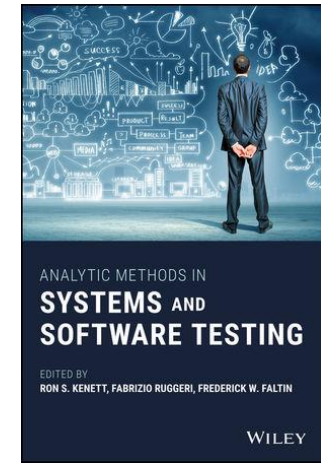
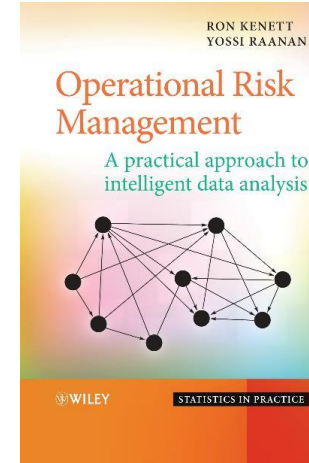
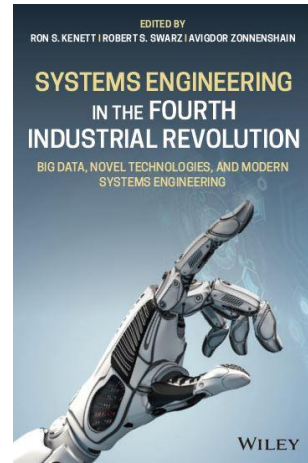
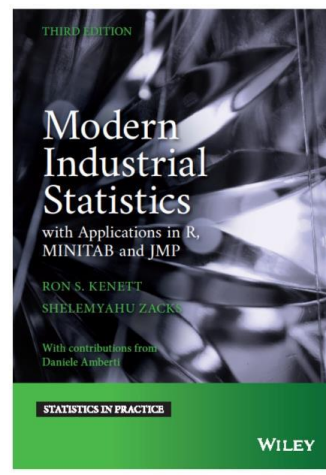
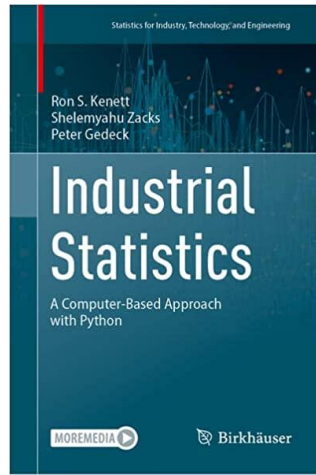
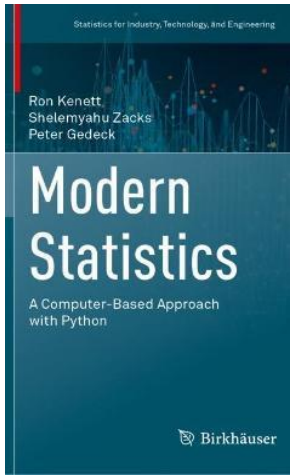


# **A Biomed Data Analyst Training Program**

## **Introduction**

**Professor Ron S. Kenett**



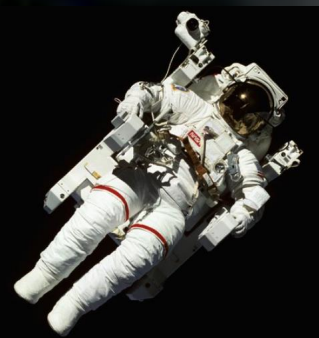
**Professor Ron Kenett**, Chairman of the *KPA Group*, Senior Research Fellow, *Samuel Neaman Institute*, Technion and chairman of the *Data Science Society*, Israel. Ron authored and co-authored over 250 papers and 18 books on topics ranging from industrial statistics, biostatistics, data science, surveys and statistical methods in healthcare. He was awarded the 2015 Greenfield Medal by the Royal Statistical Society (RSS), the 2018 Box Medal by the European Network for Business and Industrial Statistics (ENBIS), life achievement award from the Israel Society for Quality (2022) and Alan Tayer Memorial Lecture, European Consortium for Mathematics in Industry (2023). He is editor in chief of *Wiley's StatsRef electronic Encyclopedia*, on the advisory board of *Data Centric Engineering* published by Cambridge University Press and the book series *Statistics for Industry, Technology and Engineering*, published by Springer. Ron is Member of the International Board of the UniSR University Centre for Statistics in the Biomedical Sciences (Cussb).



The  
application

The  
theory

The  
need





## The data analyst competencies - 1

1. Have a genuine desire to solve real problems and help others to solve problems.
2. Be able to help investigators formulate their problem in quantifiable terms.
3. Be able to listen carefully and to ask probing questions.
4. Have a broad knowledge and true understanding of statistical and scientific methods.
5. Be able to adapt existing statistical procedures to novel environments.
6. Be able to locate or develop good statistical procedures in a timely fashion.
7. Be able to keep abreast of developments in statistics.
8. Be willing to meet deadlines, even if it requires substantial extra effort.
9. Be able to understand something about the clients' subject matter and speak a bit of the clients' language.



## The data analyst competencies - 2



10. Be a good teacher—much success in consulting depends on being able to help others understand statistical tools, and their strengths and weaknesses.
11. Be willing to settle for a reasonably correct approximate solution, then go on to the next problem.
12. Be able to identify important problems (and thus avoid spending too much time on projects of little significance).
13. Have the confidence to use as simple a procedure as will get the job done, be it design or analysis.
14. Be able to convince others of the validity of a solid solution and see to it that proper action is taken.
15. Be able to use computers effectively and direct others in their use.

# The data analyst competencies - 3



16. Be a good problem solver.
17. Be willing to meet clients regularly on their home ground, and take the responsibility to meet and communicate with all members of the working team.
18. Be diplomatic and know when to bend, when to stand firm, and how to help smooth over conflicts among other team members.
19. Be willing to get some experience in the actual collection of the data.
20. Be willing to take the time to check and double-check procedures and results.
21. Be able to communicate effectively in writing as well as orally (this often includes helping clients write their reports as well).
22. Be able to make a good estimate of how much effort will be required to solve the problem without actually having to solve the problem itself.

# Prof. Ron Kenett – Short Course in Data Science



Statistical Software for  
Insight-Driven Improvement  
Seeing is understanding.

JMP software combines interactive visualization with  
powerful statistics.

[Download Free Trial](#)

[www.jmp.com](http://www.jmp.com)

RON S. KENETT | THOMAS C. REDMAN

THE REAL WORK OF  
DATA SCIENCE

**Lecture Series in Analytics (Sala Laurea)** [Materials : *Kenett – Analytics*]

22/01 10.30-13.30

23/01 9.30-12.30

24/01 10.30-13.30

**Lecture Series in Causality (Sala Laurea)** [Materials : *Kenett – Causality*]

28/01 9.30-12.30

29/01 10.30-13.30

**Seminar on 'Statistics at a Crossroad'** [Materials : *Kenett – Seminar*]

10.45-11.45 Via Santa Sofia, 9 – aula M203

**List of references**

Kenett, R. and Zacks, S. (2021) Modern Industrial Statistics: With Applications in R, MINITAB, and JMP, 3rd Edition, Wiley, UK, <https://www.wiley.com/en-il/Modern+Industrial+Statistics%3A+With+Applicationsin+R%2C+MINITAB%2C+and+JMP%2C+3rd+Edition-p-9781119714903>



<https://www.linkedin.com/pulse/statistics-crossroad-generating-information-quality-ron-s-kenett/>



RON S. KENETT | THOMAS C. REDMAN

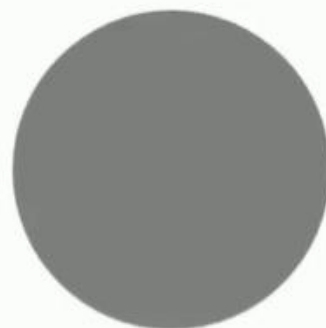
# THE REAL WORK OF DATA SCIENCE

TURNING DATA INTO INFORMATION,  
BETTER DECISIONS, AND STRONGER ORGANIZATIONS



WILEY

<https://www.youtube.com/watch?v=gHoeuuwCps>



*The Real Work of Data Science: How to  
turn data into information, better  
decisions, and stronger organizations*

Play (k)

Ron S. Kenett



0:03 / 59:45



<https://cran.r-project.org/web/packages/mistat>

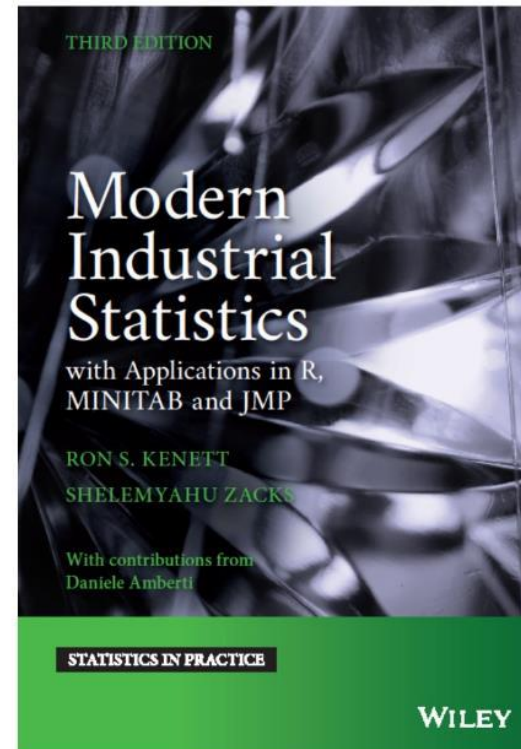
**mistat: Data Sets, Functions and Examples from the Book: "Modern Industrial Statistics" by Kenett, Zacks and Amberti**

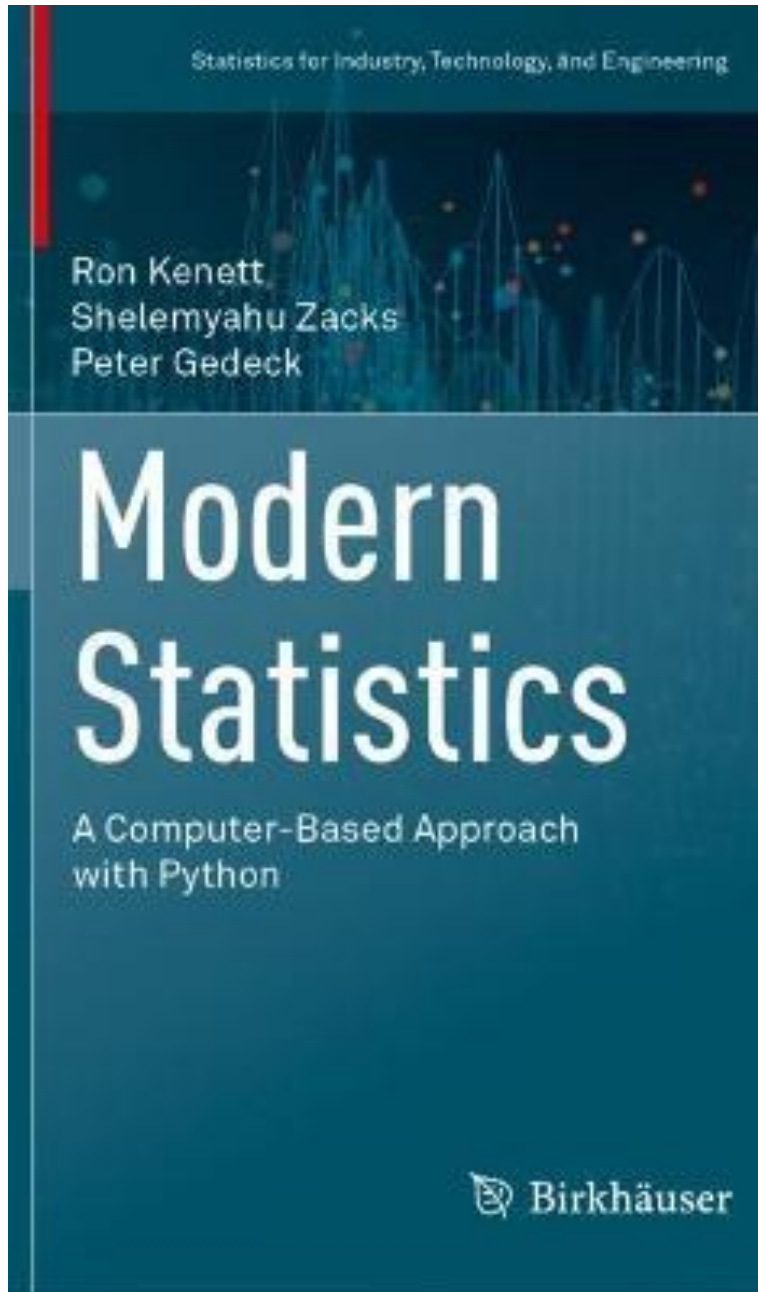
Provide all the data sets and statistical analysis applications used in "Modern Industrial Statistics: with applications in R, MINITAB and JMP" by R.S. Kenett and S. Zacks with contributions by D. Amberti, John Wiley and Sons, 2021, which is a third revised and expanded revision of "Modern Industrial Statistics: Design and Control of Quality and Reliability", R. Kenett and S. Zacks, Duxbury/Wadsworth Publishing, 1998.

Version: 2.0.4  
Depends: R ( $\geq 3.5$ )  
Imports: graphics, methods, stats, utils  
Suggests: [e1071](#), [mvtnorm](#), [AcceptanceSampling](#), [boot](#), [car](#), [Dodge](#), [tseries](#), [qcc](#), [DoE.base](#), [FrF2](#), [rsm](#), [LearnBayes](#), [ggplot2](#), [grid](#), [DiceEval](#), [DiceKriging](#), [DiceDesign](#), [DiceView](#), [lhs](#), [survival](#), [rpart](#), [fdapace](#), [randomForest](#), [xgboost](#)  
Published: 2023-02-17  
Author: Daniele Amberti  
Maintainer: Daniele Amberti <daniele.amberti at gmail.com>  
License: [GPL-2](#) | [GPL-3](#) [expanded from: GPL ( $\geq 2$ )]  
NeedsCompilation: no  
Materials: [README](#) [ChangeLog](#)  
CRAN checks: [mistat results](#)

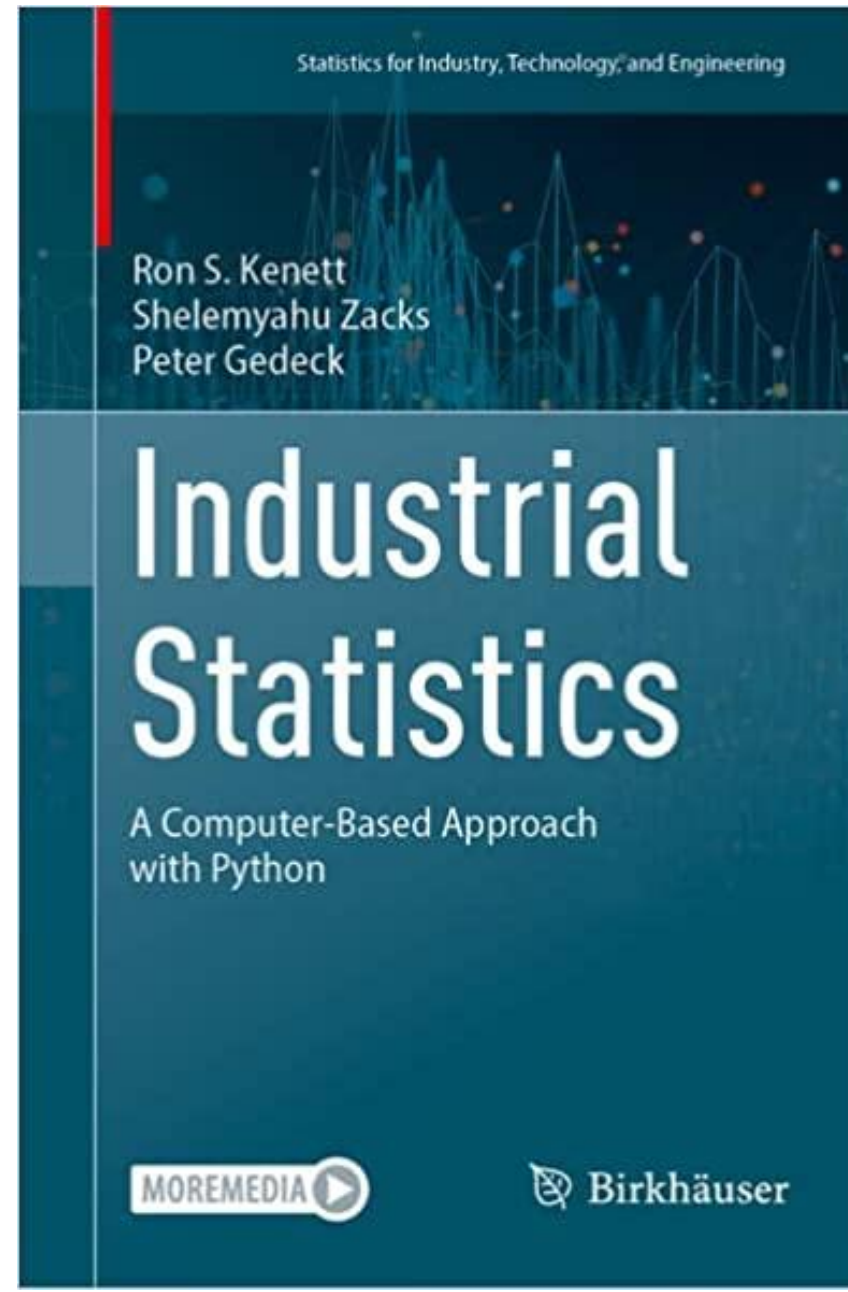
**Documentation:**

Reference manual: [mistat.pdf](#)







<https://gedeck.github.io/mistat-code-solutions/Modern Statistics/>




<https://gedeck.github.io/mistat-code-solutions/Industrial Statistics/>

 Home

 Environments

 Learning

 Community

A full Python IDE directly

DOCUMENTATION

ANACONDA BLOG



Applications on

base (root)

Channels



JupyterLab

3.3.2

An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.

Launch



Notebook

[↗ 6.4.8](#)

Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

Launch



Log in to Overleaf - Overleaf, Onl x mistat-code-solutions | Code rep x

gedeck.github.io/mistat-code-solutions/ModernStatistics/

Email link RonLinks Facebook LinkedIn News Ynet הארץ בנק הפועלים - לקו... Demo: Industry 4.0 קורונה - לוח בקרה Wiley StatsRef: Stati... Other bookmarks

View on GitHub

# Modern Statistics

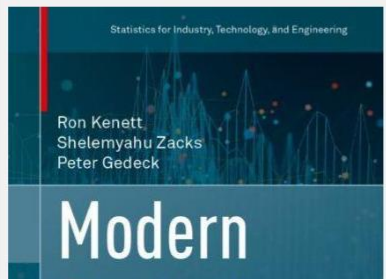
A Computer Based Approach with Python

Home > README.md

<https://gedeck.github.io/mistat-code-solutions/ModernStatistics/>

Test passing launch binder

## Code repository



Modern Statistics: A Computer Based Approach with Python  
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: Springer International Publishing;  
1st edition (September 15, 2022)  
ISBN-13: 978-3-031-07565-0 (print)  
ISBN-13: 978-3-031-07566-7 (eBook).  
Buy at Springer, Amazon, Barnes & Noble

### Python notebooks

Modern Statistics: A Computer Based Approach with Python  
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: Springer International Publishing; 1st edition (September 15, 2022)  
ISBN-13: 978-3031075650

(c) 2022 Ron Kenett, Shelemyahu Zacks, Peter Gedeck

The code needs to be executed in sequence.

Python packages and Python itself change over time. This can cause warnings or errors. We "Warnings" are f execution and need to be fixed in order to get results.

If you come across an issue with the code, please follow these steps

- Check the repository (<https://gedeck.github.io/mistat-code-solutions/>) to see if the code has been upgr
- Check the errata at (<https://gedeck.github.io/mistat-code-solutions/ModernStatistics/>) and the issue tra to see if the issue is known and a solution available.
- Report the problem using the issue tracker at <https://github.com/gedeck/mistat-code-solutions/issues>



Starting repository: `gedeck/mistat-code-solutions/binder-modern-statistics`

Your session is taking longer than usual to start!  
Check the log messages below to see what is happening.

Thanks to Google Cloud, OVH, GESIS Notebooks and the Turing Institute for supporting us!

Donate to mybinder.org!

<https://mybinder.org/v2/gh/gedeck/mistat-code-solutions/binder-modern-statistics?labpath=ModernStatistics%2Fnotebooks%2Findex.ipynb>



Starting repository: `gedeck/mistat-code-solutions/binder-modern-statistics`

The tool that powers this page is called BinderHub. It is an open source tool that you can deploy yourself.



<https://gedeck.github.io/mistat-code-solutions/ModernStatistics/>

<https://gedeck.github.io/mistat-code-solutions/IndustrialStatistics/>

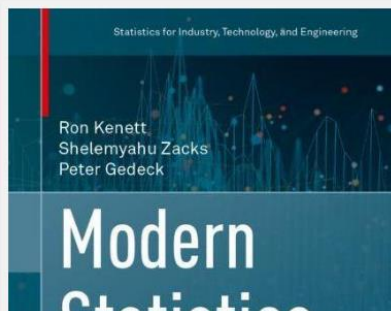
# Modern Statistics

A Computer Based Approach with Python

Home >README.md

Test passing launch binder

## Code repository



Modern Statistics: A Computer Based Approach with Python  
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: [Springer International Publishing](#); 1st edition (September 15, 2022)

ISBN-13: 978-3-031-07565-0 (hardcover)

ISBN-13: 978-3-031-07568-1 (softcover)

ISBN-13: 978-3-031-28482-3 (eBook).

Buy at [Springer](#), [Amazon](#), [Barnes & Noble](#)

# Industrial Statistics

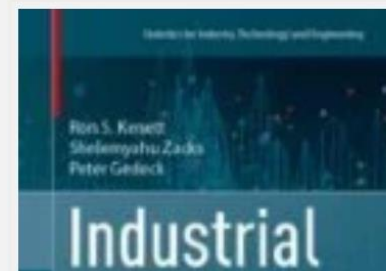
A Computer Based Approach with Python

Home >README.md

Test passing

launch binder

## Code repository



Industrial Statistics: A Computer Based Approach with Python  
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: [Springer International Publishing](#); 1st edition (August 5, 2023)

ISBN-13: 978-3-031-28481-6 (hardcover)

ISBN-13: 978-3-031-28484-7 (softcover)

ISBN-13: 978-3-031-28482-3 (eBook).

Buy at [Springer](#), [Amazon](#).



<https://gedeck.github.io/mistat-code-solutions/ModernStatistics/>

<https://gedeck.github.io/mistat-code-solutions/IndustrialStatistics/>

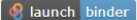
Modern Statistics: A Computer Based Approach with Python is a companion volume to the book [Industrial Statistics: A Computer Based Approach with Python](#).

This part of the repository contains:

- `notebooks` : Python code of individual chapters in [Jupyter notebooks](#) - [download all as notebooks.zip](#)
- `code` : Python code for solutions as plain [Python files](#) - [download all as code.zip](#)
- `solutions_manual` : [Solutions\\_Modernstatistics.pdf](#): solutions of exercises
- `solutions` : Python code for solutions in [Jupyter notebooks](#) - [download all as solutions.zip](#)
- `all` : zip file with all files combined - [download all as all.zip](#)
- `datafiles` : zip file with all data files - [download all as data\\_files.zip](#) - the `mistat` package gives you already access to all datafiles, you only need to download this file if you want to use it with different software

All the Python applications referred to in this book are contained in a package called `mistat` available for installation from the Python package index <https://pypi.org/project/mistat/>. The `mistat` packages is maintained in a GitHub repository at <https://github.com/gedeck/mistat>.

## Try the code

You can explore the code on [Binder](#) 

## Installation instructions

Instructions on installing Python and required packages are [here](#).

These Python packages are used in the code examples of *Modern Statistics*:

- `mistat` (for access to data sets and additional functionality)
- `numpy`
- `scipy`
- `scikit-learn`

Industrial Statistics: A Computer Based Approach with Python is a companion volume to the book [Modern Statistics: A Computer Based Approach with Python](#).

This part of the repository contains:

- `notebooks` : Python code of individual chapters in [Jupyter notebooks](#) - [download all as notebooks.zip](#)
- `code` : Python code for solutions as plain [Python files](#) - [download all as code.zip](#)
- `solutions_manual` : [Solutions\\_IndustrialStatistics.pdf](#): solutions of exercises
- `solutions` : Python code for solutions in [Jupyter notebooks](#) - [download all as solutions.zip](#)
- `all` : zip file with all files combined - [download all as all.zip](#)
- `datafiles` : zip file with all data files - [download all as data\\_files.zip](#) - the `mistat` package gives you already access to all datafiles, you only need to download this file if you want to use it with different software

All the Python applications referred to in this book are contained in a package called `mistat` available for installation from the Python package index <https://pypi.org/project/mistat/>. The `mistat` packages is maintained in a GitHub repository at <https://github.com/gedeck/mistat>.

## Table of contents (with sample excerpts from chapters)

Chapter 1: Introduction to Industrial Statistics ([sample 1](#))

Chapter 2: Basic Tools and Principles of Process Control ([sample 2](#))

Chapter 3: Advanced Methods of Statistical Process Control ([sample 3](#))

Chapter 4: Multivariate Statistical Process Control ([sample 4](#))

<https://gedeck.github.io/mistat-code-solutions/ModernStatistics/>

## Table of contents

- [Chapter 1: Analyzing Variability: Descriptive Statistics](#)
- [Chapter 2: Probability Models and Distribution Functions](#)
- [Chapter 3: Statistical Inference and Bootstrapping](#)
- [Chapter 4: Variability in Several Dimensions and Regression Models](#)
- [Chapter 5: Sampling for Estimation of Finite Population Quantities](#)
- [Chapter 6: Time Series Analysis and Prediction](#)
- [Chapter 7: Modern analytic methods: Part I](#)
- [Chapter 8: Modern analytic methods: Part II](#)

## Introductory videos

### Chapter 1: Analyzing Variability: Descriptive Statistics

The chapter focuses on statistical variability and on various methods of analyzing random data. Random results of experiments are illustrated with distinction between deterministic and random components of variability. The difference between accuracy and precision is explained. Frequency distributions are defined to represent random phenomena. Various characteristics of location and dispersion of frequency distributions are defined. The elements of exploratory data analysis are presented.

<https://gedeck.github.io/mistat-code-solutions/IndustrialStatistics/>

## Table of contents (with sample excerpts from chapters)

- [Chapter 1: Introduction to Industrial Statistics \(sample 1\)](#)
- [Chapter 2: Basic Tools and Principles of Process Control \(sample 2\)](#)
- [Chapter 3: Advanced Methods of Statistical Process Control \(sample 3\)](#)
- [Chapter 4: Multivariate Statistical Process Control \(sample 4\)](#)
- [Chapter 5: Classical Design and Analysis of Experiments \(sample 5\)](#)
- [Chapter 6: Quality by Design \(sample 6\)](#)
- [Chapter 7: Computer Experiments \(sample 7\)](#)
- [Chapter 8: Cybermanufacturing and Digital Twins \(sample 8\)](#)
- [Chapter 9: Reliability Analysis \(sample 9\)](#)
- [Chapter 10: Bayesian Reliability Estimation and Prediction \(sample 10\)](#)
- [Chapter 11: Sampling Plans for Batch and Sequential Inspection \(sample 11\)](#)

## Installation instructions

Instructions on installing Python and required packages are [here](#).

These Python packages are used in the code of *Industrial Statistics*:

- mistat (for access to data sets and additional functionality)
- numpy
- pandas
- scipy
- statsmodels
- matplotlib==3.6.0
- seaborn
- pingouin
- lifelines

[https://user-images.githubusercontent.com/8720575/180794703-c6f05f40-eefd-4e1a-93f9-42cb78e6a6b4.mp4?fbclid=IwAR08kOITzOhBUIYZhyclkrh\\_p7Ad9q53rnrwUicOmiUBHsbd4TwJgzkhEpo](https://user-images.githubusercontent.com/8720575/180794703-c6f05f40-eefd-4e1a-93f9-42cb78e6a6b4.mp4?fbclid=IwAR08kOITzOhBUIYZhyclkrh_p7Ad9q53rnrwUicOmiUBHsbd4TwJgzkhEpo)



<b>1</b>	<b>Analyzing Variability: Descriptive Statistics</b> .....	<b>1</b>			
1.1	Random Phenomena and the Structure of Observations .....	1			
1.2	Accuracy and Precision of Measurements.....	6			
1.3	The Population and the Sample .....	8			
1.4	Descriptive Analysis of Sample Values.....	9			
1.4.1	Frequency Distributions of Discrete Random Variables ...	9			
1.4.2	Frequency Distributions of Continuous Random Variables.....	14			
1.4.3	Statistics of the Ordered Sample .....	17			
1.4.4	Statistics of Location and Dispersion .....	19			
1.5	Prediction Intervals .....	23			
1.6	Additional Techniques of Exploratory Data Analysis .....	25			
1.6.1	Density Plots .....	25			
1.6.2	Box and Whiskers Plots .....	27			
1.6.3	Quantile Plots .....	29			
1.6.4	Stem-and-Leaf Diagrams .....	30			
1.6.5	Robust Statistics for Location and Dispersion.....	31			
1.7	Chapter Highlights.....	34			
1.8	Exercises .....	34			
<b>2</b>	<b>Probability Models and Distribution Functions</b> .....	<b>39</b>			
2.1	Basic Probability.....	39			
2.1.1	Events and Sample Spaces: Formal Presentation of Random Measurements .....	39			
2.1.2	Basic Rules of Operations with Events: Unions and Intersections .....	41			
2.1.3	Probabilities of Events .....	44			
2.1.4	Probability Functions for Random Sampling.....	46			
2.1.5	Conditional Probabilities and Independence of Events ....	49			
2.1.6	Bayes' Theorem and Its Application.....	51			
2.2	Random Variables and Their Distributions .....	54			
			2.2.1	Discrete and Continuous Distributions.....	55
			2.2.1.1	Discrete Random Variables .....	55
			2.2.1.2	Continuous Random Variables .....	56
			2.2.2	Expected Values and Moments of Distributions.....	59
			2.2.3	The Standard Deviation, Quantiles, Measures of Skewness, and Kurtosis .....	62
			2.2.4	Moment Generating Functions .....	65
		2.3	Families of Discrete Distribution .....	66	
		2.3.1	The Binomial Distribution.....	66	
		2.3.2	The Hypergeometric Distribution .....	69	
		2.3.3	The Poisson Distribution .....	72	
		2.3.4	The Geometric and Negative Binomial Distributions .....	74	
		2.4	Continuous Distributions .....	78	
		2.4.1	The Uniform Distribution on the Interval $(a, b)$ , $a < b$ ....	78	
		2.4.2	The Normal and Log-Normal Distributions .....	79	
			2.4.2.1	The Normal Distribution.....	79
			2.4.2.2	The Log-Normal Distribution .....	84
		2.4.3	The Exponential Distribution.....	85	
		2.4.4	The Gamma and Weibull Distributions .....	88	
		2.4.5	The Beta Distributions .....	92	
		2.5	Joint, Marginal, and Conditional Distributions.....	93	
		2.5.1	Joint and Marginal Distributions .....	93	
		2.5.2	Covariance and Correlation.....	96	
		2.5.3	Conditional Distributions .....	99	
		2.6	Some Multivariate Distributions .....	102	
		2.6.1	The Multinomial Distribution .....	102	
		2.6.2	The Multi-Hypergeometric Distribution .....	104	
		2.6.3	The Bivariate Normal Distribution .....	105	
		2.7	Distribution of Order Statistics.....	108	
		2.8	Linear Combinations of Random Variables .....	111	
		2.9	Large Sample Approximations.....	117	
		2.9.1	The Law of Large Numbers .....	117	
		2.9.2	The Central Limit Theorem .....	117	
		2.9.3	Some Normal Approximations .....	119	
		2.10	Additional Distributions of Statistics of Normal Samples .....	120	
		2.10.1	Distribution of the Sample Variance .....	121	
		2.10.2	The "Student" $t$ -Statistic.....	122	
		2.10.3	Distribution of the Variance Ratio.....	123	
		2.11	Chapter Highlights.....	125	
		2.12	Exercises .....	126	
		<b>3</b>	<b>Statistical Inference and Bootstrapping</b> .....	<b>139</b>	
		3.1	Sampling Characteristics of Estimators .....	139	
		3.2	Some Methods of Point Estimation.....	141	
		3.2.1	Moment Equation Estimators .....	142	



3.2.2	The Method of Least Squares .....	144	3.11.5.2	Comparing Several Means: The One-Way Analysis of Variance .....	201	4.10	Chapter Highlights .....	290
3.2.3	Maximum Likelihood Estimators .....	146	3.12	Bootstrap Tolerance Intervals .....	204	4.11	Exercises .....	291
3.3	Comparison of Sample Estimates .....	149	3.12.1	Bootstrap Tolerance Intervals for Bernoulli Samples .....	204	<b>5</b>	<b>Sampling for Estimation of Finite Population Quantities</b> .....	299
3.3.1	Basic Concepts .....	149	3.12.2	Tolerance Interval for Continuous Variables .....	205	5.1	Sampling and the Estimation Problem .....	299
3.3.2	Some Common One-Sample Tests of Hypotheses .....	152	3.12.3	Distribution-Free Tolerance Intervals .....	206	5.1.1	Basic Definitions .....	299
3.3.2.1	The Z-Test: Testing the Mean of a Normal Distribution, $\sigma^2$ Known .....	152	3.13	Non-Parametric Tests .....	208	5.1.2	Drawing a Random Sample from a Finite Population .....	301
3.3.2.2	The $t$ -Test: Testing the Mean of a Normal Distribution, $\sigma^2$ Unknown .....	155	3.13.1	The Sign Test .....	208	5.1.3	Sample Estimates of Population Quantities and Their Sampling Distribution .....	302
3.3.2.3	The Chi-Squared Test: Testing the Variance of a Normal Distribution .....	156	3.13.2	The Randomization Test .....	210	5.2	Estimation with Simple Random Samples .....	305
3.3.2.4	Testing Hypotheses About the Success Probability, $p$ , in Binomial Trials .....	158	3.13.3	The Wilcoxon Signed-Rank Test .....	211	5.2.1	Properties of $\bar{X}_n$ and $S_n^2$ Under RSWR .....	306
3.4	Confidence Intervals .....	160	3.14	Chapter Highlights .....	214	5.2.2	Properties of $\bar{X}_n$ and $S_n^2$ Under RSWOR .....	310
3.4.1	Confidence Intervals for $\mu$ ; $\sigma$ Known .....	161	3.15	Exercises .....	215	5.3	Estimating the Mean with Stratified RSWOR .....	314
3.4.2	Confidence Intervals for $\mu$ ; $\sigma$ Unknown .....	162	<b>4</b>	<b>Variability in Several Dimensions and Regression Models</b> .....	225	5.4	Proportional and Optimal Allocation .....	316
3.4.3	Confidence Intervals for $\sigma^2$ .....	162	4.1	Graphical Display and Analysis .....	226	5.5	Prediction Models with Known Covariates .....	320
3.4.4	Confidence Intervals for $p$ .....	163	4.1.1	Scatterplots .....	226	5.6	Chapter Highlights .....	324
3.5	Tolerance Intervals .....	166	4.1.2	Multiple Boxplots .....	229	5.7	Exercises .....	325
3.5.1	Tolerance Intervals for the Normal Distributions .....	166	4.2	Frequency Distributions in Several Dimensions .....	230	<b>6</b>	<b>Time Series Analysis and Prediction</b> .....	329
3.6	Testing for Normality with Probability Plots .....	169	4.2.1	Bivariate Joint Frequency Distributions .....	231	6.1	The Components of a Time Series .....	330
3.7	Tests of Goodness of Fit .....	173	4.2.2	Conditional Distributions .....	234	6.1.1	The Trend and Covariances .....	330
3.7.1	The Chi-Square Test (Large Samples) .....	173	4.3	Correlation and Regression Analysis .....	235	6.1.2	Analyzing Time Series with Python .....	331
3.7.2	The Kolmogorov-Smirnov Test .....	175	4.3.1	Covariances and Correlations .....	236	6.2	Covariance Stationary Time Series .....	336
3.8	Bayesian Decision Procedures .....	176	4.3.2	Fitting Simple Regression Lines to Data .....	237	6.2.1	Moving Averages .....	337
3.8.1	Prior and Posterior Distributions .....	177	4.3.2.1	The Least Squares Method .....	239	6.2.2	Auto-Regressive Time Series .....	338
3.8.2	Bayesian Testing and Estimation .....	181	4.3.2.2	Regression and Prediction Intervals .....	243	6.2.3	Auto-Regressive Moving Average Time Series .....	343
3.8.2.1	Bayesian Testing .....	181	4.4	Multiple Regression .....	245	6.2.4	Integrated Auto-Regressive Moving Average Time Series .....	344
3.8.2.2	Bayesian Estimation .....	184	4.4.1	Regression on Two Variables .....	246	6.2.5	Applications with Python .....	345
3.8.3	Credibility Intervals for Real Parameters .....	185	4.4.2	Partial Regression and Correlation .....	251	6.3	Linear Predictors for Covariance Stationary Time Series .....	346
3.9	Random Sampling from Reference Distributions .....	186	4.4.3	Multiple Linear Regression .....	254	6.3.1	Optimal Linear Predictors .....	346
3.10	Bootstrap Sampling .....	189	4.4.4	Partial- $F$ Tests and the Sequential SS .....	260	6.4	Predictors for Non-stationary Time Series .....	349
3.10.1	The Bootstrap Method .....	189	4.4.5	Model Construction: Step-Wise Regression .....	263	6.4.1	Quadratic LSE Predictors .....	349
3.10.2	Examining the Bootstrap Method .....	190	4.4.6	Regression Diagnostics .....	265	6.4.2	Moving Average Smoothing Predictors .....	351
3.10.3	Harnessing the Bootstrap Method .....	192	4.5	Quantal Response Analysis: Logistic Regression .....	268	6.5	Dynamic Linear Models .....	352
3.11	Bootstrap Testing of Hypotheses .....	192	4.6	The Analysis of Variance: The Comparison of Means .....	271	6.5.1	Some Special Cases .....	353
3.11.1	Bootstrap Testing and Confidence Intervals for the Mean .....	192	4.6.1	The Statistical Model .....	271	6.5.1.1	The Normal Random Walk .....	353
3.11.2	Studentized Test for the Mean .....	193	4.6.2	The One-Way Analysis of Variance (ANOVA) .....	271	6.5.1.2	Dynamic Linear Model With Linear Growth ...	354
3.11.3	Studentized Test for the Difference of Two Means .....	195	4.7	Simultaneous Confidence Intervals: Multiple Comparisons .....	275	6.5.1.3	Dynamic Linear Model for ARMA(p,q) .....	355
3.11.4	Bootstrap Tests and Confidence Intervals for the Variance .....	197	4.8	Contingency Tables .....	279	6.6	Chapter Highlights .....	358
3.11.5	Comparing Statistics of Several Samples .....	199	4.8.1	The Structure of Contingency Tables .....	279	6.7	Exercises .....	359
3.11.5.1	Comparing Variances of Several Samples .....	200	4.8.2	Indices of association for contingency tables .....	282	<b>7</b>	<b>Modern Analytic Methods: Part I</b> .....	361
			4.8.2.1	Two Interval-Scaled Variables .....	282	7.1	Introduction to Computer Age Statistics .....	361
			4.8.2.2	Indices of Association for Categorical Variables .....	284	7.2	Data Preparation .....	362
			4.9	Categorical Data Analysis .....	288	7.3	The Information Quality Framework .....	363
			4.9.1	Comparison of Binomial Experiments .....	288			



## Analyzing Variability: Descriptive Statistics

Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Pages 1-38

---

## Probability Models and Distribution Functions

Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Pages 39-138

---

## Statistical Inference and Bootstrapping

Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Pages 139-223

---

## Variability in Several Dimensions and Regression Models

Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Pages 225-297

---

## Sampling for Estimation of Finite Population Quantities

Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Pages 299-327

## Time Series Analysis and Prediction

Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Pages 329-360

---

## Modern Analytic Methods: Part I

Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Pages 361-393

---

## Modern Analytic Methods: Part II

Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Pages 395-419

---

# Chapter 6

Modern Statistics: A Computer Based Approach with Python  
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: Springer International Publishing; 1st edition (September 15, 2022)  
ISBN-13: 978-3031075650

(c) 2022 Ron Kenett, Shelemyahu Zacks, Peter Gedeck

The code needs to be executed in sequence.

## Time Series Analysis and Prediction

Ron Kenett, Shelemyahu Zacks, Peter Gedeck  
Pages 329-360

```
In [1]: import os
os.environ['OUTDATED_IGNORE'] = '1'
import warnings
from outdated import OutdatedPackageWarning
warnings.filterwarnings('ignore', category=FutureWarning)
warnings.filterwarnings('ignore', category=OutdatedPackageWarning)
```

## Time Series Analysis and Prediction

```
In [2]: import datetime
import statsmodels.formula.api as smf
from statsmodels.tools.sm_exceptions import ValueWarning
import pandas as pd

import random
import numpy as np
import pingouin as pg
from scipy import stats
import matplotlib.pyplot as plt
import mistat
```



# Chapter 7

Modern Statistics: A Computer Based Approach with Python  
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: Springer International Publishing; 1st edition (September 15, 2022)  
ISBN-13: 978-3031075650

(c) 2022 Ron Kenett, Shelemyahu Zacks, Peter Gedeck

The code needs to be executed in sequence.

```
In [1]: import warnings
import os
os.environ['OUTDATED_IGNORE'] = '1'
from outdated import OutdatedPackageWarning
warnings.filterwarnings('ignore', category=FutureWarning)
warnings.filterwarnings('ignore', category=OutdatedPackageWarning)
```

## [Modern Analytic Methods: Part I](#)

Ron Kenett, Shelemyahu Zacks, Peter Gedeck  
Pages 361-393

## Modern analytic methods: Part I

```
In [2]: import warnings
import random
import pandas as pd
import numpy as np
from sklearn.ensemble import RandomForestClassifier
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score
from sklearn.impute import SimpleImputer
from sklearn.neural_network import MLPClassifier
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
```

## Chapter 8

Modern Statistics: A Computer Based Approach with Python  
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: Springer International Publishing; 1st edition (September 15, 2022)  
ISBN-13: 978-3031075650

(c) 2022 Ron Kenett, Shelemyahu Zacks, Peter Gedeck

The code needs to be executed in sequence.

```
In [1]: import os
os.environ['OUTDATED_IGNORE'] = '1'
import warnings
from outdated import OutdatedPackageWarning
warnings.filterwarnings('ignore', category=FutureWarning)
warnings.filterwarnings('ignore', category=OutdatedPackageWarning)
```

## Modern analytic methods: Part II

```
In [2]: import networkx as nx

import statsmodels.api as sm
from statsmodels.tsa.stattools import grangercausalitytests
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import mistat
```

### [Modern Analytic Methods: Part II](#)

Ron Kenett, Shelemyahu Zacks, Peter Gedeck  
Pages 395-419

# A Biomed Data Analyst Training Program

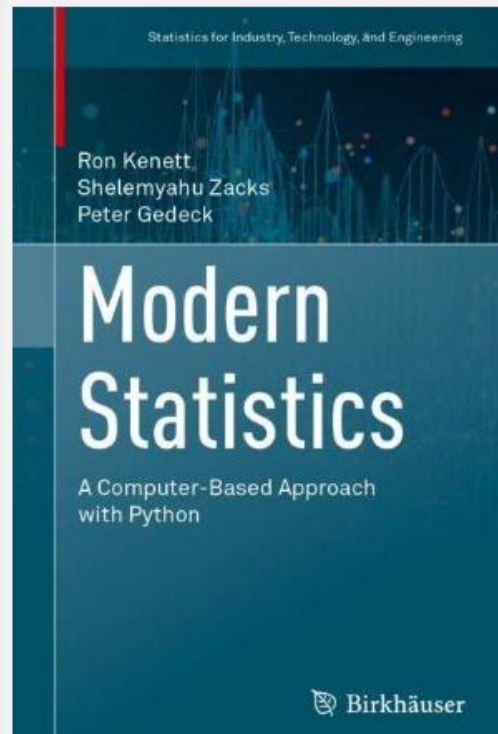
Modern Statistics: A Computer Based Approach with Python

Home >[README.md](#)

 Test  passing  launch  binder

[https://gedeck.github.io/mistat-code-solutions/BioMed\\_DataAnalyst\\_Course/](https://gedeck.github.io/mistat-code-solutions/BioMed_DataAnalyst_Course/)

## A Biomed Data Analyst Training Program



Modern Statistics: A Computer Based Approach with Python  
by Ron Kenett, Shelemyahu Zacks, Peter Gedeck

Publisher: [Springer International Publishing](#); 1st edition (September 2022)

ISBN-13: 978-3-031-07565-0 (hardcover)

ISBN-13: 978-3-031-07568-1 (softcover)

ISBN-13: 978-3-031-28482-3 (eBook).

Buy at [Springer](#), [Amazon](#), [Barnes & Noble](#)

Errata: [See known errata here](#)

# Slides

1. Introduction
2. Data types and data integration
3. Supervised learning
4. Model performance
5. Time series
6. Data visualization
7. Causality and experimental design

[https://gedeck.github.io/mistat-code-solutions/BioMed\\_DataAnalyst\\_Course/](https://gedeck.github.io/mistat-code-solutions/BioMed_DataAnalyst_Course/)

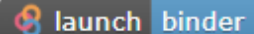
## Code and data files

This part of the repository contains:

- `notebooks` : Python code of individual chapters in [Jupyter notebooks - download all as notebooks.zip](#)

The Python package `mistat` contains datafiles and utility functions referred to in the [Modern Statistics](#) book. It is available for installation from the Python package index <https://pypi.org/project/mistat/>. The `mistat` packages is maintained in a GitHub repository at <https://github.com/gedeck/mistat>.

## Try the code

You can explore the code on [Binder](#) .



