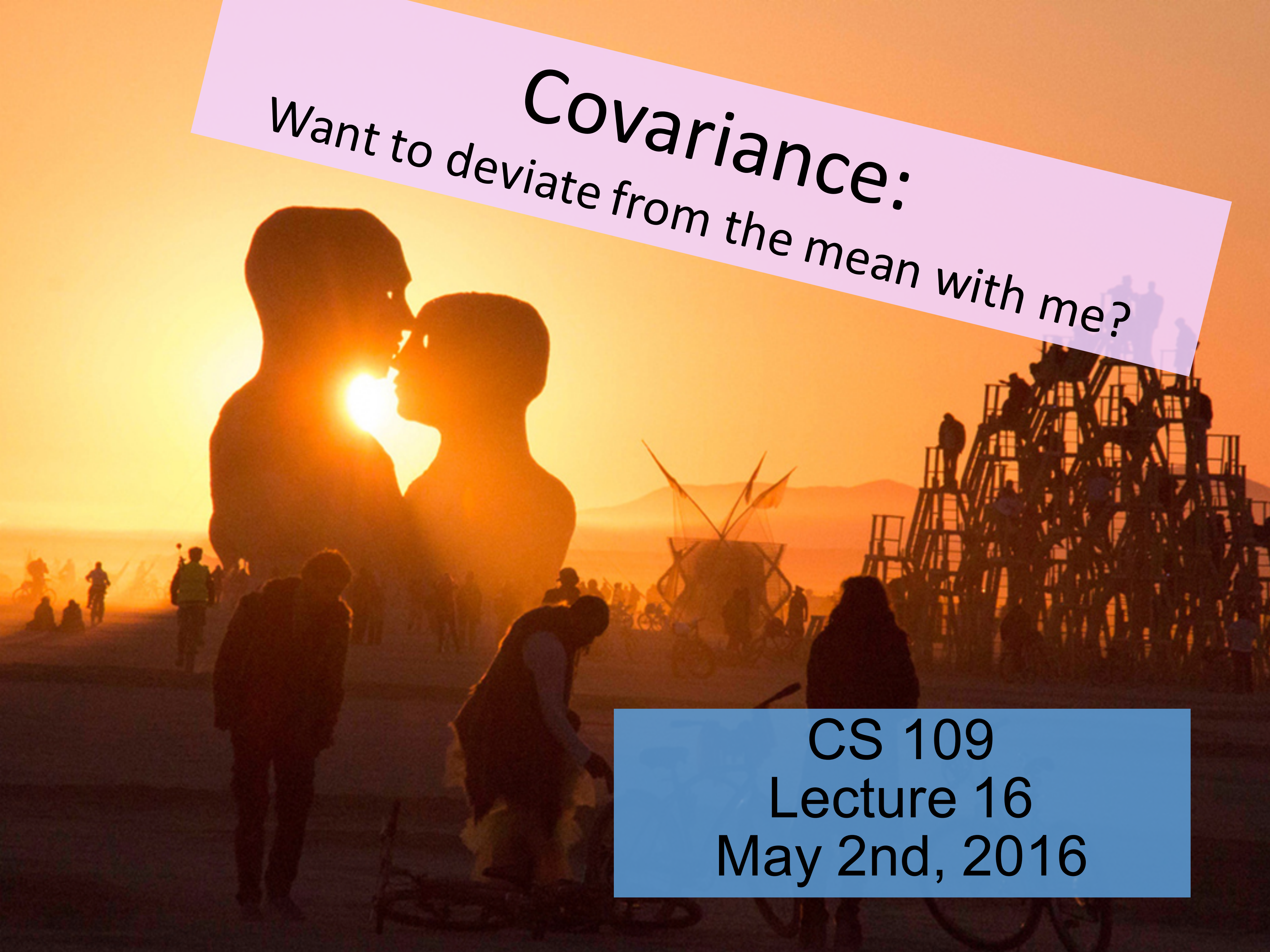
- 1 handout
- Midterm Tomorrow
- “hope”

Covariance:
Want to deviate from the mean with me?



CS 109
Lecture 16
May 2nd, 2016

Review

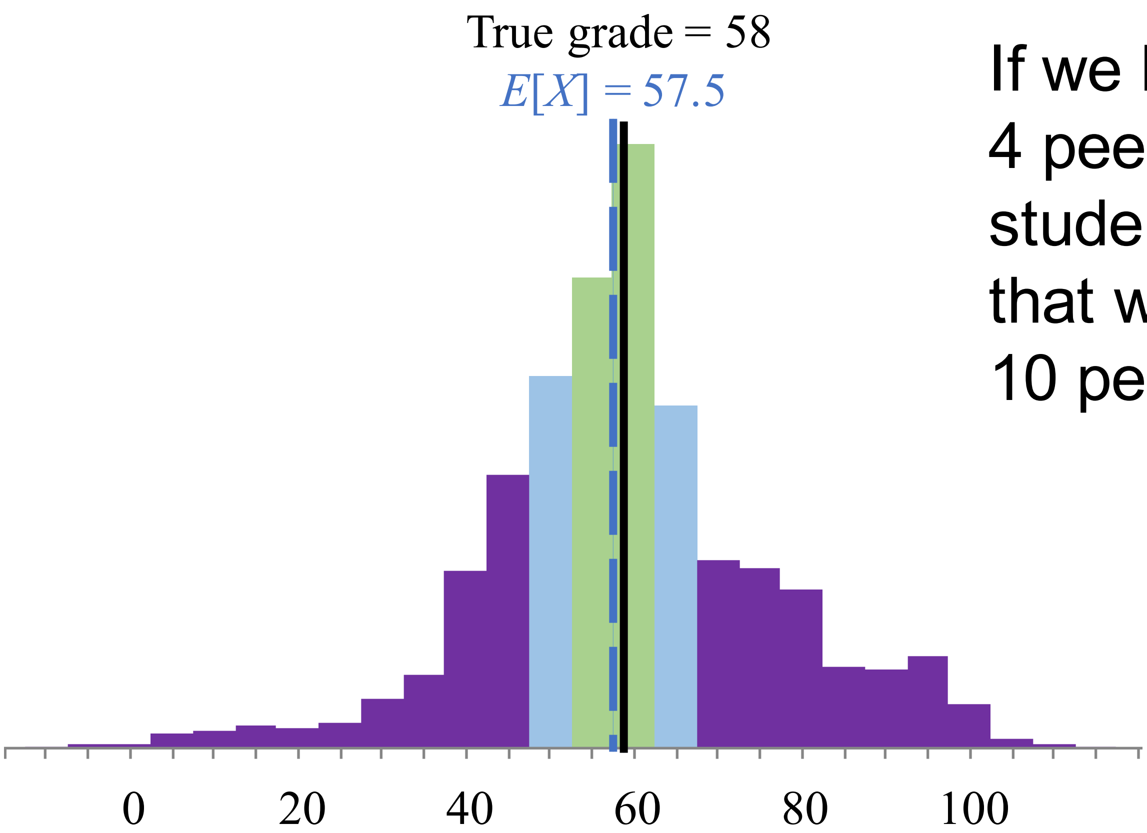
Expectation and Variance

The two most important descriptors of a distribution, a random variable or a dataset

Peer Grades in Coursera HCI

Let X be a random variable that represents a peer grade

$E(X)$ = weighted average

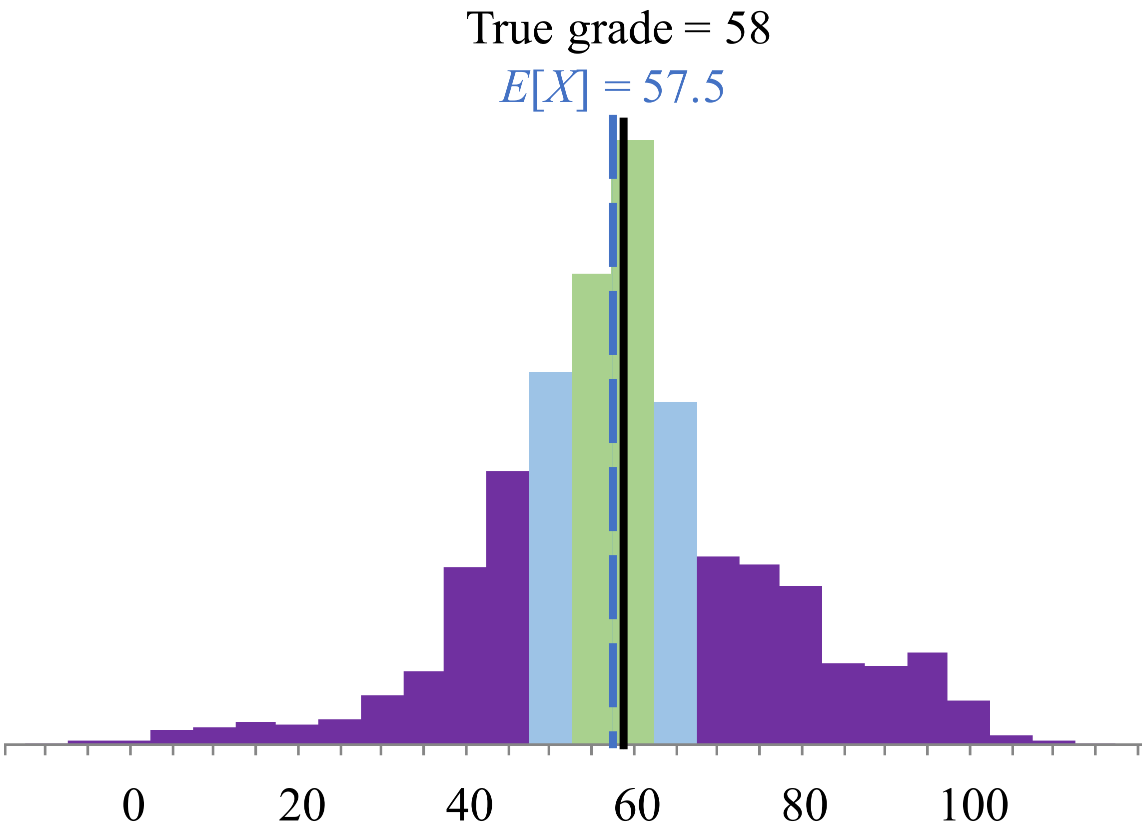


If we base a final score off 4 peer-grades, 19% of students would get a score that was off by more than 10 percentage points

Peer Grades in Coursera HCI

Let X be a random variable that represents a peer grade

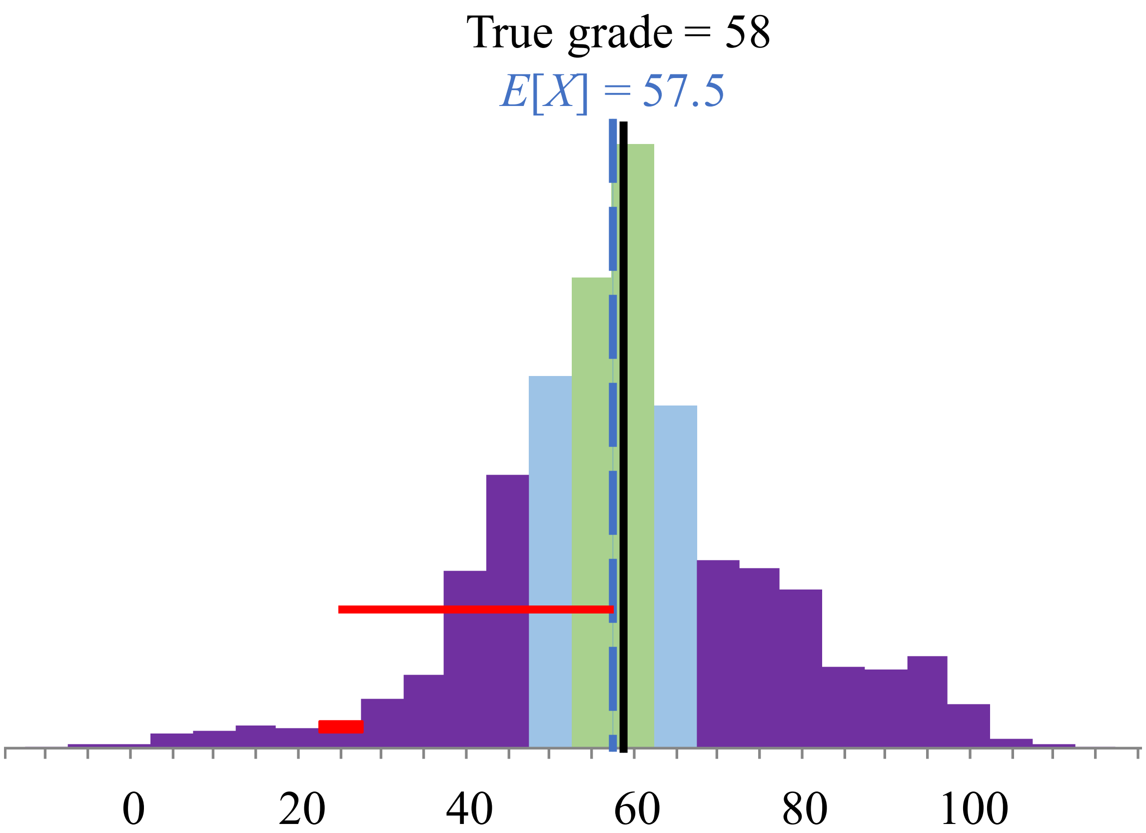
$$\mathrm{Var}(X) = E[(X-\mu)^2]$$



Peer Grades in Coursera HCI

Let X be a random variable that represents a peer grade

$$\mathrm{Var}(X) = E[(X-\mu)^2]$$

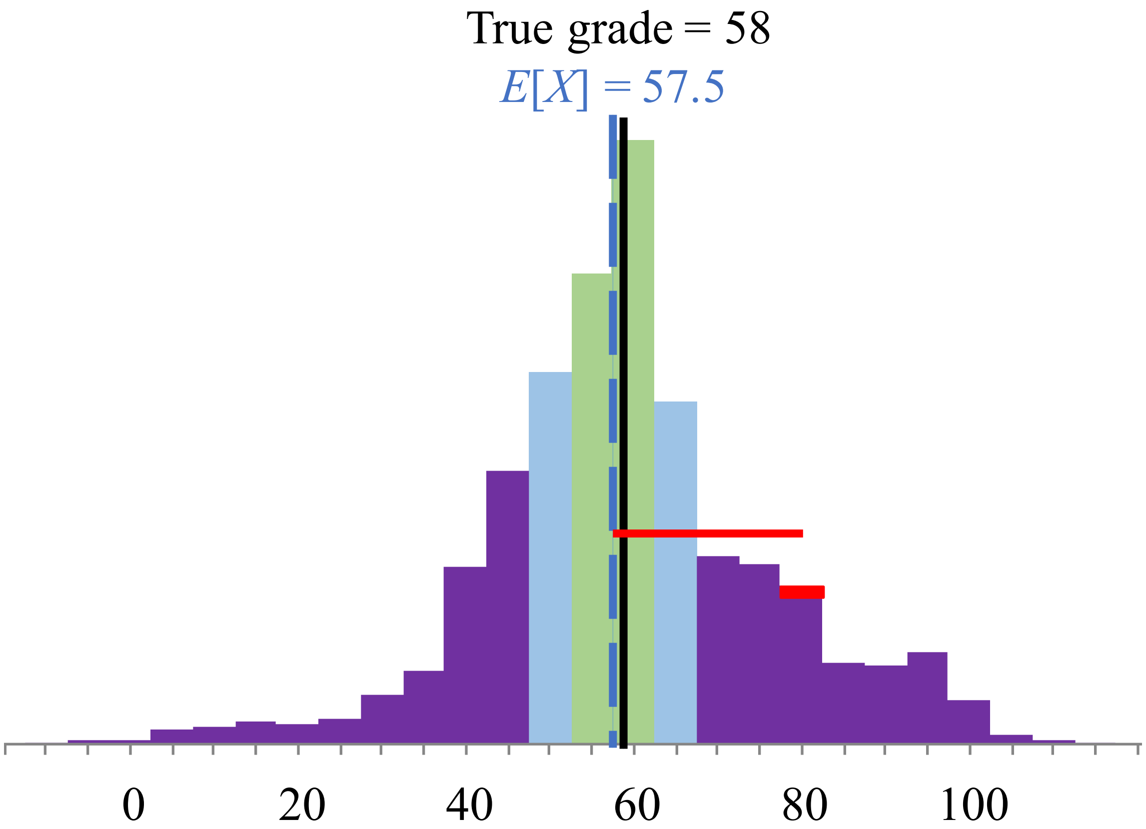


X	$(X-\mu)^2$
25 points	1056 points^2

Peer Grades in Coursera HCI

Let *X* be a random variable that represents a peer grade

$$\mathrm{Var}(X) = E[(X - \mu)^2]$$

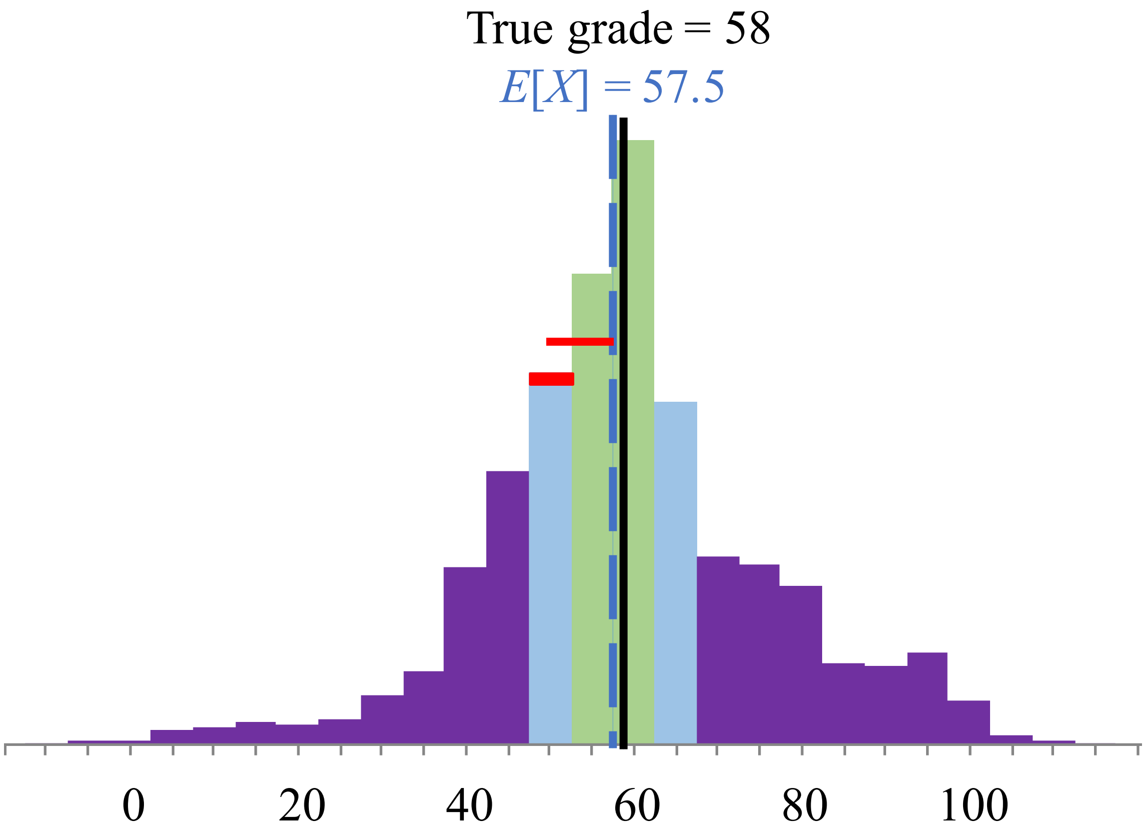


X	$(X - \mu)^2$
25 points	1056 points^2
80 points	506 points^2

Peer Grades in Coursera HCI

Let X be a random variable that represents a peer grade

$$\text{Var}(X) = E[(X - \mu)^2]$$

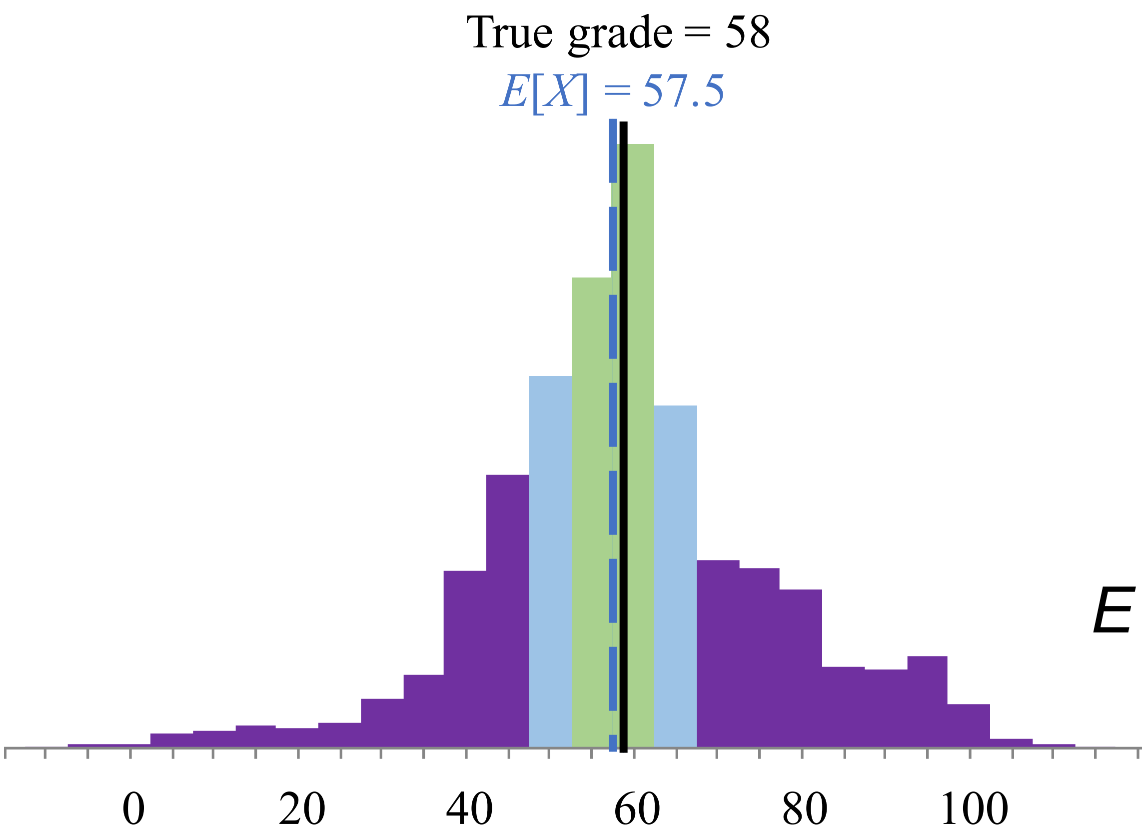


X	$(X - \mu)^2$
25 points	1056 points^2
80 points	506 points^2
50 points	56 points^2

Peer Grades in Coursera HCI

Let *X* be a random variable that represents a peer grade

$$\text{Var}(X) = E[(X - \mu)^2]$$



X	$(X - \mu)^2$
25 points	1056 points^2
80 points	506 points^2
50 points	56 points^2

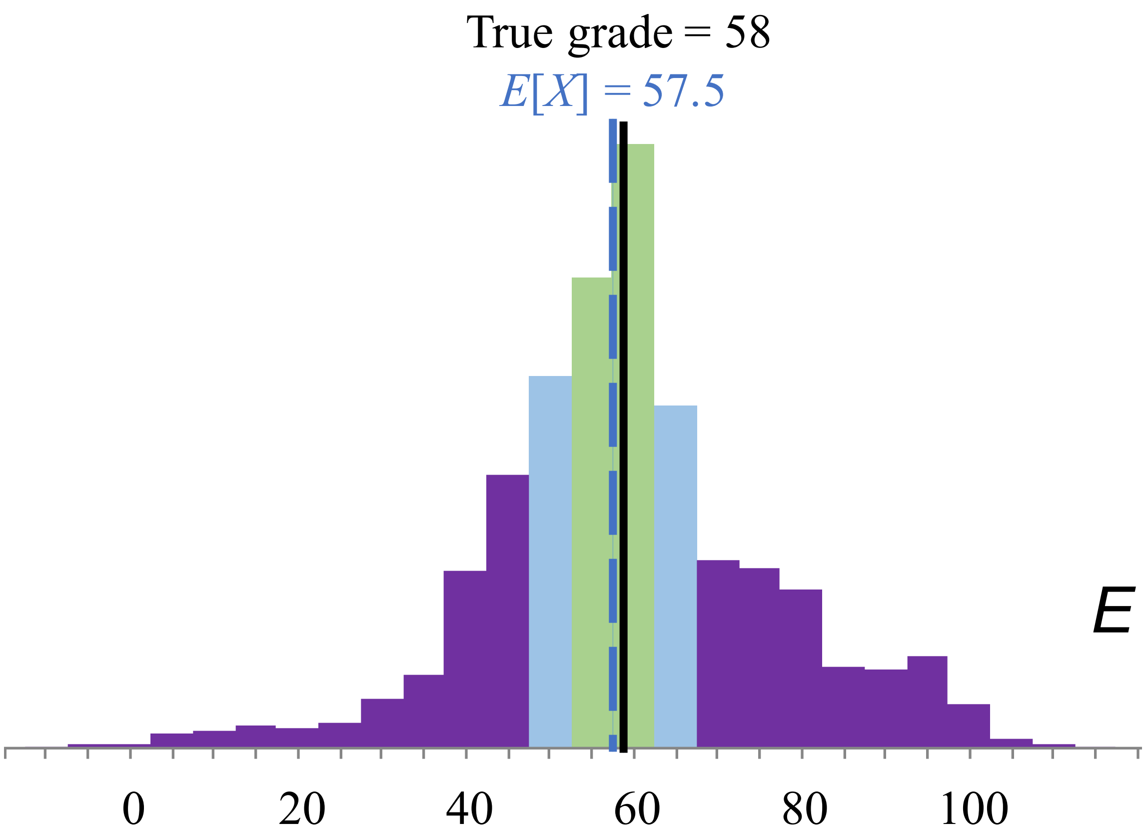
…

$$E\,[(X - \mu)^2] = 52 \text{ points}^2$$

Peer Grades in Coursera HCI

Let X be a random variable that represents a peer grade

$$\mathrm{Var}(X) = E[(X - \mu)^2]$$



X	$(X - \mu)^2$
25 points	1056 points^2
80 points	506 points^2
50 points	56 points^2

…

$$E[(X - \mu)^2] = 52 \text{ points}^2$$

$$SD(X) = 7.2 \text{ points}$$

Expected Values of Sums

$$E[X + Y] = E[X] + E[Y]$$

Generalized: $$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Holds regardless of dependency between X_i's

Expectation of Functions?

Expected Values of Functions

$$E[g(X)] = \sum_x g(x)p(x)$$

$$E[g(X,Y)] = \sum_{x,y} g(x,y)p(x,y)$$

For example:

$$E\left[\sum_{i=1}^{n} \frac{X_i}{n}\right]$$

$$E\left[\binom{X}{2}\right]$$

Product of Expectations

- Say X and Y are <u>independent</u> random variables, and $g(\bullet)$ and $h(\bullet)$ are real-valued functions

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

  - Proof:

$$E[g(X)h(Y)] = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} g(x)h(y) f_{X,Y}(x,y)\, dx\, dy$$

$$= \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} g(x)h(y) f_X(x) f_Y(y)\, dx\, dy$$

$$= \int_{x=-\infty}^{\infty} g(x) f_X(x)\, dx \cdot \int_{y=-\infty}^{\infty} h(y) f_Y(y)\, dy$$

$$= E[g(X)]E[h(Y)]$$

Variance from events

Indicators: Now with pair-wise flavor!

- Recall I_i is indicator variable for event A_i when:

$$I_i = \begin{cases} 1 & \text{if } A_i \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

  - Let X = # of events that occur:

$$\text{Var}(X) = 2\sum_{i<j} P(A_i A_j) + E[X] - (E[X])^2$$

$$= 2\sum_{i<j} P(A_i A_j) + \sum_{i=1}^{n} P(A_i) - \left(\sum_{i=1}^{n} P(A_i)\right)^2$$

Let's Try it With the Binomial

- X ~ Bin(n, p) $$E[X] = \sum_{i=1}^{n} P(A_i) = np$$
  - Each trial: X_i ~ Ber(p) $$E[X_i] = p$$
  - Let event A_i = trial *i* is success (i.e., $X_i = 1$)

$$\sum_{i<j} P(A_i A_j) = \sum_{i<j} p^2 = \binom{n}{2} p^2 = \frac{n(n-1)}{2} p^2$$

$$\text{Var}(X) = 2\sum_{i<j} P(A_i A_j) + E[X] - (E[X])^2$$

$$\text{Var}(X) = 2\frac{n(n-1)}{2}p^2 + np - (np)^2$$

Substitute in for each term

$$= n^2p^2 - np^2 + np - n^2p^2$$

$$= np - np^2 = np(1-p)$$

Expand and simplify

End Review

Computer Cluster Utilization

- Computer cluster with *k* servers
  - Requests independently go to server *i* with probability p_i
  - Let event A_i = server *i* receives no requests
  - X = # of events $A_1, A_2, \ldots A_k$ that occur
  - Y = # servers that receive ≥ 1 request = $k - X$
  - E[Y] after first *n* requests?
  - Since requests independent: $P(A_i) = (1 - p_i)^n$

$$E[X] = \sum_{i=1}^{k} P(A_i) = \sum_{i=1}^{k} (1 - p_i)^n$$

$$E[Y] = k - E[X] = k - \sum_{i=1}^{k} (1 - p_i)^n$$

when $p_i = \frac{1}{k}$ for $1 \le i \le k$, $E[Y] = k - \sum_{i=1}^{k} (1 - \frac{1}{k})^n = k\left(1 - (1 - \frac{1}{k})^n\right)$

amazon

amazon web services™

* 52% of Amazon’s Profits

**As profitable as Amazon’s North America commerce operations

http://www.zdnet.com/article/amazons-finds-its-profit-horse-in-aws-why-its-so-disruptive-to-its-old-guard/

Computer Cluster Utilization

- Computer cluster with *k* servers
  - Requests independently go to server *i* with probability p_i
  - Let event A_i = server *i* receives no requests
  - X = # of events $A_1, A_2, \ldots A_k$ that occur
  - Y = # servers that receive ≥ 1 request = *k* – X
  - Var(Y) after first *n* requests? (= $(-1)^2$ Var(X) = Var(X))
  - Independent requests: $P(A_i A_j) = (1 - p_i - p_j)^n, \quad i \neq j$

$$\mathrm{Var}(X) = 2\sum_{i<j} P(A_i A_j) + E[X] - (E[X])^2$$

$$E[X] = \sum_{i=1}^{k} (1 - p_i)^n$$

$$\mathrm{Var}(X) = 2\sum_{i<j} (1 - p_i - p_j)^n + E[X] - (E[X])^2$$

$$= 2\sum_{i<j} (1 - p_i - p_j)^n + \sum_{i=1}^{k} (1 - p_i)^n - \left(\sum_{i=1}^{k} (1 - p_i)^n \right)^2 = \mathrm{Var}(Y)$$

Dance of Covariance

The Dance of the Covariance

- Say X and Y are arbitrary random variables
- Covariance of X and Y:

$$\mathrm{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$

x	y	(x – E[X])(y – E[Y])p(x,y)
Above mean	Above mean	Positive
Bellow mean	Bellow mean	Positive
Bellow mean	Above mean	Negative
Above mean	Bellow mean	Negative

The Dance of the Covariance

- Say X and Y are arbitrary random variables
- Covariance of X and Y:

$$\mathrm{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$

- Equivalently:

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E[XY - E[X]Y - XE[Y] + E[Y]E[X]] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y]\end{aligned}$$

  - X and Y independent, E[XY] = E[X]E[Y] → Cov(X,Y) = 0
  - But Cov(X,Y) = 0 does **not** imply X and Y independent!

Independence and Covariance

- X and Y are random variables with PMF:

Y \ X	-1	0	1	$p_Y(y)$
0	1/3	0	1/3	2/3
1	0	1/3	0	1/3
$p_X(x)$	1/3	1/3	1/3	1

$$Y = \begin{cases} 0 & \text{if } X \neq 0 \\ 1 & \text{otherwise} \end{cases}$$

  - E[X] = -1(1/3) + 0(1/3) + 1(1/3) = 0
  - E[Y] = 0(2/3) + 1(1/3) = 1/3
  - Since XY = 0, E[XY] = 0
  - Cov(X, Y) = E[XY] – E[X]E[Y] = 0 – 0 = 0
- But, X and Y are clearly dependent!

Example of Covariance

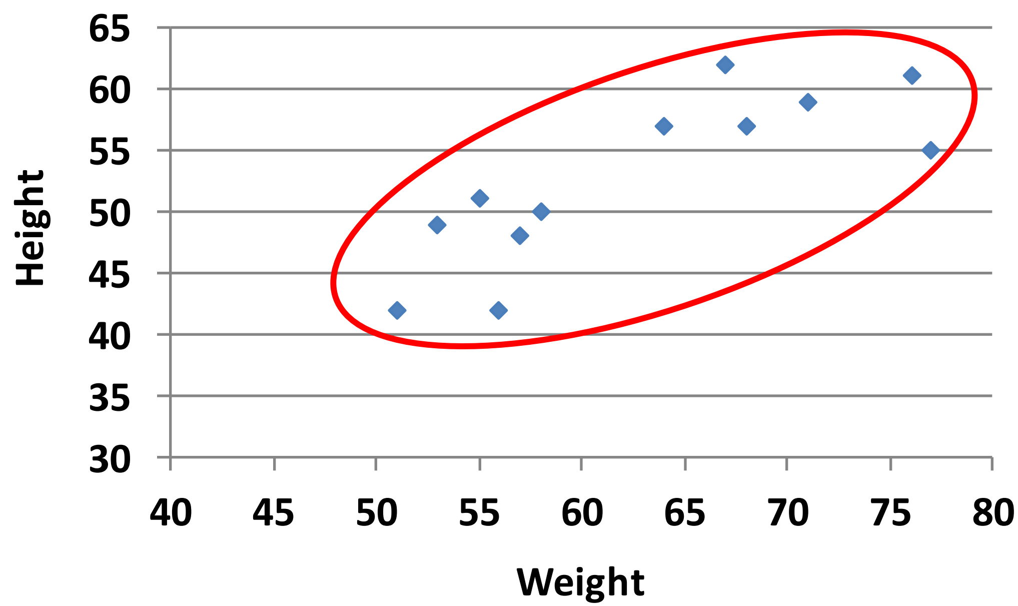
- Consider rolling a 6-sided die
  - Let indicator variable X = 1 if roll is 1, 2, 3, or 4
  - Let indicator variable Y = 1 if roll is 3, 4, 5, or 6
- What is Cov(X, Y)?
  - E[X] = 2/3 and E[Y] = 2/3
  - E[XY] $= \sum_x \sum_y xy\ p(x, y)$

    = (0 * 0) + (0 * 1/3) + (0 * 1/3) + (1 * 1/3) = 1/3
  - Cov(X, Y) = E[XY] – E[X]E[Y] = 1/3 – 4/9 = -1/9
  - Consider: P(X = 1) = 2/3 and P(X = 1 | Y = 1) = 1/2
    - Observing Y = 1 makes X = 1 *less* likely

Another Example of Covariance

- Consider the following data:

Weight	Height	Weight * Height
64	57	3648
71	59	4189
53	49	2597
67	62	4154
55	51	2805
58	50	2900
77	55	4235
57	48	2736
56	42	2352
51	42	2142
76	61	4636
68	57	3876
E[W] = 62.75	E[H] = 52.75	E[W*H] = 3355.83



Cov(W, H) = E[W*H] – E[W]E[H]

= 3355.83 – (62.75)(52.75)

= 45.77

Properties of Covariance

- Say X and Y are arbitrary random variables
  - $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
  - $\mathrm{Cov}(X,X) = E[X^2] - E[X]E[X] = \mathrm{Var}(X)$
  - $\mathrm{Cov}(aX + b, Y) = a\mathrm{Cov}(X,Y)$
- Covariance of sums of random variables
  - $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ are random variables
  - $\mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{Cov}(X_i, Y_j)$

Variance of Sum of Variables

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \mathrm{Cov}(X_i, X_j)$$

- Proof:

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \mathrm{Cov}\left(\sum_{i=1}^{n} X_i,\ \sum_{j=1}^{n} X_j\right)$$

Note: $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)$$

Variance of Sum of Variables

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \text{Cov}(X_i, X_j)$$

	X_1	X_2	X_3	X_4
X_1	$\text{Cov}(X_1,X_1)$	$\text{Cov}(X_1,X_2)$	$\text{Cov}(X_1,X_3)$	$\text{Cov}(X_1,X_4)$
X_2	$\text{Cov}(X_2,X_1)$	$\text{Cov}(X_2,X_2)$	$\text{Cov}(X_2,X_3)$	$\text{Cov}(X_2,X_4)$
X_3	$\text{Cov}(X_3,X_1)$	$\text{Cov}(X_3,X_2)$	$\text{Cov}(X_3,X_3)$	$\text{Cov}(X_3,X_4)$
X_4	$\text{Cov}(X_4,X_1)$	$\text{Cov}(X_4,X_2)$	$\text{Cov}(X_4,X_3)$	$\text{Cov}(X_4,X_4)$

Variance of Sum of Variables

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)$$

	X_1	X_2	X_3	X_4
X_1	$\mathrm{Cov}(X_1,X_1)$	$\mathrm{Cov}(X_1,X_2)$	$\mathrm{Cov}(X_1,X_3)$	$\mathrm{Cov}(X_1,X_4)$
X_2	$\mathrm{Cov}(X_2,X_1)$	$\mathrm{Cov}(X_2,X_2)$	$\mathrm{Cov}(X_2,X_3)$	$\mathrm{Cov}(X_2,X_4)$
X_3	$\mathrm{Cov}(X_3,X_1)$	$\mathrm{Cov}(X_3,X_2)$	$\mathrm{Cov}(X_3,X_3)$	$\mathrm{Cov}(X_3,X_4)$
X_4	$\mathrm{Cov}(X_4,X_1)$	$\mathrm{Cov}(X_4,X_2)$	$\mathrm{Cov}(X_4,X_3)$	$\mathrm{Cov}(X_4,X_4)$

Variance of Sum of Variables

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)$$

	X_1	X_2	X_3	X_4
X_1	**Var(X_1)**	Cov(X_1,X_2)	Cov(X_1,X_3)	Cov(X_1,X_4)
X_2	Cov(X_2,X_1)	**Var(X_2)**	Cov(X_2,X_3)	Cov(X_2,X_4)
X_3	Cov(X_3,X_1)	Cov(X_3,X_2)	**Var(X_3)**	Cov(X_3,X_4)
X_4	Cov(X_4,X_1)	Cov(X_4,X_2)	Cov(X_4,X_3)	**Var(X_4)**

Variance of Sum of Variables

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)$$

	X_1	X_2	X_3	X_4
X_1	**Var(X_1)**	Cov(X_1,X_2)	Cov(X_1,X_3)	Cov(X_1,X_4)
X_2	Cov(X_2,X_1)	**Var(X_2)**	Cov(X_2,X_3)	Cov(X_2,X_4)
X_3	Cov(X_3,X_1)	Cov(X_3,X_2)	**Var(X_3)**	Cov(X_3,X_4)
X_4	Cov(X_4,X_1)	Cov(X_4,X_2)	Cov(X_4,X_3)	**Var(X_4)**

Variance of Sum of Variables

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)$$

	X_1	X_2	X_3	X_4
X_1	**$\mathrm{Var}(X_1)$**	$\mathrm{Cov}(X_1,X_2)$	$\mathrm{Cov}(X_1,X_3)$	$\mathrm{Cov}(X_1,X_4)$
X_2	$\mathrm{Cov}(X_2,X_1)$	**$\mathrm{Var}(X_2)$**	$\mathrm{Cov}(X_2,X_3)$	$\mathrm{Cov}(X_2,X_4)$
X_3	**$\mathrm{Cov}(X_1,X_3)$**	$\mathrm{Cov}(X_3,X_2)$	**$\mathrm{Var}(X_3)$**	$\mathrm{Cov}(X_3,X_4)$
X_4	$\mathrm{Cov}(X_4,X_1)$	$\mathrm{Cov}(X_4,X_2)$	$\mathrm{Cov}(X_4,X_3)$	**$\mathrm{Var}(X_4)$**

Variance of Sum of Variables

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)$$

	X_1	X_2	X_3	X_4
X_1	**Var(X_1)**	Cov(X_1,X_2)	Cov(X_1,X_3)	Cov(X_1,X_4)
X_2	**Cov(X_1,X_2)**	**Var(X_2)**	Cov(X_2,X_3)	Cov(X_2,X_4)
X_3	**Cov(X_1,X_3)**	**Cov(X_2,X_3)**	**Var(X_3)**	Cov(X_3,X_4)
X_4	**Cov(X_1,X_4)**	**Cov(X_2,X_4)**	**Cov(X_3,X_4)**	**Var(X_4)**

Variance of Sum of Variables

- $$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \mathrm{Cov}(X_i, X_j)$$

  - Proof:

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \mathrm{Cov}\left(\sum_{i=1}^{n} X_i,\ \sum_{j=1}^{n} X_j\right)$$

Note: $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \mathrm{Cov}(X_i, X_j)$$

By symmetry:
$\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$

$$= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \mathrm{Cov}(X_i, X_j)$$

  - If all X_i and X_j independent ($i \neq j$): $\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$

Three Ways of Computing Variance of a Sum!

Hola Amiga: La Distribución Binomial

- Let $Y \sim \text{Bin}(n, p)$
  - *n* independent trials
  - Let $X_i = 1$ if *i*-th trial is "success", 0 otherwise
  - $X_i \sim \text{Ber}(p)$ $\quad E[X_i] = p$
  - $\text{Var}(Y) = \text{Var}(X_1) + \text{Var}(X_2) + \ldots + \text{Var}(X_n)$
  - $\text{Var}(X_i) = E[X_i^2] - (E[X_i])^2$
    $= E[X_i] - (E[X_i])^2$ since $X_i^2 = X_i$
    $= p - p^2 = p(1 - p)$
  - $\text{Var}(Y) = n\text{Var}(X_i) = np(1 - p)$

Any Indication of Similar Funk?

- Let I_A and I_B be indicators for events A and B

$$I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases} \qquad I_B = \begin{cases} 1 & \text{if } B \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

  - $E[I_A] = P(A)$, $E[I_B] = P(B)$, $E[I_A I_B] = P(AB)$
  - $\text{Cov}(I_A, I_B) = E[I_A I_B] - E[I_A]\,E[I_B]$
    $= P(AB) - P(A)P(B)$
    $= P(A \mid B)P(B) - P(A)P(B)$
    $= P(B)[P(A \mid B) - P(A)]$
  - $\text{Cov}(I_A, I_B)$ determined by $P(A \mid B) - P(A)$
  - $P(A \mid B) > P(A) \Rightarrow \rho(I_A, I_B) > 0$
  - $P(A \mid B) = P(A) \Rightarrow \rho(I_A, I_B) = 0$ (and $\text{Cov}(I_A, I_B) = 0$)
  - $P(A \mid B) < P(A) \Rightarrow \rho(I_A, I_B) < 0$

Egg Samples

Machine Learning Example

- You want to know the true mean and variance of happiness in Buthan
  - But you can't ask everyone.
  - Randomly sample 200 people.
  - Your data looks like this:

Happiness = {72, 85, 79, 91, 68, … , 71}

  - The mean of all of those numbers is 83. Is that the true average happiness of Bhutanese people?

Sample Mean

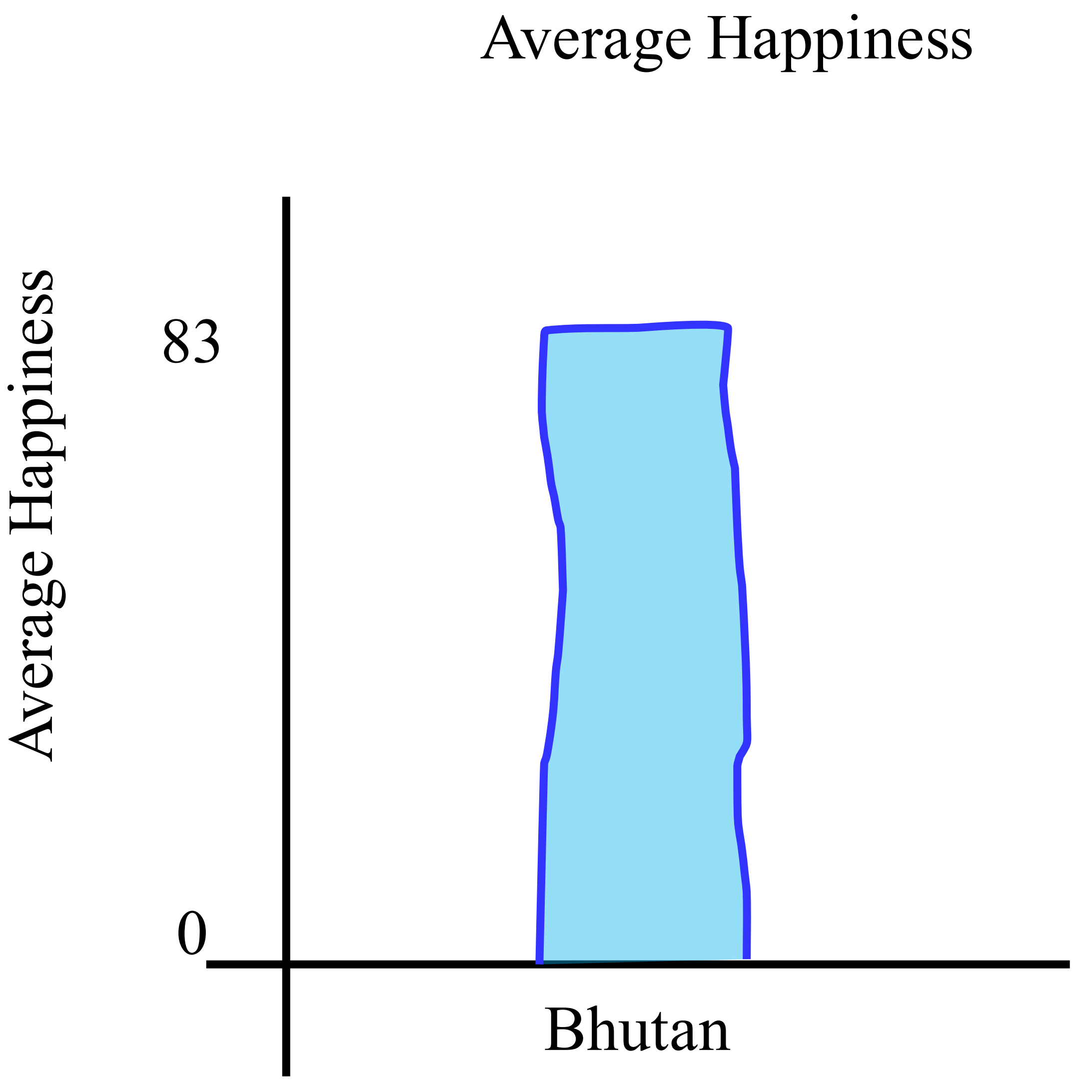
- Consider *n* random variables $X_1, X_2, \ldots X_n$
  - X_i are all independently and identically distributed (I.I.D.)
  - Have same distribution function *F* and $E[X_i] = \mu$
  - We call sequence of X_i a **sample** from distribution *F*
  - Sample mean: $\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$
  - Compute $E[\bar{X}]$

$$E[\bar{X}] = E\left[\sum_{i=1}^{n} \frac{X_i}{n}\right] = \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} n\mu = \mu$$

  - $\bar{X}$ is “unbiased” estimate of μ $(E[\bar{X}] = \mu)$

Sample Mean



Sample Variance

- Consider *n* I.I.D. random variables $X_1, X_2, \dots X_n$
  - X_i have distribution *F* with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$
  - We call sequence of X_i a **<u>sample</u>** from distribution *F*
  - Recall sample mean: $\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ where $E[\bar{X}] = \mu$
  - Sample deviation: $\bar{X} - X_i$ for i = 1, 2, ..., n
  - Sample variance: $S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$

  - What is $E[S^2]$?
  - $E[S^2] = \sigma^2$
  - We say S^2 is "unbiased estimate" of σ^2

Proof that E[S^2] = σ² (just for reference)

$$E[S^2] = E\left[\sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}\right] \Rightarrow (n-1)E[S^2] = E\left[\sum_{i=1}^{n} (X_i - \bar{X})^2\right]$$

$$(n-1)E[S^2] = E\left[\sum_{i=1}^{n} (X_i - \bar{X})^2\right] = E\left[\sum_{i=1}^{n} ((X_i - \mu) + (\mu - \bar{X}))^2\right]$$

$$= E\left[\sum_{i=1}^{n} (X_i - \mu)^2 + \sum_{i=1}^{n} (\mu - \bar{X})^2 + 2\sum_{i=1}^{n} (X_i - \mu)(\mu - \bar{X})\right]$$

$$= E\left[\sum_{i=1}^{n} (X_i - \mu)^2 + n(\mu - \bar{X})^2 + 2(\mu - \bar{X})\sum_{i=1}^{n} (X_i - \mu)\right]$$

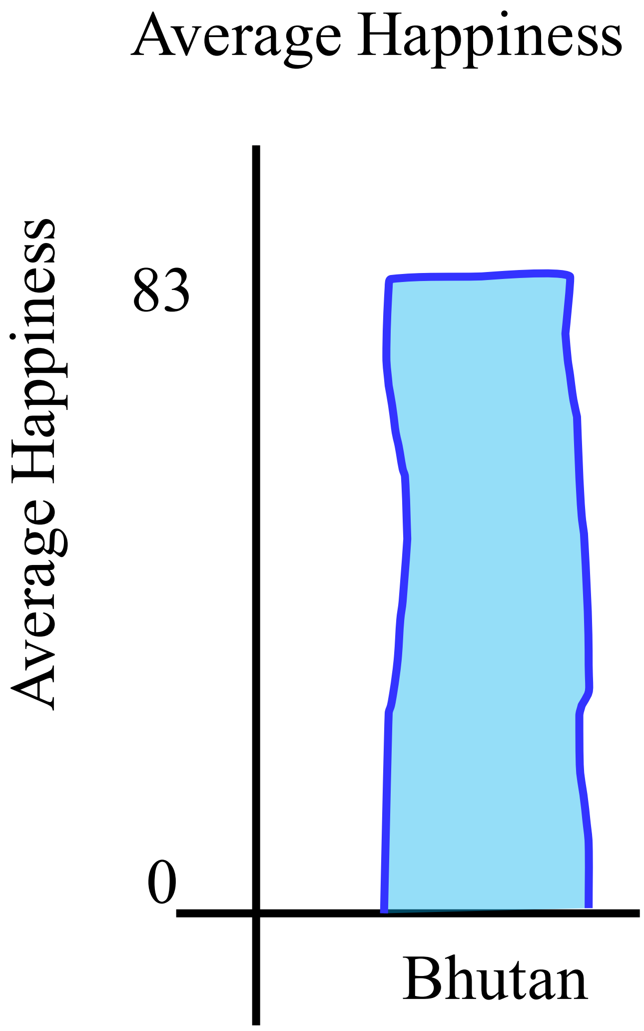
$$= E\left[\sum_{i=1}^{n} (X_i - \mu)^2 + n(\mu - \bar{X})^2 + 2(\mu - \bar{X})n(\bar{X} - \mu)\right]$$

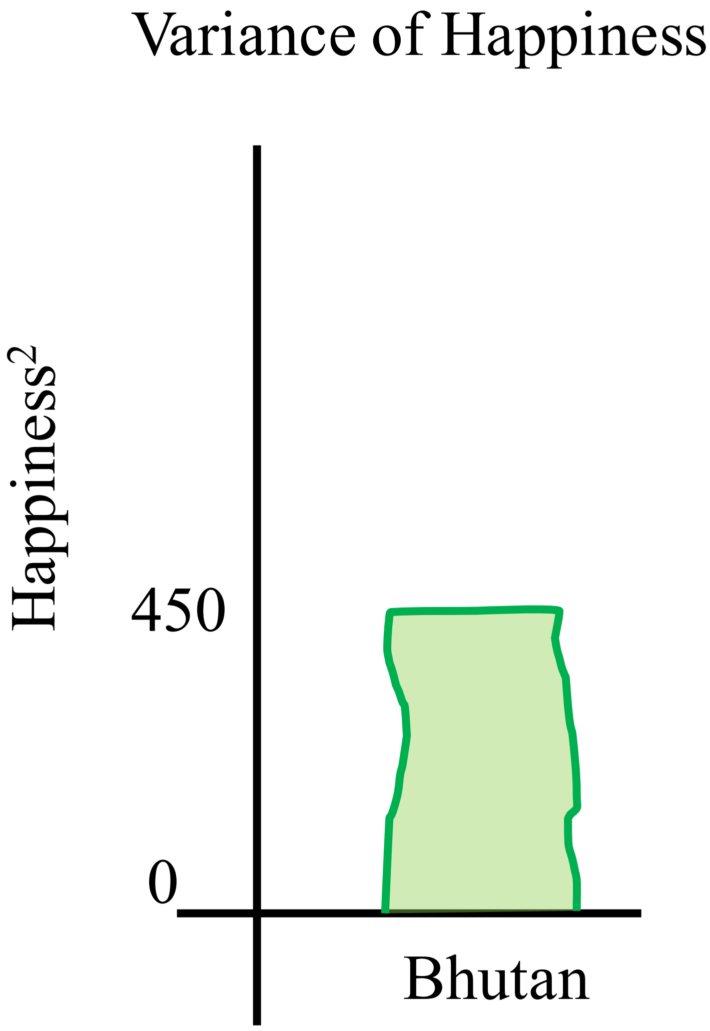
$$= E\left[\sum_{i=1}^{n} (X_i - \mu)^2 - n(\mu - \bar{X})^2\right] = \sum_{i=1}^{n} E[(X_i - \mu)^2] - nE[(\mu - \bar{X})^2]$$

$$= n\sigma^2 - n\text{Var}(\bar{X}) = n\sigma^2 - n\frac{\sigma^2}{n} = n\sigma^2 - \sigma^2 = (n-1)\sigma^2$$

- So, E[S^2] = σ^2

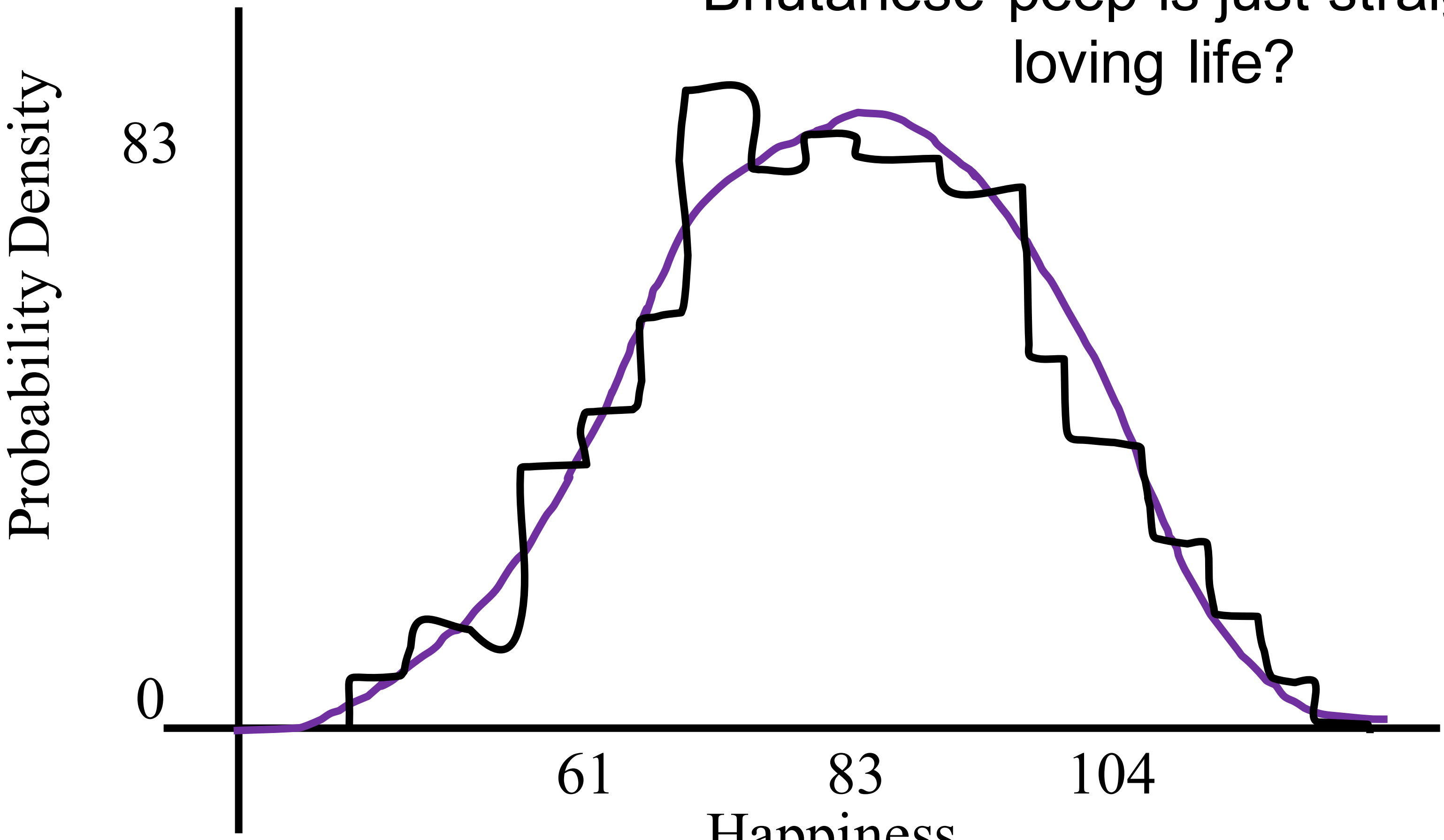
Sample Mean


Average Happiness
Average Happiness
83
0
Bhutan

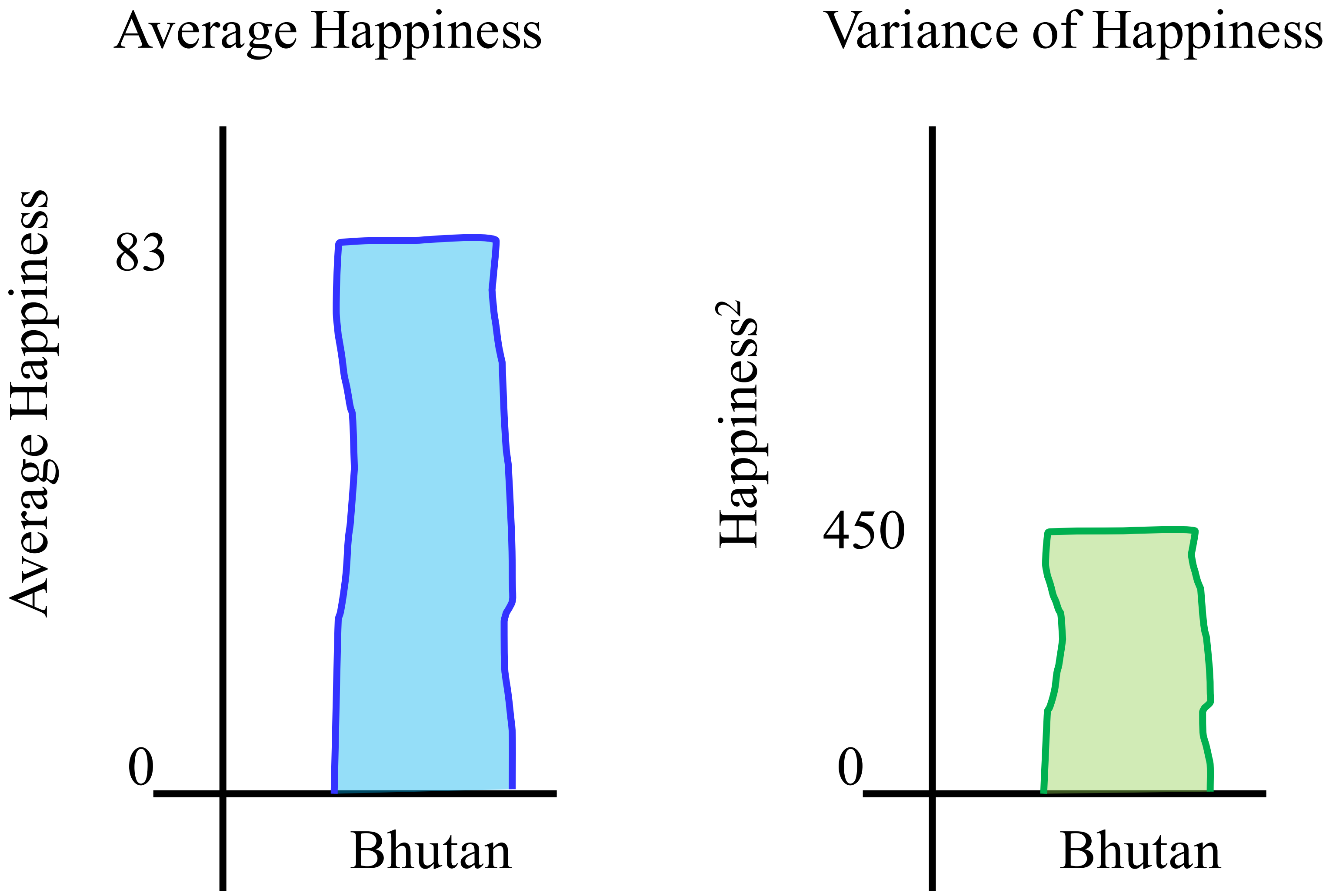

Variance of Happiness
Happiness^2
450
0
Bhutan

Happiness of Bhutan

What is the probability that a Bhutanese peep is just straight up loving life?



Sample Mean


Average Happiness
Variance of Happiness
Average Happiness
83
0
Bhutan
Happiness²
450
0
Bhutan

No Error Bars ☹

Variance of Sample Mean

- Consider n I.I.D. random variables $X_1, X_2, \ldots X_n$
  - X_i have distribution F with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$
  - We call sequence of X_i a **sample** from distribution F
  - Recall sample mean: $\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ where $E[\bar{X}] = \mu$
  - What is $\mathrm{Var}(\bar{X})$?

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \left(\frac{1}{n}\right)^2 \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)$$

$$= \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \mathrm{Var}(X_i) = \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \sigma^2 = \left(\frac{1}{n}\right)^2 n\sigma^2$$

$$= \frac{\sigma^2}{n}$$

Standard Error of the Mean

$$\text{Var}(\bar{X}) = \text{Var}\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \left(\frac{1}{n}\right)^2 \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{\sigma^2}{n}$$

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

$$= \frac{S^2}{n}$$

Since S_2 is an unbiased estimate

$$\text{Std}(\bar{X}) = \sqrt{\frac{S^2}{n}}$$

Change variance to standard deviation

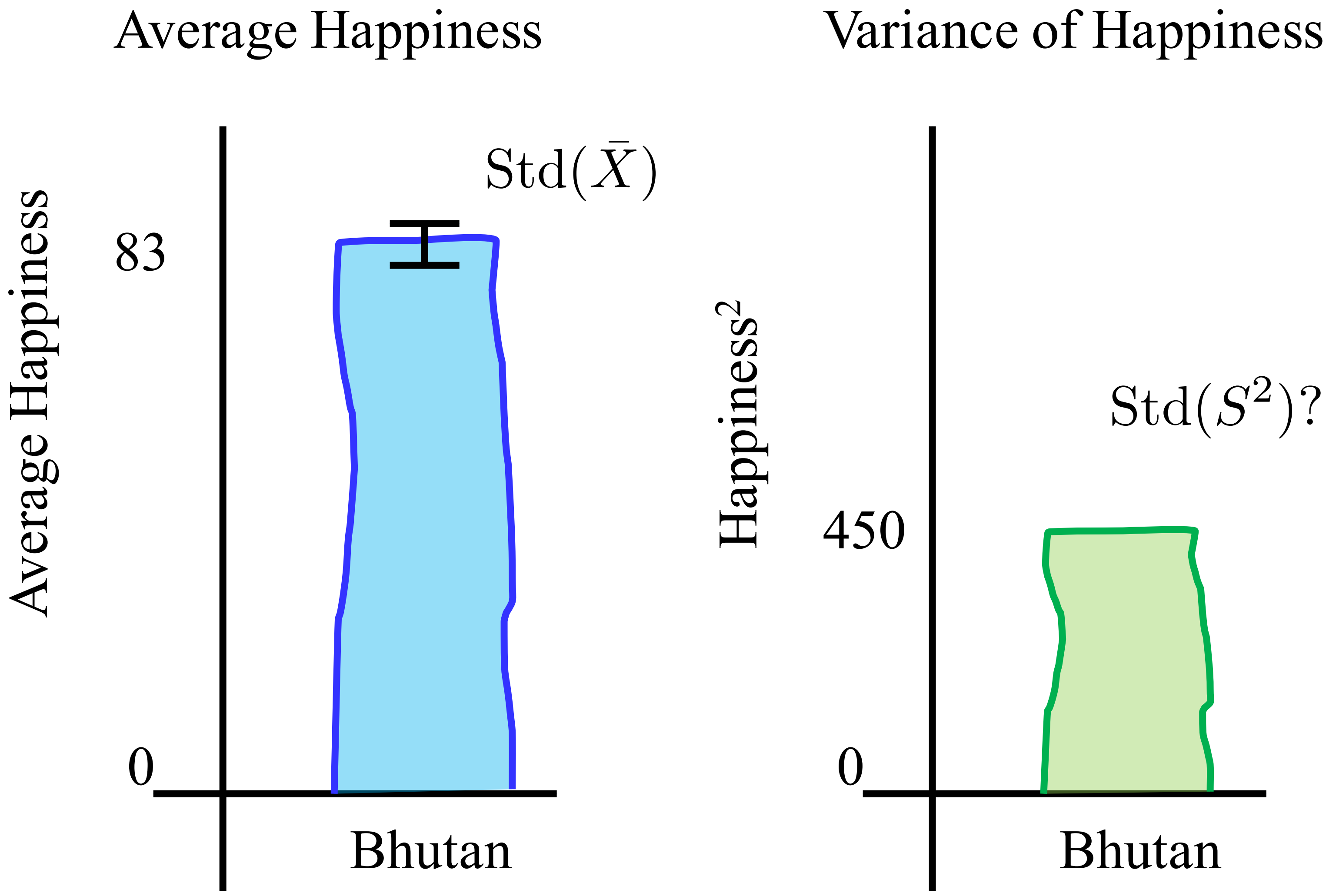
$$= \sqrt{\frac{450}{200}}$$

The numbers for our Bhutanese poll

$$= 1.5$$

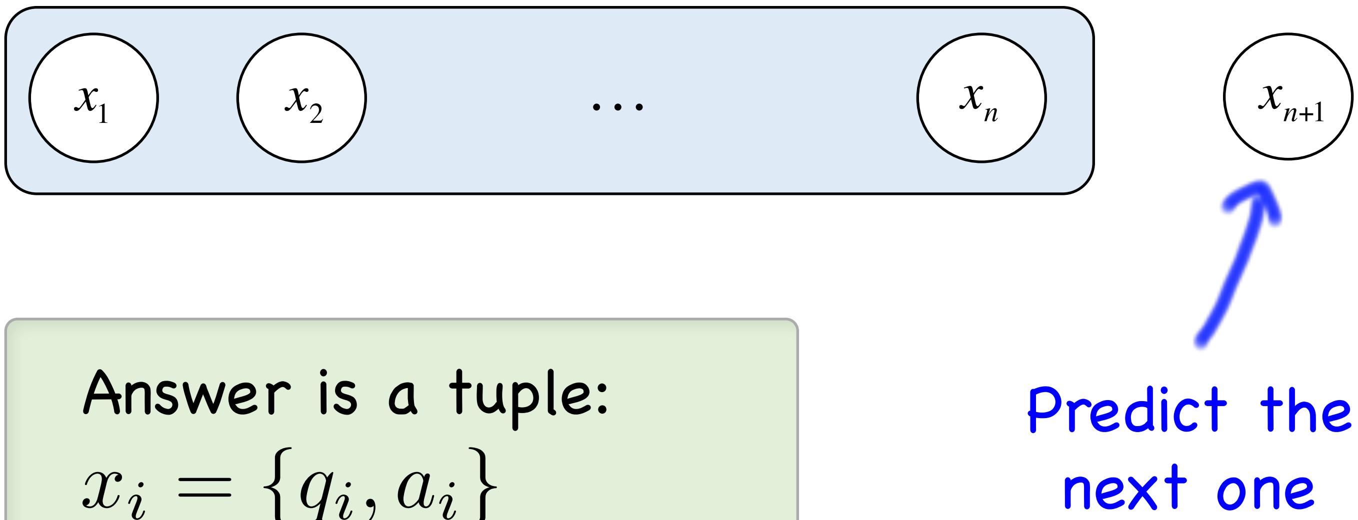
Bhutanese standard error of the mean

Sample Mean


Average Happiness
Variance of Happiness
Average Happiness
83
0
$\mathrm{Std}(\bar{X})$
Bhutan
Happiness^2
450
0
$\mathrm{Std}(S^2)$?
Bhutan

Knowledge Tracing

Given *n* historical answers:



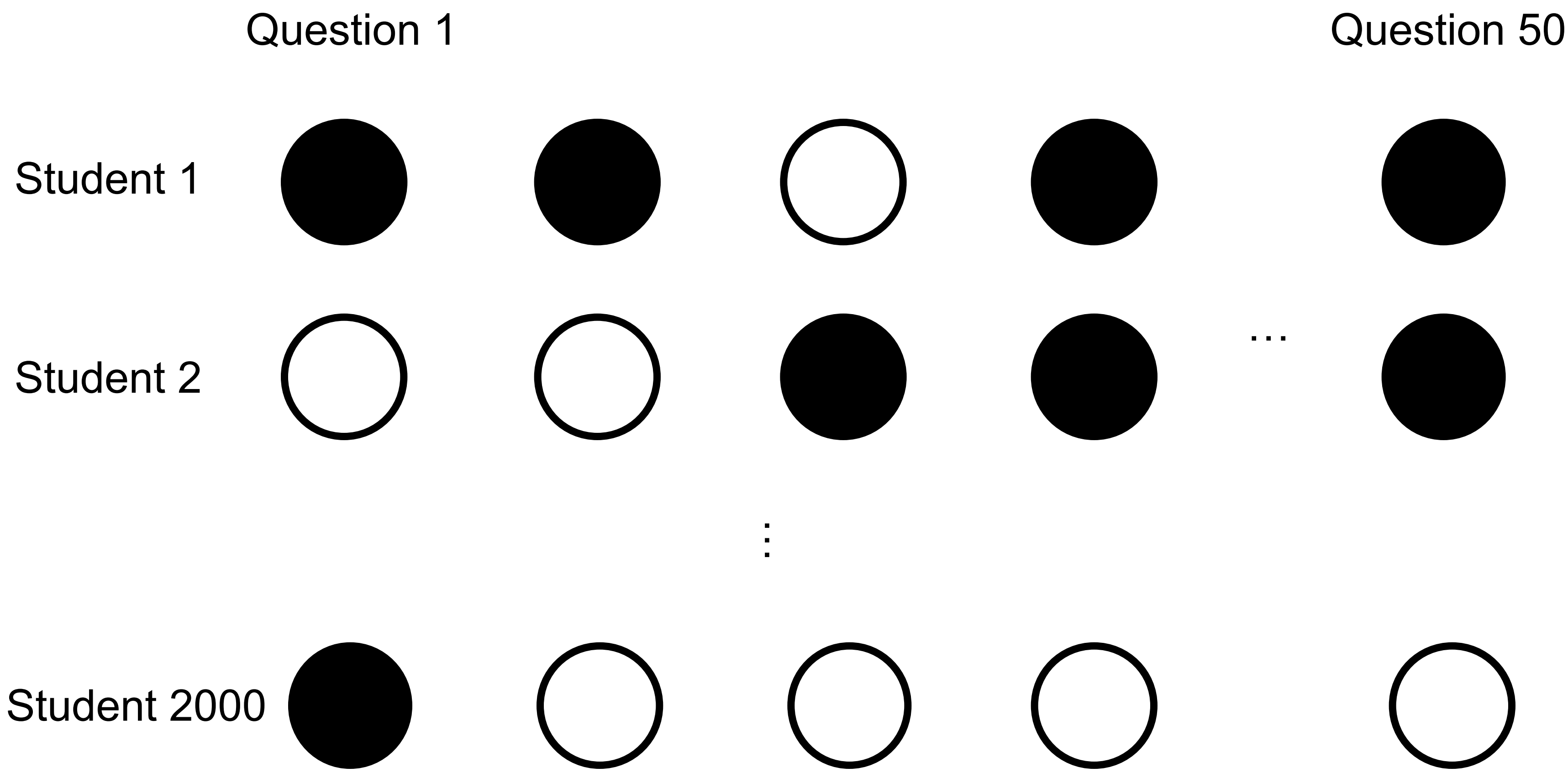
Answer is a tuple:

$$x_i = \{q_i, a_i\}$$

Question id

Correct or not

Synthetic Data



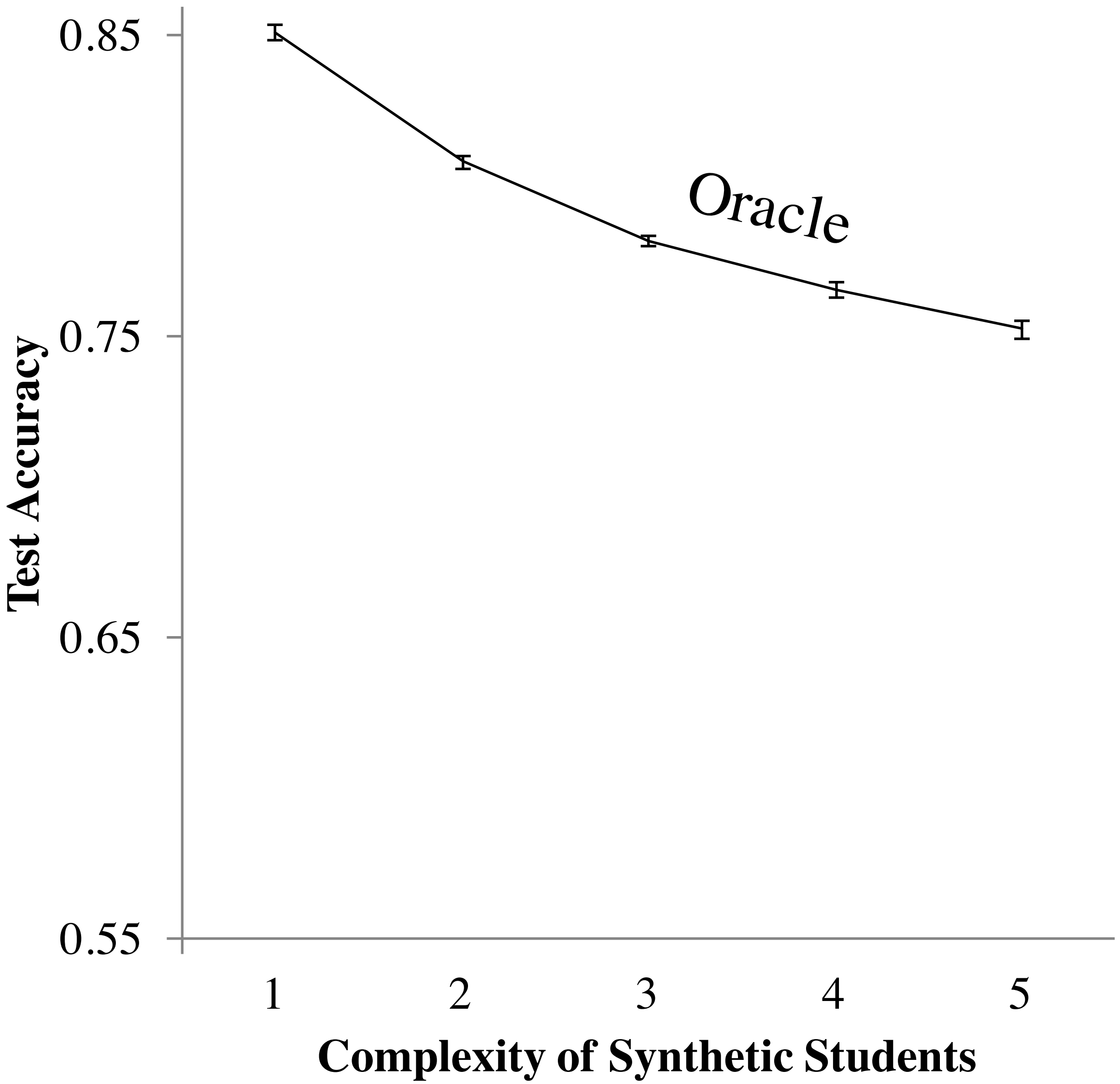
Each question corresponds to one of *n* concepts. Results based on IRT model

Machine Learning Example

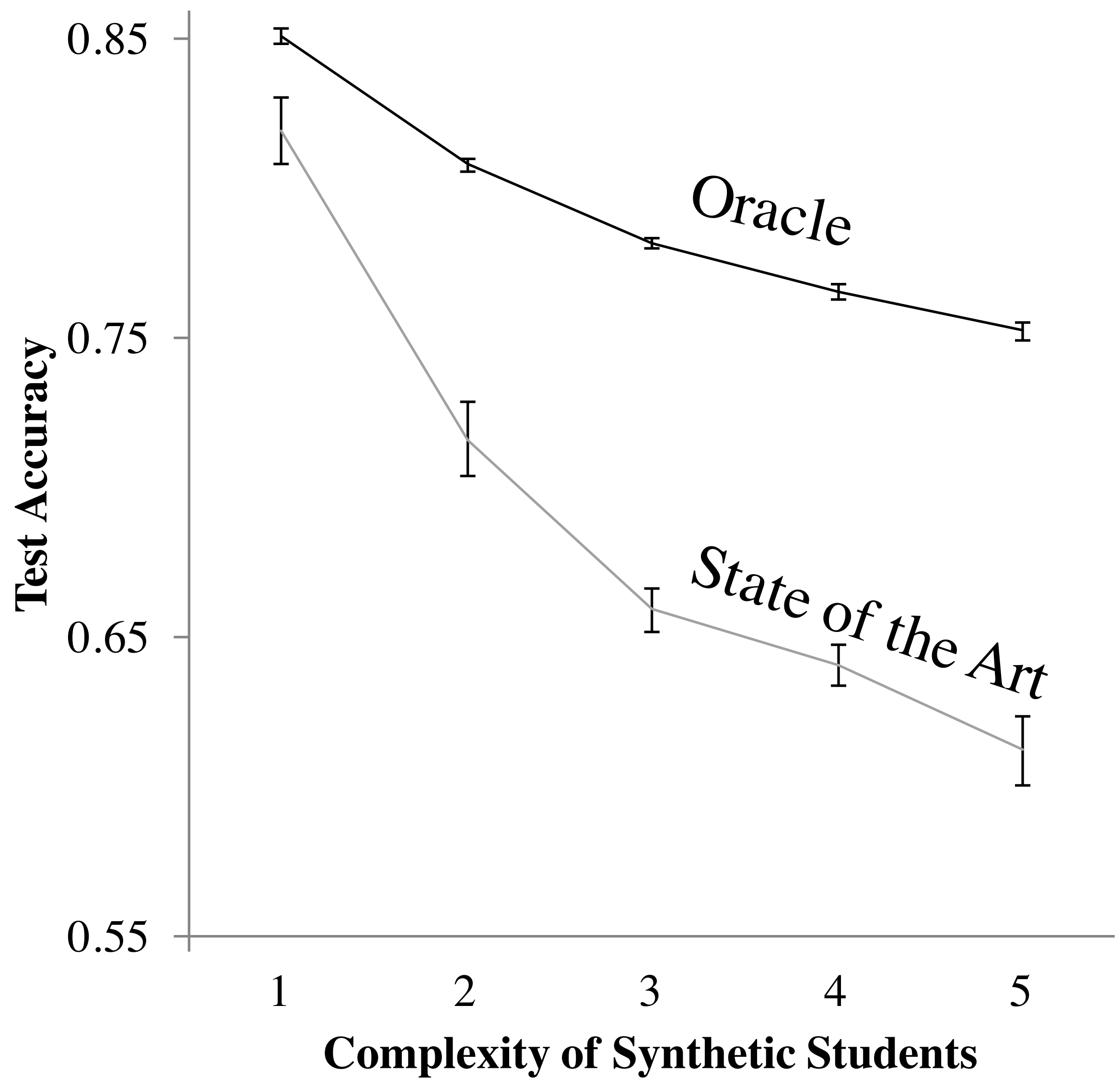
- You have a machine learning algorithm.
  - You can randomly generate input dataset
  - For each input dataset you can measure accuracy
  - Do you have to try infinite experiments to know your true average accuracy?

Accuracies = {0.72, 0.85, 0.79, 0.91, 0.68, 0.71}

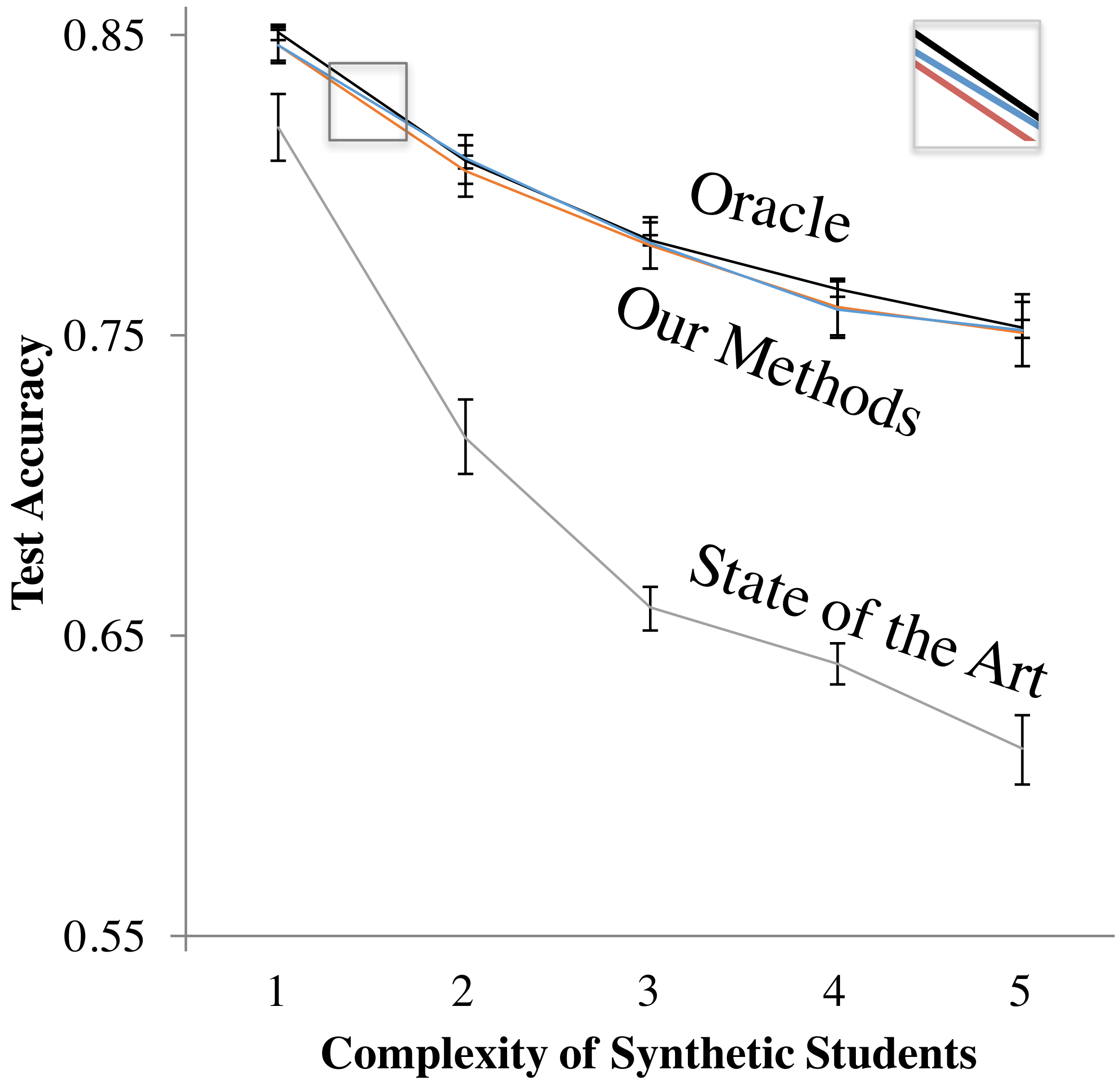
Item Response Theory Data



Item Response Theory Data



Item Response Theory Data



Correlation (given enough time)

Viva La Correlatión

- Say X and Y are arbitrary random variables
  - Correlation of X and Y, denoted $\rho(X, Y)$:

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var(X)Var(Y)}}}$$

  - Note: $-1 \leq \rho(X, Y) \leq 1$
  - Correlation measures <u>linearity</u> between X and Y
  - $\rho(X, Y) = 1$ $\Rightarrow$ Y = aX + b where $a = \sigma_y/\sigma_x$
  - $\rho(X, Y) = -1$ $\Rightarrow$ Y = aX + b where $a = -\sigma_y/\sigma_x$
  - $\rho(X, Y) = 0$ $\Rightarrow$ absence of <u>linear</u> relationship
    - But, X and Y can still be related in some other way!
  - If $\rho(X, Y) = 0$, we say X and Y are "uncorrelated"
    - Note: Independence implies uncorrelated, but <u>**not**</u> vice versa!

Que te vayas bien