

## Review Question from Old Midterm

A group of men have average height 69 inches and SD 3 inches. A group of women have average height 63 inches and SD 3 inches.

Is the percentage of men with heights between the two averages greater, about the same as, or less than the percentage of women with heights between the two averages?

What if the SD of the women were 2 inches?

1

---

---

---

---

---

---

---

---

---

---



---

---

---

---

---

---

---

---

---

---



---

---

---

---

---

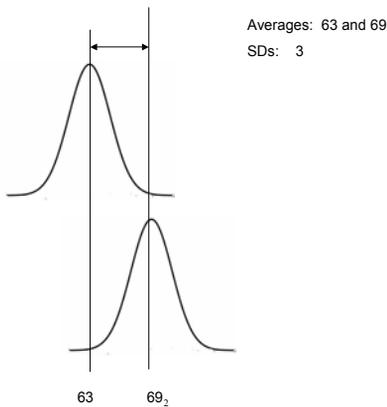
---

---

---

---

---



## The Butterfly Ballot in Palm Beach

Voters in Palm Beach, Florida, claimed that they were confused by the ballot structure and inadvertently voted for Buchanan when in fact they intended to vote for Gore.

### Confusion over Palm Beach County ballot

Although the Democrats are listed second in the column on the left, they are the third hole on the ballot.

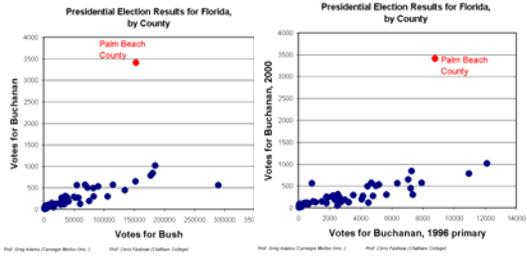
Punching the second hole casts a vote for the Reform Party.

(REPUBLICAN)	3	(REFORM)	4
GEORGE W. BUSH, PRESIDENT		PAT BUCHANAN, PRESIDENT	
DICK CHENEY, VICE PRESIDENT		EZELA FOSTER, PRESIDENT	
(DEMOCRATIC)	5	(SOCIALIST)	6
AL GORE, PRESIDENT		SAVIO BAREYANLOS, PRESIDENT	
JOE LIEBERMAN, VICE PRESIDENT		MARY CAL HELLIS, VICE PRESIDENT	
(LIBERTARIAN)	7	(CONSTITUTION)	8
HARRY BROWNE, PRESIDENT		HOWARD PHILLIPS, PRESIDENT	
ART DEWEER, VICE PRESIDENT		J. CHRIS FRASER, VICE PRESIDENT	
(GREEN)	9	(OWNERS WORLD)	10
RALPH NADER, PRESIDENT		MONICA MADOREAD, PRESIDENT	
WENDY LADLOW, VICE PRESIDENT		GLORIA LA NIÑA, VICE PRESIDENT	
(SOCIALIST WORKERS)	11		
JAMES HARRIS, PRESIDENT			
MARGARET TROWE, VICE PRESIDENT			
(NATURAL LAW)	12		
JOHN HAZDEW, PRESIDENT			
NAT SCHUMASER, VICE PRESIDENT			

WRITE IN CANDIDATE  
To vote for a write-in candidate, follow the directions on the long strip of your ballot card.

Sam Senneth/Graphic/Charles Helms

## Scatter Plots Tell the Story



4

---

---

---

---

---

---

---

---

## Regression

5

---

---

---

---

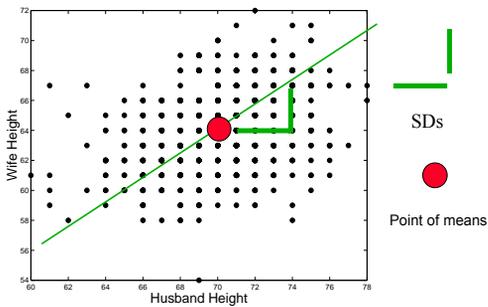
---

---

---

---

Heights of Husbands and Wives: The Point of Means and the SD Line  
 $r = .34$



6

---

---

---

---

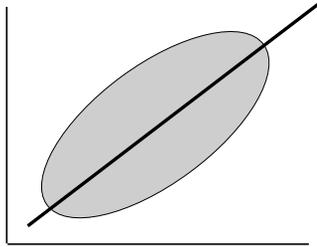
---

---

---

---

# The SD Line and the Football



7

---

---

---

---

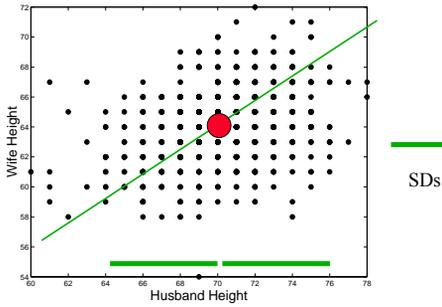
---

---

---

---

If a husband is 2 SDs shorter than average, would you predict his wife to be 2 SDs shorter than the average wife?



8

---

---

---

---

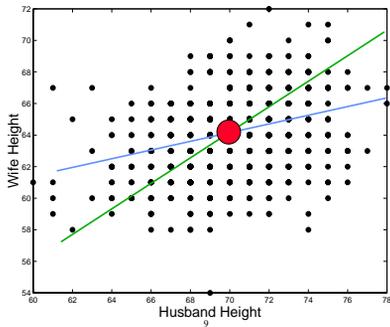
---

---

---

---

**The Regression Line:** associated with a 1 SD increase in husband's height is a  $r \times$  SD increase in wife's height. Compare it to the **SD line**.



9

---

---

---

---

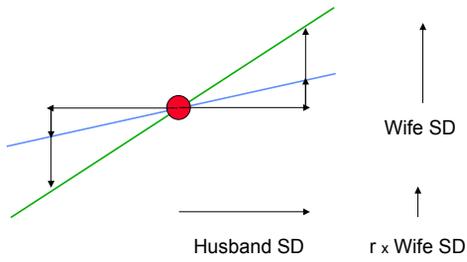
---

---

---

---

## Regression Line and SD Line



10

---

---

---

---

---

---

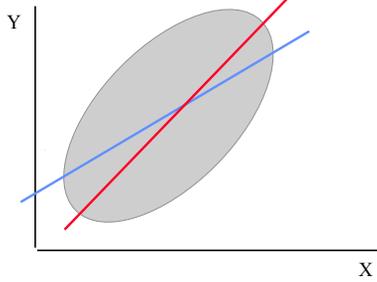
---

---

---

---

## The SD and Regression Lines on the Football



11

---

---

---

---

---

---

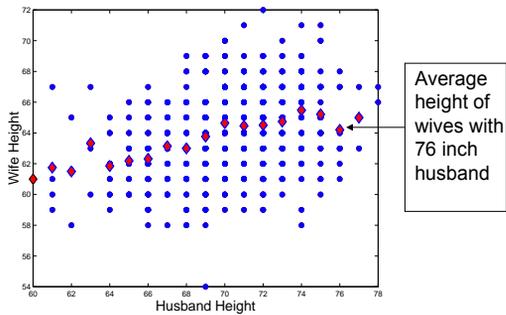
---

---

---

---

## The Graph of Averages



12

---

---

---

---

---

---

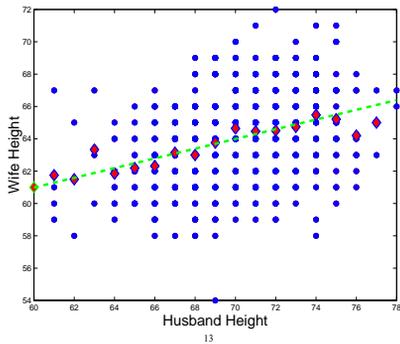
---

---

---

---

# The Regression Line and the Graph of Averages



---

---

---

---

---

---

---

---

The regression line is a smoothed version of the graph of averages.

If the graph of averages is a straight line, that line is the regression line.

---

---

---

---

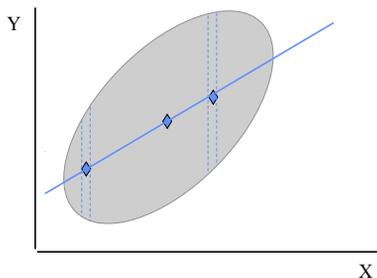
---

---

---

---

An Idealized Scatter Diagram, Graph of Averages, and Regression Line: Predicting Y by X



---

---

---

---

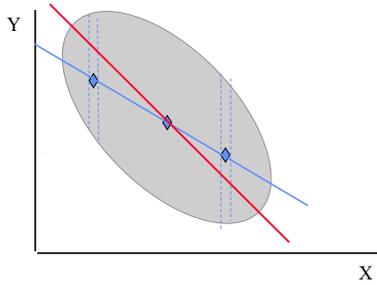
---

---

---

---

Negative Correlation, SD Line, Graph of Averages, and Regression Line: Predicting Y by X



16

---

---

---

---

---

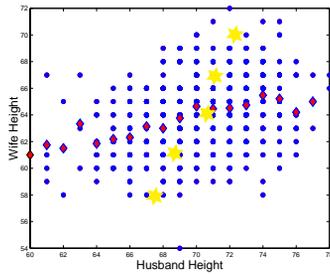
---

---

---

There are *two* graphs of averages

\*



17

---

---

---

---

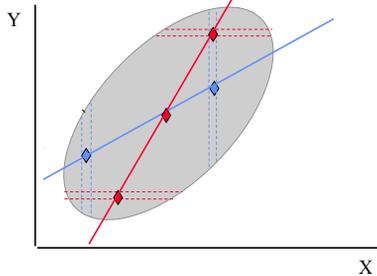
---

---

---

---

Two Regression Lines:  
Predicting Y from X  
Predicting X from Y



18

---

---

---

---

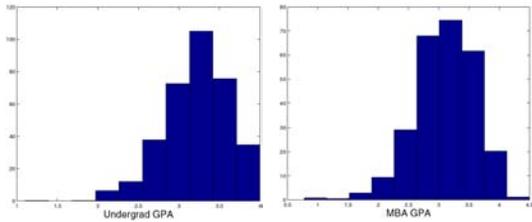
---

---

---

---

## Undergrad GPA and MBA GPA



Average = 3.2

SD = .40

Average = 3.1

SD = .49

19

---

---

---

---

---

---

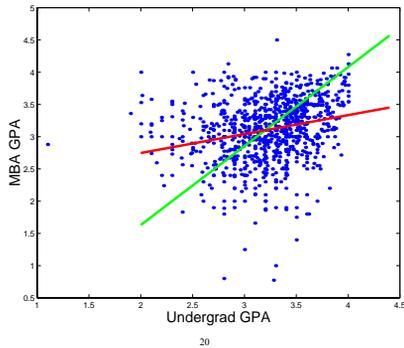
---

---

---

---

## The Regression Line and the SD Line ( $r = .24$ )



20

---

---

---

---

---

---

---

---

---

---

**Q:** What is the predicted MBA GPA if undergrad GPA = 3.5?

	Average	SD
Undergrad	3.2	.40
MBA	3.1	.49
$r = .24$		

21

---

---

---

---

---

---

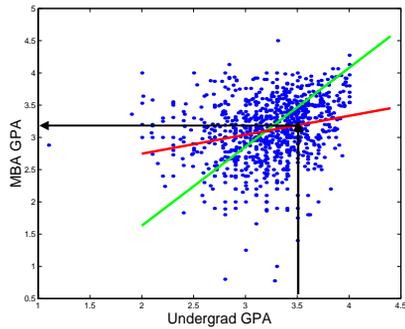
---

---

---

---

Undergrad GPA = 3.5  $\Rightarrow$  Predicted MBA GPA = 3.19



22

---

---

---

---

---

---

---

---

---

---

**Q:** An UG GPA of 3.5 predicts MBA GPA = 3.19  
 What is the predicted undergrad GPA if MBA GPA = 3.19?

	Average	SD
Undergrad	3.2	.40
MBA	3.1	.49
$r = .24$		

23

---

---

---

---

---

---

---

---

---

---

## Standard Units

If X is 2 SD's above the mean, the predicted value of Y is  $r \times 2$  SDs greater than the mean of Y



If the X score in standard units equals 2, then the predicted Y score in standard units equals  $r \times 2$

*If the X score is expressed in standard units, the predicted Y score, in standard units, is  $r \times$  the X score in standard units.*

24

---

---

---

---

---

---

---

---

---

---

## The Three Step Method

Undergrad GA = 3.5. Predicted MBA GPA?

	Average	SD
Undergrad	3.2	.40
MBA	3.1	.49

$r = .24$

1. *Convert to standard units:* 3.5 is how many SDs above average?  $(3.5 - 3.2)/.40 = .75$  SDs
2. *Multiply by r:*  $.75 \times .24 = .18$   
This is the estimated MBA GPA in standard units.
3. *Convert to regular units:*  $(.18 \times .49) + 3.1 = 3.19$

25

---

---

---

---

---

---

---

---

---

---

**Q:** An undergrad's GPA is 2 SDs above the average. How many SDs above the average is his MBA GPA predicted to be?

**A:** It is predicted to be

$$r \times 2 \times \text{SD} = .48 \text{ SD}$$

above the average MBA GPA. A much better than average undergrad is predicted to do only a little better than average as an MBA.

26

---

---

---

---

---

---

---

---

---

---

If GPAs follow the normal curve, 2 SDs above the average is the 95th percentile. And .48 SDs above the average is the 69th percentile.

By symmetry, an undergrad who was 2 SDs below average, at the 5th percentile, would be predicted to be at the 31st percentile in the MBA program.

They both "**regress toward the mean.**"



This term was coined by the English scientist Francis Galton, b. Feb. 16, 1822, d. Jan. 17, 1911, a cousin of Charles Darwin, who made significant contributions to genetics, psychology, and statistics

27

---

---

---

---

---

---

---

---

---

---

## The Regression Fallacy

Suppose a special program is set up to help low achievers in the first year of the MBA program. Suppose their percentile ranks improve. Is this evidence that the special program is effective?

28

---

---

---

---

---

---

---

---

## The Regression Effect

In test-retest situations, the bottom group on the first test will improve and the top group will do more poorly.

29

---

---

---

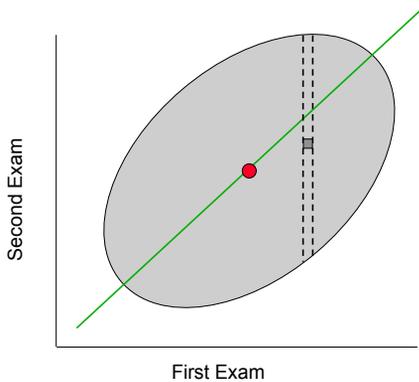
---

---

---

---

---



30

---

---

---

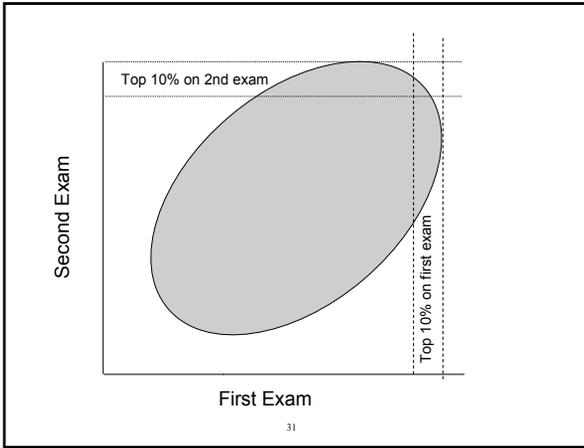
---

---

---

---

---




---

---

---

---

---

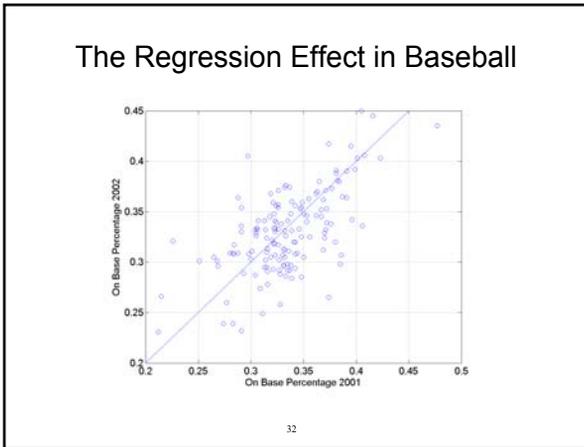
---

---

---

---

---




---

---

---

---

---

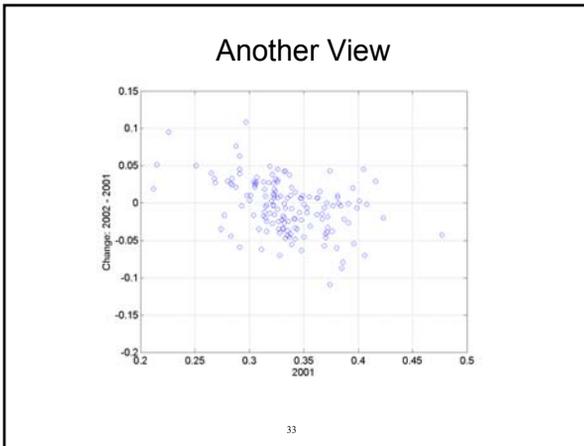
---

---

---

---

---




---

---

---

---

---

---

---

---

---

---



## Summary

- The regression line can be found from the two means, the two SDs and the correlation.
- It is a smooth version of the graph of averages. It can be used for predicting one variable from the other. For every 1 SD increase in X, average Y increases  $r$  SDs
- There are two regression lines.
- The regression effect: in test-retest situations, the bottom group on the first test will improve and the top group will do more poorly.

34

---

---

---

---

---

---

---

---

---

---

? A doctor is in the habit of measuring blood pressures twice. He notices that patients who are unusually high on the first reading tend to be lower on the second. He concludes that patients are more relaxed on the second reading. A colleague disagrees, pointing out that patients who are unusually low on the first reading tend to have higher second readings, suggesting they get more nervous.

Which doctor is right? Or perhaps both are wrong?

35

---

---

---

---

---

---

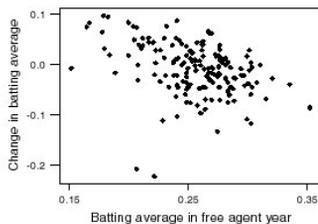
---

---

---

---

? Baseball free agency: do teams get what they pay for? The graph shows that those players who had high batting averages before they changed teams did worse after changing while those who had low batting averages improved. How can this be explained?




---

---

---

---

---

---

---

---

---

---

? The scatter diagram of scores on a midterm and final is football shaped and the correlation is .6. A student is on the 70<sup>th</sup> percentile of the midterm. Estimate the student's percentile rank on the final.

---

---

---

---

---

---

---

---