

Review"Homework" #5 optionalSubmit before midnight Wed 4/29  
any missing homeworks

$$\text{Score}_{\text{HWS}} = \frac{\# \text{ homework submitted}}{21}$$

Final Exam

cumulative

Ch 2

Ch 3-6

Ch 8-12

conforming variable

Ch 7

Simpson's Paradox

Design of Experiments

Ch 13-15

Ch 16-21, 23

Ch 2

~ 2x

100-minute

2 hrs

length exam or Z paper

"50-minute"

3

Canvas

TentativeProposed schedule

Tuesday 5/5 12:00 - 4:00pm

ReviewRegression★ Z-scores ← Standardized value of a variable

Ch 13/14/15

Variable income, educ.level,

ht, wt, SAT score, lifetime lightbulb, Alcohol content of beer

measurements on a population (people or objects)

chance selection is built in

units: inches, cm, kg, lbs, SAT pts, hours, % \$, yrs,

2 main summary numbers:

average, SD

$$X_{\text{standardized}} = \frac{X_{\text{original}} - \text{AVE}(X)}{SD(X)}$$

Standardized variable  $X$ = # standard units the value  $X_{\text{original}}$  is from average.

long right tail not normal unit less

Sometimes  $Z = X_{\text{standardized}} = \frac{X_{\text{original}} - \text{AVE}(X)}{SD(X)}$ 

wt, ht, SAT

Area 68%  $-1.0 \quad 0 \quad +1.0$ 

Z-Scores again Regression Chapters

Z-scores again Ch 17 and to the end random (a lot) draws

box of tix sum (draws) ave (draws)  $\approx 1$  (draws)  $(0-1 \text{ box})$ 

all normally dist. Central Limit Theorem

 $z = \frac{(x - \text{ave})}{SE(\frac{\sum}{n} \text{ave})}$ 

Example regression + some normal calculation.

Suppose given variables  $X$  and  $Y$ 

are both approximately normally distributed, with linear association and a homoscedastic scatter diagram.

The summary statistics are:

$$\text{AVE}(X) = 50 \quad \text{AVE}(Y) = 80$$

$$SD(X) = 5 \quad SD(Y) = 10$$

$$\rho = .6$$

(a) Estimate the average  $Y$  value for data with  $X$  value near 57.Estimate the average percentile rank of  $Y$  data with  $X$  percentile rank near 40th percentile.(b)  $z_X = \frac{57 - 50}{5} = \frac{7}{5} = 1.4$ slope =  $\rho \cdot \frac{SD(Y)}{SD(X)} = .6(1.4) = .84$ slope-run = rise =  $.6(1.4) = .84$ Ans (a):  $(.84)10 + 80 = 88.4 = ?$ (b)  $z_X = \frac{40 - 50}{5} = -2$ slope =  $\rho \cdot \frac{SD(Y)}{SD(X)} = .6(-2) = -1.2$ slope-run = rise =  $.6(-2) = -1.2$ Ans (b):  $(-1.2)10 + 80 = 68 = ?$ 

OK correct unnecessary

Convert to  $Y_{\text{orig}}$ 

$$(-.16)(10) + 80 \approx 78.4$$

Ans (b): about 78.4

100-12 ≈ 44%

12

-1.6

Ans (b): about 44th percentile

The value 78.4 is about 44th percentile