COURSE SYLLABUS

Statistics and Data Analysis

Course code: 220141

1. General information

Course type		*Number of credits*	*Number of learning periods*
General	□	Theory: 02 Exercise: 00 Practice: 01	Theory: 30 Exercise: 00 Practice: 30
Basic	☑		
Specialized	□		
Required	☑		
Elective	□		

***Learners*:**

Level	Bachelor
Discipline	Information Technology

Course requirements:

Prerequisites	Linear Algebra
Parallels	None
Other requirements	None

2. Learning resources:

Books	[1] Christian Heumann, Michael Schomaker and Shalabh (2016). *Introduction to Statistics and Data Analysis*. Springer. [2] Nguyễn Văn Tuấn (2014). *Phân tích dữ liệu với R.* NXB Tổng hợp TP Hồ Chí Minh.
References	[3] Hadley Wickham, Garrett Grolemund (2016). *R for Data Science*. O'Reilly Media.

	[4] Hoàng Trọng, Chu Nguyễn Mộng Ngọc (2010). *Thống kê ứng dụng trong kinh tế - Xã hội.* NXB Lao động - Xã hội.
Other learning materials	Website: https://r4ds.had.co.nz/ Website: http://makemeanalyst.com/r-programming/

3. Course description

The course provides students basic knowledge on probability and statistics for analyzing research data. The course also aims to provide opportunities to practice professional skills including designing research, collecting data, managing data, and describing data using statistical characteristics and visualization techniques, and testing statistical hypotheses using R language. The course is the foundation of statistics-related courses such as Artificial Intelligence, Data mining, etc. Additionally, the course develops students' appropriate awareness and attitudes on the importance of statistics and data analysis in science as well as in social life.

4. Course learning outcomes (CLOs)

After finishing the course, students will be able to:

		Satisfy LOs of the program	***Satisfy LOs of the ABET***
❖ ***Topic 1: Disciplinary Knowledge and Reasoning***			B.1.1
L1.	Present basic concepts of probability and statistics.	1.1.3	
L2.	Present the role of statistics in scientific research.		
L3.	Analyze data using visualization techniques with R.		
L4.	Perform hypothesis testing using statistics with R.		
L5.	Perform variance and correlation analysis, regression analysis and time series data analysis with R.		
❖ ***Topic 2: Personal and Professional Skills and Attributes***			
L6.	Problem Identification and Formulation	2.1.1	
L7.	Modeling	2.1.2	
L8.	Estimation and Qualitative Analysis	2.1.3	

L9.	Hypothesis Formulation	2.2.1	
L10.	Survey of Print and Electronic Literature	2.2.2	
L11.	Hypothesis Test and Defense	2.2.4	
❖ ***Topic 3: Interpersonal Skills: Teamwork and Communication***			
L12.	Technical and Multidisciplinary Teaming	3.1.5	
❖ ***Topic 4: Conceiving, Designing, Implementing and Operating Systems in The Enterprise, Societal and Environmental Context – The Innovation Process***			
L13.	Utilization of Knowledge in Design	4.3.3	

5. Course content

Course content	CLOs	Number of learning periods		
		Theory	Practice	Others
Chapter 1. Basic concepts of probability	L1	3	5	
1.1. Complementary on combinatorial analysis				
1.2. Event and relationship between events				
1.3. Probability and its formula				
1.4. Bernoulli trial				
1.5. Random variable and probability distribution				
1.6. Common distributions (Binomial distribution, Poisson distribution, Standard distribution, Student distribution, F distribution, and χ^2 distribution)				
1.7. Introduction to R and programming in R				
☐ ***Personal and Professional Skills and Attributes***	L6-L9 (T)			
☐ ***Interpersonal Skills: Teamwork and Communication***	L12 (U)			
☐ ***CDIO in the enterprise, societal and environmental context***	L13 (U)			

Chapter 3. Descriptive statistics	L2, L3	2	5	
2.1. Population and sample				
2.2. Category variable, binary variable and continuous variable				
2.3. Measures of descriptive analysis: mean, median, standard deviation, variance, standard error				
2.4. Data managing with R				
2.5. Tabular data describing				
2.6. Categorical variable describing				
2.7. Continuous variable describing				
☐ ***Personal and Professional Skills and Attributes***	L6-L9 (T)			
☐ ***Interpersonal Skills: Teamwork and Communication***	L12 (U)			
☐ ***CDIO in the enterprise, societal and environmental context***	L13 (U)			
Chapter 3. Data analysis using graphs	L3	1	3	
3.1. Plot window in R				
3.2. Naming plot axis				
3.3. Color and markers				
3.4. Bar plot and pie plot				
3.5. Histogram				
3.6. Boxplot				
3.7. Scatter plot				
☐ ***Personal and Professional Skills and Attributes***	L6-L9 (T)			
☐ ***Interpersonal Skills: Teamwork and Communication***	L12 (U)			
☐ ***CDIO in the enterprise, societal and environmental context***	L13 (U)			

Chapter 4. Statistical hypothesis testing	L4	3	5	
4.1. Introduction to statistical hypothesis testing				
4.2. Type I error and Type II error				
4.3. Testing model of significance				
4.4. Testing model of hypothesis				
4.5. Mixture model				
4.6. P-value and problems of p-value				
4.7. Testing of standard distribution				
4.8. T-test				
4.9. Wilcoxon testing for two samples				
4.10. Scale test				
4.11. Chi-square test				
4.12. Fisher test				
☐ ***Personal and Professional Skills and Attributes***	L6-L11 (T)			
☐ ***Interpersonal Skills: Teamwork and Communication***	L12 (U)			
☐ ***CDIO in the enterprise, societal and environmental context***	L13 (U)			
Chapter 5. Correlation analysis and Regression	L5	3	4	
5.1. Relationship between two random variables				
5.2. Correlation coefficient				
5.3. Linear regression model				
5.4. Multivariable linear regression				
5.4. Polynomial regression analysis				
☐ ***Personal and Professional Skills and Attributes***	L6-L11 (T)			
☐ ***Interpersonal Skills: Teamwork and Communication***	L12 (U)			

☐ ***CDIO in the enterprise, societal and environmental context***	L13 (U)			
Chapter 6. Variance analysis	L5	1	3	
6.1. Simple variance analysis - ANOVA				
6.2. Two-side variance analysis				
6.3. Covariance analysis - ANCOVA				
☐ ***Personal and Professional Skills and Attributes***	L6-L11 (T)			
☐ ***Interpersonal Skills: Teamwork and Communication***	L12 (U)			
☐ ***CDIO in the enterprise, societal and environmental context***	L13 (U)			
Chapter 7. Logistic regression analysis	L5	1	3	
7.1 Logistic regression analysis				
7.2 Multivariable logistic regression and model selection				
☐ ***Personal and Professional Skills and Attributes***	L6-L11 (T)			
☐ ***Interpersonal Skills: Teamwork and Communication***	L12 (U)			
☐ ***CDIO in the enterprise, societal and environmental context***	L13 (U)			
Chapter 8. Time series data analysis	L5, L9	2	2	
8.1. Components of time series data				
8.2. Prediction models for time series data				
☐ ***Personal and Professional Skills and Attributes***	L6-L11 (T)			
☐ ***Interpersonal Skills: Teamwork and Communication***	L12 (U)			
☐ ***CDIO in the enterprise, societal and environmental context***	L13 (U)			
Summary of skills in course level				

☐ ***Personal and Professional Skills and Attributes***	L6-L11 (T)
☐ ***Interpersonal Skills: Teamwork and Communication***	L12 (U)
☐ ***CDIO in the enterprise, societal and environmental context***	L13 (U)

6. Teaching and learning methods

ID	Teaching method/technique		Description
M1.	Lecturing	☒	
M2.	Questions – Answers	☐	
M3.	Group-based Learning	☒	
M4.	Problem-based Learning	☐	
M5.	Project-based Learning	☐	
M6.	Case studies	☒	
M7.	Role play	☐	
M8.	Demo	☒	
M9.	Simulations	☐	
M10.	Debate	☒	
M11.	Game	☐	
M12.	Brainstorming	☐	
M13.	Think-Pair-Share	☐	

7. Course assessment

ID	Assessment activity		Quantity	Weight	LOs assessed
T1.	Text-based midterm exam	☒	1	25%	*L1, L2, L4*
T2.	Text-based final exam	☐			
T3.	Practice midterm exam	☐			
T4.	Practice final exam	☒	*1*	*25%*	*L3, L5*
T5.	Report	☐			
T6.	In-class exercises	☐			
T7.	Homework assignments	☐			
T8.	Question – Answer	☐			
T9.	Term Project	☐			
T10.	Final Exam	☒		50%	*L1, L2, L3, L4, L5*
Formula for Overall score	**T1*25% + T2*25% + T10*50%**				

8. Course requirements and expectations

8.1. Requirements on attendance

- Students are responsible for attending in all classes. In case of absence due to force majeure circumstances, there must be sufficient and reasonable evidence.
- Students who do not attend more than 20% of the class sections, whether for reason or not, are deemed not to have completed the course and must re-enroll in the following semester.

8.2. Requirements and expectations on student behaviors

- Students must show their respects for teachers and other learners.
- Students must be on time. Students who are late more than five minutes will not be allowed to attend the class.
- Students should not make noise and interfere with others in the learning process.
- Students should not eat, chew gum, and use devices such as cell phones, music players during class hours.
- Laptops and tablets can only be used in class for the purpose of learning.
- Students who violate the above principles will be asked to leave the class and considered absent from the class.

8.3. Requirements on learning issues

Issues related to applying for score reservation, scoring complaints, scoring, exam disciplines are done according to the Learning Regulation of Tra Vinh University.

9. Tentative course instructor

Tram Hoang Nam

DEAN	DEPARTMENT HEAD	LECTURER
		Tram Hoang Nam