1/22/15 263 Lecture 3

Announcements:

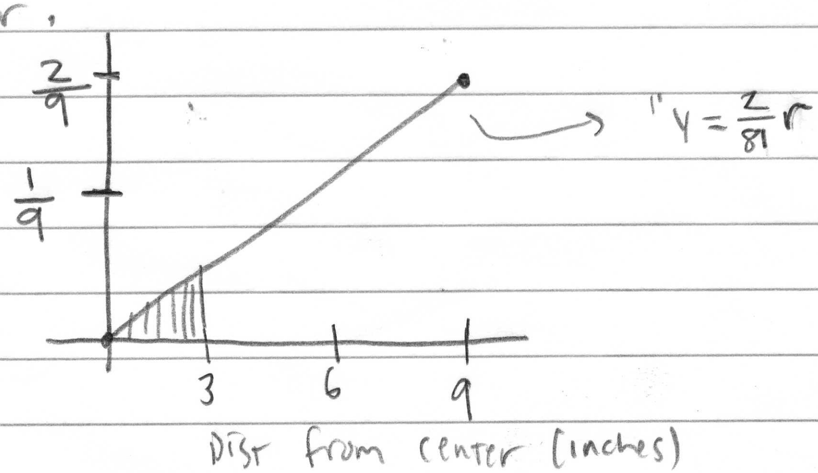
- office hrs today 2-3:30 (i'll be there until later but 2-3:30 is your priority)
- Excel HW1 DUE today
- Quiz @ end of class

* - NOTE: Excel computes Q1, Q3 Funny!

Last time

A distribution curve is a "continuous" graph that describes the proportion of a (quant) variable which appears in some range.

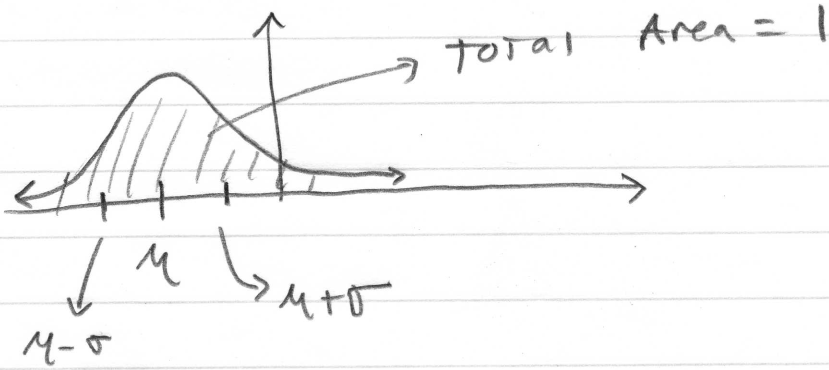
EX Darts are thrown at a dart board at random. The following distribution curve shows the proportion of darts hitting some distance r from the center:



- Always between 0 & 1
- Total Area is 1
- What % of darts are thrown between 0 & 3 in from center? Find shaded area!

Important Note: Can only find areas for a range of vals. % of darts that are exactly 3 inches from center is 0

Normal distributions are symmetric, bell-shaped, & described by just mean & std dev.

$N(\mu, \sigma) =$ 

* Unimodal, no skew, ...
* 68-95-99.7 rule: 68% of the distribution lies between $\mu - \sigma$ & $\mu + \sigma$ "1 std dev from mean". 95% is 2σ from μ, 99.7% is 3σ from μ.

* Z scores: convert units of horizontal axis to "standard deviations"

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

EX A student takes the SAT and gets 580 Reading, 595 Math, 575 Writing. Which score was their "best"? Assume Normal distr. for the population, with:

	R	W	M
μ	501	493	515
σ	112	111	116

Solution: Convert to z-scores!

reading: $z = \dfrac{580-501}{112} \approx 0.7054$

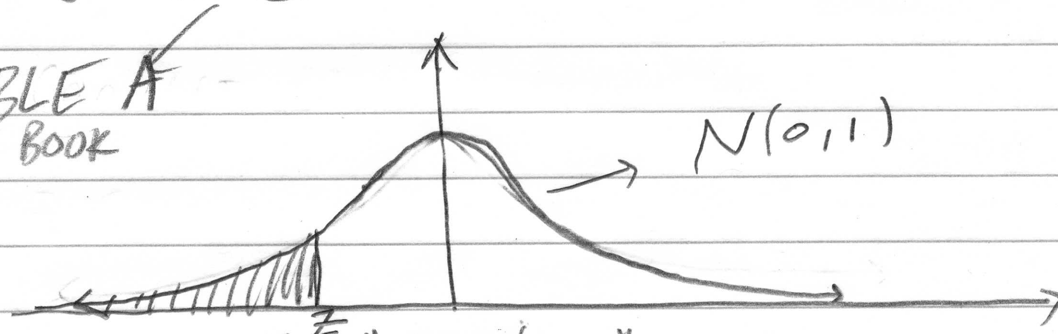
writing: $z = \dfrac{575-493}{111} \approx 0.7388$

Math: $z = \dfrac{595-515}{116} \approx 0.6897$

So writing was their "best" score!

Using z-Tables to find Proportions:

TABLE A in BOOK



The "Standard" Normal

A z-table tells you the proportion of the Standard Normal distribution that lies to the left of a particular z-value.

z	.00	.01
−0.6	0.2743	0.2709
−0.5	0.3085	0.3050
0.5	0.6915	0.6950

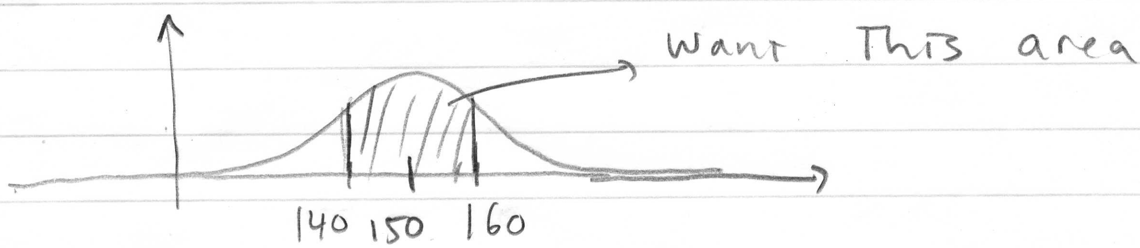
EX: % of z-values less than 0.51 is 69.5%.

EX Suppose the # of bacteria on a culture is normally distributed w/ $\mu = 150$ & $\sigma = 15$. What prop of samples will have between 140 & 160 bacteria?

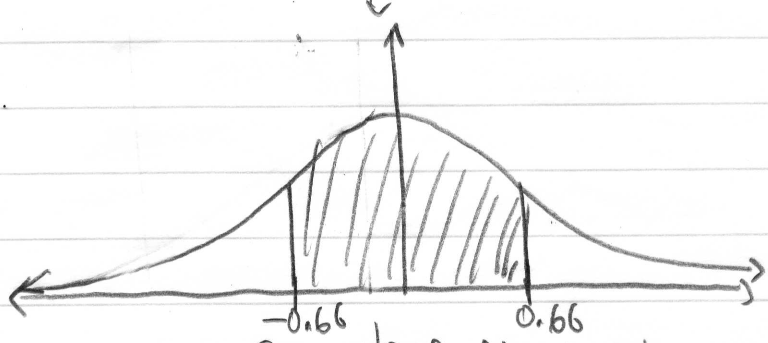
1) Convert to Z-scores

$$Z_{140} = \frac{140-150}{15} = -\frac{2}{3} \simeq -0.66$$

$$Z_{160} = \frac{160-150}{15} = \frac{2}{3} \simeq 0.66$$



Distribution of # of bacteria



Standard Normal

This will be (area left of 0.66) − (area left of −0.66)

$$\simeq 0.7454 - 0.2546$$

$$= 0.4908 \text{ or } \boxed{49.08\%}$$

4

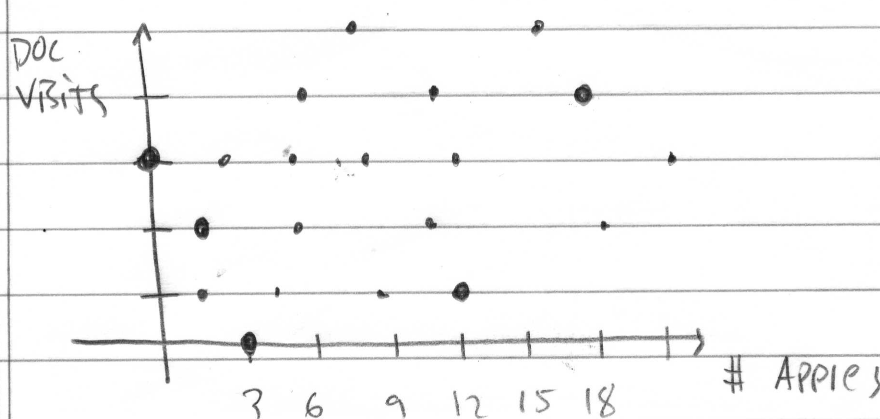
Chapter 2: Relationships

(Scatter plots, correlation, least-squares regression)

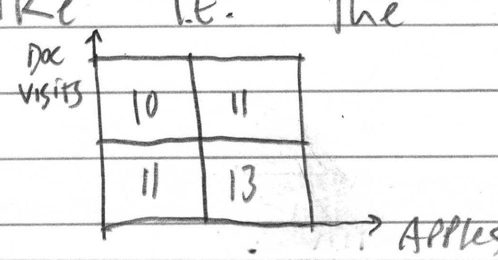
One of the main goals of stats is establishing meaningful relationships between variables.

<u>EX</u> Want to know if eating apples reduces doctor's visits. Form a study:

Person	# apples / mo	Doctor's Visits / mo
1	3	0
2	1	2
3	12	1
4	0	3
5	17	4
...	...	...



Is there a trend? No - "joint" distribution is very "blob"-like i.e. the "2D histogram" would look like:



Scatter Plots display:

- form, direction & strength of relationship

linear vs. nonlinear

Positive or negative association

strong or weak.

Example:

Smoking vs. eating fruits/veggies

Source: CDC / Book:

State	F/V (%)	Smoke (%)
Al	20.3	18.6
AZ	24.1	13.1
CA	27.7	8.6
DC	31.5	12.8
Fl	24.4	13.5
ID	24.6	12.4
KS	18.6	16.9
MS	16.8	18.8
MT	25.7	16.4
NV	23.7	18.0
NM	23.2	13.5
OR	26.3	14.4
RI	26.1	14.4
VT	29.3	14.7
WV	16.2	23.8
	"X"	"Y"

* Make a Scatter Plot on Calculator.

* Do you see an association? Strong or weak? linear or nonlinear?

* Does X cause Y?

Correlation

A way to measure the extent to which a pair of variables exhibit a linear relationship is with correlation;

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

$\bar{x}$ = mean of $\{x_i\}$ $\qquad \frac{x_i - \bar{x}}{s_x}$ ~ z-score for x_i

$\bar{y}$ = mean of $\{y_i\}$ $\qquad \frac{y_i - \bar{y}}{s_y}$ ~ z-score for y_i

s_x, s_y ~ std dev's

Interpretation:

- Make a scatter plot
- "re-center" the axes so the origin is at $(\bar{x}, \bar{y})$
- Scale the axes to make them unitless

$\boxed{\text{I.E. compute z-scores!}}$

- The product of 2 z-scores gives a positive or negative "area"