Math III Unit 1: STATISTICS

Carole Barber, FVHS
Lauren Winstead, Heritage High School

Main topics of instruction:
1) Normal Distributions
2) Sampling and Study Design
3) Estimating Population Parameters
4) Expected Value and Fair Game (honors only)

Day 1: NORMAL DISTRIBUTIONS

low std. dev. = tight data
high std. dev. = loose data
ask: in this room, what data would be close together? (age) far apart? (shoes in your closet)

First, a reminder…

What is standard deviation (σ)?

A measure of the amount of variation from the mean.

Characteristics of a normal distribution:

1) unimodal - only one mode (or peak)
2) symmetric - equal amt. of data on each side
3) asymptotic - approaches but never touches x-axis (data is infinite)
4) mean, median, and mode are all equal.
(average, middle, most frequent)

The Standard Normal Distribution: has a mean of 0 and std. dev. of 1

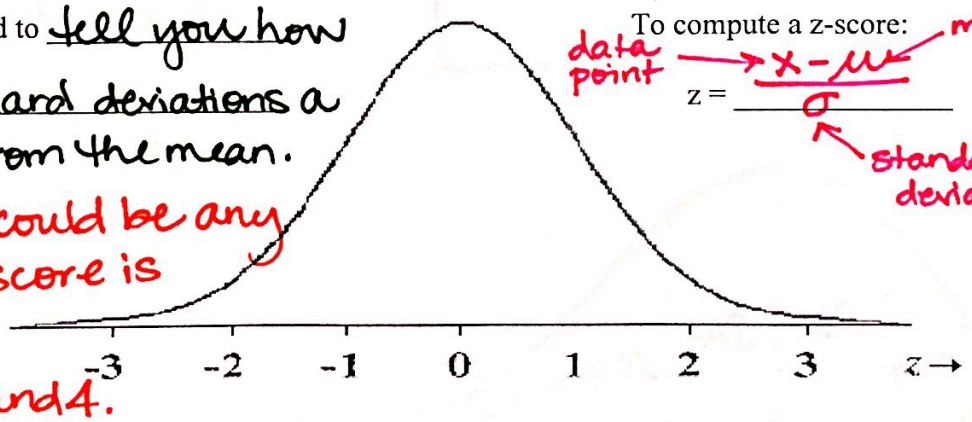
This is important because: it's used to handle problems involving any normal distribution.

A z-score is used to tell you how many standard deviations a number is from the mean.

To compute a z-score:

$$z = \frac{x - \mu}{\sigma}$$

data point → x; mean → μ; standard deviation → σ



* Std.-dev. could be any number. z-score is usually between 0 and 4.

Example 1: The mean score on the SAT is 1500, with a standard deviation of 240. The ACT, a different college entrance examination, has a mean score of 21 with a standard deviation of 6. If Bobby scored 1740 on the SAT and Kathy scored 30 on the ACT, who scored higher?

Use z-scores to compare!

Bobby: $z = \frac{1740 - 1500}{240} = 1$

Kathy: $z = \frac{30 - 21}{6} = 1.5$

Kathy scored higher

You try! Sarah enters a three-point shooting contest at her school. On average, students in her school can shoot 10 three-pointers in one minute with a standard deviation of 2. If Sarah's z-score is -1.4, approximately how many three-pointers can Sarah make in one minute?

$-1.4 = \frac{x-10}{2}$ $\quad -2.8 = x - 10$ $\quad \boxed{7.2x}$

Sarah can shoot approximately 7 three-pointers in one minute.

The Empirical Rule: If data is normally distributed, then

68% of the data will fall between one standard deviation of the mean
95% of the data will fall between two standard deviations of the mean
99.7% of the data will fall between three standard deviations of the mean

This means that, because the area under the curve is 1, there is a 68% probability that an individual data point will fall between $\pm 1\sigma$.

How to use the Empirical Rule:

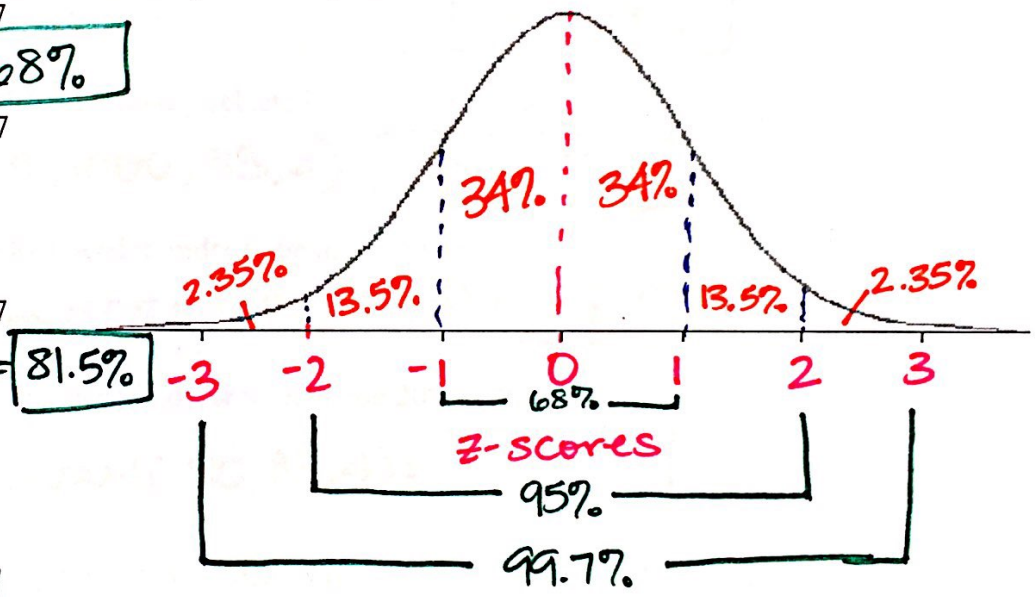
Example 2: The scores on the Math 3 midterm were normally distributed. The mean is 82 with a standard deviation of 5. Find the probability that a randomly selected person:

a. scored between 77 and 87

$34\% + 34\% = \boxed{68\%}$

b. scored between 82 and 87

$\boxed{34\%}$



Try these on your own!

c. scored between 72 and 87

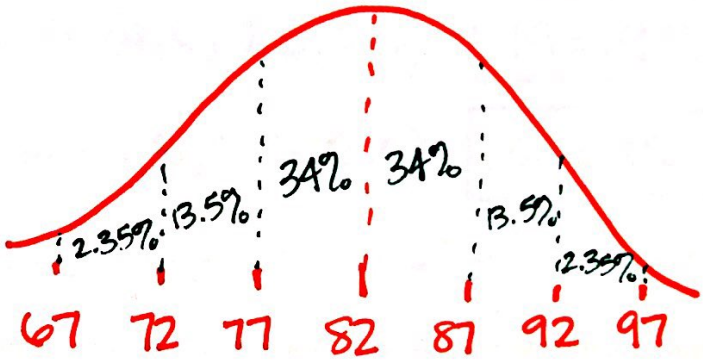
$13.5\% + 34\% + 34\% = \boxed{81.5\%}$

d. scored higher than 92

$\frac{100\% - 95\%}{2} = \boxed{2.5\%}$

e. scored less than 77

$\frac{100\% - 68\%}{2} = \boxed{16\%}$



Day 2: FINDING PROBABILITIES ON THE NORMAL CURVE USING TECHNOLOGY

If you are looking for values that are not exactly on the standard deviation lines, there are two ways to find the probability – you can use a normal distribution chart OR you can use a calculator. For the purpose of this course, we will use a calculator exclusively.

data to percentage/probability

On the TI-84: $\mu = 83, \sigma = 4$

To find the probability: 2nd - DISTR, NORMALCDF - normalcdf (lower bound, upper bound, μ, σ)

To find the boundary: 2nd – DISTR, INVNORM – invnorm(area to the left of your boundary, μ, σ)

percentage/prob to datapt.

Example 1: What's the probability that a randomly selected student scored between 80 and 90?

normalcdf (80,90,83,4) = .73 = 73%

Example 2: What's the probability that a randomly selected student scored below 70?

normalcdf (0,70,83,4) = .00058 = .058%

Example 3: What's the probability that a randomly selected student scored above 79?

normalcdf (79,1000,83,4) = .84 = 84%

Example 4: What score would a student need in order to be in the 90th percentile?

invNORM(.90, 83, 4) = 88.13

Example 5: What score would a student need in order to be in top 20% of the class?

invNORM(.80, 83, 4) = 86.37

Now try this one: The average waiting time at Walgreen's drive-through window is 7.6 minutes, with a standard deviation of 2.6 minutes. When a customer arrives at Walgreen's, find the probability that he will have to wait

a) between 4 and 6 minutes — normalcdf (4,6,7.6,2.6) = .19 = 19%

b) less than 3 minutes — normalcdf (0,3,7.6,2.6) = .04 = 4%

c) more than 8 minutes — normalcdf (8,1000,7.6,2.6) = .44 = 44%

d) If only 8% of customers wait longer than Mrs. Sickalot, find how long Mrs. Sickalot has to wait.

invNORM(.9992, 7.6, 2.6)

= 15.8 min.

ay 3: SAMPLING AND STUDY DESIGN

There are three ways to collect data:

1. Observational Studies
2. Experiments
3. Surveys

was a treatment given? → yes: experiment; no: obs. study

Experiment or Observational Study?

1. Fifty people with clinical depression were divided into two groups. Over a 6 month period, one group was given a traditional treatment for depression while the other group was given a new drug. The people were evaluated at the end of the period to determine whether their depression had improved.

experiment

2. To determine whether or not apples really do keep the doctor away, forty patients at a doctor's office were asked to report how often they came to the doctor and the number of apples they had eaten recently.

observational study

3. To determine whether music really helped students' scores on a test, a teacher who taught two U. S. History classes played classical music during testing for one class and played no music during testing for the other class.

experiment

Types of Sampling:

1. **Simple Random Sample:** all individuals in the population have the same probability of being selected.
2. **Stratified Random Sample:** the researcher divides the target pop. into different subgroups, then randomly selects a few from each group.
3. **Systematic Random Sample:** the researcher selects a number at random and then selects every n^{th} individual for the study.
4. **Convenience Sample:** subjects are taken from a group that is conveniently accessible to a researcher. (ex. first 100 people to enter a movie theater)
5. **Cluster Sample:** a sampling technique where the entire population is divided into groups, and a random sample of clusters are selected. ALL individuals in the cluster are surveyed.

Name that sample:

1. The names of 70 contestants are written on 70 cards, the cards are placed in a bag, the bag is shaken, and three names are picked from the bag.

simple random sample

2. To avoid working late, the quality control manager inspects the last 10 items produced that day.

convenience sample

3. A researcher for an airline interviews all of the passengers on five randomly selected flights.

cluster sample

4. A researcher randomly selects and interviews fifty male and fifty female teachers.

stratified random sample

5. Every fifth person boarding a plane is searched thoroughly.

systematic random sample

Types of Bias:

1. Question wording bias – wording is confusing or misleading
2. Undercoverage bias – the pop. is not accurately represented in the sample
3. Response bias – people give false or misleading answers to questions
4. Nonresponse bias – people refuse to respond to survey
5. Voluntary response bias – people are asked to call in or respond to a survey by mail.

On the twelfth anniversary of the death of Elvis Presley, a Dallas record company sponsored a national call-in survey. Listeners of over 1000 radio stations were asked to call a 1-900 number (at a charge of $2.50) to voice an opinion concerning whether or not Elvis was really dead. It turned out that 56% of the callers felt Elvis was alive.

voluntary response bias

In 1936, Literary Digest magazine conducted the most extensive (to that date) public opinion poll in history. They mailed out questionnaires to over 10 million people whose names and addresses they had obtained from telephone books and vehicle registration lists. More than 2.4 million people responded, with 57% indicating that they would vote for Republican Alf Landon in the upcoming Presidential election. Incumbent Democrat Franklin Roosevelt won the election, carrying 63% of the popular vote.

undercoverage

For the following survey questions, identify the bias and rewrite the question so that it is free of bias.

1. *Do you think the city should risk an increase in pollution by allowing expansion of the Northern Industrial Park?*

Do you think the city should allow expansion of the Northern Industrial Park?

2. *If you found a wallet with $100 in it on the street, would you do the honest thing and return it to the person or would you keep it?*

If you found a wallet, would you return it to the owner or keep it?

Day 4: ESTIMATING POPULATION PARAMETERS: SAMPLES AND PARAMETERS DEFINED

Parameter – a value that represents a population

Statistic – a value that is taken from a sample and used to estimate a parameter.

the symbols change as you move from a statistic to a parameter – students need to know the names of these greek symbols:

	Statistic	Parameter
Mean	$\bar{x}$	μ
Standard Deviation	s	σ
Proportion	$\hat{p}$	p

Make sure your students understand the difference between a population and a sample!

Take 2 minutes to answer the following questions:

* count up correct answers

1. What state is called the Sioux State? North Dakota
2. What is the capital of Alabama? Montgomery
3. Who was the youngest US President to die in office? John F. Kennedy
4. What is the postal abbreviation for Alaska? AK
5. What type of animal was Babe in the film of the same name? Pig
6. Which city has an area called Haight-Ashbury? San Francisco
7. Which city is the home of jazz? New Orleans

Assuming that this group is a population, find μ: 2.25

From a sample of 5, find $\bar{x}$: 3

nding a Margin of Error for a sample proportion:

Margin of error – "cushion" around a statistic

$$ME = \frac{1}{\sqrt{n}}, \text{where } n \text{ is the sample size}$$

larger sample size = smaller margin of error

Example 1: Suppose that 900 American teens were surveyed about their favorite event of the Winter Olympics. Ski jumping was the favorite for 20% of those surveyed. This result can be used to predict the true interval of the proportion of American teens who favor ski jumping.

$ME = \frac{1}{\sqrt{900}} = \frac{1}{30} = .03$

true interval = $20\% \pm 3\% = [17\%, 23\%]$

You try! 200 Heritage High School students were surveyed about their favorite event of the year, and 15% of them answered that Prom was their favorite. Predict the true interval of the percentage of students whose favorite event is Prom.

$ME = \frac{1}{\sqrt{200}} = .07 = 7\%$

$15\% \pm 7\% = [8\%, 22\%]$

Example 2: If your sample size is 400 and you wish to cut the margin of error in half, what will your new sample size be?

$ME = \frac{1}{\sqrt{400}} = \frac{1}{20} = .05 = 5\%$

half of $\frac{1}{20} \rightarrow \frac{1}{20} \cdot \frac{1}{2} = \frac{1}{40}$ $\quad \frac{1}{40} = \frac{1}{\sqrt{n}} \rightarrow \sqrt{n} = 40$

$n = 1600$

You try! a) You want to cut your margin of error from Example 1 by 75%. What will your new sample size be?

$\frac{1}{30} \times \frac{1}{4} = \frac{1}{120}$ $\quad \frac{1}{120} = \frac{1}{\sqrt{n}} \rightarrow \sqrt{n} = 120$

$n = 14{,}400$

b) If you want your margin of error to be 5%, what size sample will you need?

$.05 = \frac{1}{20} = \frac{1}{\sqrt{n}}$

$20 = \sqrt{n}$

$n = 400$

ay 5: ESTIMATING POPULATION PARAMETERS: SIMULATIONS

Simulations: Ways to model random events, such that simulated outcomes closely match real-world outcomes.

Example 1: Sam decides to order pizza for his Student Government meeting. There are six toppings to choose from: pepperoni, mushrooms, sausage, peppers, olives, and onions. If he can only order pizzas with two or fewer toppings and he chooses his toppings randomly, how many pizzas will he have to buy in order to include every topping?

Let's set up our simulation:

1: pepperoni
2: mushrooms
3: sausage
4: peppers
5: olives
6: onions
7-9: no 2nd topping

Random Digit Table:

8 pizzas

101	19223	95034	05756	28713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056
104	52711	38889	93074	60227	40011	85848	48767	52573
105	95592	94007	69971	91481	60779	53791	17297	59335
106	68417	35013	15529	72765	85089	57067	50211	47487

What are the results of these trials? You must order 8 pizzas in order to receive every topping.

What predictions can be made based on these results? On average, 8 pizzas will need to be ordered in order to get every topping if toppings are chosen randomly.

Example 2: On any given day, Goodberry's has three flavors available: chocolate, vanilla, and the Flavor of the Day. Goodberry's claims that, on a normal day, 52% of people choose vanilla, 33% choose chocolate, and 15% choose the Flavor of the Day. If you run a simulation of the first 20 customers to arrive at Goodberry's on a random day, can you validate these claims? (start on line 102)

00-51: vanilla
52-84: chocolate
85-99: Flavor of the Day

vanilla	chocolate	FOTD
𝍸 I	𝍸 III	II
10/20 = 50%	8/20 = 40%	2/20 = 10%

You try! On a certain day the blood bank needs 4 donors with type O blood. How many donors, on average, would they have to see to get exactly four donors with type O blood, assuming that 45% of the population has type O blood? Use random digit table: MATH → PRB → randInt(0,99)

0-44: Type O
45-99: non-Type O

51, 40, 73, 4, 33, 99, 20

They will need to see 7 donors before they will have 4 donations of Type O blood, on average.

Day 6: EXPECTED VALUE AND FAIR GAMES

Expected value: the weighted average of all possible values that the variable can take on

Probability distribution: all of the values that the variable takes on and their respective probabilities.

The expected value, E(x), of a distribution is the mean of that distribution, μ.

Using a TI-84, put the x values in L1 and the probabilities of each value in L2.
go to STAT, 1 VAR STAT – then enter L1, L2 before pressing ENTER. The expected value is $\bar{x}$.

The Law of Large Numbers states that in repeated, independent trials with the same probability p of success in each trial, the experimental probability will approach the theoretical probability as the number of trials n goes to infinity.

Example 1: At Tucson Raceway Park, your horse, Secretariat, has a probability of 1/20 of coming in first place, a probability of 1/10 of coming in second place, and a probability of ¼ of coming in third place. First place pays $4,500 to the winner, second place $3,500 and third place $1,500. Is it worthwhile to enter the race if it costs $1,000?

.	1st	2nd	3rd	other
x	$3,500	$2,500	$500	-$1,000
P(x)	0.05	0.10	0.25	0.60

$3500(.05)+2500(.10)+500(.25)$
$+(-1000)(.60)$
$=-\$50$

Nobody actually loses $50, but on average, the race track earns $50 for every horse that races. This is your average loss.

In a fair game, the expected value is 0

You try! You play a game in which you roll one fair die. If you roll a 6 on the first roll, you win $5. If you roll a 1 or a 2, you win $2. If you roll anything else, you lose. Create a probability model for this game:

x	6	1, 2	3, 4, 5
payout	$5	$2	$0
P(x)	1/6	1/3	1/2

What would you be willing to pay to play this game? cost = x

$$\left(\tfrac{1}{6}\right)(\$5-x)+\left(\tfrac{1}{3}\right)(\$2-x)+\left(\tfrac{1}{2}\right)(\$0-x)=0$$

$$\tfrac{5}{6}-\tfrac{1}{6}x+\tfrac{2}{3}-\tfrac{1}{3}x+0-\tfrac{1}{2}x=0$$

$$\tfrac{3}{2}-x=0$$

$$\tfrac{3}{2}=x$$

$$\$1.50=x$$

Challenge Problem!

Chris and Jack are playing a board game with a fair coin and a fair number cube (die) numbered 1 to 6. They take turns tossing the coin and die and then calculate their scores.
Scoring is as follows:

- If the coin lands on heads, the score is twice the number on the die.
- If the coin lands on tails, the score is two more than the number on the die.

If the score is a prime number, Chris moves two squares on a game board.
If the score is not a prime number, Jack moves one square on a game board.

Is this a fair game? If not, how could you make it fair?

Heads

2, 4, 6, 8, 10, 12

↑ prime

Tails

3, 4, 5, 6, 7, 8

↑ prime ↑ prime ↑ prime

$\frac{4}{12} = \frac{1}{3}$ are prime

Chris's perspective: $\frac{1}{3}(2) + \frac{2}{3}(-1) = 0$

FAIR!

Jack's perspective: $\frac{2}{3}(1) + \frac{1}{3}(-2) = 0$

FAIR!