

# Lecture 1: Introduction, logistics, review

## STATS 101: Foundations of Statistics

Linh Tran

ThetaHat.AI@gmail.com

November 14, 2019

# About me



University of California  
San Francisco



Theta Hat



# About me



University of California  
San Francisco



Theta Hat

Cerebral Palsy • Vegetative State • Autism • Down's  
***Life Expectancy Project***

Traumatic Brain Injury • Spinal Cord Injury



**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  $\langle$ Next hype term here $\rangle$

**Question:** What is statistics?

**Question:** What is statistics?

**Origin:** Derived from Latin statisticum collegium ( “council of state” )

**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.

**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.



**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.

# Why statistics?

## 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



## Current plan

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

## Current plan

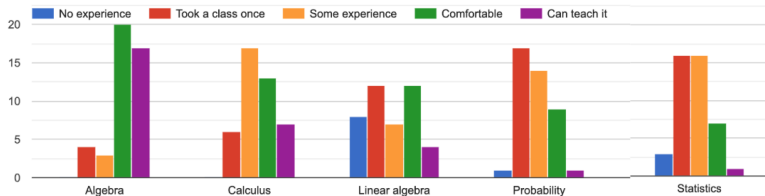
- ▶ STATS 101: Foundations of Statistics
- ▶ STATS 102: Introduction to Data Analysis
- ▶ STATS 103: Introduction to Statistical Learning

## Wishful thinking

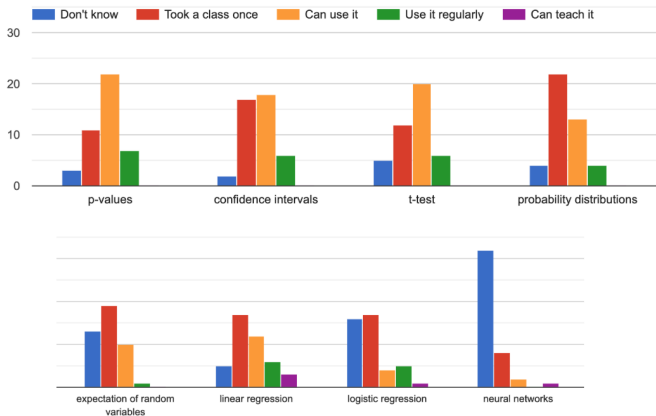
- ▶ STATS 204, 205, ...: Machine Learning, Deep Learning, etc.
- ▶ STATS 214, 215, ...: Survival analysis, clinical trial design, genomics, etc.
- ▶ STATS 224, 225, ...: Statistical theory, Semi-parametric efficiency, etc.

# Preliminary survey

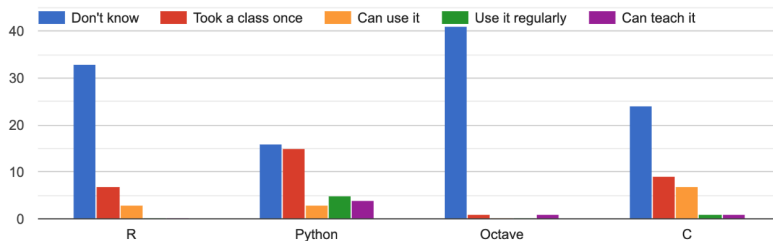
## Math background



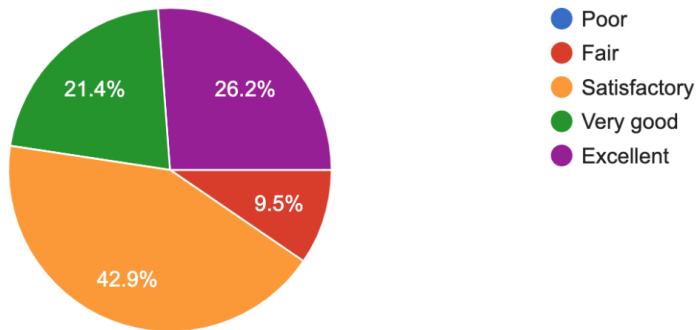
## Statistical experience



## Coding experience



## Effort desired





## Summary

### ▶ **Math:**

- ▶ Some haven't really used algebra/calculus; many have
- ▶ About half are weak on linear algebra

## Summary

### ▶ **Math:**

- ▶ Some haven't really used algebra/calculus; many have
- ▶ About half are weak on linear algebra

### ▶ **Statistics:**

- ▶ Most are weak to uncomfortable with statistical concepts

## Summary

### ▶ **Math:**

- ▶ Some haven't really used algebra/calculus; many have
- ▶ About half are weak on linear algebra

### ▶ **Statistics:**

- ▶ Most are weak to uncomfortable with statistical concepts

### ▶ **Coding:**

- ▶ Very few people code

### ▶ **Effort:**

- ▶ Most are willing to commit (at least reasonably)

## Summary

### ► Math:

- Some haven't really used algebra/calculus; many have
- About half are weak on linear algebra

### ► Statistics:

- Most are weak to uncomfortable with statistical concepts

### ► Coding:

- Very few people code

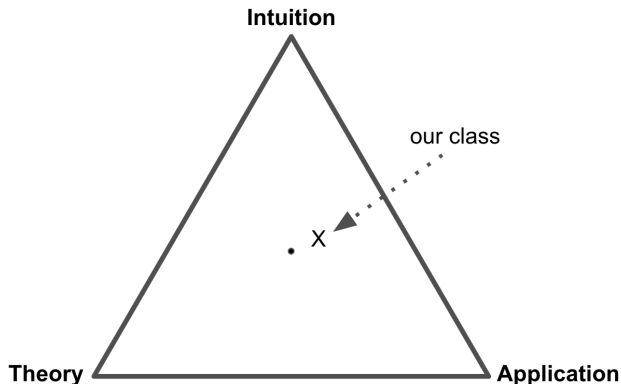
### ► Effort:

- Most are willing to commit (at least reasonably)

**Conclusion:** Split the class into two

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

## Proportion of concepts targeted



- ▶ **Class website:** *[talisstats.github.io](https://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	Β	Γ	Δ	Ε	Ζ
Alpha	Beta	Gamma	Delta	Epsilon	Zeta
Η	Θ	Ι	Κ	Λ	Μ
Eta	Theta	Iota	Kappa	Lambda	Mu
Ν	Ξ	Ο	Π	Ρ	Σ
Nu	Xi	Omicron	Pi	Rho	Sigma
Τ	Υ	Φ	Χ	Ψ	Ω
Tau	Upsilon	Phi	Chi	Psi	Omega

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

†	\dag	§	\S	©	\copyright
‡	\ddag	¶	\P	£	\pounds
...	\ldots	...	\cdots	⋮	\vdots
⋯	\ddots	/	\prime	∇	\nabla
ℵ	\aleph	∅	\emptyset	∃	\exists
ℏ	\hbar	▽	\nabla	¬	\neg
ℓ	\imath	✓	\surd	ℬ	\mathbb{B}
ℓ	\jmath	⊤	\top	ℕ	\mathbb{N}
ℓ	\ell	⊥	\bot	#	\sharp
ℓ	\wp	\	\backslash	∠	\angle
ℓ	\Re	∂	\partial	∞	\infty
ℓ	\Im	∞	\infty	∅	\emptyset
◇	\Diamond	△	\triangle	♣	\clubsuit
◇	\diamondsuit	♥	\heartsuit	♠	\spadesuit
ℒ	\mathcal{L}	ℒ	\mathcal{L}		

*Example:*

$$\hat{K}(P_n) \triangleq \arg \min_k \mathbb{E}_{B_n} \int L(o, \hat{\Psi}_k(P_{n,B_n}^0)) \partial P_{n,B_n}^1(o) \quad (1)$$

# Review: algebra

Deals with equations and unknown quantities, e.g.

$$-x^2(x + 2)(x + 3) \quad (2)$$



# Review: algebra

Deals with equations and unknown quantities, e.g.

$$-x^2(x+2)(x+3) \quad (2)$$

Sometimes written as a polynomial, i.e.

$$-x^2(x+2)(x+3) = -x^2(x^2+5x+6) \quad (3)$$

$$= -x^4 - 5x^3 - 6x^2 \quad (4)$$

# Review: algebra

Deals with equations and unknown quantities, e.g.

$$-x^2(x+2)(x+3) \quad (2)$$

Sometimes written as a polynomial, i.e.

$$-x^2(x+2)(x+3) = -x^2(x^2+5x+6) \quad (3)$$

$$= -x^4 - 5x^3 - 6x^2 \quad (4)$$

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

$$f(x) = -x^2(x+2)(x+3) \quad (5)$$

# Review: algebra

Deals with equations and unknown quantities, e.g.

$$-x^2(x+2)(x+3) \quad (2)$$

Sometimes written as a polynomial, i.e.

$$-x^2(x+2)(x+3) = -x^2(x^2+5x+6) \quad (3)$$

$$= -x^4 - 5x^3 - 6x^2 \quad (4)$$

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

$$f(x) = -x^2(x+2)(x+3) \quad (5)$$

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

$$y = -x^2(x+2)(x+3) \quad (6)$$

## System of equations

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

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

## System of equations

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

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

Implies that certain values hold true for both equations

## System of equations

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

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

Implies that certain values hold true for both equations

Can solve them by combining/substituting equations, e.g.

$$x^2 + y^2 = 25 \quad (9)$$

$$x^2 + (x + 1)^2 = \quad (10)$$

$$2x^2 + 2x + 1 = \quad (11)$$

## System of equations

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

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

Implies that certain values hold true for both equations

Can solve them by combining/substituting equations, e.g.

$$x^2 + y^2 = 25 \quad (9)$$

$$x^2 + (x + 1)^2 = \quad (10)$$

$$2x^2 + 2x + 1 = \quad (11)$$

Solving for  $x$  gives us

$$2x^2 + 2x - 24 = 0 \quad (12)$$

$$x^2 + x - 12 = \quad (13)$$

$$(x + 4)(x - 3) = \quad (14)$$

$$\implies x = -4 \text{ or } x = 3 \quad (15)$$

## Logarithms

Deals with bases and exponentials, e.g.

$$\log_b a = x \quad (16)$$



## Logarithms

Deals with bases and exponentials, e.g.

$$\log_b a = x \quad (16)$$

Read as a power of, i.e.

$$\log_b a = x \implies a = b^x \quad (17)$$

## Logarithms

Deals with bases and exponentials, e.g.

$$\log_b a = x \quad (16)$$

Read as a power of, i.e.

$$\log_b a = x \implies a = b^x \quad (17)$$

Has lots of properties, e.g.

$$\log_b(x \cdot y) = \log_b(x) + \log_b(y) \quad (18)$$

$$\log_b(x/y) = \log_b(x) - \log_b(y) \quad (19)$$

$$\log_b(x^y) = y \cdot \log_b(x) \quad (20)$$

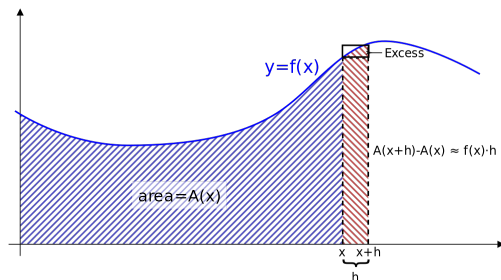
$$\log_b(c) = 1/\log_c(b) \quad (21)$$

$$\log_b(x) = \log_c(x)/\log_c(b) \quad (22)$$

$$f(x) = \log_b(x) \implies f'(x) = 1/(x \log(b)) \quad (23)$$

## 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.

$$f(x, y) = x^2y + \sin(y) \quad (24)$$

## Derivatives and integrals

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

$$f(x, y) = x^2y + \sin(y) \quad (24)$$

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

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

## Derivatives and integrals

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

$$f(x, y) = x^2y + \sin(y) \quad (24)$$

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

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

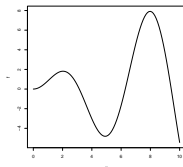
We can take an integral, e.g. wrt  $x$

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

# Review: calculus

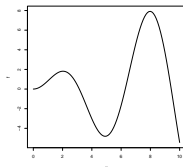
## Example:

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



## 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)

$$\frac{\partial}{\partial x} f(x) = \frac{\partial}{\partial x}(x) \cdot \sin(x) + x \cdot \frac{\partial}{\partial x}(\sin(x)) \quad (27)$$

$$= \sin(x) + x \cos(x) \quad (28)$$



**Example:**  $f(x) = x \sin(x)$

What about acceleration?

**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).

$$\frac{\partial}{\partial x} f'(x) = \frac{\partial}{\partial x} \sin(x) + x \cos(x) \quad (29)$$

$$= \cos(x) + (\cos(x) - x \sin(x)) \quad (30)$$

$$= 2 \cos(x) - x \sin(x) \quad (31)$$

## Gradients

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

$$f(x, y) = x^2y + \sin(y) \quad (32)$$

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

$$\nabla f(x, y) = \begin{bmatrix} \frac{\partial}{\partial x} f(x, y) \\ \frac{\partial}{\partial y} f(x, y) \end{bmatrix} \quad (33)$$

$$= \begin{bmatrix} 2xy \\ x^2 + \cos(y) \end{bmatrix} \quad (34)$$

## The Jacobian matrix

Given multiple multivariate functions, e.g.

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

$$f_2(x, y) = x^2 \sin(y) \quad (36)$$

$$(37)$$

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

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

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

## The Jacobian matrix

Given multiple multivariate functions, e.g.

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

$$f_2(x, y) = x^2 \sin(y) \quad (36)$$

$$(37)$$

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

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

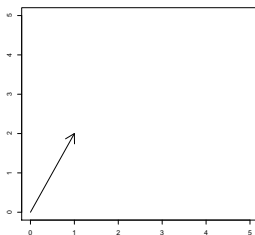
$$= \begin{bmatrix} 2xy & x^2 + \cos(y) \\ 2x \sin(y) & x^2 \cos(y) \end{bmatrix} \quad (39)$$

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

# Review: linear algebra

Deals with vectors and spaces, e.g.

$$V = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad (40)$$



# Review: linear algebra

Vectors can be added together, e.g.

$$V = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, W = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (41)$$

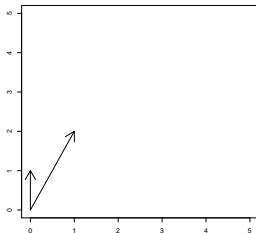
$$V + W = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad (42)$$

# Review: linear algebra

Vectors can be added together, e.g.

$$V = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, W = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (41)$$

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



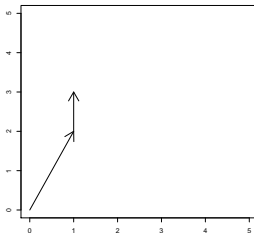
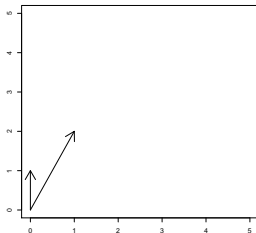


# Review: linear algebra

Vectors can be added together, e.g.

$$V = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, W = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (41)$$

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

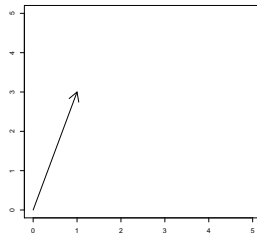
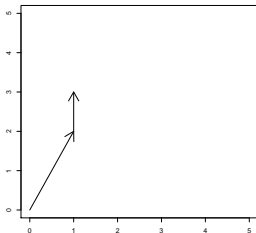
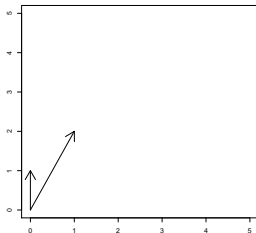


# Review: linear algebra

Vectors can be added together, e.g.

$$V = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, W = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (41)$$

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



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

$$V = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, W = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (43)$$

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

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

$$V = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, W = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (43)$$

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

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

## Determinants:

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

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \quad (45)$$

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

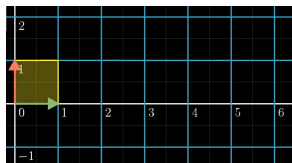
# Review: linear algebra

## Determinants:

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

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \quad (45)$$

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



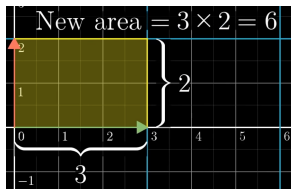
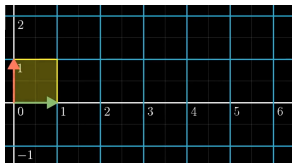
# Review: linear algebra

## Determinants:

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

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \quad (45)$$

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



# Review: linear algebra

## Determinants:

How to calculate determinants:

For a 2x2 matrix:

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

For a 3x3 matrix:

$$|A| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} \quad (47)$$

$$= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \quad (48)$$

$$= aei + bfg + cdh - ceg - bdi - afh \quad (49)$$



# 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