Stat 202-2015S W3-~~Mon~~ ~~Tues~~ Wed (Pg 1)

Review

A transformation is a function that transforms an old variable into a new variable.

$$X_{new} = f(X_{old})$$

A transformation is just another name for a function: thought of as transforming variables.

Examples X_{old} = distance traveled in km
X_{new} = distance traveled in miles

$$X_{new} = .62 X_{old} = f(X_{old})$$

Linear transformation - A transformation whose graph is a line.

Most transformations we will deal with are linear. The only examples I ~~can~~ can think of at the moment where non-linear transformations are used involve changing from a linear to a log scale.

Energy of Earthquake $\xrightarrow{f}$ Value on Richter Scale

Stat 202-2015S W3-~~Mon~~ ~~Tues~~ Wed (Pg2)

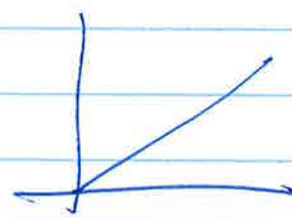
Many examples we will see are changes in units:

Km → miles
feet → meters
miles/hr → meters/sec
e.t.c

Unit conversions ~~are~~ tend to be linear of the form

$$X_{new} = b X_{old}$$

with zero vertical intercept



The only example of a unit conversion that I can think of at the moment with a non zero intercept is conversion between Farenhiegtit and Celcius or vice versa

$$\cancel{X_{new} = a + b X_{old}}$$

$$\cancel{X_{new} = 5/9\, X_{old} - 32 \cdot 5/9}$$

Stat 202-2015S-W3-~~Mon~~ ~~tues~~ wed

Pg 3

X_{old} = temp in °F

X_{new} = temp in °C

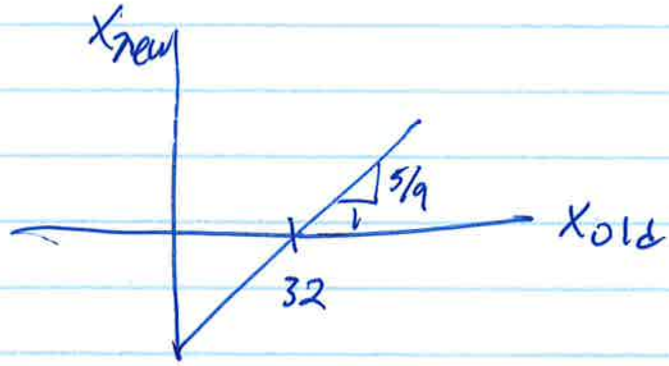
$$X_{new} = 5/9\,(X_{old} - 32)$$

$$= 5/9\, X_{old} - 32 \cdot 5/9$$

$$= -32 \cdot 5/9 + 5/9\, X_{old}$$

$$= a + b\, X_{old}$$

$$a = -32 * 5/9 \qquad b = 5/9$$



When $X_{old} = 32$

$X_{new} = 0$

Stat 202-2015S W3 - ~~Mon~~ ~~Thurs~~ wed (pg 4)

Density Curves - A smooth approximation of a histogram. More precisely, it is the theoretical limit of ~~[illegible]~~ the shape of a histogram as the number of data points gets very large and the width of the bins gets very small.

↳ histograms must be plotted with density for the vertical scale to workout.

A density curve ~~[illegible]~~ is such that
(1) it is always on or above the horizontal axis
(2) The area between the horizontal axis and the density curve is always one

Any curve that satisfies (1) and (2) is the density curve for some distribution

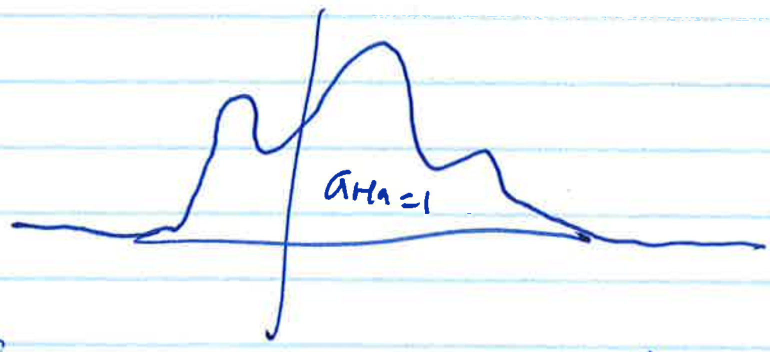
Reporting ~~[illegible]~~ the density curve for a distribution tells you everything there is to know about a distribution.

But to be sure you know the density curve you need an infinite amount of data.

Stat 202-2015S W3-~~Tues~~ wed Pg 5

There are infinitely many density curves

Any curve you can draw above or on the horizontal axis and with area 1 between it and the horizontal axis is one



This one probably doesn't have a name

A few have names—they are the

Standard distributions

Of which the most important is the

Normal distribution whose density curve

is a bell curve.

Stat 202-2015S-W3-~~Tues~~ wed

New

Measuring center and spread for density curves

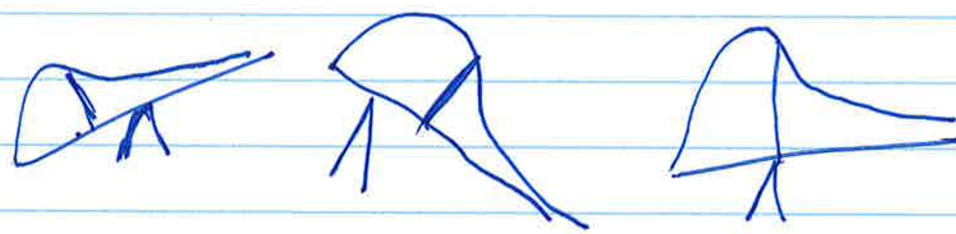
A mode of a distribution is a peak point of its density curve.

The median is the point with half the total area on each side (total area = 1 so area ½ on each side)

The quartiles are found by dividing area into quarters

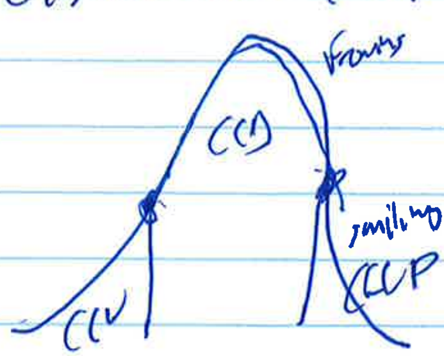
μ

The mean of a distribution is the balance point of its density curve



σ

The standard deviation doesn't have an easy locate-by-eye trick for general distributions but it does for normal distribution (bell curves



One standard deviation from the mean is the inflection points

The 68-95-99.7 Rule

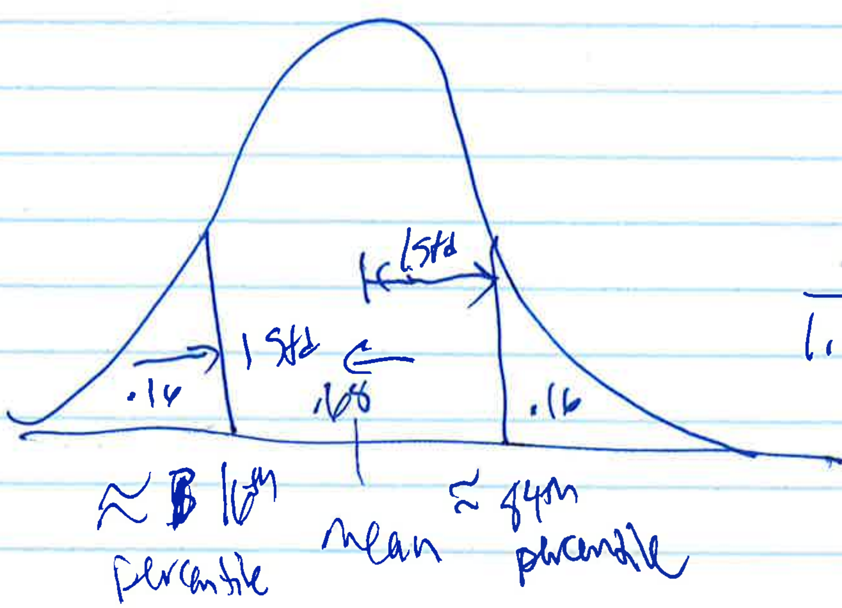
Regardless of the mean and standard deviation of a normal curve

Normal ~~distributions~~ only

Approximately 68% of the area under its density curve lies within 1 standard deviation from the mean.

Its not exactly 68% the area equals an irrational number which is approximatly 0.68...

This means that approximatly 68% of observations fall within 1 standard deviation of the mean and approximatly 32% of observations fall farther than one standard deviation from mean.

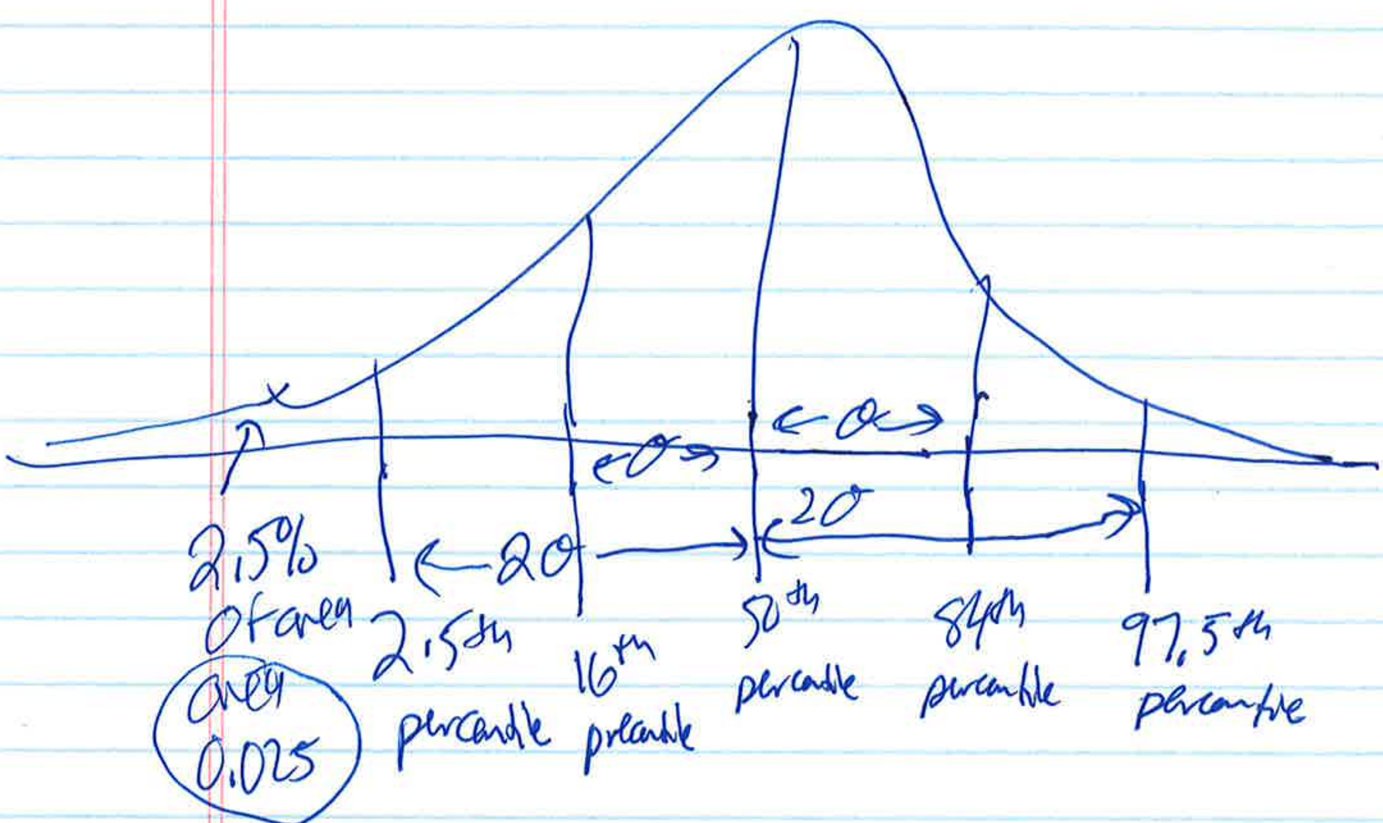


.16
.68
.16
1.00

Normal Intuition

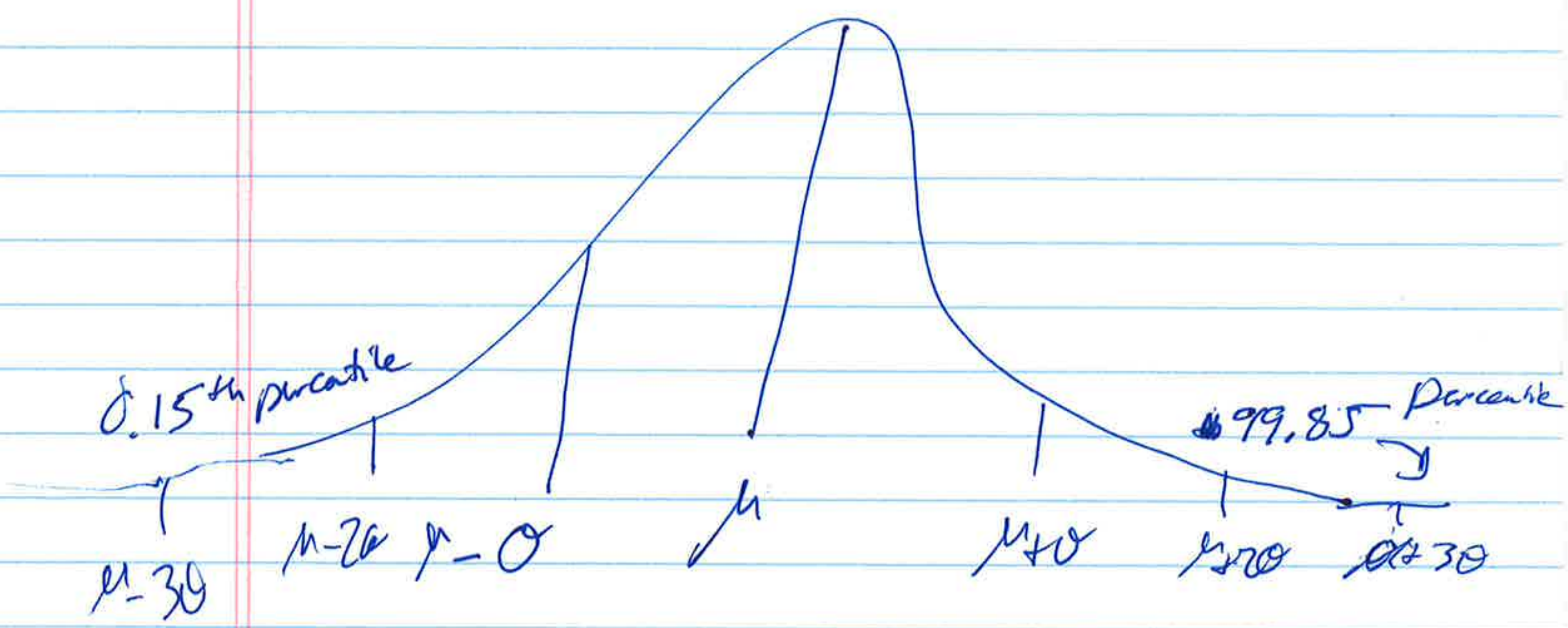
Approx 95% of obs fall within 2 standard deviations of mean (for a ~~std.~~ normal distribution)

That means 5% fall more than two std from mean



Normal Only — Approx 99.7 % of obs fall within 3σ of mean of a normal distribution.

0.3 % of obs fall ~~within~~ **beyond** 3σ of mean



The 68-95-99.7 Rule is the combination of the above three facts.

68%	within	1 std dev	Normal Only
95%	"	2 "	
99.7%	"	3 "	

As suggested by this rule all that matters for a Normal distribution is how close you are to mean.

We can actually transform (with a linear transformation) a normal variable with any mean and std dev, into one with the "standard" mean (zero) and the "standard" standard-deviation (one)

This is called **standardizing the observations**

The "new" variable is called the Z-score and is written with the letter Z.

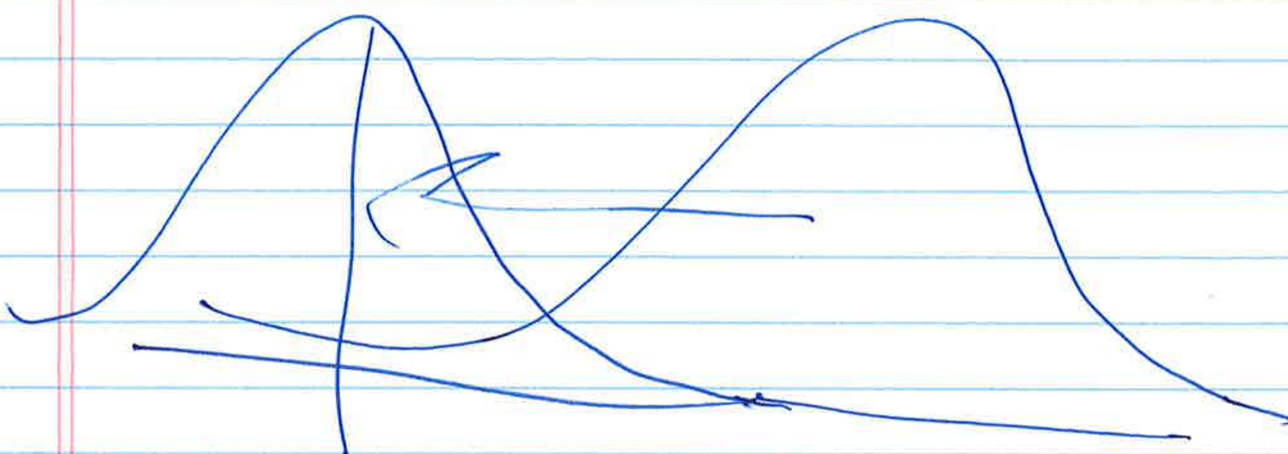
All Z-scores have mean 0 and standard dev 1.

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

The transformation

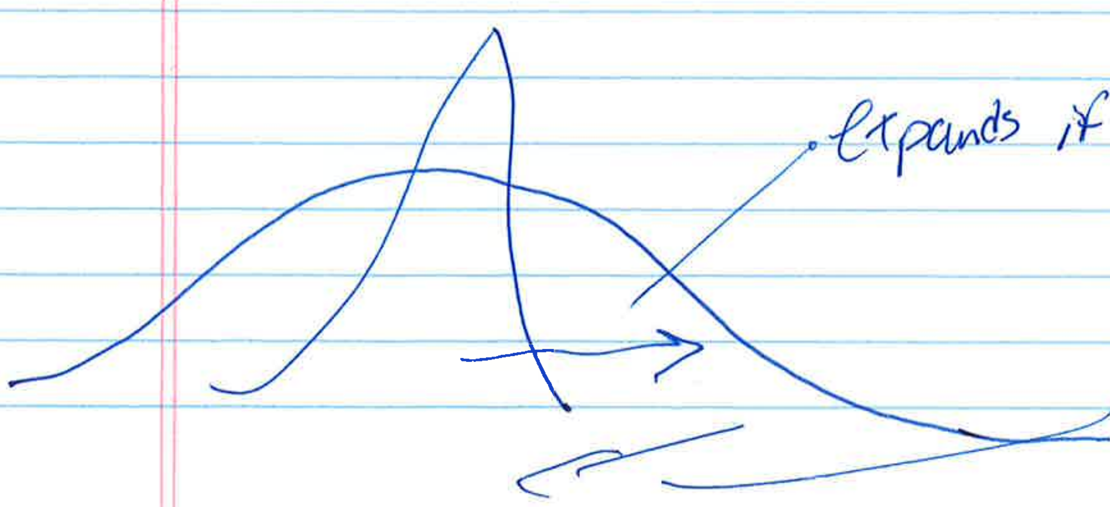
$$\boxed{X_{new} = X - \mu}$$

shifts the density curve so that the mean is zero



The transformation

$Z = b X_{new}$ scales the ~~standa~~ distribution about zero



expands if $b > 1$

shrinks if $b <$

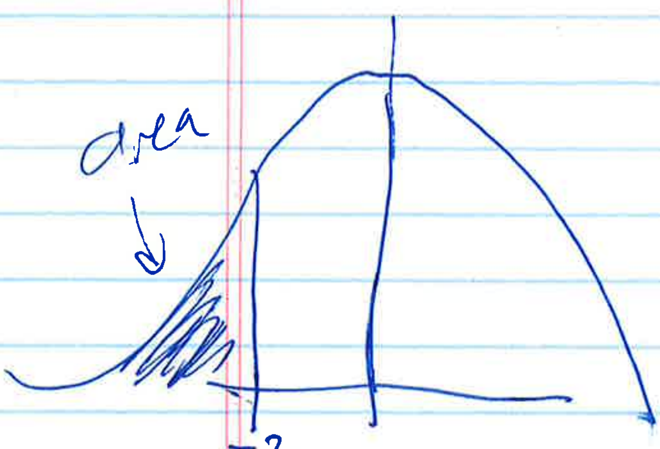
Normal distribution can be represented symbollically. To know everything about a Normal distribution we only need to know its mean and standard deviation — μ and σ.

The normal distribution with mean μ and std dev σ is denoted $N(\mu, \sigma)$

The standard normal is denoted $N(0, 1)$

$N(1201, 320)$ is a Normal distribution with mean 1201 and std dev 320.

A table for the Normal distribution will give you z scores and corresponding areas



Z-score

book has

One table for negative z-score

One table for positive z-scores