Lecture 20: Data fitting

People make all kind of statistics. Here is one about the number of kills in the Rambo series.

	I: "First Blood" (1982)	II: "Rambo: First Blood Part II" (1985)	III: "Rambo III" (1988)	IV: "Rambo" (2008)
Number of bad guys killed by Rambo with his shirt on	1	12	33	83
Number of bad guys killed by Rambo with his shirt off	0	46	45	0
Number of bad guys killed by Rambo no matter how attired	1	58	78	83
Number of bad guys killed by accomplices of Rambo acting on their own	0	10	17	40
Number of good guys killed by bad guys	0	1	37	113
Total number of people killed	1	69	132	236
Number of people killed per minute	0.01	0.72	1.30	2.59
Time at which the first person is killed (mins:secs)	29:31	33:34	41:9	3:22
Number of people killed per minute from that point until the end of the film (not including the ending credits)	0.02	1.18	2.39	3.04
Sequences in which Rambo is shot at without significant result	12	24	38	2
Number of sequences in which good guys are tortured by bad guys	2	5	7	3
Number of sex scenes	0	0	0	0

JOHN MUELLER *holds the Woody Hayes Chair of National Security Studies at Ohio State University. Among his books are "The Remnants of War" and "Overblown: How Politicians and the Terrorism Industry Inflate National Security Threats, and Why We Believe Them."*

Assume we would have want to know how many bad good guys will be killed by bad guys in Rambo IV and we know the data for Rambo I,II,III. To find out, we have to find the best linear fit through the points $(0,0),(1,1),(2,37)$.

$$\begin{aligned} 1r + 1s &= 0 \\ 2r + 1s &= 1 \\ 3r + 1s &= 37 \end{aligned}$$

$\vec{x} = (a, b)$. The best solution $\vec{x}^*$ to $A\vec{x} = \vec{b}$. is $\vec{x}^* = (A^T A)^{-1} A^T \vec{b}$.

Write down the matrix A and A^T.

What is $\vec{b}$ and $A^T\vec{b}$.

Find $(A^T A)$ and its inverse.

Compute $\vec{x}^* = (A^T A)^{-1} A^T \vec{b}$

What line $rx + s$ best fits the data?

What would the predicted loss of good guys for Rambo V be?