Analytical Skills for Business (WS 2025/26)

Business Administration (M. A.)

© Benjamin Gross

17.09.2025 11:25

This document holds the course material for the Analytical Skills for Business course in the Master of Business Administration program. It discusses version control systems such as Git and GitHub for efficient team collaboration, offers an overview of no-code and low-code tools for data analytics including Tableau, Power BI, QlikView, makeML, PyCaret, RapidMiner, and KNIME, and introduces key programming languages such as R, Python, and SQL alongside essential programming concepts like syntax, libraries, variables, functions, objects, conditions, and loops. In addition, it covers working with modern development environments, including Unix-like systems, containers, APIs, Jupyter, and RStudio, and sets expectations for project submissions and evaluation.

Table of contents

1 Introduction

Computer science is the study of computers and computation, spanning theoretical and algorithmic foundations, the design of hardware and software, and practical uses of computing to process information. It encompasses core areas such as

- algorithms and data structures
- computer architecture
- programming languages and software engineering
- databases and information systems
- networking and communications
- graphics and visualization
- human-computer interaction
- intelligent systems.

The field draws on mathematics and engineering—using concepts like

- binary representation
- Boolean logic
- complexity analysis

to reason about what can be computed and how efficiently.

Emerging in the 1960s as a distinct discipline, computer science now sits alongside computer engineering, information systems, information technology, and software engineering within the broader computing family. Its reach is inherently interdisciplinary, intersecting with domains from the natural sciences to business and the social sciences. Beyond technical advances, the discipline engages with societal and professional issues, including

- reliability
- security
- privacy
- intellectual property

in a networked world (Britannica, 2025).

1.1 Implementing version control systems

Version control systems are essential tools for managing code, tracking changes, and facilitating collaborative development in modern development projects (Çetinkaya-Rundel & Hardin, 2021). These systems enable teams to work efficiently on shared codebases while maintaining a complete history of all modifications, ensuring reproducibility and accountability in data analysis workflows.

1.1.1 Core Concepts

Version control systems provide systematic approaches to managing changes in documents, programs, and other collections of information:

- **Repository**: A central storage location containing all project files and their complete revision history

- **Commit**: A snapshot of the project at a specific point in time, representing a set of changes
- **Branch**: An independent line of development allowing parallel work on different features
- **Merge**: The process of integrating changes from different branches back together

Figure 1: **Source:** *https://uidaholib.github.io/get-git/1why.html*
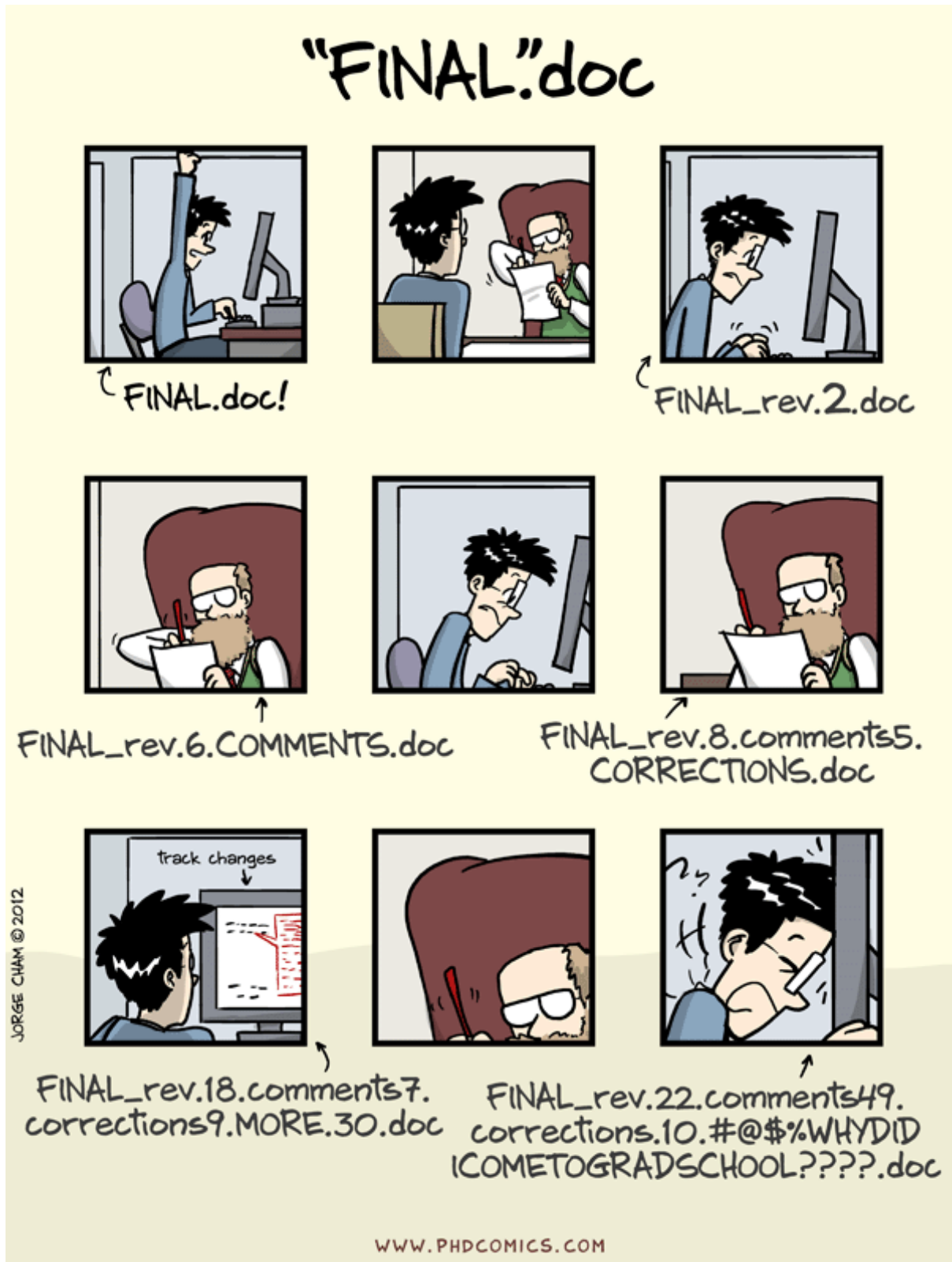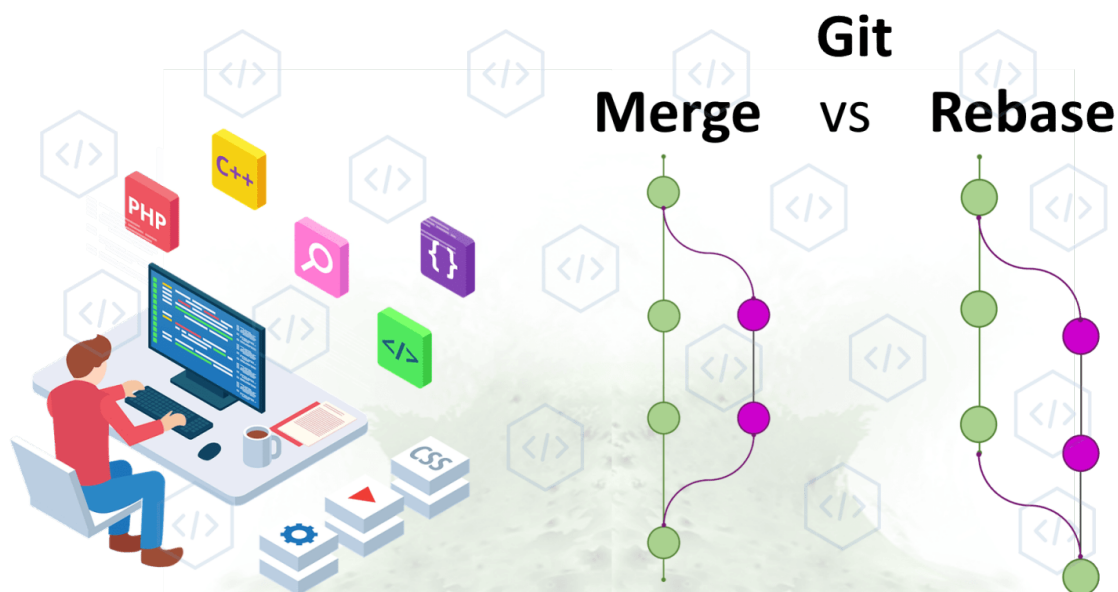
1.1.2 Git: Distributed Version Control

Git is a distributed version control system that tracks changes in files and coordinates work among multiple contributors. It was created by **Linus Torvalds** (creator of Linux) in 2005 and has since become the de facto standard for version control in software development. Key characteristics include:

Local Repository: Each user maintains a complete copy of the project history, enabling offline work and faster operations.

Staging Area: An intermediate area where changes are prepared before being committed to the repository.

Branching and Merging: Lightweight branching allows for experimental development without affecting the main codebase. Merging integrates changes from different branches. In Open Source projects often Pull Requests are used to propose and discuss changes before merging. In the coporate world often Merge Requests are used. There is a difference between merging and rebasing. Merging creates a new commit that combines the histories of two branches, while rebasing rewrites the commit history to create a linear sequence of changes as you see on the figure below.

Figure 2: **Source:** *https://i0.wp.com/digitalvarys.com/wp-content/uploads/2020/03/Git-Merge-and-Rebase.png?resize=1536%2C843&ssl=1*

Distributed Workflow: No single point of failure, as every user has a complete backup of the project.

1.1.3 GitHub: Cloud-Based Collaboration Platform

GitHub is a web-based hosting service for Git repositories that adds collaboration features and project management tools:

- **Remote Repositories**: Centralized storage accessible from anywhere with internet connectivity.
- **Pull Requests**: Structured code review process for integrating changes.
- **Issue Tracking**: Built-in project management for tracking bugs and feature requests.
- **Actions and CI/CD**: Automated workflows for testing and deployment.
- **Documentation**: Integrated wiki and README support for project documentation.

The combination of Git and GitHub creates a powerful ecosystem for collaborative analytics projects, ensuring code quality, facilitating peer review, and maintaining comprehensive project documentation (G. GeeksforGeeks, 2024).

See also Collaborating with Git and GitHub by Prof. Dr. Huber about using git and Github for collaboration.

For students Github offers a free educational plan with additional features! Included you will find access to Github Copilot, a AI based code completion tools like. These are powerful gadgets to support your coding activities, but not a replacement for learning programming and coding by yourself as you can see on this image:

Figure 3: **Source:** *LinkedIn*

	Vibe Coding	Vibe Engineering
Goal	Make it run now	Solve real problems at scale. Monetize it. Make it reliable and safe.
Prompting	Casual: "make X"	Structured: roles, security, components, data rules, background jobs
Tech skills	Not needed	System design, APIs, tools, and services, e.g., Netlify, RLS, XSS, DevTools, MCP, and agents
Coding skills	Not needed	Just enough to review & question AI's work: logical expressions, functions, etc.
Risks	Hidden bugs, data leaks, fragile code	More effort upfront, but fewer crises later
Feels like	Using cool tech, copy-pasting whatever AI spits out	Designing flows, managing the context, testing, defining KPIs and alerts, monitoring logs
Best for	Demos, experiments, prototypes	Real products, sensitive data, customer use
Career impact	Might just delay being replaced by AI	Helps you capture new opportunities and become the one who controls AI

productcompass.pm

1.1.4 Comparison of Git and GitHub

On this image you can see the integration of Git in GitHub:

Figure 4: **Source:** *https://marce10.github.io/ciencia_reproducible/intro_a_git_y_github.html*
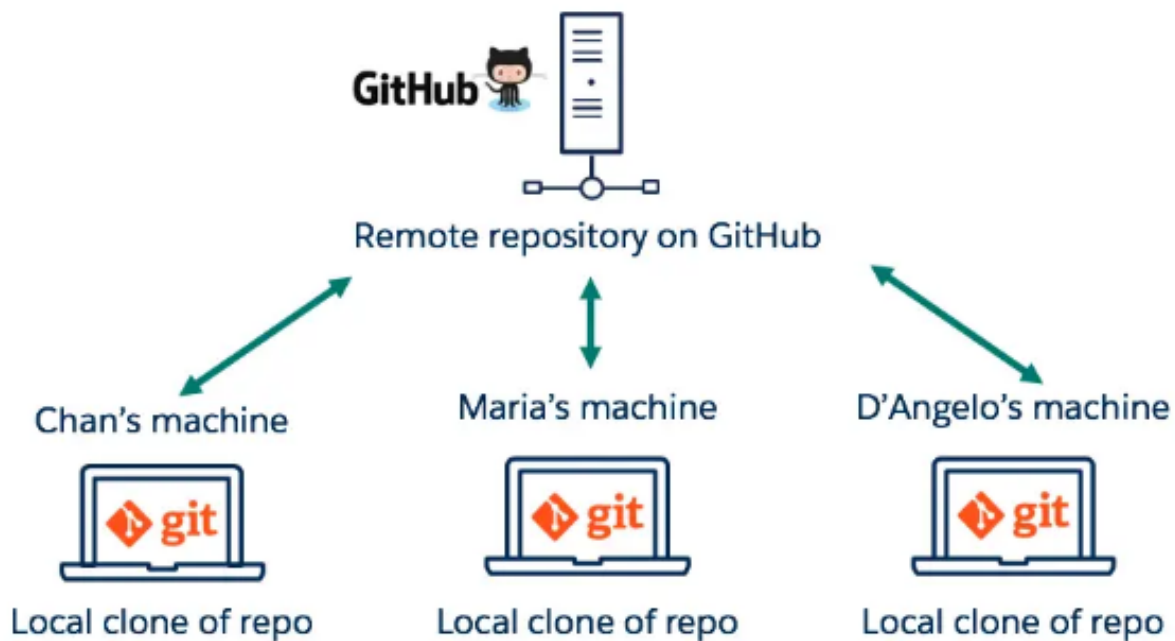
1.1.5 Business Analytics Applications

In business analytics contexts, version control systems provide:

- **Reproducible Analysis**: Complete tracking of analytical scripts and data processing steps
- **Collaborative Research**: Multiple analysts can work simultaneously on different aspects of projects
- **Model Versioning**: Systematic management of machine learning models and their evolution
- **Data Governance**: Audit trails for compliance and regulatory requirements
- **Backup and Recovery**: Protection against data loss and accidental modifications

1.1.6 A lot of documentation available

For more coverage of version control concepts, implementation strategies, and best practices, see:

- **Theoretical Foundation**: Introduction to Modern Statistics provides context on reproducible research practices and collaborative analytics (Çetinkaya-Rundel & Hardin, 2021)
- **Git Resources**: Git Cheat Sheet offers quick reference for common Git commands
- **GitHub Documentation**: GitHub Manual contains detailed guidance on platform features
- **Online Resources**:
    - GeeksforGeeks Git vs GitHub Guide provides practical comparisons and use cases (G. GeeksforGeeks, 2024)
    - Official GitHub Documentation offers authoritative guidance on getting started (GitHub, 2024)

Understanding version control systems is fundamental for modern business analytics, enabling collaborative development, ensuring reproducibility, and maintaining professional standards in data science projects.

1.2 Overview on Programming languages

- **R:** R is a programming language and free software environment used for statistical computing and graphics supported by the R Foundation for Statistical Computing. It is widely used among statisticians and data miners for developing statistical software and data analysis.
- **Python:** Python is a versatile programming language widely used in data science and analytics. It has a rich ecosystem of libraries such as Pandas, NumPy, and Matplotlib that facilitate data manipulation, analysis, and visualization. It can also be used to develop software as it is a general-purpose programming language which utilizes an object-oriented programming paradigm.
- **SQL:** SQL (Structured Query Language) is the standard language for managing and querying relational databases. It is essential for data extraction, transformation, and loading (ETL) or extraction, loading and transformation (ELT) processes in analytics workflows. Modern databases like PostgreSQL, MySQL, and SQLite use SQL for data manipulation and retrieval and just have slightly different dialects.

1.3 Elements of programming languages

- **Syntax:** the set of rules that defines the combinations of symbols that are considered to be correctly structured programs in that language.
- **Libraries:** collections of pre-written code that users can call upon to save time and effort.
- **Variables:** named storage locations in a program that hold values.
- **Functions:** reusable blocks of code that perform a specific task.
- **Objects:** instances of classes that encapsulate data and behavior.
- **Conditions:** statements that control the flow of execution based on certain criteria.
- **Loops:** constructs that repeat a block of code multiple times.

1.4 Development environments

- **Unix-like systems:** Most popular Unix like system is Linux. It is understood as a family of open source Unix-like operating systems based on the Linux kernel and mostly used in form of distributions such as Ubuntu, Debian, Fedora, CentOS and Alpine Linux. The latter is known for its simplicity and efficiency and is often used as the base image for Docker containers (container host os).
- **Containers:** Docker is a production ready containerization service. On hub.docker.com you can store images that could be pulled. It acts as the main public registry for Docker images. Container are a standard unit of software that packages up code and all its dependencies so the application runs quickly and reliably from one computing environment to another. A Docker container image is a lightweight, standalone, executable package of software that includes everything needed to run an application: code, runtime, system tools, system libraries and settings.
- **APIs:** (Application Programming Interface): An API is a set of definitions and protocols for building and integrating application software. It allows different software systems to communicate with each other. APIs are used to enable the integration of different systems,

allowing them to share data and functionality. Examples include RESTful APIs, SOAP APIs, and GraphQL APIs.

- **Jupyter:** Jupyter is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. It supports various programming languages, including Python, R, and Julia. Jupyter is widely used for data analysis, machine learning, and scientific computing.
- **IDEs:** (Integrated Development Environments): IDEs are software applications that provide comprehensive facilities to computer programmers for software development. They typically include a code editor, a debugger, and build automation tools. Examples of popular IDEs include:
  - **RStudio:** RStudio is an integrated development environment (IDE) for R, a programming language for statistical computing and graphics. RStudio provides a user-friendly interface for writing and debugging R code, as well as tools for data visualization and reporting.
  - **Visual Studio Code (VS Code):** VS Code is a free source-code editor made by Microsoft for Windows, Linux and macOS. It includes support for debugging, embedded Git control, syntax highlighting, intelligent code completion, snippets, and code refactoring. It is highly customizable, allowing users to change the theme, keyboard shortcuts, preferences, and install extensions that add additional functionality.

1.5 Overview on no-code and low-code tools for data analytics

- n8n
  - Follow this short introduction to n8n:
    1. Create an account on n8n.cloud
    2. Create a new workflow
    3. Add a new node and select "HTTP Request"
    4. Configure the node to make a GET request to `https://minio.seriousbenentertainment.org:9%202025.csv`
    5. Add a new node and select "Convert to File"
    6. See what other possibilities n8n offers you to wrangle the data You can also import the n8n workflow with the context menu on the top right of the workflow page (…) - `Import from File...` or `Import from URL...`.
- Tableau
- Snowflake
- Power BI
- QlikView
- Rapidminer
- KNIME

2 Descriptive statistics

Descriptive statistics summarizes and presents the main features of a dataset so you can understand what the data look like before modeling or inference. It organizes raw values into clear numerical summaries and visuals, without making probabilistic claims about a wider population. In analytics projects, this first pass helps you validate data quality, spot outliers, and communicate patterns to stakeholders.

What we typically summarize:

- Central tendency: mean, median, and mode capture a typical or "center" value
- Variability: range, variance, standard deviation, and interquartile range describe spread
- Distribution shape: skewness and kurtosis characterize symmetry, tails, and outliers
- Frequencies and percentiles: counts, proportions, quantiles (e.g., quartiles, deciles)

Common methods:

- Numerical methods: compute summary metrics and assemble frequency tables
- Graphical methods: histograms (continuous data), bar or pie charts (categorical data), box plots (median, quartiles, outliers), and scatter plots (bivariate relationships)

Business analytics context:

- For monthly revenue by region, the mean signals typical performance, the standard deviation shows volatility, a box plot quickly flags outliers, and a bar chart compares regions. These summaries guide prioritization (e.g., regions with high variability may require deeper investigation) and set baselines for forecasting and experimentation.

For an accessible overview of types, methods, and examples, see ResearchMethod.net.

2.1 Hypothesis testing

Hypothesis testing is a fundamental statistical method used to make inferences about population parameters based on sample data (Illowsky & Dean, 2018). It provides a systematic framework for evaluating claims about populations using sample evidence, enabling data-driven decision making in business contexts.

2.1.1 Core Concepts

A statistical hypothesis test involves formulating two competing hypotheses:

- **Null hypothesis (H)**: The status quo or default position, typically representing no effect or no difference
- **Alternative hypothesis (H or H)**: The research hypothesis representing the effect or difference we seek to detect

The process involves calculating a test statistic from sample data and comparing it to a critical value or determining a p-value to make decisions about rejecting or failing to reject the null hypothesis (GeeksforGeeks, 2024).

2.1.2 Key Components

Test Statistics: Standardized measures that quantify how far sample data deviates from what would be expected under the null hypothesis.

Significance Level ( ): The probability threshold for rejecting the null hypothesis, commonly set at 0.05 (5%).

P-value: The probability of observing test results at least as extreme as those obtained, assuming the null hypothesis is true.

Critical Region: The range of values for which the null hypothesis is rejected.

2.1.3 Types of Errors

- **Type I Error ( )**: Rejecting a true null hypothesis (false positive)
- **Type II Error ( )**: Failing to reject a false null hypothesis (false negative)
- **Statistical Power (1-)**: The probability of correctly rejecting a false null hypothesis

2.1.4 Reference Materials

For comprehensive coverage of hypothesis testing concepts, methodologies, and applications, consult:

- **Theoretical Foundation**: Introductory Statistics provides detailed explanations of hypothesis testing principles and procedures (Illowsky & Dean, 2018)
- **Visual Guide**: Hypothesis Testing Overview offers a visual representation of key concepts
- **Detailed Methodology**: Hypothesis Testing Documentation contains comprehensive methodological information
- **Online Resource**: Additional perspectives on hypothesis testing applications can be found in the GeeksforGeeks guide (GeeksforGeeks, 2024)

Understanding hypothesis testing is essential for making informed business decisions based on data analysis, forming the foundation for advanced statistical inference and predictive analytics in business contexts.

3 Literature

All references for this course.

3.1 Essential Readings

Bruce, P. and A. Bruce (2020). *Practical Statistics for Data Scientists, 2nd Edition.* https://learning.oreilly.com/library/view/practical-statistics-for/9781492072935/preface01.html.

Çetinkaya-Rundel, M. and J. Hardin (2021). *Introduction to Modern Statistics.* https://www.open-intro.org/book/ims/. https://github.com/DrBenjamin/Analytical-Skills-for-Business/blob/491a9a84dd0227aab44e0a6db7e6330830a05a6b/literature/Introduction_to_Modern_Statistics_2e.pdf/?raw=true.

Stephenson, P. (2023). *Data Science Practice.* https://datasciencepractice.study/.

3.2 Further Readings

Békés, G. and G. Kézdi (2021). *Resources for Data Analysis for Business, Economics, and Policy.* Instructor resources: https://www-cambridge-org.eux.idm.oclc.org/highereducation/books/data-analysis-for-business-economics-and-policy/D67A1B0B56176D6D6A92E27F3F82AA20/. https://github.com/DrBenjamin/Analytical-Skills-for-Business/blob/c2ec1b2061c7dc36200977cfd58daf6020c1c774/literature/B%C3%A9k%C3%A9s_Data%20Analysis%20for%20Business%2C%20Economics%2C%20and%20Policy_2021_First%20Day%20of%20Class%20Slides.pdf/?raw=true.}

Britannica (2025). *Computer science | Definition, Types, & Facts | Britannica.* https://www.britannica.com/science/computer-science.

Dougherty, J. and I. Ilyankou (2025). *Hands-On Data Visualization.* https://handsondataviz.org/.

Evans, J. R. (2020). "Business Analytics'".

GeeksforGeeks (2024a). *Understanding Hypothesis Testing.* Online resource for software testing and statistical concepts. https://www.geeksforgeeks.org/software-testing/understanding-hypothesis-testing/.

GeeksforGeeks, G. (2024b). *Difference between Git and GitHub.* Comprehensive comparison of Git and GitHub for version control. https://www.geeksforgeeks.org/git/difference-between-git-and-github/.

GitHub (2024). *About GitHub and Git.* Official GitHub documentation on Git and GitHub fundamentals. https://docs.github.com/en/get-started/start-your-journey/about-github-and-git.

Illowsky, B. and S. L. Dean (2018). *Introductory Statistics.* OpenStax, Rice University, p. 905. ISBN: 1938168208. https://github.com/DrBenjamin/Analytical-Skills-for-Business/blob/c2ec1b2061c7dc36200977cfd58daf6020c1c774/literature/Introductory%20Statistics.pdf/?raw=true.}

Irizarry, R. A. (2024). "Advanced Data Science: Statistics and Prediction Algorithms Through Case Studies'".

http://rafalab.dfci.harvard.edu/dsbook-part-2.

Kumar, U. D. (2017). "Business Analytics: The Science of Data-Driven Decision Making'".

Pochiraju, B. and S. Seshadri, ed. (2019). *Essentials of Business Analytics.* Vol. 264. https://link.springer.com/10.1007/978-3-319-68837-4. Cham: Springer International Publishing. ISBN: 978-3-319-68836-7. DOI: 10.1007/978-3-319-68837-4. https://github.com/DrBenjamin/Analytical-Skills-for-Business/blob/c2ec1b2061c7dc36200977cfd58daf6020c1c774/literature/Essentials%20of%20Business%20Analytics.pdf/?raw=true.}

Vaughan, D. (2020). "Analytical Skills for AI and Data Science'".

https://learning.oreilly.com/library/view/analytical-skills-for/9781492060932/preface01.html#idm46388898852872.