Lecture 18: Human Factors in IR (2)

Jiepu Jiang
Department of Computer Science, Virginia Tech
2018/10/18
Outline

• Related theories
  • Anomalous State of Knowledge (ASK)
  • Stratified model of relevance interactions
• Laboratory user studies and experimental design
• Statistical analysis & examples
  • Correlation analysis, e.g., Pearson, Spearman
  • Regression analysis, e.g., linear & logistic regression
  • ANOVA and non-parametric variants
  • Our focus: when to use & how to read and interpret results
Regression Analysis

Correlation analysis

• Only quantifies the strength of the relationship; not predictive
• Only two variables

Regression analysis

• Predictive models
• Independent variables (IV): one or multiple predictors (factors)
• Dependent variables (DV): things being predicted (output)
• A standard tool for quantifying the influence of IVs on DV

Linear regression

• x (factors): k independent variables
• y (outcome): dependent variable
• b: coefficients
• b_0: intercept; fixed term  \[ y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k \]
Regression Analysis: Differences

LR in behavioral analysis

• Quantify the relationship between IVs and DV
  • Looking into the value and sign of the coefficients ($b$), and $p$ values
  • With other variables being controls (comparing to correlation)

• Can be predictive (but usually not the primary focus)
  • You can probably find better prediction models in ML ...

• For example:
  • Does an IV have a significant positive effect on the DV?

LR in machine learning

• Prediction & performance

• For example:
  • How accurate is my prediction model?
Linear Regression: Model Fitting

Purpose
- To estimate $b$ from data (ML: training)

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$$

Ordinary least squares (OLS)
- $y_i$: observed DV value; $\hat{y}_i$: predicted DV value
- Find $b$ that minimizes $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$
- The "default" approach for parameter estimation in LR
Linear Regression: Model Fitness

$R^2$ (R-squared; Coefficient of determination)
- Quantifies how well you model explains the data
  - Compared with a baseline model that explains just using the mean of $y$
- (Practically) ranges from 0 to 1; below 0?
- 1: (perfectly explains the data)
- 0: (cannot explain the data)

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$
Linear Regression: Adjusted $R^2$

$R^2$

• Quantifies how well you model explains the data
• But $R^2$ will be inflated when the number of IVs increase

Adjusted $R^2$

• Takes into account the number of IVs (predictors)
• Each additional IV should make a large enough contribution in $R^2$
  • Otherwise adjusted $R^2$ get penalized
• $p$: the number of IVs (does not include the bias term);
• $n$: the number of data points (sample size)

\[
\text{Adjusted } R^2 = 1 - \left(1 - R^2\right) \cdot \frac{n - 1}{n - p - 1} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \cdot \frac{n - 1}{n - p - 1}
\]
Linear Regression: F-test

Purpose

- To compare two regression models M1 and M2
  - Two-tail: Whether two models explain DV significantly different?
  - One-tail: Whether one model explains DV significantly better than another?
- Requirement
  - One model is a reduced version of another model

Usage 1:

- Test whether your regression model is statistically significant
- Compare the model with a “null” model using F-test
- The “null” model (only a fixed term): \( y = b_0 \)

Usage 2:

- Test whether some IVs significantly improves the model
- Compare M1 (without the IVs) with M2 (with the IVs)
Linear Regression: F-test (example 1)

- RQ: what makes users believe a result is useful?
- DV: the usefulness of a search result (usef)
- IV: novelty (nov); effort spent on reading it (effort); understandability (under); reliability/credibility of the information (relia)
- $R^2 = 0.391$, Adjusted $R^2 = 0.388$, F-test $p$-value $< 0.001$
- The model significantly explains the usefulness of a search result.

Linear regression model (robust fit):
usef ~ 1 + nov + effort + under + relia

Estimated Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tStat</th>
<th>pValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.34171</td>
<td>0.32334</td>
<td>-1.0568</td>
<td>0.29094</td>
</tr>
<tr>
<td>nov</td>
<td>0.42583</td>
<td>0.035567</td>
<td>11.972</td>
<td>2.7355e-30</td>
</tr>
<tr>
<td>effort</td>
<td>0.12231</td>
<td>0.049757</td>
<td>2.4581</td>
<td>0.014199</td>
</tr>
<tr>
<td>under</td>
<td>0.15881</td>
<td>0.045018</td>
<td>3.5276</td>
<td>0.00044554</td>
</tr>
<tr>
<td>relia</td>
<td>0.32186</td>
<td>0.041967</td>
<td>7.6692</td>
<td>5.5306e-14</td>
</tr>
</tbody>
</table>

Number of observations: 736, Error degrees of freedom: 731
Root Mean Squared Error: 1.64
R-squared: 0.391, Adjusted R-Squared 0.388
F-statistic vs. constant model: 118, p-value = 2.16e-77
Linear Regression: F-test (example 2)

- **Compare two models:**
  - M1: usefulness ~ 1 + novelty + effort + understandability + reliability
    
    Linear regression model (robust fit):
    
    $$\text{usef} \sim 1 + \text{nov} + \text{effort} + \text{under} + \text{relia}$$
    
    Number of observations: 736, Error degrees of freedom: 731
    Root Mean Squared Error: 1.64
    \textbf{R-squared: 0.391, Adjusted R-Squared: 0.388}
    
    F-statistic vs. constant model: 118, p-value = 2.16e-77
  - M2: usefulness ~ 1 + novelty + understandability + reliability
    
    Linear regression model (robust fit):
    
    $$\text{usef} \sim 1 + \text{nov} + \text{under} + \text{relia}$$
    
    Number of observations: 736, Error degrees of freedom: 732
    Root Mean Squared Error: 1.64
    \textbf{R-squared: 0.386, Adjusted R-Squared: 0.384}
    
    F-statistic vs. constant model: 154, p-value = 3.2e-77

- M1 and M2 got very close R² and adjusted R²
- Need to test whether the difference is significant using an F-test
  - \( p = 0.012 \) \( \Rightarrow \) M1 and M2 are sig. different (in terms of explaining DV)
  - Effort (the additional IV) is indeed helpful for explaining usefulness
Linear Regression: Coefficients

- Coefficients
  - $b$
  - *While other conditions being the same*, how much $y$ would change if an $x_i$ changes by one unit
    \[ y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k \]

- Significance
  - Each coefficient will have a p-value
  - indicates whether the IV has an significant effect on the DV
    - You will almost always get a $b \neq 0$; indicate some influence of the IV
    - But to which extent the influence is statistically significant?
    - The p-value tells you how statistically significant the influence is (instead of a random effect)
Linear Regression: Coefficients (example)

• The influence of the four factors on usefulness
  • Usef = 0.42 nov + 0.12 effort + 0.15 under + 0.32 relia – 0.34
  • Interpret the effect of Novelty on Usef
    • b > 0 and p << 0.001; our model suggests a significant positive effect of novelty on usefulness
    • b = 0.42; while other conditions being equal, our model suggests that one unit increase in user’s novelty rating increases the usefulness rating by 0.42

Linear regression model (robust fit):
usef ~ 1 + nov + effort + under + relia

Estimated Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>tStat</th>
<th>pValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.34171</td>
<td>0.32334</td>
<td>-1.0568</td>
<td>0.29094</td>
</tr>
<tr>
<td>nov</td>
<td>0.42583</td>
<td>0.035567</td>
<td>11.972</td>
<td>2.7355e-30</td>
</tr>
<tr>
<td>effort</td>
<td>0.12231</td>
<td>0.049757</td>
<td>2.4581</td>
<td>0.014199</td>
</tr>
<tr>
<td>under</td>
<td>0.15881</td>
<td>0.045018</td>
<td>3.5276</td>
<td>0.00044554</td>
</tr>
<tr>
<td>relia</td>
<td>0.32186</td>
<td>0.041967</td>
<td>7.6692</td>
<td>5.5306e-14</td>
</tr>
</tbody>
</table>
Linear Regression: Multicollinearity

Multicollinearity
• IVs are highly correlated
• Violates the assumption of using OLS linear regression
• Causes issues for model fitting

Examine multicollinearity
• Compute the correlation matrix of IVs
• Check using some measures such as tolerance and VIF
• Tolerance (T)
  • With $T < 0.1$ there might be multicollinearity in the data and with $T < 0.01$ there certainly is.
 • VIF (variance inflation factor)
  • Values of 10-30 indicate a mediocre multicollinearity in the linear regression variables, values $> 30$ indicate strong multicollinearity.
• Remove or merge variables to solve multicollinearity
Linear Regression: Data Assumptions

Linear
• The IV and DV should have a linear relationship
• Can draw scatterplots to double check

Normality
• All variables need to be multivariate normal
• Can check using a Q-Q plot or using the Kolmogorov-Smirnov test

No multicollinearity
• Multicollinearity: IVs are highly correlated

Can we still use linear regression?
• No data perfectly satisfies the assumptions
• It is always worthwhile to double check
• Be cautious in interpreting the results (particularly when data assumptions are violated)
Logistic Regression

For binary dependent variable (DV)

• $y$ is either 1 or 0

• e.g., whether or not to click on a search result

$$P(y = 1) = \frac{1}{1 + e^{-z}}$$

$$z = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$$

Most of the things are similar to linear regression

• Except for how to interpret the coefficients of IVs

• $b_i$ indicates how would changes in $x_i$ influence the log probability ratio of $y = 1 / y = 0$

$$\log \frac{P(y = 1)}{P(y = 0)} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$$
Outline

• Related theories
  • Anomalous State of Knowledge (ASK)
  • Stratified model of relevance interactions
• Laboratory user studies and experimental design
• Statistical analysis & examples
  • Correlation analysis, e.g., Pearson, Spearman
  • Regression analysis, e.g., linear & logistic regression
  • ANOVA and non-parametric variants
  • Our focus: when to use & how to read and interpret results
ANOVA

• Limitation of t-test
  • Only compare two groups’ means
  • Only one factor; no factorial design

• Example
  • Ask users to use three systems in two different types of tasks
    • Rate their satisfaction after each condition
  • Research questions
    • Does the choice of search system influence user satisfaction?
    • Does the type of task influence user satisfaction?
  • Difficult to draw conclusions using t-test

<table>
<thead>
<tr>
<th></th>
<th>Lucene</th>
<th>Galago</th>
<th>Indri</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>3.40</td>
<td>3.72</td>
<td>3.65</td>
</tr>
<tr>
<td>Task 2</td>
<td>3.56</td>
<td>3.48</td>
<td>3.82</td>
</tr>
</tbody>
</table>
One-way ANOVA

- Only one factor
  - $k$ groups, e.g., $(k = 3)$ Lucene vs. Galago vs. Indri
  - We ask N users to use and rate the three systems
  - We do not look into task difference so far
  - We observe the mean ratings as follows

<table>
<thead>
<tr>
<th></th>
<th>Lucene</th>
<th>Galago</th>
<th>Indri</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>3.40</td>
<td>3.82</td>
<td>3.60</td>
</tr>
</tbody>
</table>

- Null hypothesis
  - The three systems have no difference (mean is the same for all)

- Alternative hypothesis
  - At least one system is different.
  - Do not assume particular relationship.
  - Research question: whether the choice of system matters.
One-way ANOVA

• The Idea
  • Examine where the variance of the observations comes from
  • Between group: the variance for data from different conditions
    • Quantify the differences caused by the factor and chance
  • Within group: the variance for data in the same condition
    • Quantify the differences caused by chance (e.g., individual difference)
  • The test statistics looks into the between and within-group variances

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>91.476</td>
<td>2</td>
<td>45.733</td>
<td>4.467</td>
<td>.021</td>
</tr>
<tr>
<td>Within</td>
<td>276.400</td>
<td>27</td>
<td>10.237</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>367.867</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
One-way ANOVA

• Results & Interpretation
  • F-test gets $p = 0.021 < 0.05$ (the threshold)
  • Reject the null hypothesis
  • Conclude that at least one system is different from others
    • Indicates that the choices of systems do influence user satisfaction
  • But we don’t know how exactly the three systems differ
    • Run post-hoc tests, e.g., Turkey’s HSD

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>91.476</td>
<td>2</td>
<td>45.733</td>
<td>4.467</td>
<td>.021</td>
</tr>
<tr>
<td>Within</td>
<td>276.400</td>
<td>27</td>
<td>10.237</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>367.867</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Two-way ANOVA

- Now look into two factors, for example:
  - Factor 1: choice of system
  - Factor 2: task
  - Ask users to each combination and rate their satisfaction

<table>
<thead>
<tr>
<th></th>
<th>Lucene</th>
<th>Galago</th>
<th>Indri</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task 1</strong></td>
<td>3.40</td>
<td>3.72</td>
<td>3.65</td>
</tr>
<tr>
<td><strong>Task 2</strong></td>
<td>3.56</td>
<td>3.48</td>
<td>3.82</td>
</tr>
<tr>
<td><strong>Task 3</strong></td>
<td>3.75</td>
<td>3.82</td>
<td>3.52</td>
</tr>
</tbody>
</table>

- Interested in knowing
  - Does the choice of system influence user satisfaction?
  - Does the choice of task influence user satisfaction?
  - Does the combination of two factors (interaction) influence user satisfaction?
Two-way ANOVA

- The Idea
  - Similar to one-way ANOVA
  - look into where the variance comes from
    - Each factor
    - Interaction

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Adj SS</th>
<th>Adj MS</th>
<th>F-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task</td>
<td>2</td>
<td>8.222</td>
<td>4.1111</td>
<td>1.71</td>
<td>0.2094</td>
</tr>
<tr>
<td>System</td>
<td>2</td>
<td>20.222</td>
<td>10.1111</td>
<td>4.20</td>
<td>0.0318</td>
</tr>
<tr>
<td>Task * System</td>
<td>4</td>
<td>46.222</td>
<td>11.5556</td>
<td>4.80</td>
<td>0.0082</td>
</tr>
<tr>
<td>Error</td>
<td>18</td>
<td>43.333</td>
<td>2.4074</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>118.000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Two-way ANOVA

- Interpretation and results
  - Similar to one-way ANOVA
  - Run F-test, draw conclusion based on p-value
    - The choice of task does not affect user satisfaction
    - The choice of system does affect
    - Interaction: does affect

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Adj SS</th>
<th>Adj MS</th>
<th>F-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task</td>
<td>2</td>
<td>8.222</td>
<td>4.1111</td>
<td>1.71</td>
<td>0.2094</td>
</tr>
<tr>
<td>System</td>
<td>2</td>
<td>20.222</td>
<td>10.1111</td>
<td>4.20</td>
<td>0.0318</td>
</tr>
<tr>
<td>Task * System</td>
<td>4</td>
<td>46.222</td>
<td>11.5556</td>
<td>4.80</td>
<td>0.0082</td>
</tr>
<tr>
<td>Error</td>
<td>18</td>
<td>43.333</td>
<td>2.4074</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>118.000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ANOVA

• **Our expectation**
  - Know when to use
  - Be able to interpret ANOVA results (when reading a paper)
  - Understand the main idea (examine the influence of factors by where the variance comes from)
  - Do not need to know all tech details

• **Other things we haven’t covered about ANOVA**
  - Similar to t-test for related and independent samples, ANOVA can be different depends on the types of observations you collected, which depends on the experiment design (e.g., within-group or between group)
  - Assumptions: independence; normality; equal variance
  - When assumptions do not hold, check non-parametric alternatives, e.g., Kruskal–Wallis
Example: Kelly et al. (2009)

• **Purpose**
  - Compare query suggestion vs. term suggestion
  - Compare suggestions using two different methods
    - System-generated: using co-occurrence of words
    - User-generated: using words in other users’ queries

• **Experimental design**
  - 2 x 2 factorial design
  - A within subjects design
  - Each subject used four different conditions to work on search tasks
  - Record the users’ behaviors and perceptions after using the system

Example: Kelly et al. (2009)

- Interface: Query suggestion

Example: Kelly et al. (2009)

- Interface: Term suggestion

Example: Kelly et al. (2009)

- Results: mean & standard deviation of ratings for each condition

<table>
<thead>
<tr>
<th>Questionnaire Item</th>
<th>TS</th>
<th>QS</th>
<th>SGS</th>
<th>UGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The suggestions made things I wanted to accomplish easier to do.</td>
<td>3.24 (1.26)</td>
<td>3.58 (1.20)</td>
<td>3.20 (1.26)</td>
<td>3.66 (1.17)</td>
</tr>
<tr>
<td>2. The suggestions helped me modify my queries.</td>
<td>2.75 (1.22)</td>
<td>2.67 (1.23)</td>
<td>3.17 (1.14)</td>
<td>2.16 (1.10)</td>
</tr>
<tr>
<td>3. The suggestions helped me think of new approaches to searching.</td>
<td>2.49 (1.20)</td>
<td>3.24 (1.36)</td>
<td>2.42 (1.29)</td>
<td>3.40 (1.18)</td>
</tr>
<tr>
<td>4. The suggestions helped me better understand the topics.</td>
<td>2.65 (1.24)</td>
<td>3.29 (1.29)</td>
<td>2.72 (1.32)</td>
<td>3.28 (1.21)</td>
</tr>
<tr>
<td>5. The suggestions were easy to use.</td>
<td>3.02 (0.95)</td>
<td>2.75 (1.21)</td>
<td>3.23 (1.08)</td>
<td>2.46 (0.95)</td>
</tr>
<tr>
<td>6. It was easy to understand how the suggestions related to the search topics.</td>
<td>3.40 (1.13)</td>
<td>2.75 (1.19)</td>
<td>3.08 (1.17)</td>
<td>3.06 (1.25)</td>
</tr>
<tr>
<td>7. The quality of the suggestions was good.</td>
<td>3.51 (1.05)</td>
<td>3.27 (1.18)</td>
<td>3.63 (1.09)</td>
<td>3.10 (1.09)</td>
</tr>
<tr>
<td>8. It was easy to find relevant documents with the system.</td>
<td>2.91 (1.08)</td>
<td>3.51 (1.01)</td>
<td>3.13 (1.07)</td>
<td>3.30 (1.11)</td>
</tr>
<tr>
<td>9. It was easy to understand why some documents were retrieved in response to my queries.</td>
<td>2.44 (0.88)</td>
<td>2.76 (1.02)</td>
<td>2.55 (0.93)</td>
<td>2.66 (1.00)</td>
</tr>
<tr>
<td>10. Overall, the system was effective in helping me complete the search tasks.</td>
<td>2.91 (1.04)</td>
<td>3.31 (1.25)</td>
<td>3.15 (1.21)</td>
<td>3.06 (1.11)</td>
</tr>
<tr>
<td>11. Overall, I was satisfied with my performance.</td>
<td>3.13 (1.04)</td>
<td>3.44 (0.99)</td>
<td>3.43 (1.05)</td>
<td>3.10 (0.97)</td>
</tr>
</tbody>
</table>

Example: Kelly et al. (2009)

- Results: effects of the two factors

<table>
<thead>
<tr>
<th>Item</th>
<th>Type (Terms or Queries)</th>
<th>Source (System or Users)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$F(1,108) = 2.29, p=.133, \text{ns}$</td>
<td>$F(1,108) = 3.88, p&lt;.05, \text{U}&gt;S$</td>
</tr>
<tr>
<td>2</td>
<td>$F(1,108) = .115, p=.735, \text{ns}$</td>
<td>$F(1,108) = 21.72, p&lt;.01, \text{S}&gt;U$</td>
</tr>
<tr>
<td>3</td>
<td>$F(1,108) = 10.75, p&lt;.01, \text{Q}&gt;T$</td>
<td>$F(1,108) = 18.48, p&lt;.01, \text{U}&gt;S$</td>
</tr>
<tr>
<td>4</td>
<td>$F(1,108) = 7.34, p&lt;.01, \text{Q}&gt;T$</td>
<td>$F(1,108) = 5.63, p&lt;.05, \text{U}&gt;S$</td>
</tr>
<tr>
<td>5</td>
<td>$F(1,108) = 2.04, p=.156, \text{ns}$</td>
<td>$F(1,108) = 15.59, p&lt;.01, \text{S}&gt;U$</td>
</tr>
<tr>
<td>6</td>
<td>$F(1,108) = 9.45, p&lt;.01, \text{T}&gt;Q$</td>
<td>$F(1,108) = .011, p=.917, \text{ns}$</td>
</tr>
<tr>
<td>7</td>
<td>$F(1,108) = 1.31, p=.254, \text{ns}$</td>
<td>$F(1,108) = 6.48, p&lt;.05, \text{S}&gt;U$</td>
</tr>
<tr>
<td>8</td>
<td>$F(1,108) = 10.18, p&lt;.01, \text{Q}&gt;T$</td>
<td>$F(1,108) = .705, p=.403, \text{ns}$</td>
</tr>
<tr>
<td>9</td>
<td>$F(1,108) = 3.38, p=.069, \text{ns}$</td>
<td>$F(1,108) = .361, p=.549, \text{ns}$</td>
</tr>
<tr>
<td>10</td>
<td>$F(1,108) = 3.84, p&lt;.05, \text{Q}&gt;T$</td>
<td>$F(1,108) = .168, p=.682, \text{ns}$</td>
</tr>
<tr>
<td>11</td>
<td>$F(1,108) = 3.17, p=.078, \text{ns}$</td>
<td>$F(1,108) = 3.05, p=.084, \text{ns}$</td>
</tr>
</tbody>
</table>

Note that...

- Note that many papers may not report the statistics in the same format you learned from your statistics courses...
  - Probably because of page limits in conference papers 😊
Summary

• Related theories
  • Anomalous State of Knowledge (ASK)
  • Stratified model of relevance interactions

• Experimental design & related issues
  • Factorial design
  • Carry over effect and counter balancing
  • Risks & ethics, e.g., IRB

• Statistical analysis & examples
  • Correlation analysis, e.g., Pearson, Spearman
  • Regression analysis, e.g., linear & logistic regression
  • ANOVA, one-way and two-way
  • Our focus: when to use & how to read and interpret results