



Department of Computer Science

UNIVERSITY OF COLORADO **BOULDER**



Linear Regression

Introduction to Data Science Algorithms

Jordan Boyd-Graber and Michael Paul

SLIDES ADAPTED FROM DAVID NEAL

Correlation Coefficient

True Correlation

$$\rho = \frac{\mu_{XY} - \mu_X \mu_Y}{\sigma_X \sigma_Y} \quad (1)$$

Sample Correlation

$$r = \frac{\bar{x}\bar{y} - (\bar{x})(\bar{y})}{\sqrt{x^2 - (\bar{x})^2} \sqrt{y^2 - (\bar{y})^2}} \quad (2)$$

- If x and y are independent, then correlation is 0.
- Great if $\rho = \pm 1$
- Can we test how good the regression is?

Statistical Test for Regression

- Null hypothesis $H_0 : \rho = 0$
- Test statistic

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (3)$$

- Follows a t -distribution with $n-2$ degrees of freedom (estimating two parameters)
- Can do either two-tailed or one-tailed test

Wrapup

- Regression: powerful tool for explaining data
- Allows you to tell stories
- Allows you to predict the future (HW4)
- Foundation for more complicated models