

Modern Statistics:

A Computer-Based Approach with Python

<https://link.springer.com/book/10.1007/978-3-031-07566-7#toc>



Book series

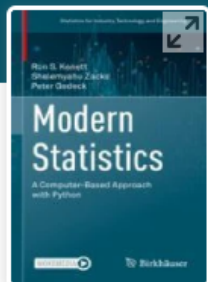
Statistics for Industry, Technology, and Engineering

 [Editorial board](#)

About this book series

The *Statistics for Industry, Technology, and Engineering* series will present up-to-date statistical ideas and methods that are relevant to researchers and accessible to an interdisciplinary audience: carefully organized authoritative presentations, numerous illustrative examples based on current practice, reliable methods, realistic data sets, and discussions of select new emerging methods and their application potential. Publications will appeal to a broad interdisciplinary readership including both researchers and practitioners in applied statistics, data science, industrial statistics, engineering statistics, quality control, manufacturing, applied reliability, and general quality improvement methods. — [show all](#)

<https://link.springer.com/book/9783031075650>



Textbook | © 2022

Modern Statistics

A Computer-Based Approach with Python

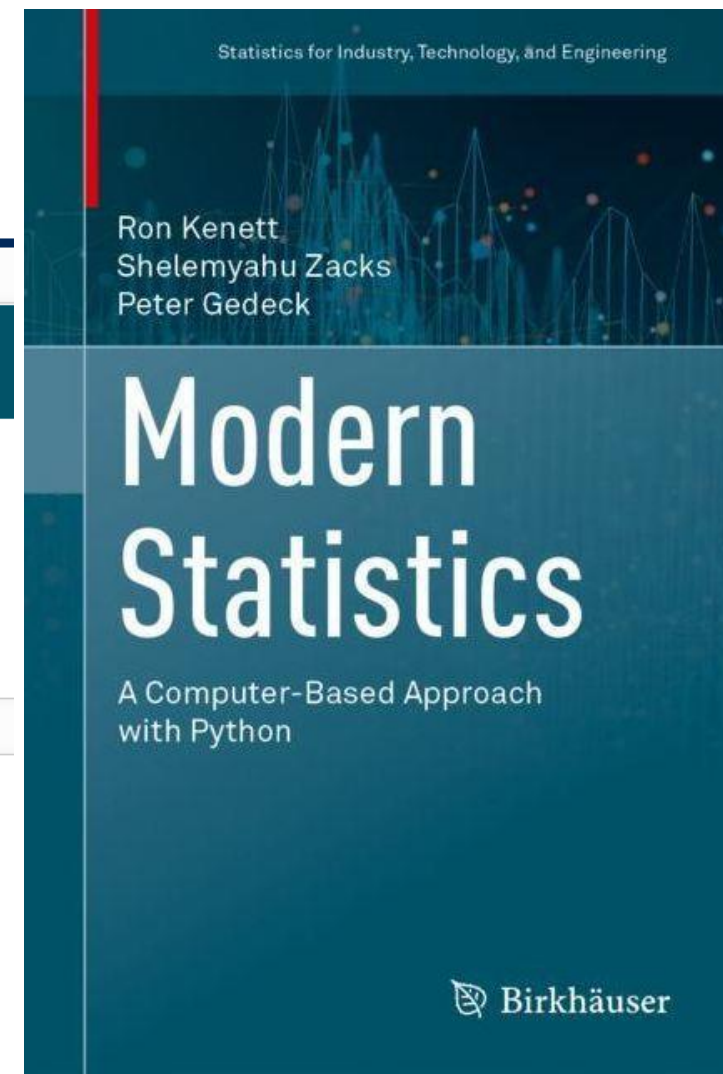
Authors: [Ron S. Kenett](#), [Shelemyahu Zacks](#), [Peter Gedeck](#)

Demonstrates how to incorporate Python into the modern statistics curriculum

Includes over 40 case studies to facilitate experiential learning

An accompanying Python package is available for download, allowing students to engage directly with the material

Part of the book series: [Statistics for Industry, Technology, and Engineering](#) (SITE)



[Table of contents](#)

[About this book](#)

[Keywords](#)

[Authors and Affiliations](#)

[About the authors](#)

[Bibliographic Information](#)



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

- `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`

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>.

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

Code repository

Installation instructions

Introductory videos

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

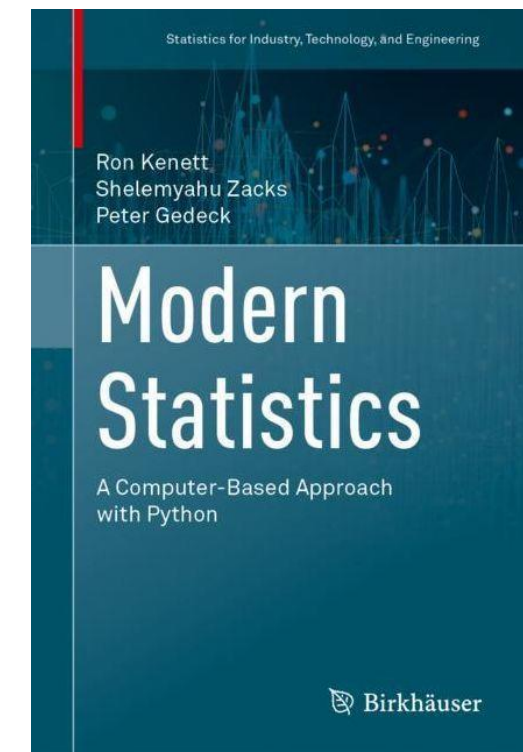
Table of contents (with sample excerpts from chapters)

- 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](#))

Companion book

- 1 Analyzing Variability: Descriptive Statistics**
 - 1.1 Random Phenomena and the Structure of Observations
 - 1.2 Accuracy and Precision of Measurements
 - 1.3 The Population and the Sample
 - 1.4 Descriptive Analysis of Sample Values
 - 1.5 Prediction Intervals
 - 1.6 Additional Techniques of Exploratory Data Analysis
 - 1.7 Chapter Highlights
 - 1.8 Exercises
- 2 Probability Models and Distribution Functions**
 - 2.1 Basic Probability
 - 2.2 Random Variables and Their Distributions
 - 2.3 Families of Discrete Distribution
 - 2.4 Continuous Distributions
 - 2.5 Joint, Marginal and Conditional Distributions
 - 2.6 Some Multivariate Distributions
 - 2.7 Distribution of Order Statistics
 - 2.8 Linear Combinations of Random Variables
 - 2.9 Large Sample Approximations
 - 2.10 Additional Distributions of Statistics of Normal Samples
 - 2.11 Chapter Highlights
 - 2.12 Exercises
- 3 Statistical Inference and Bootstrapping**
 - 3.1 Sampling Characteristics of Estimators
 - 3.2 Some Methods of Point Estimation
 - 3.3 Comparison of Sample Estimates
 - 3.4 Confidence Intervals
 - 3.5 Tolerance Intervals
 - 3.6 Testing for Normality with Probability Plots
 - 3.7 Tests of Goodness of Fit
 - 3.8 Bayesian Decision Procedures
 - 3.9 Random Sampling From Reference Distributions
 - 3.10 Bootstrap Sampling
 - 3.11 Bootstrap Testing of Hypotheses
 - 3.12 Bootstrap Tolerance Intervals
 - 3.13 Non-Parametric Tests
 - 3.14 Chapter Highlights

- 4 Variability in Several Dimensions and Regression Models**
 - 4.1 Graphical Display and Analysis
 - 4.2 Frequency Distributions in Several Dimensions
 - 4.3 Correlation and Regression Analysis
 - 4.4 Multiple Regression
 - 4.5 Quantal Response Analysis: Logistic Regression
 - 4.6 The Analysis of Variance: The Comparison of Means
 - 4.7 Simultaneous Confidence Intervals: Multiple Comparisons
 - 4.8 Contingency Tables
 - 4.9 Categorical Data Analysis
 - 4.10 Chapter Highlights
 - 4.11 Exercises
- 5 Sampling for Estimation of Finite Population Quantities**
 - 5.1 Sampling and the Estimation Problem
 - 5.2 Estimation with Simple Random Samples
 - 5.3 Estimating the Mean with Stratified RSWOR
 - 5.4 Proportional and Optimal Allocation
 - 5.5 Prediction Models with Known Covariates
 - 5.6 Chapter Highlights
 - 5.7 Exercises
- 6 Time Series Analysis and Prediction**
 - 6.1 The Components of a Time Series
 - 6.2 Covariance Stationary Time Series
 - 6.3 Linear Predictors for Covariance Stationary Time Series
 - 6.4 Predictors for Non-Stationary Time Series
 - 6.5 Dynamic Linear Models
 - 6.6 Chapter Highlights
 - 6.7 Exercises
- 7 Modern analytic methods: Part I**
 - 7.1 Introduction to Computer Age Statistics
 - 7.2 Data Preparation
 - 7.3 The Information Quality Framework
 - 7.4 Determining Model Performance
 - 7.5 Decision Trees
 - 7.6 Ensemble Models
 - 7.7 Naïve Bayes Classifier
 - 7.8 Neural Networks
 - 7.9 Clustering Methods
 - 7.10 Chapter Highlights
 - 7.11 Exercises



- 8 Modern analytic methods: Part II**
 - 8.1 Functional Data Analysis
 - 8.2 Text Analytics
 - 8.3 Bayesian Networks
 - 8.4 Causality Models
 - 8.5 Chapter Highlights
 - 8.6 Exercises

- A Introduction to Python**
- B List of Python packages**
- C Code Repository and Solution Manual**
- D Bibliography**
- Index**

Modern Statistics: A Computer Based Approach with Python

S

Solutions

Import required modules and define required functions

```
import math
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib
from scipy import stats

def trim_std(data, alpha):
    """ Calculate trimmed standard deviation """
    data = np.array(data)
    data.sort()
    n = len(data)
    low = int(n * alpha) + 1
    high = int(n * (1 - alpha))
    return data[low:(high + 1)].std()
```

Solution 1.1 `random.choices` selects k values from the list using sampling with replacement.

```
import random
random.seed(1)
values = random.choices([1, 2, 3, 4, 5, 6], k=50)
```

Counter counts the number of occurrences of a given value in a list.

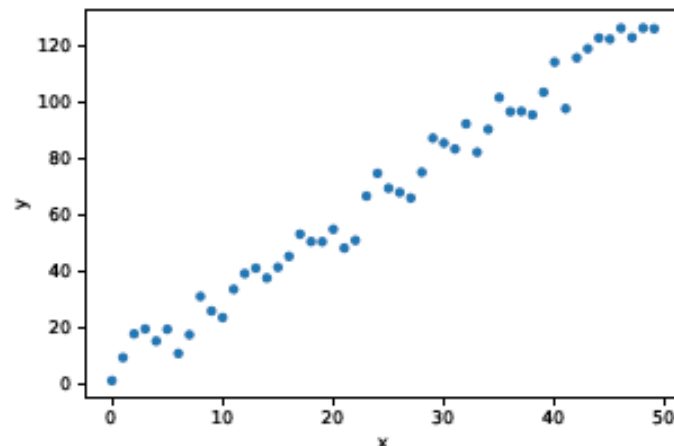
```
from collections import Counter
Counter(values)

Counter({1: 9, 6: 9, 5: 8, 2: 10, 3: 10, 4: 4})
```

The expected frequency in each cell, under randomness is $50/6 = 8.3$. You will get different numerical results, due to randomness.

Solution 1.2 The Python function `range` is an iterator. As we need a list of values, we need to explicitly convert it.

```
x = list(range(50))
y = [5 + 2.5 * xi for xi in x]
y = [yi + random.uniform(-10, 10) for yi in y]
pd.DataFrame({'x': x, 'y': y}).plot.scatter(x='x', y='y')
plt.show()
```



Solution 1.3 In Python

```
from scipy.stats import binom
np.random.seed(1)

for p in (0.1, 0.3, 0.7, 0.9):
    X = binom.rvs(1, p, size=50)
    print(p, sum(X))
```

```
0.1 4
0.3 12
0.7 33
0.9 43
```

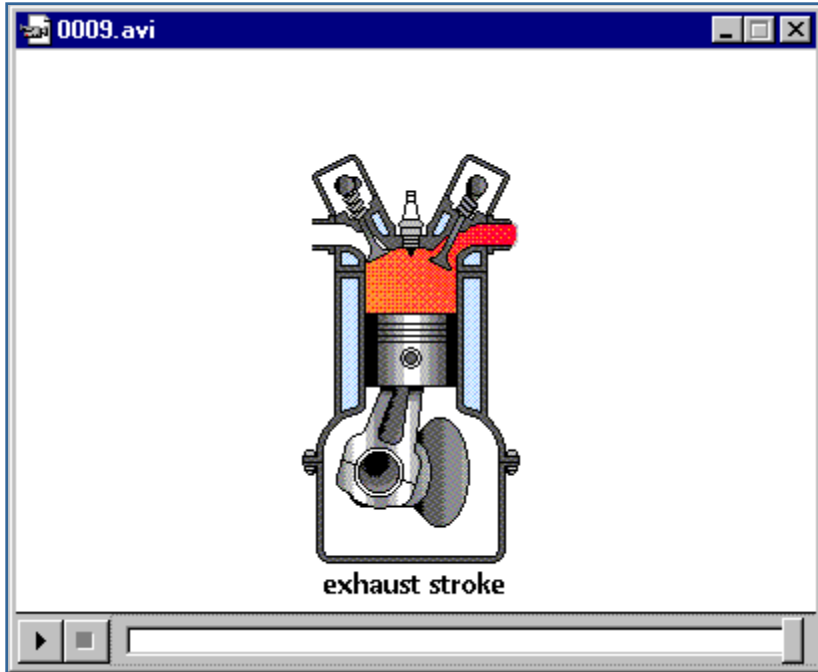
Notice that the expected values of the sums are 5, 15, 35 and 45.

Solution 1.4 We can plot the data and calculate mean and standard deviation.

```
inst1 = [9.490950, 10.436813, 9.681357, 10.996083, 10.226101, 10.253741,
10.458926, 9.247097, 8.287045, 10.145414, 11.373981, 10.144389,
11.265351, 7.956107, 10.166610, 10.800805, 9.372905, 10.199018,
9.742579, 10.428091]
inst2 = [11.771486, 10.697693, 10.687212, 11.097567, 11.676099,
10.583907, 10.505690, 9.958557, 10.938350, 11.718334,
11.308556, 10.957640, 11.250546, 10.195894, 11.804038,
11.825099, 10.677206, 10.249831, 10.729174, 11.027622]
ax = pd.Series(inst1).plot(marker='o', linestyle='none',
fillstyle='none', color='black')
pd.Series(inst2).plot(marker='+', linestyle='none', ax=ax,
fillstyle='none', color='black')
```

The Piston Simulator

1



D: Spring Coefficient (N/m)

E: Atmospheric Pressure (N/m²)

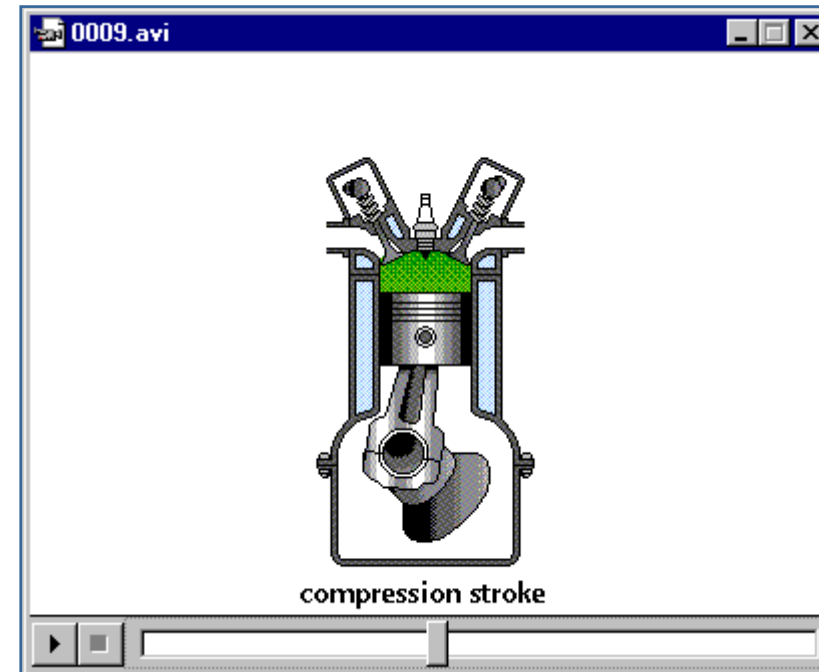
F: Ambient Temperature (°K)

G: Gas Temperature (°K)

A: Piston Weight (Kg)

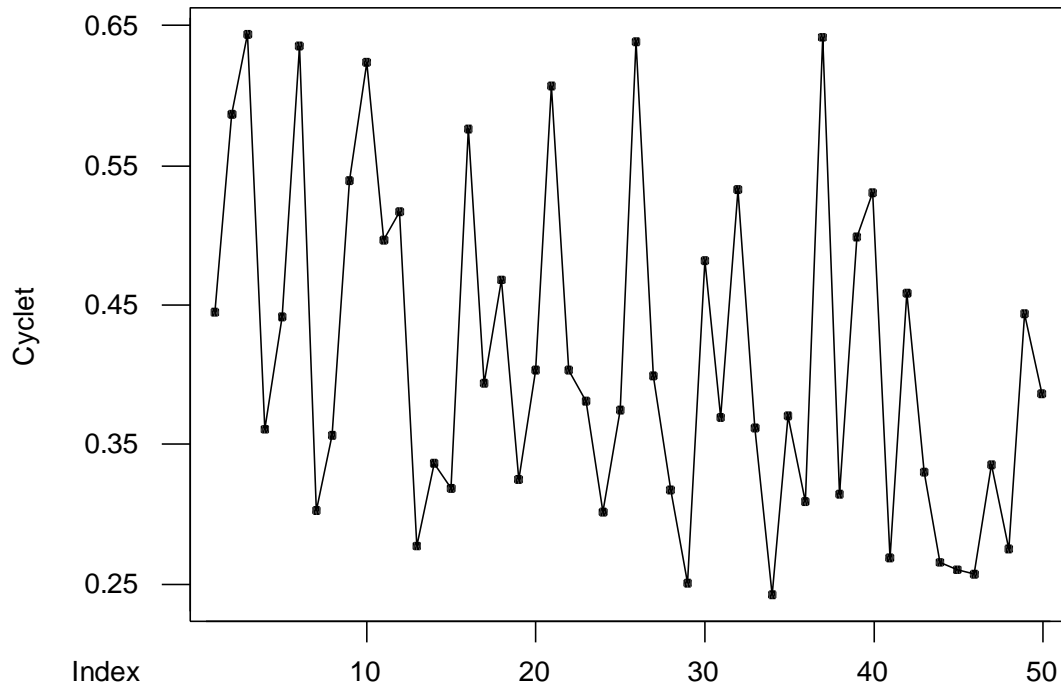
B: Piston Surface Area (m²)

C: Initial Gas Volume (m³)



The Piston Simulator

1



A run chart of 50 cycle times, in seconds, with all factors set at their **maximum value**. The cycle times in this sequence have an average of **0.413** seconds and a standard deviation of **0.121** seconds

1.1 Random Phenomena and the Structure of Observations

Many phenomena which we encounter are only partially predictable. It is difficult to predict the weather or the behavior of the stock market. In this book we focus on industrial phenomena, like performance measurements from a product which is being manufactured, or the sales volume in a specified period of a given product model. Such phenomena are characterized by the fact that measurements performed on them are often not constant but reveal a certain degree of variability. Variability is also a reflection of uncertainty. For a comprehensive treatment of uncertainty in engineering applications see [del Rosario and Iaccarino \(2022\)](#). The objectives of this chapter is to present methods for analyzing this variability, in order to understand the variability structure and enhance our ability to control, improve and predict future behavior of such phenomena. We start with a few simple examples.

Example 1.1 A piston is a mechanical device that is present in most types of engines. One measure of the performance of a piston is the time it takes to complete one cycle. We call this measure **cycle time**. In Table [1.1](#) we present 50 cycle times of a piston operating under fixed operating conditions (a sample data set is stored in file **CYCLT.csv**). We provide with this book code in Python for running a piston software simulation. If you installed Python, install the `mistat` package using the Python package installer `pip`.

You will get access to a piston simulator with 7 factors that you can change. We will also use this simulator when we discuss Statistical Process Control (Chapters 2, 3 and 4 in the Industrial Statistics book) and the Design of Experiments (Chapters 5, 6 and 7 in the Industrial Statistics book). We continue at a pedestrian pace by recreating Table 1.1 using Python. All the Python applications referred to in this book are contained in a package called `mistat` available from the Python package index and on GitHub at <https://github.com/gedeck/mistat>. The following Python commands will import the `mistat` package, read the cycle time data and print the first five values on your monitor:

```
import mistat
data = mistat.load_data('CYCLT')
print(data.head())
```

```
0    1.008
1    1.098
2    1.120
3    0.423
4    1.021
Name: CYCLT, dtype: float64
```

Notice that functions in Python have parenthesis. The `import` statement imports the `mistat` package and makes its functionality available. `mistat.load_data` is a function that loads the CYCLT data set as a *Pandas* data series. A *Pandas* data series is a simple vector of values.

Example 3.26 Let us determine $(0.95, 0.95)$ -tolerance interval for samples of size $n = 100$, of piston cycle times. Use the sample in the data file **CYCLT.csv**. The original sample is of size $n_0 = 50$. Since future samples are of size $n = 100$, we draw from the original sample RSWR of size $n = 100$.

In Python, the calculations from Example 3.26 are straightforward:

```

cyclt = mistat.load_data('CYCLT')
cyclt = [*cyclt, *cyclt] # create a dataset of size 100 by duplication

def getQuantile(x, p):
    return np.quantile(x, p)

B_025 = pg.compute_bootci(cyclt, func=lambda x: getQuantile(x, p=0.025),
                          n_boot=500, seed=1, return_dist=True)
B_975 = pg.compute_bootci(cyclt, func=lambda x: getQuantile(x, p=0.975),
                          n_boot=500, seed=1, return_dist=True)
print('0.025%', np.quantile(B_025[1], 0.025))
print('0.975%', np.quantile(B_975[1], 0.975))

```

```

0.025% 0.175
0.975% 1.141

```

The bootstrap $(.95, .95)$ -tolerance interval for $n = 100$ piston cycle times was estimated as $(0.175, 1.141)$. ■

Bootstrapping

M-test

1. Compute the expected number of returns in category i ,
 $E_i = Np_i,$ $i = 1, \dots, K.$
2. Compute
 $S_i = \text{SQRT} [Np_i(1-p_i)],$ $i = 1, \dots, K.$
3. Compute the adjusted residuals
 $Z_i = (n_i - E_i)/S_i,$ $i = 1, \dots, K.$
4. Determine the critical value C of Z_i from the M-Test table.

N = Total number of returns

K = Number of categories

n_i = Number of returns in category i

p_i = proportion of category i in sampling frame

M-test

K	10%	5%	1%
4	1.95	2.24	2.81
5	2.05	2.32	2.88
6	2.12	2.39	2.93
7	2.18	2.44	2.99
8	2.23	2.49	3.04
9	2.28	2.53	3.07
10	2.32	2.57	3.10
20	2.57	2.81	3.30
30	2.71	2.94	3.46

If all adjusted residuals, Z_i , are smaller, in absolute value, than C , no significant bias is declared.

Cells with values of Z_i , above C or below $-C$ are declared significantly different and a follow up effort is initiated.

Bonferroni
correction



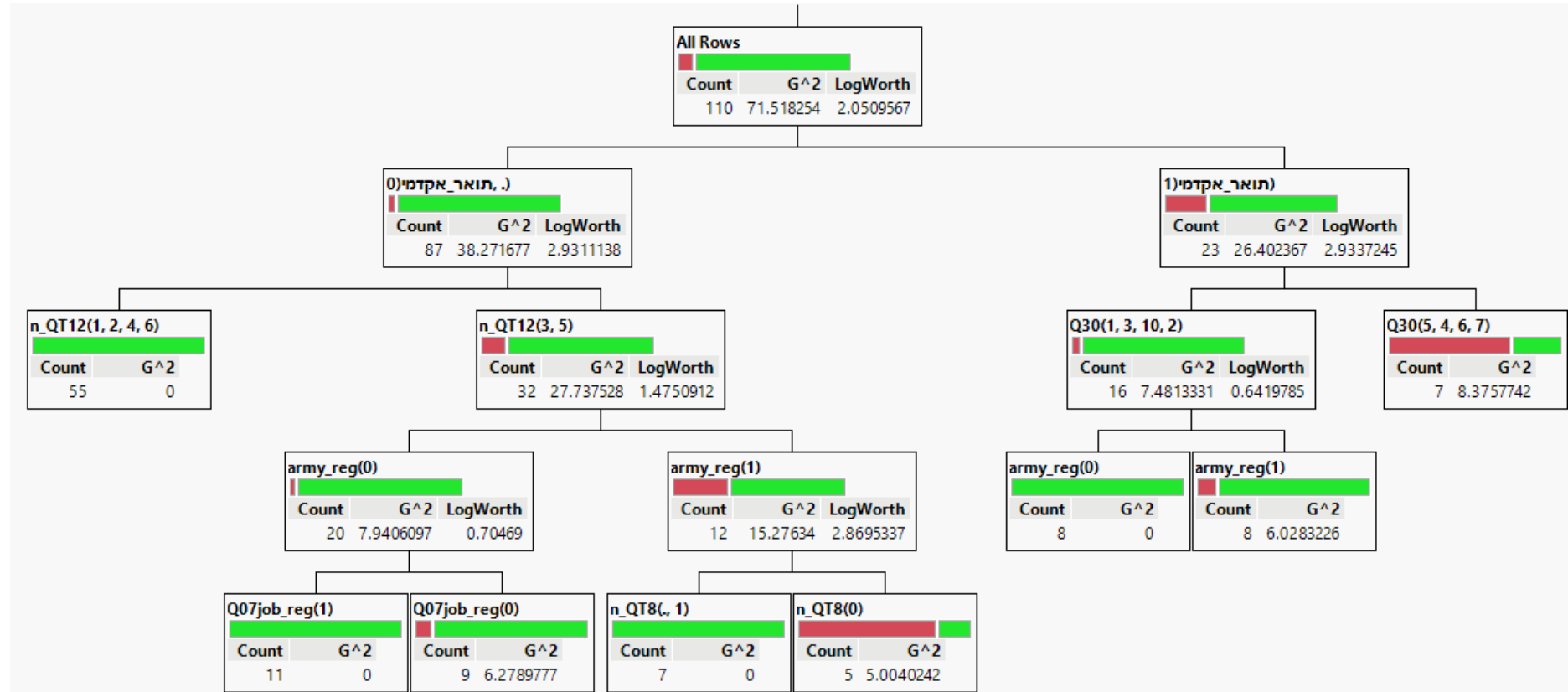
Carlo Bonferroni
1892 - 1960

"A Test for Detecting Outlying Cells in the Multinomial Distribution and Two-Way Contingency Tables", Fuchs and Kenett, *Journal of the American Statistical Association*, 75, pp. 395-398, 1980.


"Two Methods for Comparing Pareto Charts", Kenett, *Journal of Quality Technology*, 23, pp. 27-31, 1991.


Decision trees


7



 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

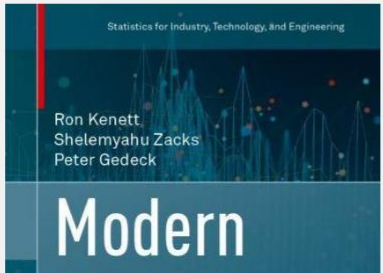
Modern Statistics

A Computer Based Approach with Python

Home > README.md

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: 978-3-031-17565-0 (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.

index.ipynb - JupyterLab

https://hub.gke2.mybinder.org/user/gedeck-mistat-code-solutions-p5gh031n/doc/tree/ModernStatistics/notebooks/index.ipynb

Email linkRonLinksFacebookLinkedInNews Ynetהארץבנק הפועלים - לקו...Demo: Industry 4.0קורונה - לוח בקרהWiley StatsRef: Stati...Other bookmarks

index.ipynb

FileEditViewRunKernelTabsSettingsHelp

DownloadGitHubBinderMarkdown

Python 3 (ipykernel)

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 for information only and can usually be ignored. "Errors" will stop 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 upgraded. This might solve the problem.
- Check the errata at (<https://gedeck.github.io/mistat-code-solutions/ModernStatistics/>) and the issue tracker at <https://github.com/gedeck/mistat-code-solutions/issues> 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>
- Paste the error message into Google and see if someone else already found a solution

Simple

0 1 Python 3 (ipykernel) | Idle Mem: 166.07 / 2048.00 MB

Mode: CommandLn 1, Col 1index.ipynb

Modern Statistics

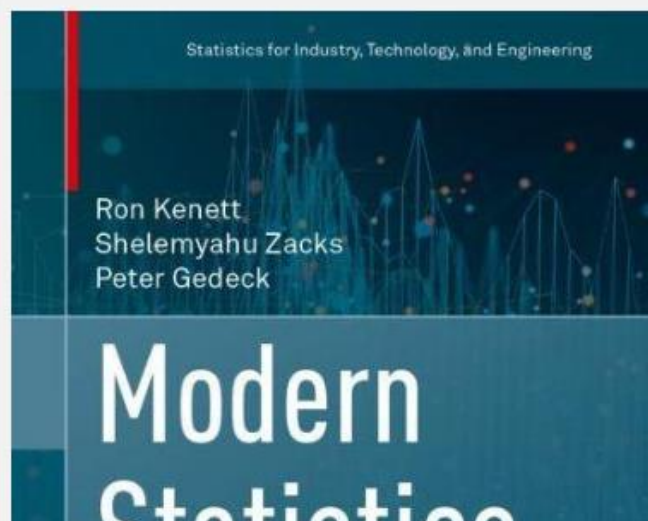
A Computer Based Approach with Python

Home >README.md



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

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](#)

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](#)  [launch 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`

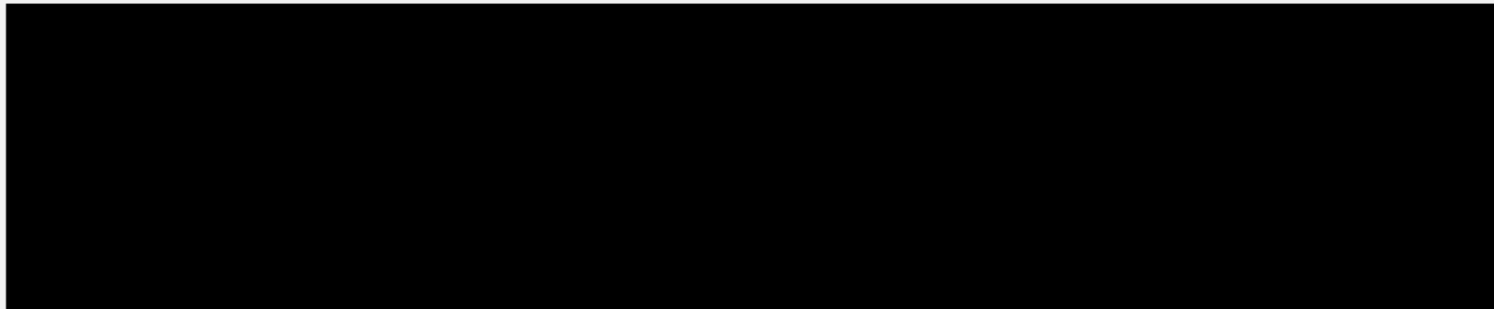
Table of contents

- 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://user-images.githubusercontent.com/8720575/180794703-c6f05f40-eefd-4e1a-93f9-42cb78e6a6b4.mp4?fbclid=IwAR08kOITzOhBUIYZhyclkrh_p7Ad9q53rnrwUicOmiUBHsbd4TwJgzkHepo

Files






Running

Clusters

Select items to perform actions on them.

☐ 0

/ Documents / Python Scripts / Modern Statistics

 Chap001.ipynb Chap002.ipynb Chap003.ipynb Chap004.ipynb Chap005.ipynb Chap006.ipynb Chap007.ipynb Chap008.ipynb

JupyterLab



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

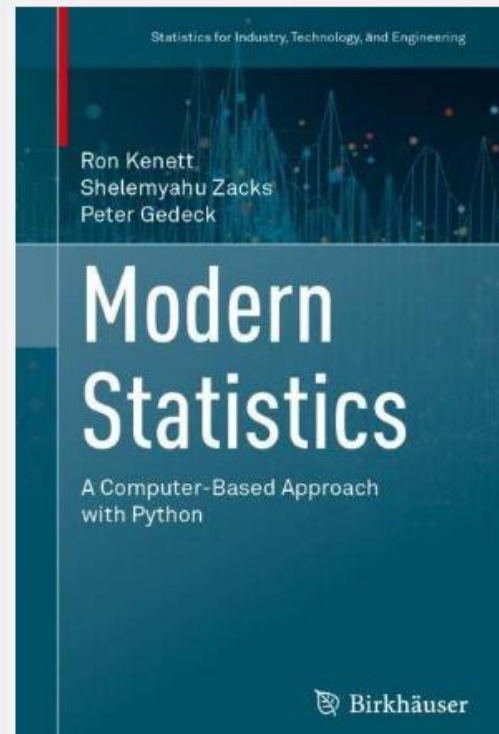
Modern Statistics: A Computer Based Approach with Python

Home >README.md



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/


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](#)  [launch binder](#).

**Thank you for
listening**