Announcements

- Unclaimed 1st & 2nd exams & homeworks (1, 2, 3, & 4) are up front
- Old Handouts & 1st Day Survey are up front
- One Handout Today
- Review Material for the final will be handed out on Friday

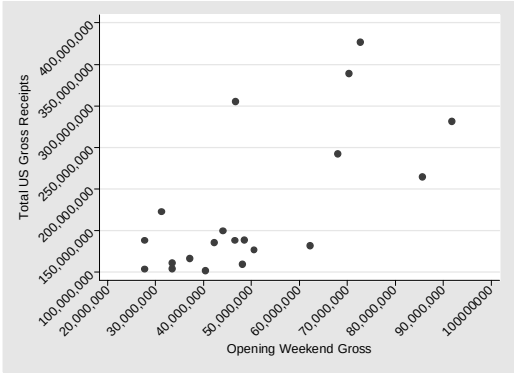
INTRODUCTION TO LINEAR REGRESSION (Ch. 8)

- some data on movies, their total domestic (U.S.) gross receipts (totusgross) and first weekend ticket sales (weekend1).
- Questions of interest:
  - what is the relationship between the total gross receipts for a movie and first weekend ticket sales?
  - Do big first weekends always generate more ticket sales?
  - How well can we predict a movie's total domestic gross receipts given that we have information on it's first weekend's ticket sales.

Movie Data: Title, Total Domestic (US) gross receipts and first weekend ticket sales (in dollars)

	title	totusgross	weekend1
1.	The Lord of the Rings: The Return of the King	377,027,325	72,629,713
2.	Finding Nemo	339,714,978	70,251,710
3.	Pirates of the Caribbean: The Curse of the Black Pearl	305,413,918	46,630,690
4.	The Matrix Reloaded	281,576,461	91,774,413
5.	Bruce Almighty	242,829,261	67,953,330
6.	X2: X-Men United	214,949,694	85,558,731
7.	Elf	173,398,518	31,113,501
8.	Terminator 3: Rise of the Machines	150,371,112	44,041,440
9.	The Matrix Revolutions	139,270,910	48,475,154
10.	Cheaper by the Dozen	138,614,544	27,557,647
11.	Bad Boys II	138,608,444	46,522,560
12.	Anger Management	135,645,823	42,220,847
13.	Hulk	132,177,234	62,128,420
14.	2 Fast 2 Furious	127,154,901	50,472,480
15.	S.W.A.T.	116,934,650	37,062,535
16.	Spy Kids 3D: Game Over	111,761,982	33,417,739
17.	Scary Movie 3	110,003,217	48,113,770
18.	American Wedding	104,565,114	33,369,440
19.	Daddy Day Care	104,297,061	27,623,580
20.	Daredevil	102,543,518	40,310,419

Some statistics and a graph



r = .68

- Means and Standard Deviations are:

Variable	Obs	Mean	Std. Dev.	Min	Max
totusgross	20	177342928	85658496	102543518	377027325
weekend1	20	50361404	18775706	27557647	91774413

Working with Linear Regression Using Z scores

- "The Incredibles" opened on the weekend of November 5th-7th and its opening weekend gross was $70,467,623.
- Predict the total US gross receipts for this new movie using Z scores. We must assume that X and Y are normal to do this.
- A line that passes through the origin (0,0) can be written with just a slope and no intercept:

  y = mx where m is the SLOPE

  (or y=bx in some textbooks)

- We can substitute Z scores for Y and X so now this looks like

  $Z_y = m*Z_x$

- As your book states (p. 138) this line won't fall perfectly on every point (if you look at the graph you can see that there is no way one straight line will be driven through all points on the graph). So generally we write $\hat{Z}_y = mZ_x$
- Which recognizes that the Z_y is some predicted value.
- It turns out that the best value for the slope of the line (m) is the correlation, r, so we can rewrite this as

  $\hat{Z}_y = rZ_x$

Expressing "The Incredibles" first weekend in terms of Z scores.

- Convert the first weekend to Z

$$Z_x = \frac{x - \mu_x}{\sigma} = \frac{70467623 - 50361404}{18775706} = 1.07$$

- where we just use the mean and standard deviation of the x variable (first weekend ticket sales) as substitutes for μ (mu) and σ (sigma).
- Using $\hat{Z}_y = rZ_x$ if we take the 1.07 Z score and multiply it by the correlation .68 we can relate it to the Y variable (total gross receipts) and get a .73 Z scores for Y.

- Recall that Z scores are just standard deviations.
- We can now give an actual dollar value prediction for Y by multiplying .73 by the standard deviation for total gross receipts of the Y variable 85,658,496:

  .73 * 85,658,496 = 62,530,702

  and add this result to the average total gross receipts

  ($177,342,928 + $62,530,702) = $239,873,630.

  So "The Incredibles" first weekend was $70,467,623, given what we know about first weekend's relationship to total gross receipts, we predict (or expect) about $239,873,630

Another Question involving Regression

- Suppose a movie is in the highest 10% of first weekend ticket sales and suppose first weekend ticket sales is normally distributed. What is it's predicted total US gross receipts?
- Recall that if you are in the top 10% of A NORMAL, this means you have a Z = +1.28 or you are 1.28 standard deviations above average
- If a movie is in the top 10% of first weekend ticket sales, its first weekend ticket sales is calculated to be 50361404 + (1.28 * 18775706) = $74,394,308

- The corresponding total US gross receipts will be will be:
  - find the product of the Standard Deviation of first weekend sales and the correlation: 1.28* .68= .87
  - multiply this result by the Standard Deviation of total US gross receipts: .87* 85,658,496= 74,522,892
  - add this value to the average total US gross receipts 74,522,892 + 177,342,928 = 251,865,820
- Therefore, if a movie is in the top 10% of first weekend ticket sales, the corresponding total US gross receipts is predicted to be $251,865,820