

**Welcome to
Applied Probability and Statistics!**

- ▶ Unit 1: Introduction to Probability
- ▶ Unit 2: Probability Distributions
- ▶ Unit 3: Statistical Inference
- ▶ Unit 4: Introduction to Linear Regression
- ▶ Unit 5: Regression Analysis
- ▶ Unit 6: Regression Modeling

Peterjohn-Richards Distinguished Professor
Area Coordinator, Decision Sciences

Saša Pekeč (Sasha Pecketch?)

Peterjohn-Richards Distinguished Professor
Area Coordinator, Decision Sciences

Peterjohn-Richards Distinguished Professor
Area Coordinator, Decision Sciences

- ▶ Research: market design, analytics
 - pricing streaming media
 - bilateral ratings in service platforms
 - combinatorial auctions
- ▶ Teaching:
 - Stats (core, MSQM, MMS, MBA, EMBA)
 - Technological Transformation of Business (core, MBA)
 - Digital Platforms (elective, EMBA)
 - Strategic Modeling (elective, MBA)
 - Data Driven Decision Making (exec. ed.)
- ▶ Other:
 - consulting (mostly strategy, some analytics, market design)
 - Supervisory Board of Atlantic Grupa, some startups (HR analytics, crypto)
 - Council of Economic Advisors to the President of Croatia, 2010-2015

Learning flow

- ▶ Before Virtual Class Session:
Watch videos in the order as outlined. (Skip supplementary/optional videos.)
Attempt/review illustrative examples.
- ▶ Virtual Class Session:
Prepare for the session.
Attend and actively participate.
- ▶ After Virtual Class Session:
Complete the assignment.

Course Logistics

Learning flow

- ▶ Before Virtual Class Session:
Watch videos in the order as outlined. (Skip supplementary/optional videos.)
Attempt/review illustrative examples.
- ▶ Virtual Class Session:
Prepare for the session.
Attend and actively participate.
- ▶ After Virtual Class Session:
Complete the assignment.

Textbook

Groebner, Shannon and Fry, not required (supplementary).
Practice problems/exercises: any business stats textbook is OK.

Software

Excel used in presenting the course material; *FSBstats* add-in in the second half.
Additionally, R code provided (optional/supplementary).
Could use any data analysis software for exercises/assignments.

Course Deliverables

- ▶ Unit 1: Individual Assignment
- ▶ Unit 2: Individual Assignment
- ▶ Unit 3: Team Assignment
- ▶ Unit 4: Individual Assignment
- ▶ Unit 5: Team Assignment
- ▶ Unit 6: No assignment due
- ▶ Final Exam

Component	Percentage
3 Individual assignments*	30%
2 Team assignments	30%
Class participation	5%
Final exam	35%

* Individual assignments scoring details

- ▶ 6 questions/points on each assignment.
- ▶ Total individual assignments score capped at 15 points.
- ▶ Can miss three questions and still earn max possible score.

Keep in mind

- ▶ Time management and balancing personal/professional/academic life.
- ▶ Focus of the course is on modeling and decision-making.
- ▶ Help is available:

Keep in mind

- ▶ Time management and balancing personal/professional/academic life.
- ▶ Focus of the course is on modeling and decision-making.
- ▶ Help is available:
 - TA support (still in the works)
 - Team Fuqua: help each other!
 - Contact instructor/program.
- ▶ Honor code.

Unit 1: Introduction to Probability

► The Basics of Probability

- quantifying uncertainty
- probability p : $0 \leq p \leq 1$

► Random Variables

- numerical quantities with uncertain outcomes
- succinct representation and modeling of (complex) uncertainties

► Summary Statistics

- expected value
- variance (std. deviation)

] ← NON-INTUITIVE

Oil Drilling: Combining Investments

An oil wildcatter owns drilling rights at an oil tract. The odds against discovering oil if a well is drilled are 9 to 1. A well costs \$100,000 to drill, and this is a total loss if no oil is found. On the other hand, if oil is discovered, rights to the oil can be sold for \$1,600,000.

- if no oil: -100000
- if oil: $-100000 + 1500000$
 $= 1500000$

odds against 9 to 1

$$P = \frac{1}{9+1} = .1$$

Oil Drilling: Combining Investments

An oil wildcatter owns drilling rights at an oil tract. The odds against discovering oil if a well is drilled are 9 to 1. A well costs \$100,000 to drill, and this is a total loss if no oil is found. On the other hand, if oil is discovered, rights to the oil can be sold for \$1,600,000.

$$p = .1$$

- a) What is the expected value and the standard deviation of the oil wildcatter's profit?

$X = \text{investor's profit (in units of \$100K)}$

• if no oil $X = -1$

• if oil $X = 15$

$$X = \begin{cases} -1 & \text{with } p = .9 \\ 15 & \text{with } p = .1 \end{cases}$$

$$E[X] = (-1)(.9) + (15)(.1) = -.9 + 1.5 = .6 = 60K$$

$$\begin{aligned} \text{Var}[X] &= (-1 - .6)^2(.9) + (15 - .6)^2(.1) \\ &= (-1.6)^2(.9) + (14.4)^2(.1) \\ &= 23.04 (\$100K)^2 \end{aligned}$$

$$\sqrt{\text{Var}[X]} = \sqrt{23.04} = 4.8 = 480K$$

(std. dev.)

Oil Drilling: Combining Investments

An oil wildcatter owns drilling rights at an oil tract. The odds against discovering oil if a well is drilled are 9 to 1. A well costs \$100,000 to drill, and this is a total loss if no oil is found. On the other hand, if oil is discovered, rights to the oil can be sold for \$1,600,000.

- a) What is the expected value and the standard deviation of the oil wildcatter's profit?

Consider n independent identical oil tracts, each owned by a different oil wildcatter. Suppose n oil wildcatters create a partnership, each partner giving rights to their own oil tract to the partnership and earning $1/n$ share of the partnership.

- b) What is the expected value and the standard deviation of the oil-wildcatter's profit in the partnership with two partners?
- c) What is the expected value and the standard deviation of the oil-wildcatter's profit in the partnership with n partners?

Oil Drilling: Combining Two Investments ($n = 2$)

- b) What is the expected value and the standard deviation of the oil-wildcatter's profit in the partnership with two partners ($n = 2$)?

$X_i = \text{profit in location } i=1,2$
 $E[X_i] = .6 \quad \text{Var}[X_i] = 23.04$

investor's profit in the partnership

$$\bar{X} = \frac{X_1 + X_2}{2}$$

$$E[\bar{X}] = E\left[\frac{X_1 + X_2}{2}\right] = \frac{1}{2} E[X_1 + X_2] = \frac{1}{2} [0.6 + 0.6] = 0.6$$

$$\text{Var}[\bar{X}] = \text{Var}\left[\frac{X_1 + X_2}{2}\right] = \left(\frac{1}{2}\right)^2 \text{Var}[X_1 + X_2] = \frac{1}{4} [2 \times 23.04] = \frac{23.04}{2}$$

$$\sqrt{\text{Var}[\bar{X}]} = \sqrt{\frac{23.04}{2}} = \frac{\sqrt{23.04}}{\sqrt{2}} = \frac{4.8}{\sqrt{2}}$$

Oil Drilling: Combining n Investments

- c) What is the expected value and the standard deviation of the oil-wildcatter's profit in the partnership with n partners?

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

$$E[\bar{X}] = \frac{1}{n} (\overbrace{0.6 + 0.6 + \dots + 0.6}^n) = \frac{1}{n} \cdot n(0.6) = 0.6$$

$$\text{Var}[\bar{X}] = \frac{1}{n^2} (\overbrace{23.04 + \dots + 23.04}^n) = \frac{1}{n^2} \cdot n(23.04) = \frac{23.04}{n}$$

$$\sqrt{\text{Var}[\bar{X}]} = \frac{\sqrt{23.04}}{\sqrt{n}} = \frac{4.8}{\sqrt{n}}$$

Risk Pooling

Combining independent risks reduces uncertainty.

- ▶ Advantages of owning $1/n$ share of a portfolio with n independent identically distributed investments (vs. owning one such investment)
 - The expected value of the profit is unchanged
 - Variance of the profit is reduced by a factor of n
 - The standard deviation of the profit is reduced by a factor of \sqrt{n}
- ▶ Examples:
 - Managing investment portfolio (uncertain returns)
 - Managing risk exposure in insurance (uncertain losses)
 - Managing supplies (uncertain demand)
- ▶ Risk pooling is a basic risk management strategy
(Another strategy is diversification; more in finance course.)

Unit 1 Individual Assignment

- ▶ Due Monday, 23:59 ET (Durham local time).
- ▶ Six questions: some multiple choice, some numeric response.
- ▶ No need to show your work, i.e., no supporting documents needed.

Assignment questions:

- ▶ Questions 1-4: use S&P1500 CEO compensation data to connect data and random variables
- ▶ Questions 5-6: similar to oil drilling (risk pooling)

Connecting data and random variables: CEO compensation data

2022 annual total compensation (in million of US dollars) as reported in SEC filings for the CEOs of S&P 1500 companies in a recent year is available in the dataset `CEOCompensation_Data`. How do we connect compensation data with the compensation random variable?

Name	Company	Compensation (\$M)
Sundar Pichai	ALPHABET INC	226
Stephen M. Scherr	HERTZ GLOBAL HOLDINGS INC	182
Michael Rapino	LIVE NATION ENTERTAINMENT	139
Timothy D. Cook	APPLE INC	99
Peter Salvatore Zaffino	AMERICAN INTERNATIONAL GROUP	75
Peter Anevski	PROGYNY INC	69
Hock E. Tan	BROADCOM INC	61
Richard Brian Handler	JEFFERIES FINANCIAL GRP INC	57
Vicente Reynal	INGERSOLL RAND INC	55
...

Connecting data and random variables: CEO compensation distribution

