Review for exam 2.

3 main topics:

1) correlation, regression, interpretation of regression (s_{yx}, ...)

2) probability & binomial distributions, normal approximation to the binomial.

3) Central limit theorem - all about "the sampling distribution of the mean"

μ_x, σ_x mean & s.d. of population

1) $\mu_{\bar{x}} = \mu_x$

2) $\sigma_{\bar{x}} = \sigma_x / \sqrt{n}$

3) means tend toward a normal distribution

Correlation and stuff

given Σx, Σy, Σxy, ... know how to calculate correlation, r

$$r = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{SS_x SS_y}} \quad \ldots$$

coeficient of variation: r^2 the proportion of variability in Y explained by X.

Regression:

pop. s.d.

Slope: $r \cdot \frac{s_y}{s_x}$

goes through $\bar{x}$, $\bar{y}$

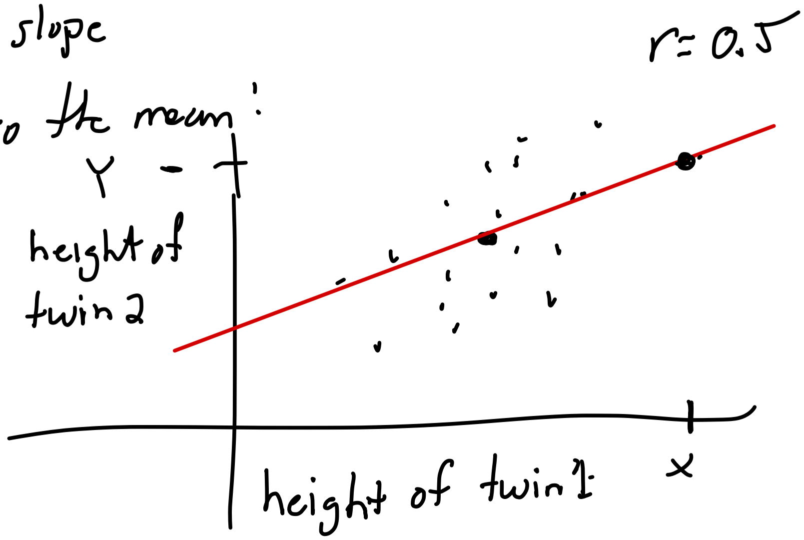
$$y - \bar{y} = r \cdot \frac{s_y}{s_x} (x - \bar{x})$$

y-intercept

point-slope: $$y = r \cdot \frac{s_y}{s_x} x + \left[\bar{y} - r \cdot \frac{s_y}{s_x} \bar{x}\right]$$

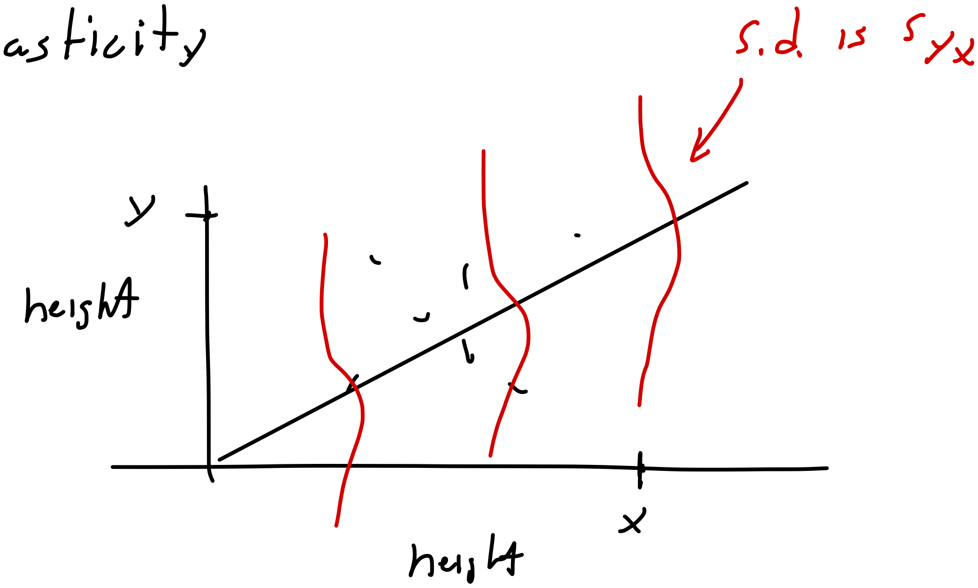
slope

Regression to the mean:



standard error of the estimate: $s_{yx} = s_y \sqrt{1-r^2}$

homoscedasticity



---

probabilities & binomial stuff

"and" + "or" rule

$\times$ $+$

binomial distribution

N = # events

k = # of "successful" outcomes out of N

p = probability of success in any one event

$Pr(k)$ probability of k out of N

Example: coin flips   probability of 4 out of 10 heads.

$N = 10$ flips

$p = .5$

$k = 4$

pdf Stats Tables
xlsx
R

} gives $p = .0251$

Normal approximation to the binomial

mean $= n \cdot p$

s.d. $\sqrt{n \cdot p(1-p)}$

Example:

Suppose a population is split 50/50 about two candidates. You poll 100 people.
What is the probability that 55 or more people will vote for one of the candidates?

$p = 0.5, \quad n = 100$

$\mu = n \cdot p = 100 \cdot .5 = 50$

$\sigma = \sqrt{n \cdot p(1-p)} = 5$

$Pr(x \geq 55) = Pr(z > ?)$

$$z = \frac{54.5 - 50}{5} = 0.9$$

$Pr(z > 0.9) = 1 - \text{pnorm}(0.9) = \underline{.1841}$

Central limit theorem

$$\mu_{\bar{x}} = \mu_x$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

normal.

Suppose we know that cats weights are distributed with a mean of 8 lbs and a standard deviation of 2 lbs. If I sample 16 cats, what is the probability of getting a mean of 10 or more lbs?

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \qquad \mu_{\bar{x}} = 8$$

$$\sigma_{\bar{x}} = \frac{2}{\sqrt{16}} = 0.5$$

$$z = \frac{10 - 8}{0.5} = \underline{4} \qquad p(z \geq 4) = .00003167$$