Chapter 2

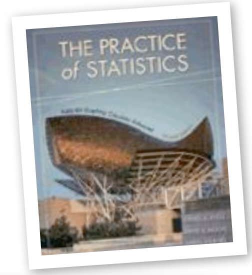
The Normal Distributions

In Chapter 1, we built a 'toolbox' of graphical and numerical tools for describing distributions of data. We now have a clear strategy for exploring data on a single variable. In this chapter, we'll discover that sometimes the overall pattern of a large number of observations is so regular it can be described by a smooth curve. Further, many distributions can be modeled by what we'll refer to as a "normal" distribution.

Normal Distributions:

- Density Curves
- Standard Normal Calculations

AP Stats Chapter 2: The Normal Distributions



"Thou, nature, art my goddess; to thy laws my services are bound..." ~ Shakespeare's King Lear {Gauss' motto}

Tentative Lesson Guide					
Date		**Stats**	**Lesson**	**Assignment**	**Done**
Mon	9/25	2.1	Normal Dist Activity	Read Intro to Ch2	
Tues	9/26	2.1	Density Curves	Rd 78-83 Do 1-4	
Wed	9/27	2.1	Normal Distributions	Rd 85-90 Do 6-9,11-15	
Thu	9/28	2.2	Standardizing, Normal Curves	Rd 93-101 Do 19-24	
Fri	9/29	2.2	z-score Calculations	Rd 101-109 Do 26, 28-34	
Mon	10/2	**Quiz**	**Review/Quiz**		
Tues	10/3	2.2	Assessing Normality	Normal Dist Practice	
Wed	10/4	Rev	Review Chapter 2	Rd 112 Do 40-42, 44-48	
Thu	10/5	Rev	Review Chapter 2	Chapter 2 Online Quiz	
Fri	10/6	***Woohoo!***	***Homecoming!***		
Mon	10/9	**Exam**	**Chapter 2 Exam**	Homework Due	

Note:
The purpose of this guide is to help you organize your studies for this chapter.
The schedule and assignments may change slightly.

Keep your homework organized and refer to this when you turn in your assignments at the end of the chapter.

Class Website:
Be sure to log on to the class website for notes, worksheets, links to our text companion site, etc.

http://web.mac.com/statsmonkey

Don't forget to take your online quiz!. Be sure to enter my email address correctly!

http://bcs.whfreeman.com/yates2e

My email address is:

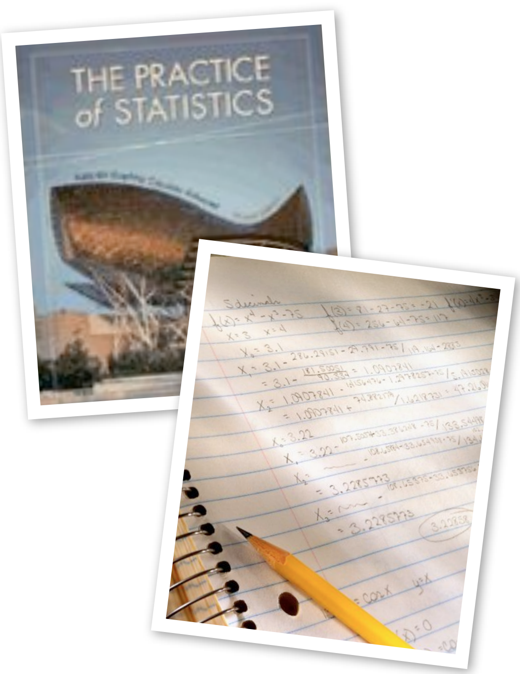
jmmolesky@isd194.k12.mn.us

Chapter 2 Objectives and Skills:

These are the expectations for this chapter. You should be able to answer these questions and perform these tasks accurately and thoroughly. Although this is not an exhaustive review sheet, it gives a good idea of the "big picture" skills that you should have after completing this chapter. The more thoroughly and accurately you can complete these tasks, the better your preparation.

DENSITY CURVES:

- ☐ What's the point of a density curve? How are they different than relative frequency histograms or other visual displays?
- ☐ What are the fundamental properties of a density curve?
- ☐ Given a density curve, can you calculate the probability of a particular event happening?
- ☐ Given a density curve, use symmetry, a bit of math, and problem-solving to find areas under a curve, quartiles, percentiles, medians, etc…
- ☐ Understand how the shape of a density curve indicates the relative positions of the mean, the median and the mode.



NORMAL DISTRIBUTIONS:

- ☐ What are the fundamental properties of a normal density curve?
- ☐ Give examples of variable which would have a normal distributions. Give an example of some variables which have non-normal distributions.
- ☐ What are z-scores? Explain them to a person who knows just a little bit of statistics. Why are they used?
- ☐ Explain the meaning of the 68-95-99.7 rule, and use it to estimate the probability of events coming from a normal distribution.
- ☐ What is the standard normal distribution?
- ☐ Give a clear, well detailed, and accurate probability calculation for problems that require the use of normal distributions. A clear, error-free path to a final answer is expected.
- ☐ Be able to calculate percentiles for normally distributed data. Again, a clear path to a final answer is expected.

ASSESSING NORMALITY:

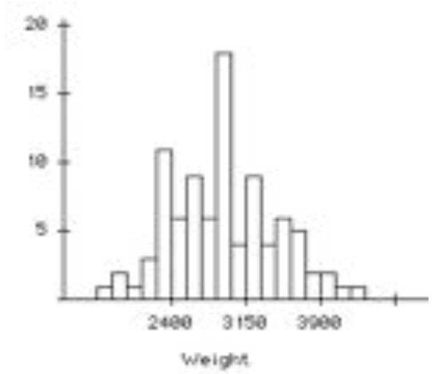
- ☐ Given a set of data, judge whether you think that they are normally distributed. You should be able to do this in at least two different ways (what are the two ways?)
- ☐ Use the TI-83 to quickly and easily calculate probabilities and percentiles from normal distributions.
- ☐ Use normal distribution tables in your textbook to perform the same skills.

2.1: Density Curves

Density Curve:

...From Histograms to Density Curves...

Here is a display of the weights of a random sample of US cars in 1993.



Describe the *SOCS* of this distribution:

Draw a smooth curve that best "idealizes" the shape of the distribution. Do this twice.

Note:The area under each curve is 1 {ie100%}.

a) Shade in the lower 60% of all car weights on the first smooth curve. Estimate the weight that corresponds to the 60^{th} percentile.

b) Shade in the area under the second curve that shows the cars with weights between 2800 and 3900. Estimate the probability that 2800 < weight < 3900. {ie. P(2800<w<3900)}

Draw a histogram that might result from taking random numbers from 0 to 5 from a UNIFORM distribution. Then draw a density curve (on a new set of axes) that "idealizes" this histogram.

a) If the area under the density curve is 100% = 1, what must the height of the curve be?

b) Find $P(0 < x < 3)$. Re-draw the density curve here, and shade in the area that corresponds to the probability.

c) Find $P(2.6 < X)$. Shade the corresponding area on the density curve.

Density Curve
A **density curve** is a curve that ☑ ☑ ☑

Are These m&m's "Normal" or Just "Plain"?

Standard Normal Calculations

We have observed variability in color distributions from bag to bag of "plain" milk chocolate m&m's. According to the m&m website, 14% of milk chocolate m&m's are yellow. Does that mean we are guaranteed 14% of the candies in each bag will be yellow? Should you be concerned if only 10% are yellow? What if all of them are? At what point would you suspect the advertised proportion? We will discuss each of these questions as we explore standard normal calculations with some sample bags of m&m's.

Background Information:

We know the proportion of yellow m&m's varies from bag to bag. Suppose these proportions follow an approximately normal distribution N(0.14, 0.05). Sketch this distribution below and note 1, 2, and 3 standard deviations above and below the mean. Interpret the Empirical (68-95-99.7) Rule in the context of this situation.

Sample Information:

Our bag of m&m's contained ______ candies. There were ________ yellow m&m's.

The sample proportion of yellow candies for our bag is ______/______ = ________.

Standard Normal Calculation:

Recall, a "z-score" is a value that tells us how many standard deviations above or below the mean a particular observation falls. To find this value, we must subtract the mean from our observation and divide the result by the standard deviation. That is,

$$z = \frac{x - \bar{x}}{s} = \frac{\square - \square}{\square} = \square$$

We can use this z-score to determine what percent of bags of m&m's (of the same size) would have a yellow proportion less than our observed proportion. Sketch two normal distributions for yellow proportions below and note our observed proportion on each curve. Using your z-table, determine the proportion of bags of the same size that would have fewer yellow candies. Shade this area on the first curve. On the second curve, shade and calculate the proportion of bags of the same size that would have more yellow candies.

Suppose we had a second bag of m&m's. We would expect about 14% of the candies in the second bag would be yellow. However, like the first bag, there is a chance that proportion will not equal 0.14 (or the proportion in the first bag, for that matter). Use the proportions from the first bag and from a new bag to determine what percent of bags of m&m's of the same size will have a yellow proportion *between* those two values.

	Bag 1	**Bag 2**
Yellow Proportion		
z-score		
% Bags Below Observed		

Sketch the two observed proportions on the normal distribution N(0.14, 0.05) and note the percent of observations we would expect to see *between* the two observed proportions.

What About Peanut Butter m&m's?

Do peanut butter m&m's follow the same color distribution as milk chocolate m&m's? If so, we would expect about 14% of the candies in a peanut butter m&m bag would be yellow. Would you be surprised if 15% were yellow? What about 20%? 30%? At what point would you suspect the color distribution for peanut butter m&m's may be different?

Open a bag of peanut butter m&m's and note the proportion of yellow: ____/____ =______

Plot this value on the N(0.14,0.05) distribution and calculate the % of observations we'd expect to be *more extreme* than this observation. Based on this %, do you feel you have evidence to suggest the color distribution of peanut butter m&m's may be different? Why or why not?

2.2 Assessing Normality

{Adapted from "Introduction to Statistics and Data Analysis" Peck, Olsen, Devore}

Before applying the empirical rule or calculating z-scores to a set of data, one should always assess the normality of the dataset. If you have been told to assume the data is approximately normal, you are good to go. If not, you'll need to do a little work first.

To determine whether or not a dataset is approximately normally distributed, you could:

- Plot the data and look for a symmetric, mound-shaped distribution (a good start, but not enough)
- Calculate the % of data within 1 std dev, 2 std dev, and 3 std dev of the mean (ugh… waaay too much work)
- Construct a **Normal Quantile (Normal Probability) Plot** and look for __________

Infertility, the inability for couples to naturally conceive affects approximately 6.1 million people in the U.S. Approximately 1/3 of infertility cases are due to some form of male infertility. Consider the following data from "Distribution of Sperm Counts in Suspected Infertile Men" (J. of Reproduction and Fertility (1983):91-96) and determine whether or not the counts are approximately normally distributed. The data are a sample of 40 "log transformed" counts of sperm concentrations (10^6/ml) taken from 1711 subjects.

1.45	1.66	1.54	1.25	1.50	1.37	1.42	1.64	1.49	1.72	1.19	1.58	1.33
1.61	1.49	1.41	1.79	1.29	1.46	1.27	1.42	1.24	1.75	1.32	1.56	1.86
1.32	1.14	1.23	1.45	1.32	1.78	1.33	1.51	1.59	1.41	1.35	1.28	1.69
1.76												

Enter the data in your calculator and construct a histogram using categories 0.1 units wide.

Sketch the Histogram: **Sketch the Normal Quantile Plot:**

Would you say these data are approximately normally distributed? Why or why not?

Calculate the mean and standard deviation of this data by using **1-Var Stat** and sketch a density curve that approximates this distribution. Label it with the appropriate notation.

Normal Calculations Practice

Assuming the distribution of transformed sperm counts is N(________,_______), answer the following. Be sure to include a sketch, z score, and appropriate calculations:

1. What percent of suspected infertile men have a transformed sperm count below 1.14?

2. What percent of suspected infertile men have a transformed count above 1.59?

3. What percent of suspected infertile men have a transformed count between 1.21 and 1.91?

4. Let's say a subject is considered infertile if their transformed count is at or below the 35th percentile. What would the cutoff point be?

5. What values would determine the middle 50% of measurements? That is, calculate the first and third quartiles for this approximately normal distribution.

Review: Normal Calculation Practice

The paper "*The Load-Life Relationship for M50 Bearings with Silicon Nitride Ceramic Balls*" (Lubric. Eng. (1984):153-159) reported the accompanying data on bearing load life (million revs). Provide evidence that we may assume normality and answer the following questions:

106	115	126	146.6	205.2
210	215.3	220.3	229	238.6
240	240.1	278	278.2	289
289.3	367	385.9	392	395

1) Provide graphical evidence that the bearing data is approximately normal.

2) Calculate the one variable descriptive statistics for the bearing data.

The bearing data can be approximated by N(_____,______)

3) Calculate the following: x=bearing load life (million revs) Include a sketch and calculations!

$P(x < 150)$ $P(x > 390)$ $P(145 < x < 260)$

4) Determine the bearing life at the 20th percentile............How about the 75th percentile?

What load lives determine the middle 50% of the distribution?

On the back: The company producing the bearings insists they will not sell a product that has a load life that differs from the mean by ±100 (million revs). What percent of bearings will be 'defective'?