TxCETP Course Component

Probability and Statistics

Written by
Ellen Moreland
Angelo State University

Table of Contents

This material is based on work supported by the National Science Foundation under Grant No. DUE 9987332. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation.

10/12/03

This material is based on work supported by the National Science Foundation under Grant No. DUE 9987332.

I. Introduction

Objectives: The teacher understands:

- How to use appropriate graphical and numerical techniques to explore data, characterize patterns, and describe departures from patterns.
- Concepts and applications of probability.
- The relationships among probability theory, sampling, and statistical inference, and how statistical inference is used in making and evaluating predictions.

Prerequisites: Algebra I, Geometry, Algebra II, Pre-Calculus

Materials: Graphing calculator
Excel

Game Plan: This is intended primarily for use in capstone courses for pre-service mathematics teachers.

Resources:

Ellen Moreland
Angelo State University
Ellen.Moreland@angelo.edu

This material is based on work supported by the National Science Foundation under Grant No. DUE 9987332.

II. Activities

1. Discuss the difference between qualitative and quantitative data and give an example of each.

2. Discuss the difference between discrete and continuous data and give an example of each.

3. Discuss the difference between nominal, ordinal, interval, and ratio measurement scales and give an example using each.

4. Each student in the class will answer the following questions. The data will be collected, organized, and displayed for all students in the class:
   Gender: Male or Female
   Classification: freshman, sophomore, junior, senior
   Height in inches to the nearest half inch
   Shoe size

   a. Using the collected data, create the following tables:

      Table 1: Breakdown of the gender of students by percentages

      Table 2: Frequency for the classification of students

      Table 3: Frequency distribution (single value grouping) for the shoe sizes of the males

      Table 4: Frequency distribution (single value grouping) for the shoe sizes of the females

      Table 5: Contingency table cross classifying shoe size and gender

      Table 6: Frequency and relative frequency distribution for heights of female students (single value grouping)

      Table 7: Frequency and relative frequency distribution for heights of male students (single value grouping)

      Table 8: Frequency and relative frequency distribution for heights of all students (use the following categories: 58" and under 60", 60"and under 62", 62" and under 64", etc.)

      Table 9: A contingency table cross classifying height (to nearest inch) of female students and their shoe sizes

   b. Create the following:

      Chart 1: A pie chart labeled *Gender of Students*

This material is based on work supported by the National Science Foundation under Grant No. DUE 9987332.

Chart 2: A bar graph labeled *Classification of Students*

Chart 3: A double bar graph labeled *Shoe Sizes of Students by Gender*

Chart 4: A frequency histogram labeled *Heights of All Students in the Class* (use table 8)

Chart 5: A relative frequency histogram labeled *Heights of All Students in the Class* (use table 8)

Chart 6: A stem and leaf plot labeled *Heights of Female Students in the Class*

Chart 7: A stem and leaf plot labeled *Heights of Male Students in the Class*

Chart 8: A box and whisker plot labeled *Heights of Female Students in the Class* (to the side of the chart, numerically identify minimum, first quartile, second quartile, third quartile, maximum, and interquartile range)

Chart 9: Assume two new students entered the class. The students are both female and one is 5'4" while the other is 6'8" tall. Create a modified box plot using all the females in the original class plus these two new students. To the side of the chart identify numerically minimum, first quartile, second quartile, third quartile, maximum, interquartile range, the numerical lower limit and upper limit, and identity the outliers.

Chart 10: A scatter plot labeled *Height of Female Students versus Shoe Size of Female Students* using the data for the original class

c. Using the data collected from the class, calculate the mean, median, and mode of the heights of the female students in the class. Discuss the skewness of the data. In your discussion, make reference to Chart 6 and to the relative positioning of the mean and median.

d. Using the data collected from the class, calculate the mean, median, and mode of the heights of the male students in the class. Discuss the skewness of the data. In your discussion, make reference to Chart 7 and to the relative positioning of the mean and median.

e. If a data distribution is left skewed, then the mean (<, =, or >) the median.

f. If a data distribution is right skewed, then the mean (<, =, or >) the median.

g. If a data distribution is symmetric, then the mean (<, =, or >) the median.

This material is based on work supported by the National Science Foundation under Grant No. DUE 9987332.

h. Explain how extreme observations affect the mean and the median. Which measure is a better measure of central tendency when extremes exist? Which measure is resistant to extreme values? Give an example of a situation when the median would be a better measure of central tendency than the mean. Explain your choice.

i. Calculate the range, variance, and standard deviation of the heights of the females in the class. Also calculate the range, variance, and standard deviation of the males in the class. Which distribution is more dispersed (larger spread) and how did you determine this? Explain why the range is not necessarily a good measure of spread.

5. Go to http://nces.ed.gov//programs/coe/2002/pdf/10_2002.pdf and download and print a copy (10 pages). This is an excerpt from the annual report on the Condition of Education published by the National Center for Education Statistics.

a. Using table 10-3 and disregarding the "other jurisdiction" category, calculate the mean and standard deviation of the average scale score in 2000 for 4^{th} graders in the 40 states included in this study.

b. Using table 10-3 and disregarding the "other jurisdiction" category, calculate the mean and standard deviation of the average scale score in 2000 for 8^{th} graders in the 40 states included in this study.

c. A z-score measures the number of standard deviations a data value is away from the mean in a data set. Calculate a z-score for Texas for the 4^{th} grade. Explain how Texas' 4^{th} graders compare to the nations' 4^{th} graders.

d. Calculate a z-score for Texas' for the 8^{th} grade. Explain how Texas' 8^{th} graders compare to the nation's 8^{th} graders.

e. Using the z-scores, make a comparison between Texas' 4^{th} graders and Texas' 8^{th} graders. Explain why the z-score is helpful in comparing these two groups of students.

f. Calculate the percentile rank of Texas' 4^{th} graders.

g. Calculate the percentile rank of Texas' 8^{th} graders.

h. What average scale score in 2000 would represent the 75^{th} percentile for 4^{th} graders?

i. If data is normally distributed, find the percentile rank of a piece of data that is one standard deviation below the mean.

j. If data is normally distributed, find the percentile rank of a piece of data that is two standard deviations below the mean.

This material is based on work supported by the National Science Foundation under Grant No. DUE 9987332.

k. If data is normally distributed, find the percentile rank of a piece of data that is three standard deviations below the mean.

6. Crunchy Cocoa Cereal is running a promotion on its cereal. The company knows that the way to get the parents to purchase the cereal is to get the children to want it. They decide to place a toy in each box of cereal. There are 6 different toys to be used in the promotion. The company randomly places the toys in boxes with the expectation that the children will keep pestering their parents to buy the cereal until they have obtained all 6 toys.

   a. About how many boxes of cereal would a family expect to have to buy to get the complete collection of toys? Use the toss of a die or a random number generator to simulate this situation. For each shopping trip, repeat the random number selection until all digits 1-6 have been obtained at least once. Record the number of cereal boxes purchased to collect the complete set of toys. Now, complete 100 of these shopping trips and calculate the mean number of boxes purchased in order to complete the set of toys. A random number generator is easy to use on a calculator, simply use *randInt (1, 6,100).*

      Make a chart similar to the one below. Use tally marks each time a number comes up in the shopping trip until all 6 numbers have appeared at least once.

Shopping Trip	1	2	3	4	5	6	# boxes
1							
2							
3							
4							
.							
.							
.							
100							

7. Describe the difference between classical, empirical, and subjective probability. Give an example of each.

8. Suppose we want to determine the probability that on the toss of a single die, the result will be at least 5.

   a. Set up a sample space and use it to calculate the probability. Is this classical, empirical, or subjective probability?

   b. Simulate an experiment using a sample of 100 tosses. Using the data from your experiment, calculate the probability that on the toss of a single die, the result will be at least a 5. Is this classical, empirical, or subjective probability?

This material is based on work supported by the National Science Foundation under Grant No. DUE 9987332.

9. Use the data for Grade 8 from table 10-3 of the Condition of Education Report that you previously printed to answer the following questions.

   a. Fill in the following contingency table.

Number of States out of the 40 Participating States-Grade 8
Change from 1990

Average Scale Score 2000	0 and under 5	5 and under 10	10 and under 15	15 and under 20
210 and under 215				
215 and under 220				
220 and under 225				
225 and under 230				
230 and under 235				
235 and under 240				

   b. A state is chosen at random from these 40 participating states. Find the probability that the state has:

      i. An average scale score of at least 225
      ii. An average scale score of at most 220
      iii. A change in scale score from 1990 of more than 10
      iv. A change in scale score from 1990 of less than 5
      v. A change in scale score from 1990 of less than 10 given an average scale score of at least 220.
      vi. An average scale score of at least 220 or a change in scale score from 1990 of at least 10
      vii. An average scale score of less than 220 or a change in scale score from 1990 of less than 10
      viii. An average scale score of at least 230 or a change in scale score from 1990 of at least 15

10. Use the data for grade 12 in 2000 from table 10-1 of the Condition of Education Report that you previously printed to answer the following questions.

   a. Draw a Venn Diagram to illustrate this situation. Use circles for all shapes in the diagram. Let the area of the smallest circle be 2 square mm. Draw to scale.

   b. Find the percentage of students that scored at the proficient level.

   c. Find the probability that a student scored at the advanced level given that he/she scored at or above the proficient level.

   d. Find the probability that a student scored at the advanced level or below the basic level.

This material is based on work supported by the National Science Foundation under Grant No. DUE 9987332.

e. Find the probability that the student scored at the advanced level or below the proficient level given that he/she scored at or above the basic level.

f. Calculate the ratio of the area of the circle representing the advanced level to the area of the circle representing at or above the proficient level. Relate this to the probability calculated in part c. Explain.

11. A dart board consists of three concentric circles. The larger circle has a radius of 25 cm, the middle circle has a radius of 15 cm, and the inner circle has a radius of 10 cm. Find the probability that a dart which lands on the board

    a. lands in the inner circle

    b. lands in the middle ring

12. A school bus is scheduled to arrive at the bus stop at 7:30 a.m. Assume arrival time is uniformly distributed between 7:25 and 7:40 a.m. What is the probability that the bus arrives early on any given day?

13. A student takes a ten question multiple choice quizzes. Each question has five possible answers of which only one is correct. The student has slept through class everyday while this material was being covered and has absolutely no knowledge of the subject matter. What is the probability that he will pass the quiz with a 70 or better given he guesses on all questions?

14. The scores on a standardized exam are normally distributed with a mean of 500 and a standard deviation of 100. Using the 68.26-95.44-99.74 Rule, find the probability that a student selected at random scored

    a. at least 600 on this exam

    b. between 300 and 500 on this exam

    c. either less than 300 or more than 700 on this exam

15. The Central Limit Theorem states that for a relatively large sample size (at least 30) the sampling distribution of the mean is approximately normally distributed, regardless of the distribution of the variable under consideration and $u_{\bar{x}} = u, s_{\bar{x}} = \frac{s}{\sqrt{n}}$. Now, let x denote scores of students on a particular math test. The mean of all scores is 75 while the standard deviation of all scores is 10. Suppose a random sample of 100 scores is to be selected.

    a. What is the distribution of $\bar{x}$?

    b. Find the mean and standard deviation of $\bar{x}$.

This material is based on work supported by the National Science Foundation under Grant No. DUE 9987332.

c. Using the 68.26-95.44-99.74 rule, find the approximate probability that $\bar{x}$ is between 65 and 95 inclusive.

16. Homework scores for a random sample of ten students from a class of thirty are given below. Find a point estimate for the mean and the standard deviation of the homework scores for all students.

    Scores: 90, 60, 80, 85, 75, 50, 100, 85, 95, 90

17. You wish to estimate the mean score of all students on a standardized exam. You use the following random sample of 30 grades obtained from the population of students taking this test.

    a. Find a point estimate for the population mean.

    b. Find a 95.44% confidence interval for the mean.

    c. Find a 99.74% confidence interval for the mean.

    d. What happens to the length of the confidence interval as the confidence level increases and the sample size remains constant?

    e. What could you do if you wished to increase the confidence level without increasing the length of the interval?

18. A teacher wishes to determine if a new method of instruction has resulted in improved student performance on a particular exam. The old method produced an average exam score of 65. The results for 30 randomly selected students taught using the new method are as follows:

    30, 40, 45, 50, 60, 60, 60, 65, 65, 65, 65, 65, 65, 70, 70,
    70, 70, 70, 70, 75, 75, 75, 80, 80, 80, 80, 80, 85, 90, 100

    At the 2.28% significance level, does the data suggest that the new method produces higher test scores?

19. A high school class plans to raise money for their mathematics club end of year trip by having a raffle. They will sell 500 raffle tickets. They will give one first prize of $50, two second prizes of $25 each, and five third prizes of $5 each. What are the club's expected earnings?

This material is based on work supported by the National Science Foundation under Grant No. DUE 9987332.

20. Divide the class into groups with each group containing two students. Each group will cut out twenty 4" by 4" squares using colored construction paper. Cut out 4 red, 4 blue, 4 green, 4 yellow, and 4 brown squares. One student in the group will be designated as "tester" while the other is "student." The tester will lay the twenty squares out on a surface in random order and out of the view of the student. The student will be given 10 seconds to look at the pattern. The tester will then pick up the squares. The student will attempt to repeat this pattern. The score will be the number of correct responses prior to the first error in repeating this pattern. This process will be repeated allowing the same student 20 seconds to repeat a new pattern. Repeat again giving the student 30 seconds, 40 seconds, and 50 seconds, 60 seconds and a new pattern each time.

    a. Complete the following table:

Number of Seconds	Score
10	
20	
30	
40	
50	
60	

    b. Draw a scatter plot for this data.

    c. Find the least squares regression line.

    d. Find the coefficient of correlation and explain what it tells you.

    e. Explain why it is important that the same student remains the "student" throughout this process.