Lecture 1: Introduction, logistics, review

STATS 101: Foundations of Statistics

Linh Tran
ThetaHat.AI@gmail.com

November 14, 2019

About me

Takeda
ONCOLOGY

REMOTE AREA MEDICAL
RAM
VOLUNTEER CORPS

Google

WORST COOKS IN AMERICA

UCSF
University of California
San Francisco

Theta Hat

Cerebral Palsy • Vegetative State • Autism • Downs
Life Expectancy Project
Traumatic Brain Injury • Spinal Cord Injury

NATIONAL REGISTRY — EMERGENCY MEDICAL TECHNICIANS

About me

I'm here to teach you statistics!

About me

I'm here to teach you statistics!

AKA Data science. AKA Big data. AKA Machine learning. AKA Artificial Intelligence. AKA ⟨`Next hype term here`⟩

Statistics

Question: What is statistics?

Statistics

Question: What is statistics?

Origin: Derived from Latin statisticum collegium ("council of state")

Statistics

Question: What is statistics?

Origin: Derived from Latin statisticum collegium ("council of state")
Merriam-Webster: a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.

Statistics

Question: What is statistics?

Origin: Derived from Latin statisticum collegium ("council of state")
Merriam-Webster: a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.
Lay impression: the collection, summary, and analysis of data of any type.

Statistics

Question: What is statistics?

Origin: Derived from Latin statisticum collegium ("council of state")
Merriam-Webster: a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.
Lay impression: the collection, summary, and analysis of data of any type.
My impression: The study of variability.

Data Scientist:

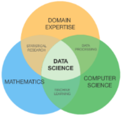
The Sexiest Job of the 21st Century

A.I. Researchers Are Making More Than $1 Million, Even at a Nonprofit

Earn a $1.5 Million Prize at 'Kaggle!' (American Applicants Only, Please.)

- **Once again, data scientist ranks as the best job in America, according to employees.**

Github tops 40 million developers as Python, data science, machine learning popularity surges



Course Series

Current plan

- STATS 101: Foundations of Statistics
- STATS 102: Introduction to Data Analysis
- STATS 103: Introduction to Statistical Learning

Course Series

Current plan

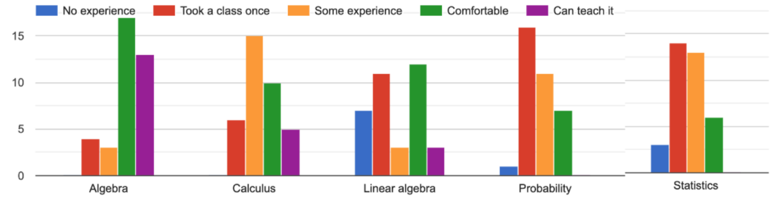
- STATS 101: Foundations of Statistics
- STATS 102: Introduction to Data Analysis
- STATS 103: Introduction to Statistical Learning

Wishful thinking

- STATS 204, 205, ...: Machine Learning, Deep Learning, etc.
- STATS 214, 215, ...: Survival analysis, clinical trial design, genomics, etc.
- STATS 224, 225, ...: Statistical theory, Semi-parametric efficiency, etc.

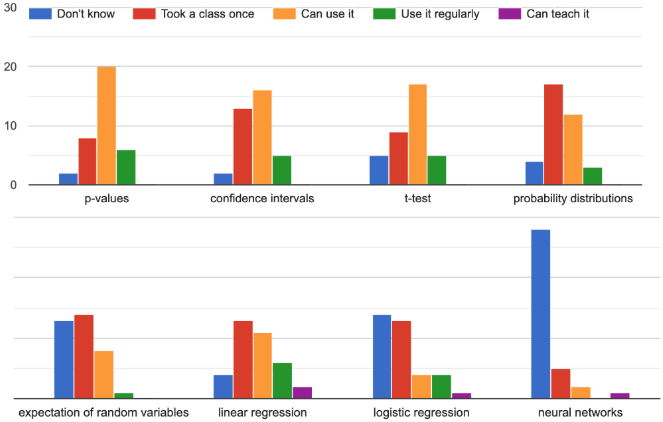
Preliminary survey

Math background



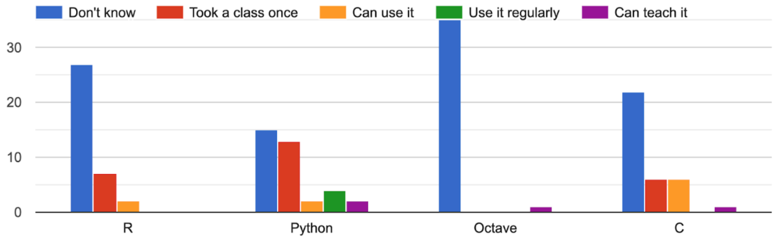
Preliminary survey

Statistical experience



Preliminary survey

Coding experience



Preliminary survey

Summary

- **Math**:
  - Some haven't really used algebra/calculus; many have
  - About half are weak on linear algebra
  - Most are not comfortable/experts at probability & statistics

Preliminary survey

Summary

- **Math**:
  - Some haven't really used algebra/calculus; many have
  - About half are weak on linear algebra
  - Most are not comfortable/experts at probability & statistics
- **Statistics**:
  - Most are weak to uncomfortable with statistical concepts

Preliminary survey

Summary

- **Math**:
  - Some haven't really used algebra/calculus; many have
  - About half are weak on linear algebra
  - Most are not comfortable/experts at probability & statistics
- **Statistics**:
  - Most are weak to uncomfortable with statistical concepts
- **Coding**:
  - Very few people code

Preliminary survey

Summary

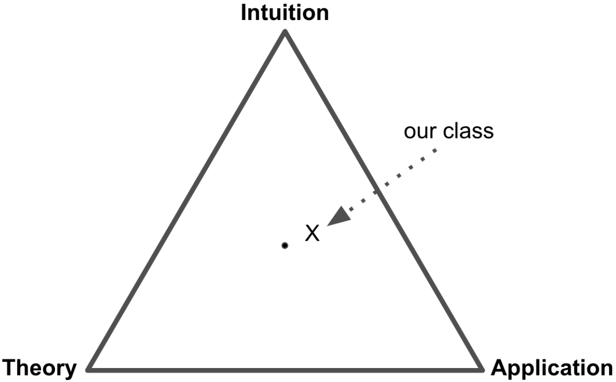
- **Math**:
  - Some haven't really used algebra/calculus; many have
  - About half are weak on linear algebra
  - Most are not comfortable/experts at probability & statistics
- **Statistics**:
  - Most are weak to uncomfortable with statistical concepts
- **Coding**:
  - Very few people code

Conclusion: Split the class into two

1. Slower & more friendly class
2. Fast & more in-depth class

Course mixture

Proportion of concepts targeted



Course Information

- **Class website:** *talisstats.github.io*
- **Textbook:** *Introduction to Probability (Grinstead and Snell)*
- **Email policy:** Please use *Piazza* for most questions.
- **Homework:** Should be submitted to *Gradescope*.
- **Exams:** No exams will be administered.

Using math language

Warning: Math uses a lot of symbols!, e.g.

Α Alpha	Β Beta	Γ Gamma	Δ Delta	Ε Epsilon	Ζ Zeta
Η Eta	Θ Theta	Ι Iota	Κ Kappa	Λ Lambda	Μ Mu
Ν Nu	Ξ Xi	Ο Omicron	Π Pi	Ρ Rho	Σ Sigma
Τ Tau	Υ Upsilon	Φ Phi	Χ Chi	Ψ Psi	Ω Omega

α Alpha	β Beta	γ Gamma	δ Delta	ε Epsilon	ζ Zeta
η Eta	θ Theta	ι Iota	κ Kappa	λ Lambda	μ Mu
ν Nu	ξ Xi	ο Omicron	π Pi	ρ Rho	σ Sigma
τ Tau	υ Upsilon	φ Phi	χ Chi	ψ Psi	ω Omega

$\dag$	\dag	$\S$	\S	©	\copyright
$\ddag$	\ddag	¶	\P	£	\pounds
$\ldots$	\ldots	$\cdots$	\cdots	$\vdots$	\vdots
$\ddots$	\ddots	$\prime$	\prime	$\forall$	\forall
$\aleph$	\aleph	$\emptyset$	\emptyset	$\exists$	\exists
$\hbar$	\hbar	∇	\nabla	$\neg$	\neg
$\imath$	\imath	$\surd$	\surd	$\flat$	\flat
$\jmath$	\jmath	$\top$	\top	$\natural$	\natural
ℓ	\ell	$\bot$	\bot	$\sharp$	\sharp
$\wp$	\wp	$\backslash$	\backslash	$\angle$	\angle
$\Re$	\Re	∂	\partial	$\mho$	\mho
$\Im$	\Im	∞	\infty	$\Box$	\Box
$\Diamond$	\Diamond	$\triangle$	\triangle	$\clubsuit$	\clubsuit
$\diamondsuit$	\diamondsuit	$\heartsuit$	\heartsuit	$\spadesuit$	\spadesuit
LaTeX	\LaTeX				

Example:

$$\hat{K}(P_n) \triangleq \arg\min_{k} \mathbb{E}_{B_n} \int L(o, \hat{\Psi}_k(P^0_{n,B_n}))\partial P^1_{n,B_n}(o) \tag{1}$$

Review: algebra

Deals with equations and unknown quantities, e.g.

$$-x^2(x+2)(x+3) \tag{2}$$

Review: algebra

Deals with equations and unknown quantities, e.g.

$$-x^2(x+2)(x+3) \tag{2}$$

Sometimes written as a polynomial, i.e.

$$\begin{aligned} -x^2(x+2)(x+3) &= -x^2(x^2+5x+6) \qquad (3) \\ &= -x^4-5x^3-6x^2 \qquad (4) \end{aligned}$$

Review: algebra

Deals with equations and unknown quantities, e.g.

$$-x^2(x+2)(x+3) \tag{2}$$

Sometimes written as a polynomial, i.e.

$$\begin{aligned} -x^2(x+2)(x+3) &= -x^2(x^2+5x+6) \qquad (3)\\ &= -x^4-5x^3-6x^2 \qquad (4)\end{aligned}$$

Many times, will be defined as a function, i.e.

$$f(x) = -x^2(x+2)(x+3) \tag{5}$$

Review: algebra

Deals with equations and unknown quantities, e.g.

$$-x^2(x+2)(x+3) \tag{2}$$

Sometimes written as a polynomial, i.e.

$$\begin{aligned} -x^2(x+2)(x+3) &= -x^2(x^2+5x+6) \quad (3) \\ &= -x^4-5x^3-6x^2 \quad (4) \end{aligned}$$

Many times, will be defined as a function, i.e.

$$f(x) = -x^2(x+2)(x+3) \tag{5}$$

Also can be defined as another variable, i.e.

$$y = -x^2(x+2)(x+3) \tag{6}$$

Review: algebra

System of equations

$$x^2 + y^2 = 25 \tag{7}$$
$$y = x + 1 \tag{8}$$

Review: algebra

System of equations

$$x^2 + y^2 = 25 \quad (7)$$

$$y = x + 1 \quad (8)$$

Implies that certain values hold true for both equations

Review: algebra

System of equations

$$
\begin{aligned}
x^2 + y^2 &= 25 && (7)\\
y &= x + 1 && (8)
\end{aligned}
$$

Implies that certain values hold true for both equations
Can solve them by combining/substituting equations, e.g.

$$
\begin{aligned}
x^2 + y^2 &= 25 && (9)\\
x^2 + (x+1)^2 &= && (10)\\
2x^2 + 2x + 1 &= && (11)
\end{aligned}
$$

Review: algebra

System of equations

$$
\begin{aligned}
x^2 + y^2 &= 25 \qquad (7)\\
y &= x + 1 \qquad (8)
\end{aligned}
$$

Implies that certain values hold true for both equations
Can solve them by combining/substituting equations, e.g.

$$
\begin{aligned}
x^2 + y^2 &= 25 \qquad (9)\\
x^2 + (x+1)^2 &= \qquad (10)\\
2x^2 + 2x + 1 &= \qquad (11)
\end{aligned}
$$

Solving for x gives us

$$
\begin{aligned}
2x^2 + 2x - 24 &= 0 \qquad (12)\\
x^2 + x - 12 &= \qquad (13)\\
(x+4)(x-3) &= \qquad (14)\\
&\implies x = -4 \text{ or } x = 3 \qquad (15)
\end{aligned}
$$

Logarithms

Deals with bases and exponentials, e.g.

$$\log_b a = x \tag{16}$$

Review: algebra

Logarithms

Deals with bases and exponentials, e.g.

$$\log_b a = x \tag{16}$$

Read as a power of, i.e.

$$\log_b a = x \implies a = b^x \tag{17}$$

Review: algebra

Logarithms

Deals with bases and exponentials, e.g.

$$\log_b a = x \tag{16}$$

Read as a power of, i.e.

$$\log_b a = x \implies a = b^x \tag{17}$$

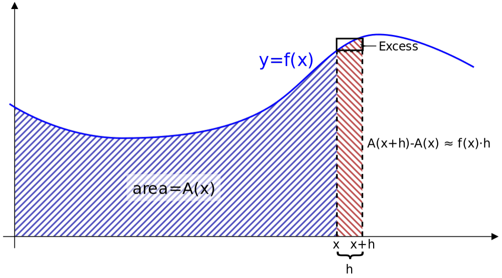
Has lots of properties, e.g.

$$\begin{aligned}
\log_b(x \cdot y) &= \log_b(x) + \log_b(y) && (18)\\
\log_b(x/y) &= \log_b(x) - \log_b(y) && (19)\\
\log_b(x^y) &= y \cdot log_b(x) && (20)\\
\log_b(c) &= 1/\log_c(b) && (21)\\
\log_b(x) &= \log_c(x)/\log_c(b) && (22)\\
f(x) &= \log_b(x) \implies f'(x) = 1/(x\log(b)) && (23)
\end{aligned}$$

Review: calculus

The fundamental theorem of calculus

1. $\int_a^b f(x)dx = F(b) - F(a)$
2. if $F(x) = \int_a^x f(t)dt$, then $F'(x) = f(x)$



Derivatives and integrals

Given a (possibly multivariate function), e.g.

$$f(x, y) = x^2 y + \sin(y) \tag{24}$$

Review: calculus

Derivatives and integrals

Given a (possibly multivariate function), e.g.

$$f(x, y) = x^2 y + \sin(y) \tag{24}$$

We can take a *partial derivative*, e.g. wrt x

$$\frac{\partial}{\partial x} f(x, y) = 2xy \tag{25}$$

Review: calculus

Derivatives and integrals

Given a (possibly multivariate function), e.g.

$$f(x,y) = x^2 y + \sin(y) \tag{24}$$

We can take a *partial derivative*, e.g. wrt x

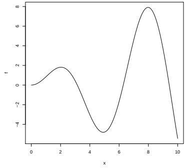
$$\frac{\partial}{\partial x} f(x,y) = 2xy \tag{25}$$

We can take an integral, e.g. wrt x

$$\int f(x,y)dx = \frac{1}{3}x^3 y + \sin(y) + C(y) \tag{26}$$

Example:

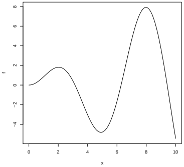
Say, $f(x) = x\sin(x)$ represents a position function, i.e.



Review: calculus

Example:

Say, $f(x) = x\sin(x)$ represents a position function, i.e.



We can calculate the velocity by $\frac{\partial}{\partial x}f(x)$ (by applying the product rule)

$$
\begin{aligned}
\frac{\partial}{\partial x}f(x) &= \frac{\partial}{\partial x}(x)\cdot\sin(x) + x\cdot\frac{\partial}{\partial x}(\sin(x)) && (27)\\
&= \sin(x) + x\cos(x) && (28)
\end{aligned}
$$

Review: calculus

Example: $f(x) = x\sin(x)$
What about acceleration?

Review: calculus

Example: $f(x) = x\sin(x)$
What about acceleration?

Take the second derivative, i.e. $\frac{\partial^2}{\partial x^2} f(x) = \frac{\partial}{\partial x} f'(x)$ (by again applying the product rule).

$$
\begin{aligned}
\frac{\partial}{\partial x} f'(x) &= \frac{\partial}{\partial x} \sin(x) + x\cos(x) && (29)\\
&= \cos(x) + (\cos(x) - x\sin(x)) && (30)\\
&= 2\cos(x) - x\sin(x) && (31)
\end{aligned}
$$

Review: calculus

Gradients

Given a (possibly multivariate function), e.g.

$$f(x, y) = x^2 y + \sin(y) \tag{32}$$

the *gradient* is a vector of partial derivatives of the function, e.g.

$$\begin{aligned} \nabla f(x, y) &= \begin{bmatrix} \frac{\partial}{\partial x} f(x, y) \\ \frac{\partial}{\partial y} f(x, y) \end{bmatrix} && (33) \\ &= \begin{bmatrix} 2xy \\ x^2 + \cos(y) \end{bmatrix} && (34) \end{aligned}$$

The Jacobian matrix

Given multiple multivariate functions, e.g.

$$f_1(x,y) = x^2y + \sin(y) \quad (35)$$
$$f_2(x,y) = x^2\sin(y) \quad (36)$$
$$\quad (37)$$

the *Jacobian matrix* is a vector of all partial derivatives of the functions, e.g.

$$\nabla \begin{bmatrix} f_1(x,y) \\ f_2(x,y) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x} f_1(x,y) & \frac{\partial}{\partial y} f_1(x,y) \\ \frac{\partial}{\partial x} f_2(x,y) & \frac{\partial}{\partial y} f_2(x,y) \end{bmatrix} \quad (38)$$

$$= \begin{bmatrix} 2xy & x^2 + \cos(y) \\ 2x\sin(y) & x^2\cos(y) \end{bmatrix} \quad (39)$$

Review: calculus

The Jacobian matrix

Given multiple multivariate functions, e.g.

$$f_1(x, y) = x^2 y + \sin(y) \quad (35)$$
$$f_2(x, y) = x^2 \sin(y) \quad (36)$$
$$(37)$$

the *Jacobian matrix* is a vector of all partial derivatives of the functions, e.g.

$$\nabla \begin{bmatrix} f_1(x, y) \\ f_2(x, y) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x} f_1(x, y) & \frac{\partial}{\partial y} f_1(x, y) \\ \frac{\partial}{\partial x} f_2(x, y) & \frac{\partial}{\partial y} f_2(x, y) \end{bmatrix} \quad (38)$$

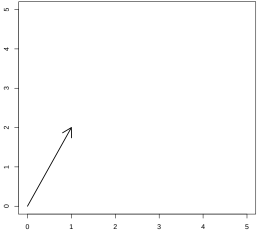
$$= \begin{bmatrix} 2xy & x^2 + \cos(y) \\ 2x \sin(y) & x^2 \cos(y) \end{bmatrix} \quad (39)$$

n.b. The matrix of second partial derivatives is called the *Hessian matrix*.

Review: linear algebra

Deals with vectors and spaces, e.g.

$$V = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \tag{40}$$



Review: linear algebra

Vectors can be added together, e.g.

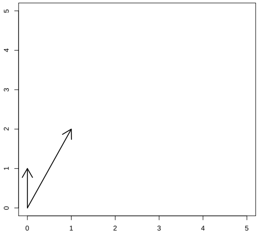
$$V = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, W = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{41}$$

$$V + W = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \tag{42}$$

Vectors can be added together, e.g.

$$V = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, W = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{41}$$

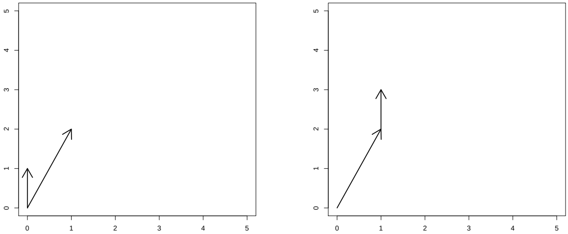
$$V + W = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \tag{42}$$



Vectors can be added together, e.g.

$$V = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, W = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{41}$$

$$V + W = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \tag{42}$$

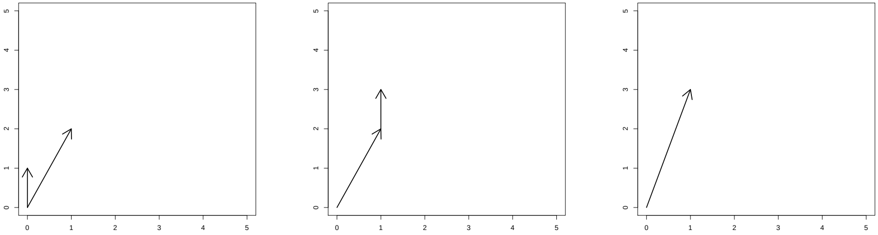


Review: linear algebra

Vectors can be added together, e.g.

$$V = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, W = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{41}$$

$$V + W = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \tag{42}$$



Vectors can be multiplied together, e.g. **dot product**

$$V = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, W = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{43}$$

$$V \cdot W = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 1 * 0 + 2 * 1 \tag{44}$$

Review: linear algebra

Vectors can be multiplied together, e.g. **dot product**

$$V = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, W = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{43}$$

$$V \cdot W = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 1 * 0 + 2 * 1 \tag{44}$$

This can be generalized to **inner product** (in functional spaces)

Review: linear algebra

Determinants:

Tell us the amount of volume scaling from transformations created by the matrix, e.g. for the matrix

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \tag{45}$$

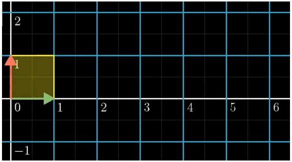
The determinant ($|A| = 6$) results in the following transformation

Determinants:

Tell us the amount of volume scaling from transformations created by the matrix, e.g. for the matrix

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \tag{45}$$

The determinant ($|A| = 6$) results in the following transformation

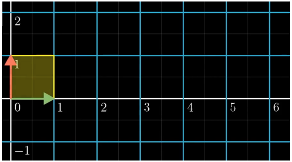


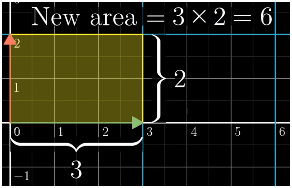
Determinants:

Tell us the amount of volume scaling from transformations created by the matrix, e.g. for the matrix

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \tag{45}$$

The determinant ($|A| = 6$) results in the following transformation





Review: linear algebra

Determinants:

How to calculate determinants:

For a 2x2 matrix:

$$|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \tag{46}$$

For a 3x3 matrix:

$$\begin{aligned} |A| &= \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} && (47) \\ &= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} && (48) \\ &= aei + bfg + cdh - ceg - bdi - afh && (49) \end{aligned}$$

Takeaway

Main points to keep in mind:

- This course establishes a feedback loop

Takeaway

Main points to keep in mind:

- This course establishes a feedback loop
- This course is separate from your work

Takeaway

Main points to keep in mind:

- This course establishes a feedback loop
- This course is separate from your work
- This course requires commitment

Takeaway

Main points to keep in mind:

- This course establishes a feedback loop
- This course is separate from your work
- This course requires commitment
- This course is meant for you