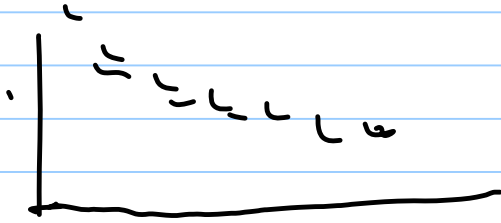


STT 200 3pm 4-14-10

CH 10 ends 7-10 CORRELATION + L.S. FIT



TRANSFORM TO ANOTHER SCALE  
 NONLINEARLY TO GET LINE.  
 THEN FIT LINE BY L.S.

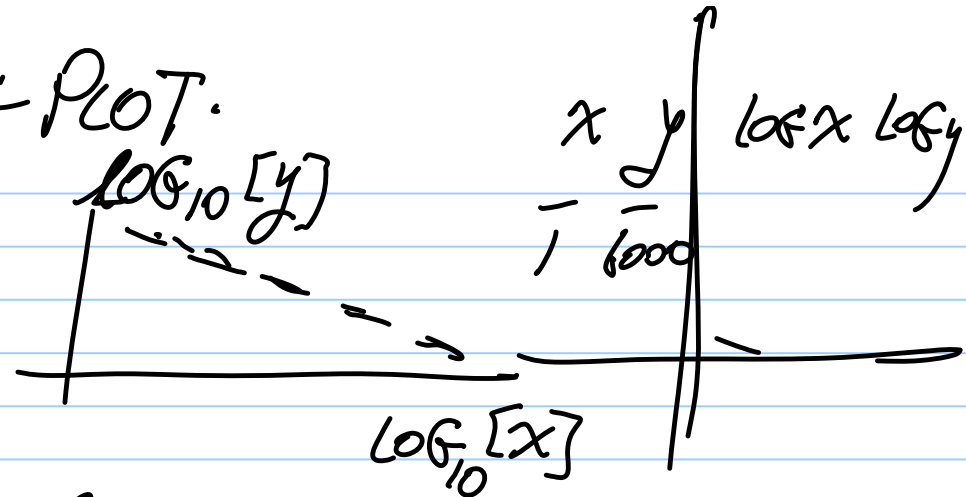
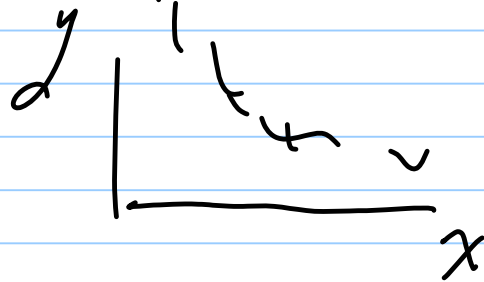
PREMIER EXAMPLE:  $y = a x^b$  LOG<sub>10</sub> 10 = 1 (check)

$$\log_{10}[y] = \underbrace{\log_{10}[a]}_{\#} + b \underbrace{\log_{10}[x]}_{\#}$$

BRINGS

LOG<sub>10</sub>[x], LOG<sub>10</sub>[y] INTO # STRAIGHT LINE RELATION.

THIS IS CALLED LOG-LOG PLOT.

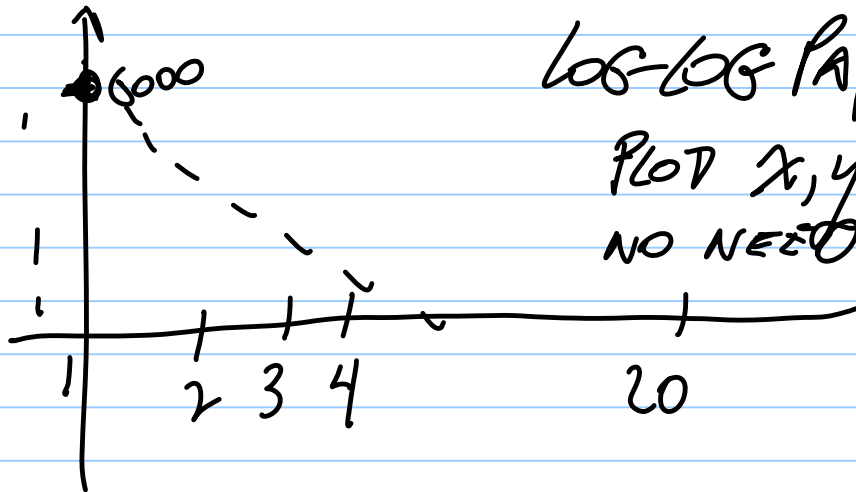


$\log y$

LOG-LOG PAPER

PLOT  $x, y$  IN DISTORTED SCALE PAPER  
NO NEED TO TRANSFORM TO LOGS.

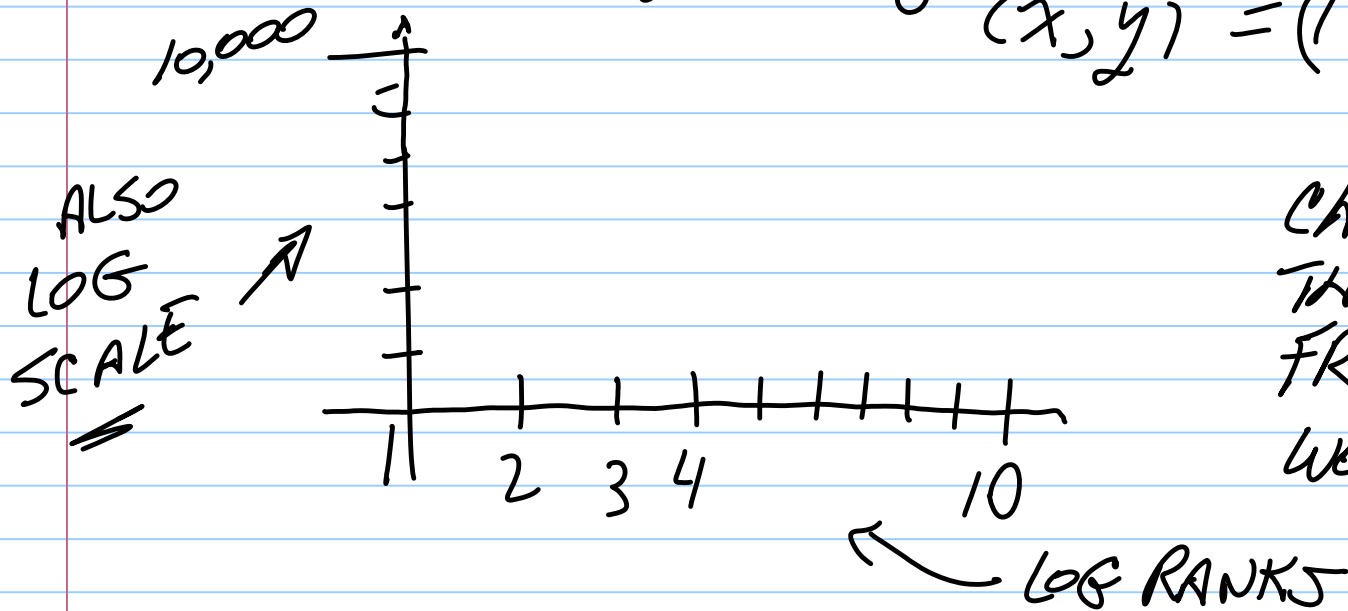
ALSO COMPRESSED SCALE



LOGS OF  $1 \rightarrow 20$ .

LEADS TO A PREMIER EXAMPLE "ZIPF'S LAW"

eg RANK 1 (MOST FREQUENT) TYPICALLY "THE" =  $x$   
FREQUENCY  $y$  eg "THE" USED 6000 TIMES  
 $(x, y) = (1, 6000)$



CHOOSE TO LOOK AT  
THE 10 MOST  
FREQUENTLY USED  
WORDS.

SEE TANUR et al.

VIEW HANS ROSLING  
(TED LECTURES)

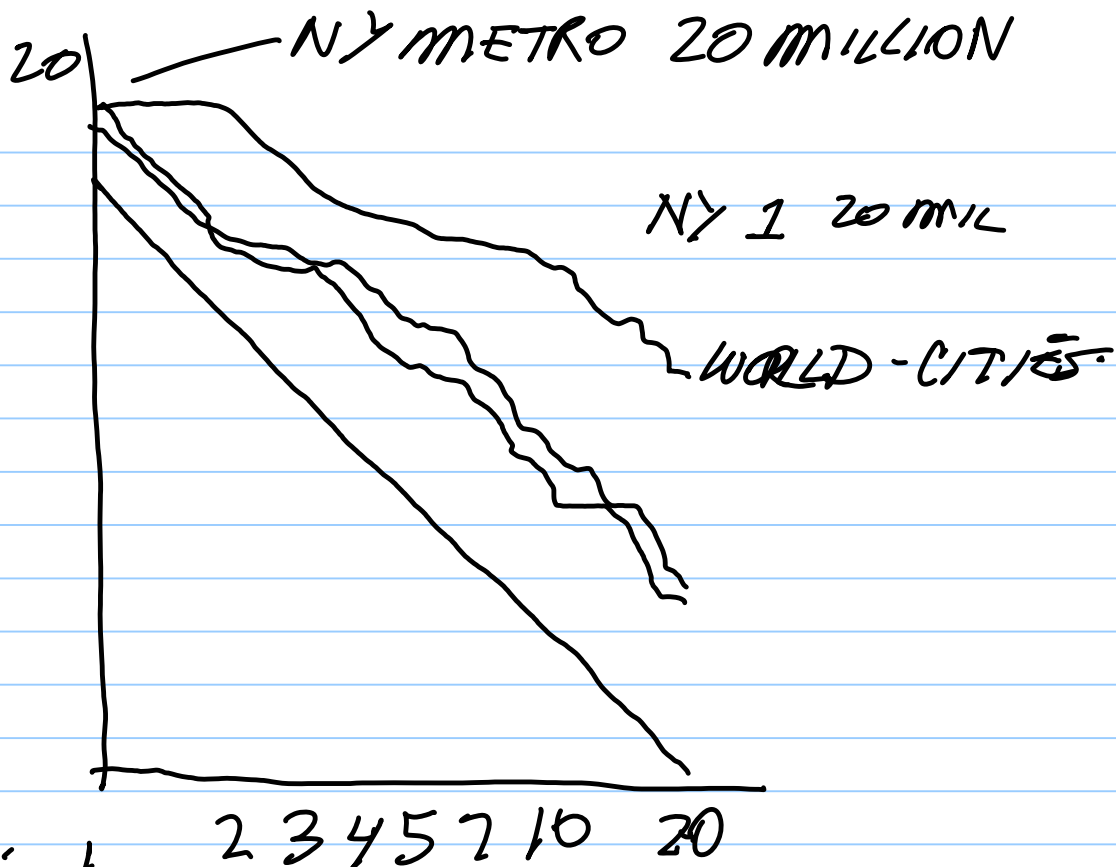
YOUTUBE

LOG LIFE EXPECTANCY

CHINA

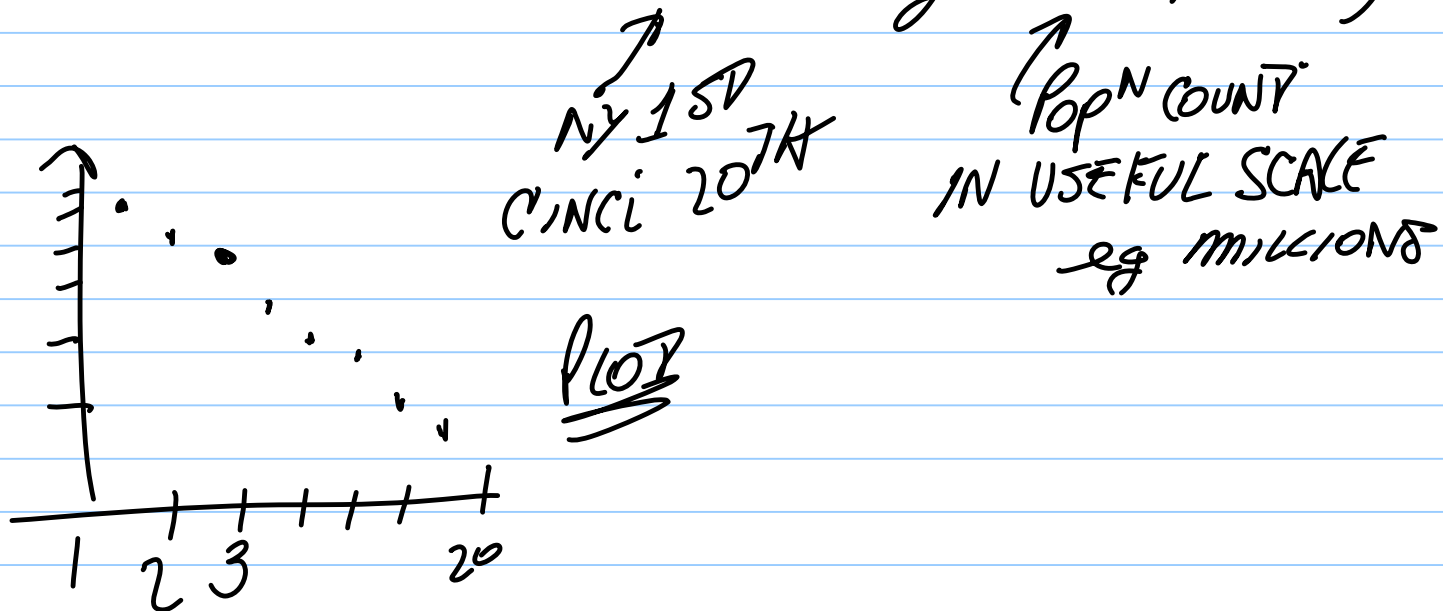
30 vs

LOG INCOME



PENDING MY GETTING YOU LOG-LOG "PAPER" (WEB POST)

- ① READ + CITE 2 WEB REF. TO ZIPF'S LAW.
- ② COME UP WITH YOUR PROPOSED APPLICATION OF ZIPF'S LAW.
- ③ PLOT DATA ( $x = \text{RANK}$ ,  $y = \text{FREQUENCY}$ )



(4) ~~#~~ FID L.S. LINE

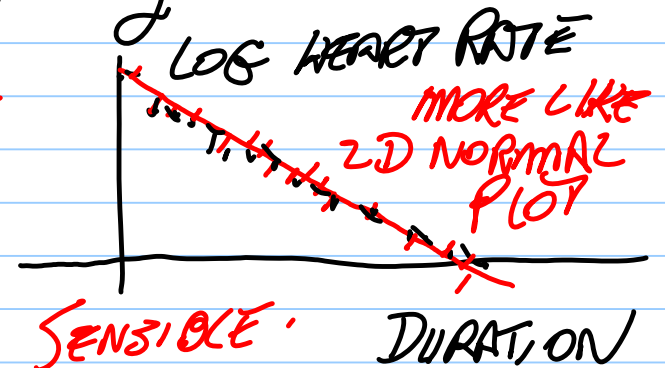
PT OF AVGS LOG(RANK)

LOG(FREQUENCY)

SLOPE FOR LOG(RANK), LOG(FREQUENCY).

CH 10. KNOW TEXT RECOMMENDS.

BE ASSURED THAT AFTER SCALING -  $\log$  LOGS.



GREAT! IF IT REALLY IS CONFIRMED  
w/ FUTURE DATA.

FUTURE REMEDIES? RANDOMIZE (IF POSSIBLE)  
DATA COLLECTION.

CH 10. NORMAL PROBABILITY PLOT OF RESIDUALS.

$$(y_i) - \left[ (x_i - \bar{x}) \frac{dy}{dx} + \bar{y} \right] = e_i \quad \text{RESIDUALS}$$

REGR PREDICTION

$y_i$   $e_i$  positive L.S.

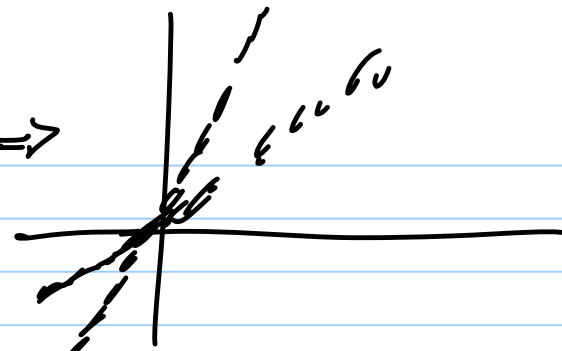
REGR PRED.

$x_i$

$$(e_1, e_2, \dots, e_n)$$

$$n = \#(x_i, y_i)$$

NORMAL  
PROB  $\Rightarrow$   
PLOT



$\bar{e} = 0$  WHEN YOU  
L.S. FIT

TEXT SUGGESTS IT'S GOOD  
IDEA FOR THIS TO BE  
 $\approx 57$  LINE.

---

BUT - NOT ABSOLUTELY NECESSARY!

CASE IN POINT: ZIPF'S LAW.

DON'T CARE ABOUT NORMAL LOOKING RESIDUALS.

NOR DO I CARE ABOUT PATTERNS IN RESIDUALS.



