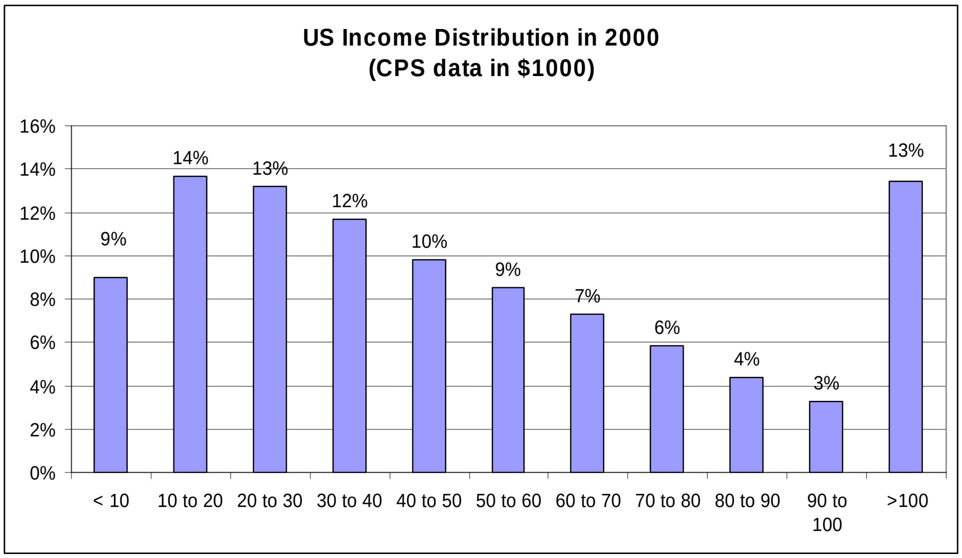
Continuous Random Variables and the Normal Distribution

Continuous Random Variable--A variable with many possible values in all intervals (eg. income, driving time, test scores)

Probability Density Function--Probabilities of values for a continuous variable expressed as a formula for specific types of distributions (eg. uniform, normal, or Chi-square)

Cumulative Probability--Adding up areas of the probability density function to get the probability of getting a range of values.



Standard score ("Z score")-

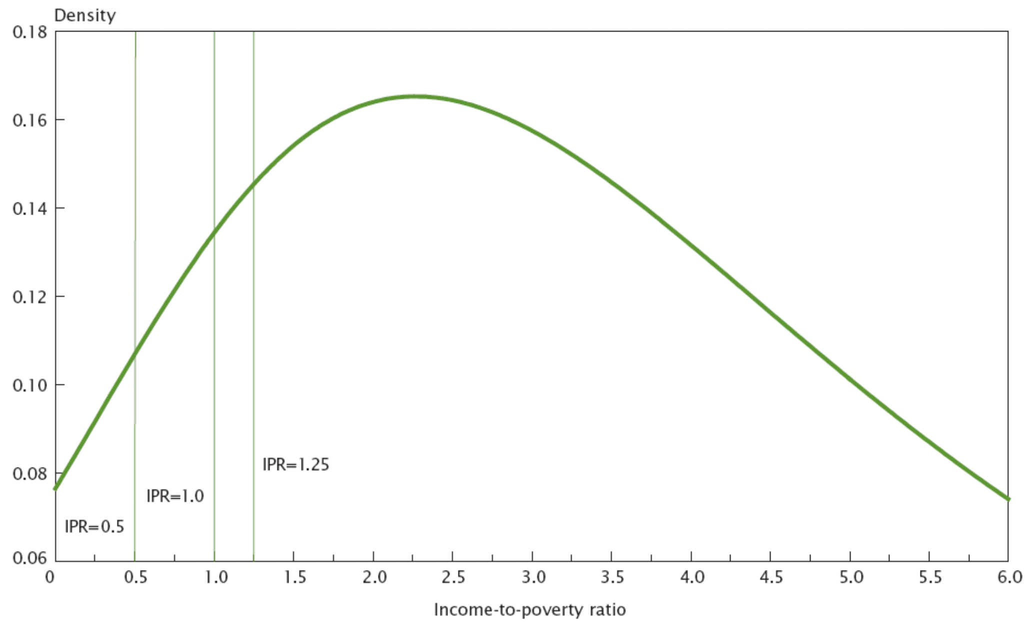
- A variable value expressed as the number of standard deviations from the mean.
- A useful tool for looking at some cumulative probabilities.

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

If the mean income in 2000 was $57,000 and the standard deviation was $59,000, what was the Z score for the poverty threshold of $17,500 for a family of 4?

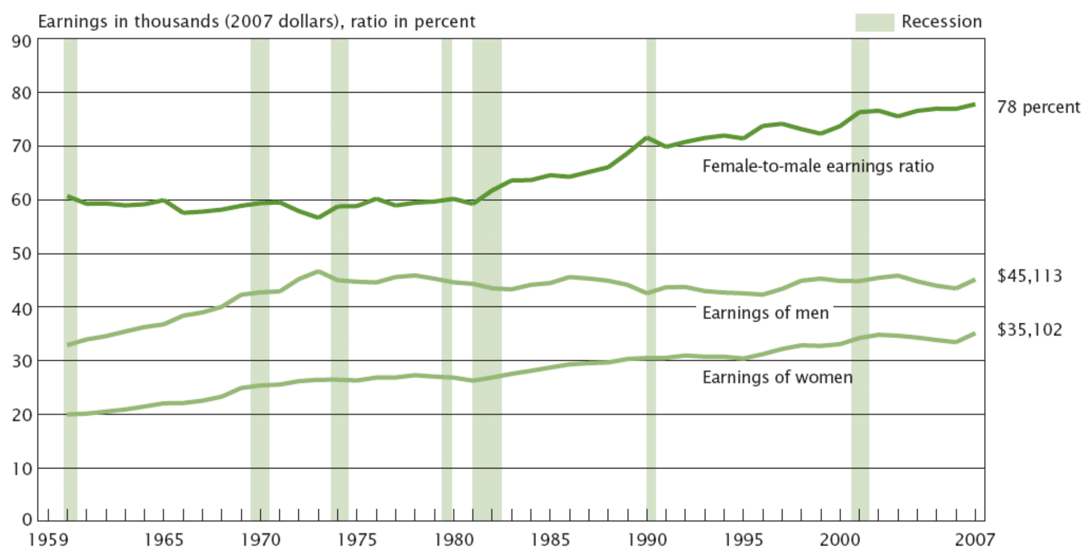
$$z = \frac{X - \mu}{\sigma} = \frac{17{,}500 - 57{,}000}{59{,}000} = -.67$$

Figure 5.
Distribution of Income-to-Poverty Ratios: 2007



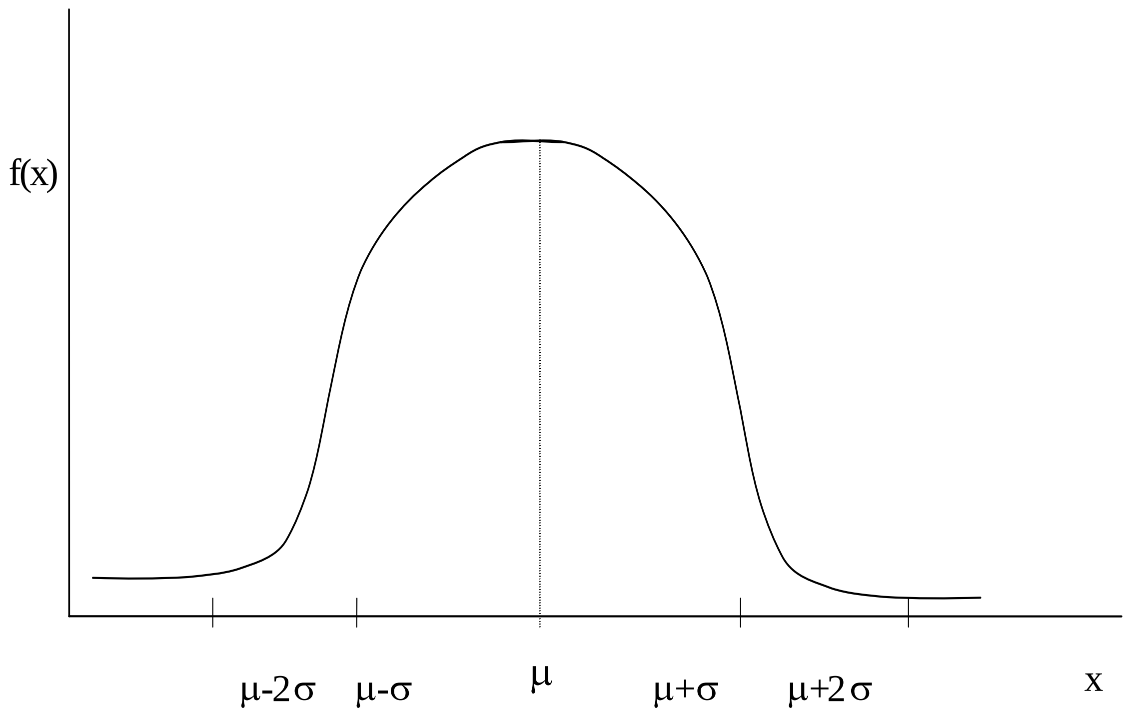
Source: U.S. Census Bureau, Current Population Survey, 2008 Annual Social and Economic Supplement.

Figure 2.
Female-to-Male Earnings Ratio and Median Earnings of Full-Time, Year-Round Workers 15 Years and Older by Sex: 1960 to 2007



Note: Data on earnings of full-time, year-round workers are not readily available before 1960. For information on recessions, see Appendix A.
Source: U.S. Census Bureau, Current Population Survey, 1961 to 2008 Annual Social and Economic Supplements.

The Normal Distribution



Normal Distribution Properties:

- It is UNIMODAL and SYMMETRIC:
- Half of “weight” below mean (because symmetrical)
- 68% of probability within 1 standard deviation of mean
- 95% of probability within 2 standard deviations
- More than 99% of probability within 3 standard deviations

Standard Normal distribution:

- Normal with mean of 0 and standard deviation of 1
- Distribution of standard scores IF original distribution is normal (ie. you can convert any normal distribution to Standard with Z scores)

Normal Probability Density Function

don't memorize this!

$$f(x)=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for } -\infty < x < \infty$$

where: x is a variable value
e=2.718...
π=3.141...
μ=mean of the random variable X
σ=standard deviation of the random variable X

Standard Normal Distribution--
A normal distribution with mean of 0 and standard deviation of 1.

- Distribution of standard scores, IF original distribution is normal.
- Use Z scores to convert any Normal distribution to standard normal distrib.

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

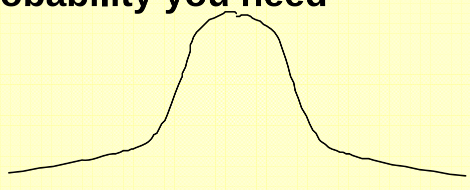
.

Table of cumulative probabilities are in book (page 794)

Suppose the distribution of students reading scores on CAT is approximately normal, with a mean of 440 and a standard deviation of 180. If students with scores of over 650 qualify for special reading classes, what proportion will qualify?

<u>To find normal probability associated with an X value:</u>

1. Draw a picture and write down the probability you need
P(X > 650)=?



2. Convert probability to standard scores.

P (X > 650) = P(Z > (**650-440)/180**)
= P (Z > **1.17)**

3. Find cumulative probability in table.

P(Z > 1.17)**= .5 - Pr (0 < Z < 1.17)**
Look up 1.17 to get probb:
=.5-.**3790**
=.121

Suppose that students with scores between 250 and 650 can benefit from regular reading classes. What proportion of students can use regular classes?

Suppose we could support 10 percent of the students with special remedial classes. What test score should we use as the upper cut-off for the classes?

<u>To find a value associated with a normal probability:</u>

1. Write down probability statement and draw picture:

P(X < __?_) = .10

➔ **P (Z<__?_)=.10**

2. Look up Z value in table

P (Z<__?_)= .5 - Pr(0 < Z < l_?_ l) = .10

Look up .40 in body of Z table to find Z cut off value for complement.

P(0 < Z < **1.28**) = **.3997**

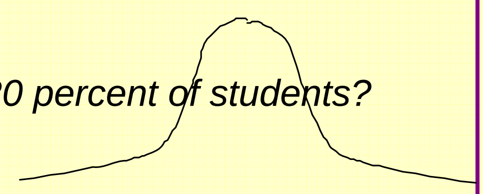
➔P(Z**<-1.28**) = .10

3. Convert Z value (SD units) to variable (X) by using mean and SD.

$$X = \mu + Z\sigma$$

X= 440 + (-1.28)(180) = **210**

What cut-off should we use if we could serve 20 percent of students?



How many students would be served if we had 1000 in our district?

Question B: ***You've just landed a prime internship with Public Health Seattle-King County. Your first task is to analyze the rates of low birth-weight babies. Babies weighting less than 5.5 pounds are defined as "low birth-weight" and are susceptible to many health problems. Some factors which appear to increase the chances of low birth-weight are: maternal smoking, drug or alcohol consumption, lack of prenatal care, teen-age mother, or African-American mother.***

a) Suppose that the weight distribution of babies in the US for 2000 was approximately normal with a mean of 8 pounds and a standard deviation of 1.6 pounds. About what proportion of babies were low birth-weight?

b) Some have suggested that all babies weighting less than the 25th percentile should be carefully monitored for health problems. What is highest weight that would be monitored under this proposal?

c) If a Seattle hospital has 300 births per year, how many would you expect to be low birth-weight? What might affect your prediction?

d) Write up your results for your hospital supervisor. Be sure to provide your numerical conclusions, as well as your educated opinions on the quality of the data and analysis. Don't forget--your supervisor is very smart, but has an aversion to statistical jargon.

Question C: ***You are the manager of a help line for families who want to find out about the Earned Income Tax Credit. Suppose that the amount of time it takes to answer a person's question has a normal distribution with a mean of 10 minutes and a standard deviation of 2 minutes.***

What range of time (in minutes) do 90 percent of the calls fall within?

What range of time do 95 percent of the calls fall within?

What is the chance the two calls would together take more than 28 minutes? What's the chance they would take less than that? [*Warning: This question goes beyond what we have covered in class in order to stretch your brain.]*