Statistics and Normal Distribution

MML 6.4-5

Marek Petrik

9/05/2023

Important Properties

Marginalization

$$\mathbb{P}[X = x] = \sum_{y \in \mathcal{T}} \mathbb{P}[X = x, Y = y]$$

Conditional probability

$$\mathbb{P}[X = x \mid Y = y] = \frac{\mathbb{P}[X = x, Y = y]}{\mathbb{P}[Y = y]}$$

Product rule

$$\mathbb{P}[X = x, Y = y] = \mathbb{P}[X = x \mid Y = y] \cdot \mathbb{P}[Y = y]$$

Independent random variables

$$\mathbb{P}[X = x, Y = y] = \mathbb{P}[X = x] \cdot \mathbb{P}[Y = y]$$

Bayes Theorem

$$\mathbb{P}[X = x \mid Y = y] = \frac{\mathbb{P}[Y = y \mid X = x] \cdot \mathbb{P}[X = x]}{\mathbb{P}[Y = y]}$$

Conditional Independence

Random variables $X, Y : \Omega \to \mathbb{R}$ are **independent conditionally** on a random variable Z if for all $x, y, z \in \mathbb{R}$:

$$\mathbb{P}[X = x, Y = y \mid Z = z] = \mathbb{P}[X = x \mid Z = z] \cdot \mathbb{P}[Y = y \mid Z = z]$$

Examples?

Continuous Probability Distributions

- **Normal**: common because of central limit theorem
- **Laplace**: Extreme weather events
- **Multivariate normal**: Height and weight

See Wikipedia for their properties

Cumulative Distribution Function

$$F_X(x) \; = \; \mathbb{P}\left[X \leq x\right] = P\left(\{\omega \in \Omega \mid X(\omega) \leq x\}\right)$$

Example: Uniform random variable on $[0, 1]$

Probability Density Function

Function $f_X : \mathbb{R} \to \mathbb{R}$ is a pdf of X if $\forall x$

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

Conditional Probability Density Function

For random variables X, Y with a joint density $f_{X,Y} \colon \mathbb{R}^2 \to \mathbb{R}$, the conditional density is (when $f_X(x) > 0$):

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

Today

1. Expected value
2. Variance, covariance, and correlation
3. Normal distribution

Expected Value (Mean)

Random variable X (discrete and continuous, see Lebesgue integrals)

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x \cdot \mathbb{P}[X = x], \qquad \mathbb{E}[X] = \int_{\Omega} X dP$$

Expected Value of Function

Random variable X (discrete and continuous) and a function g

$$\mathbb{E}\left[g(X)\right] = \sum_{g \in \mathcal{G}} g \cdot \mathbb{P}\left[g(X) = g\right], \qquad \mathbb{E}\left[X\right] = \int_{\Omega} g(X) dP$$

Law of the Unconscious Statistician proves that:

Expected Value: Properties

Assume real-valued random variables X and Y

$$\mathbb{E}[c \cdot X] = c \cdot \mathbb{E}[X]$$

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

When X and Y are **independent**:

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Conditional Expectation

https://en.wikipedia.org/wiki/Conditional_expectation

Conditioning on an event $A \in \mathcal{F}$ (discrete r.v.)

$$\mathbb{E}[X \mid A] \quad = \quad \sum_{x \in \mathcal{X}} x \cdot \mathbb{P}[X = x \mid A]$$

Conditioning on a random variable Y:

$$\mathbb{E}[X \mid Y] : \Omega \to \mathbb{R}$$

defined as (discrete)

$$\mathbb{E}[X \mid Y](\omega) \quad = \quad \sum_{x \in \mathcal{X}} x \cdot \mathbb{P}[X = x \mid Y = Y(\omega)]$$

Expected Utility Theory

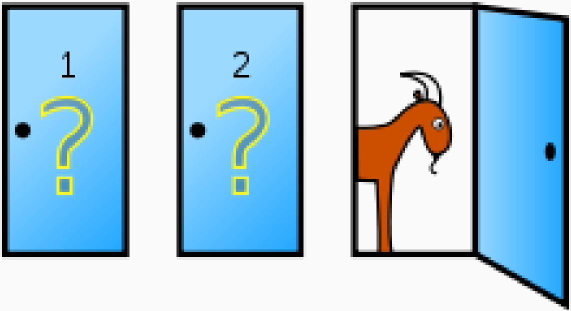
Two possible investments with profits X_1 and X_2

Choose investment one if

$$\mathbb{E}[X_1] \geq \mathbb{E}[X_2]$$

Monty Hall Problem

https://en.wikipedia.org/wiki/Monty_Hall_problem



Two Envelopes Paradox

https://en.wikipedia.org/wiki/Two_envelopes_problem

I have two envelopes, one has X dollars, another has $2 \cdot X$ dollars

After choosing one, would you want to switch?

Variance and Standard Deviation

Variance

$$\mathbb{V}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

Standard deviation

$$\text{sd}[X] = \sqrt{\mathbb{V}[X]} = \sqrt{\mathbb{E}\left[X - \mathbb{E}[X]^2\right]}$$

Why standard deviation?

Variance: Another Representation

$$\mathbb{V}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$

Covariance

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

$$\mathrm{Cov}\left[X, Y\right] =$$

Variance and Covariance Properties

$$\mathrm{Cov}\left[X, Y\right] = 0 \quad \text{when} \quad X, Y \text{ are independent}$$

$$\mathrm{Cov}\left[a \cdot X, Y\right] = \mathrm{Cov}\left[X, a \cdot Y\right] = a \cdot \mathrm{Cov}\left[X, Y\right]$$

$$\mathbb{V}\left[a \cdot X\right] = a^2 \cdot \mathbb{V}\left[X\right]$$

$$\mathbb{V}\left[\sum_{i \in \mathcal{I}} X_i\right] = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}} \mathrm{Cov}\left[X_i, X_j\right]$$

Variance of Independent Variables

Suppose that X and Y are independent random variables.
Compute

$$\mathbb{V}[X+Y] =$$

Covariance and Correlation

Covariance

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

Between $(-\infty, \infty)$

Correlation

$$\mathrm{corr}[X, Y] = \frac{\mathrm{Cov}[X, Y]}{\mathrm{sd}[X] \cdot \mathrm{sd}[Y]}$$

Between $[-1, +1]$

Correlation Coefficient

Correlation coefficient $\text{corr}[X, Y]$ is between $[-1, 1]$

- **0:** Variables are not related
- **1:** Variables are perfectly related (same)
- **-1:** Variables are negatively related (different)

Covariance and Correlation: Properties

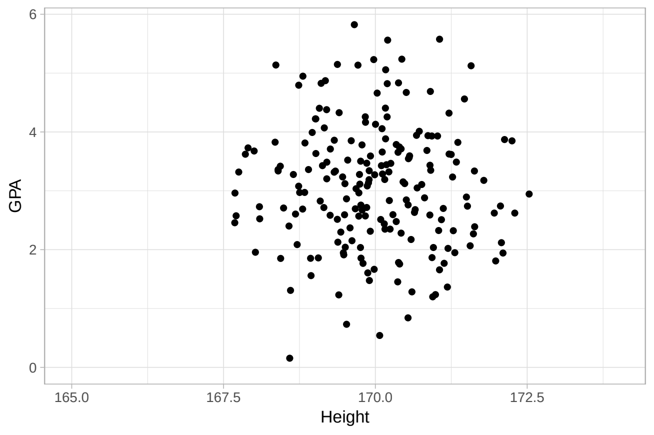
$$\text{Cov}[X, X] =$$

$$\text{corr}[X, X] =$$

$$\text{corr}[X, Y] = \qquad \text{when } X, Y \text{ are independent}$$

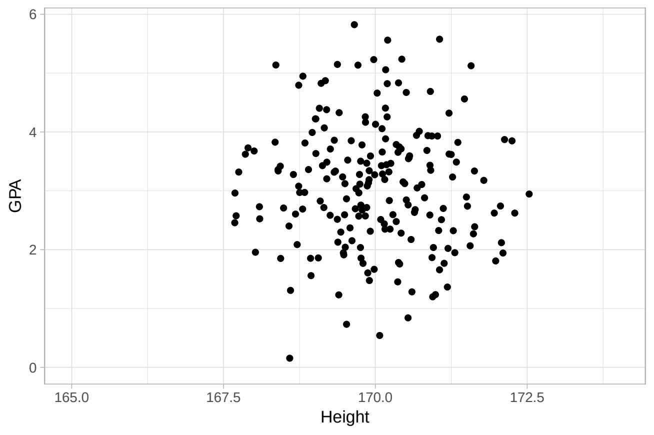
Correlation Example

Plotting code in `slides/figs/class2/plots.R`



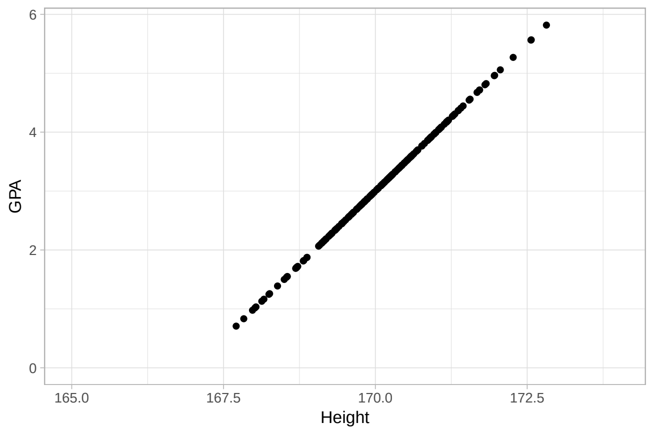
Correlation Example

Plotting code in `slides/figs/class2/plots.R`

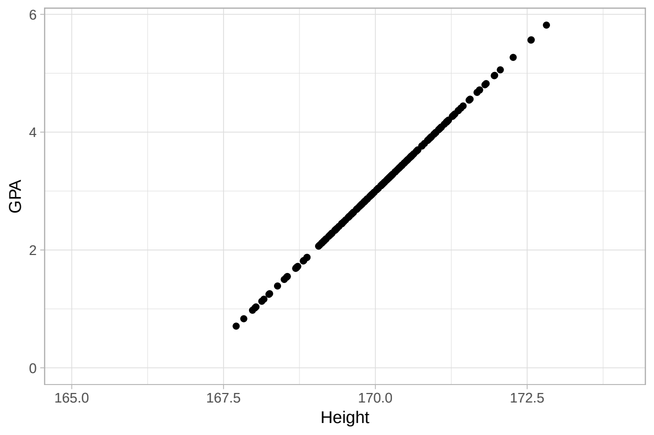


Correlation: 0

Correlation Example

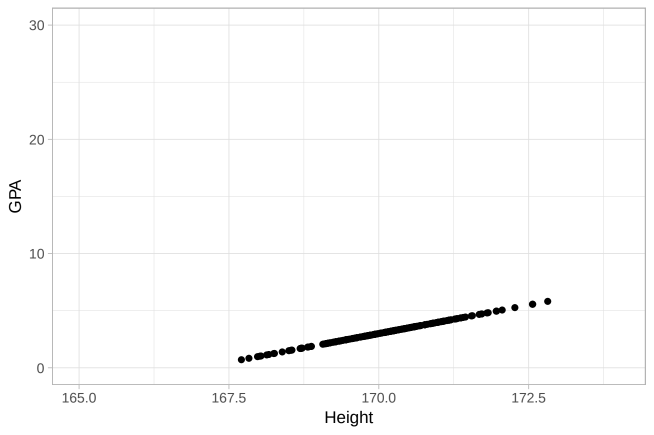


Correlation Example

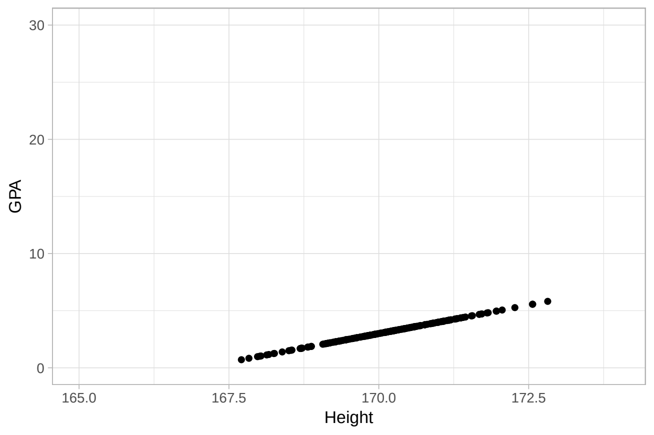


Correlation: 1

Correlation Example

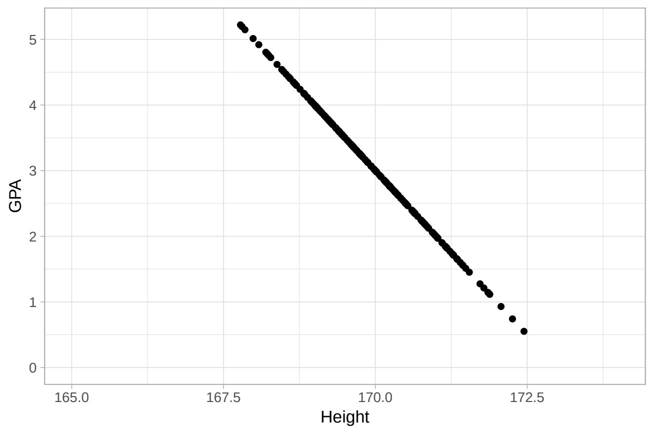


Correlation Example

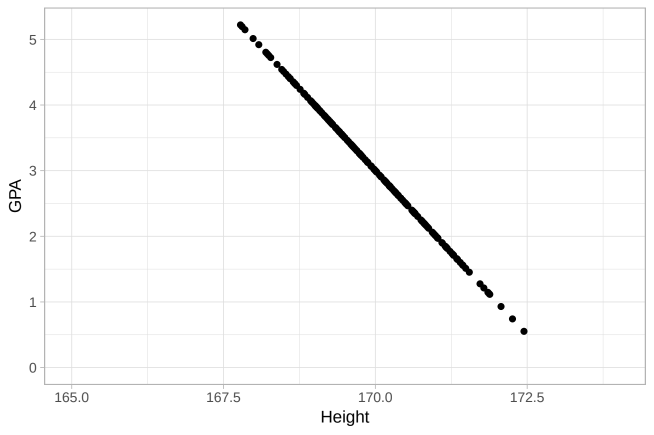


Correlation: 1

Correlation Example

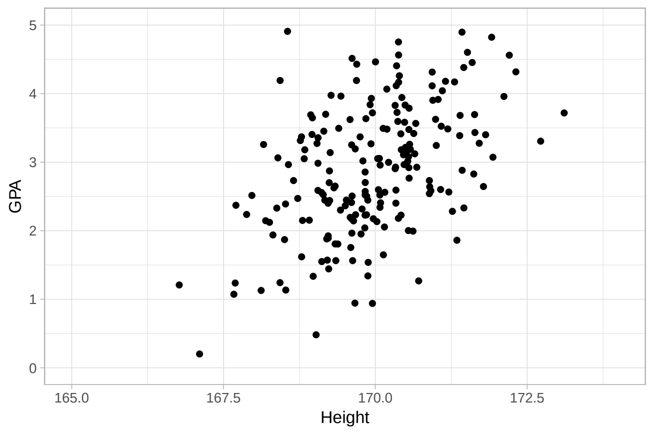


Correlation Example

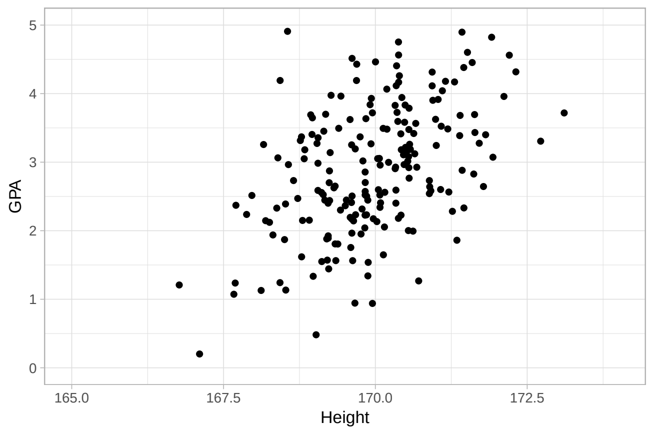


Correlation: -1

Correlation Example

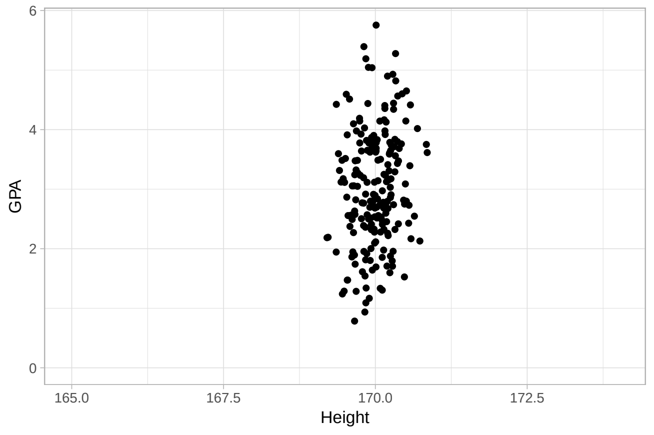


Correlation Example

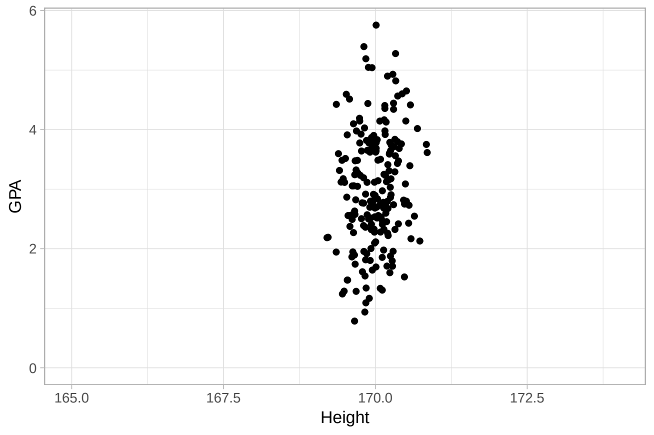


Correlation: 0.5

Correlation Example



Correlation Example

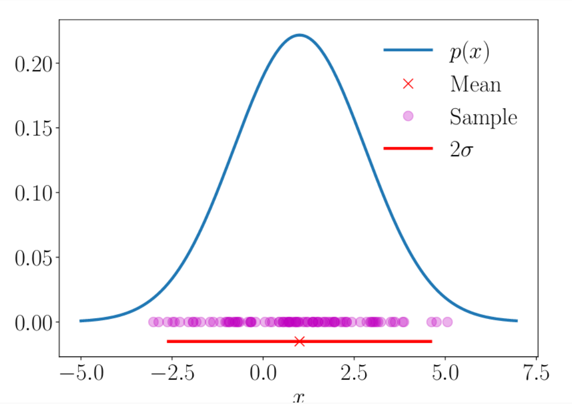


Correlation: 0.0

Univariate Normal Distribution

Density for $X \sim \mathcal{N}(\mu, \sigma^2)$

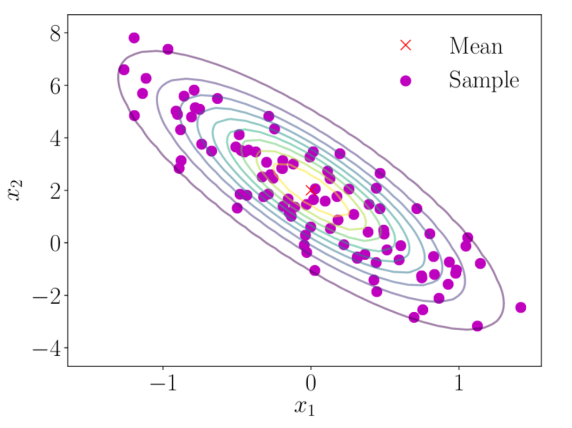
$$p(x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right)$$



Multivariate Normal Distribution

Joint probability density function

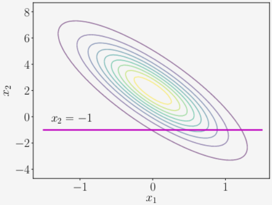
$$p(x, y \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \dots$$



Multivariate Normal Distribution

Joint probability density function

$$p(x, y \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \ldots$$



(a) Bivariate Gaussian.