows us to evaluate the chance that a statistic occurred;
    (2) it allows us to know truth;
    (3) it allows us to make claims based on facts;
    (4) not enough information to answer the question.

20. Suppose a student has a *z score* of –1. That means she
    (1) scored well above the mean;
    (2) had a negative raw score;
    (3) scored 1 standard deviation below the mean;
    (4) had the lowest score in the group.

Short-Answer Questions

1. Distinguish between theoretical and empirical distributions.

2. Write a paragraph describing the normal curve.

3. Suppose a small college has 500 freshman, 400 sophomores, 300 juniors, and 200 seniors and that you randomly pick one student. What is the probability that the student will be
    a. a freshman;
    b. a senior;

c. a sophomore or a junior;
d. not a senior.

4. Your text has some questions about hobbits, mythical creatures in J. R. R. Tolkien's books. Hobbits have furry feet and love to play games. Suppose some practical joker shaved the feet of the hobbits and reduced their height by one inch. What effect would this have on the mean and standard deviation? Drawing two pictures of the distribution of heights "before" and "after" will help you conceptualize this problem.

5. What is the difference between a binomial distribution and a normal curve?

Problems

1. Identify each of the following distributions as theoretical or empirical.
   a. Twelve quarters were tossed in the air 500 times. Each time they landed, the number of heads was recorded.
   b. At Collegiate University, the Registrar recorded the grade point average of every freshman for the years 1975, 1985, 1995, and 2005.
   c. For all Saturdays since the college began, the proportion of rainy days was determined from official weather records.
   d. Each offspring of a single fruit fly was classified as red-eyed or white-eyed.
   e. A normal-shaped distribution was found when 1000 needles from a white pine tree were measured.
   f. When pilots in the Japanese Air Force were weighed, the distribution was positively skewed.

2. Pygmies live in central Africa in a region called Ituri (after the river by that name). Colin Turnbull, who lived with a group of pygmies during the early 1950s, wrote a delightful book called *The Forest People* (1961) telling of his experiences. Turnbull reports that pygmies are less than 4 ½ feet tall. For each of the questions below assume that the height of pygmies is normally distributed with a mean of 4 ft., 3 in. and a standard deviation of 2 inches.
   a. Pygmies live in huts made up of a framework of branches covered with leaves. Materials can be gathered and the hut constructed in an

afternoon. If the opening is 4 feet high, what proportion of the pygmies will have to duck to enter?

b. Suppose only those pygmies who were between 4 ft., 2 in. and 4 ft., 6 in. tall were allowed to sing in a molino ceremony. What proportion would be left out?

c. In the group Turnbull lived with, there was a main group of about 20 families (65 people) plus a subsidiary group of about 4 families (15 people) led by Cephu. If one person from the camp were chosen at random, what is the probability that the person would be from Cephu's subgroup?

3. A classical experiment in extrasensory perception (ESP) consists of asking a participant to tell, without looking, the suit of each card in a deck of Zener cards. There are five suits in the deck so the probability of a chance match between the guess and the card is 1/5 or .20. If only chance is operating, you would expect a participant to get 20 matches in 100 guesses ($.20 \times 100 = 20$). The standard deviation for this mean of 20 is 4.

   a. What is the probability of a participant making 23 or more matches in 100 guesses?

   b. What is the probability of 27 or more matches in 100 guesses?

   c. Suppose a friend of a friend claimed to have ESP and agreed to sit just one time and guess at 100 cards. Suppose she made 36 matches. Calculate the probability of this many, or more, matches if only chance is at work, and carefully write a conclusion.

4. R. B. D'Agnostino (1973) wrote a delightful article about the weight of a 40-pound box of bananas. The problem facing a banana shipper is a rule that a 40-pound box must weigh *at least* 40 pounds upon arrival. Suppose a shipper knows from past experience that, when boxes are packed to have 40 pounds, the standard deviation is four ounces and that on the average a box loses eight ounces in transit.

   a. What mean weight should the shipper establish so that only one-fourth of 1% (.0025) of the boxes will arrive with less than 40 pounds?

   b. Suppose the shipper adopts the weight you established and ships out 5 million boxes. How many will arrive with less than 40 pounds of bananas?