Introduction to Biostatistics and R Software Workshop

02.02.1400 – 06.03.1400
(Thursdays 9:00-12:00)

Topic	Time (week-minutes)
1. **Introduction**	
1.1. What is (Bio)statistics and Why It is Important?	**W1- 5 min**
1.2. Sampling	**W1- 15 min**
1.3. Observational Studies and Experiments	**W1- 10 min**
1.4. Data Exploration and Analysis	**W1- 5 min**
1.4.1. Cross-Sectional, Longitudinal, and Time Series Data	**W1- 5 min**
1.1. Statistical Inference	**W1- 5 min**
1.1. Computation	
1.1.1. Why R?	
1.1.1. The R project website	
1.1.2. Downloading and Installing R	
1.1.3. Working with R Using Rstudio	
1.1.4. R Package management	**W1- 20 min**
1.1.5. Repositories: CRAN, Bioconductor, Neuroconductor, GitHub...	
1.1.6. Installing Packages	
1.1.7. Getting Help	
1.1.8. Basic Data Structures: Vectors, Factors, Matrices, Data Frames and list	**W1- 20 min**
1.1.1. Importing Data Files	**W1- 10 min**
1.1.1. Basic Data manipulation (select variables, filter or subset cases; reshape data)	**W1- 25 min**
2. **Data Exploration I**	
2.1. Data Visualization and Summary Statistics	**W1- 10 min**
1.1. Variable Types	
1.1. Exploring Categorical Variables	
1.1.1. Relative Frequency and Percentage 1.1.2. Bar Graph 1.1.3. Pie Chart	**W1- 25 min**
1. Exploring Numerical Variables	**W1- 25 min**

Introduction to Biostatistics and R Software Workshop

02.02.1400 – 06.03.1400
(Thursdays 9:00-12:00)

1.1.1.Histograms	
1.1.2.Mean and Median	
1.1.3.Variance and Standard Deviation	
1.1.4.Coefficient of Variation	
1.1.5.Quantiles	
1.1.6.Boxplots	
2. **Data Preprocessing**	
2.1. Missing Data	**W2- 5 min**
2.2. Outliers	
2.3. Data transformation	**W2- 10 min**
2.4. Creating New Variable Based on Two or More Existing Variables	
2.5. Scaling and Shifting Variables	
2.6. Variable Standardization	
2.7. Creating Categories for Numerical Variables	**W2- 15 min**
3. **Data Exploration II (Exploring Relationships)**	
3.3 Relationships Between Two Numerical Random Variables	**W2- 30 min**
3.4 Relationships Between Categorical Variables	**W2- 15 min**
3.5 Relationships Between Numerical and Categorical Variables	**W2- 10 min**
4. **Probability**	
4.1. Probability as a Measure of Uncertainty	**W2- 15 min**
4.2. The Sample Space	
4.3. Conditional Probabilities	
4.4. Bayes' Theorem	**W2- 35 min**
4.4.1.Application of Bayes' Theorem in Medical Diagnosis	
4.5. Bayesian Statistics (Introduction)	
5. **Random Variables and Probability Distributions**	
5.1. Random Variables	**W2- 5 min**
5.2. Distributions and Their Shapes	**W2- 15 min**
6. **Estimation**	
6.1. Parameter Estimation	**W2- 10 min**
6.1.1.Point Estimation	
6.2. Population Mean	**W2- 15 min**
6.3. Population Variance	
6.4. Sampling Distribution	**W3- 15 min**

Introduction to Biostatistics and R Software Workshop

02.02.1400 – 06.03.1400
(Thursdays 9:00-12:00)

6.5. Confidence Intervals for the Population Mean	**W3- 20 min**
6.5.1.Confidence Interval When the Population Variance Is Unknown	
6.6. Using Central Limit Theorem for Confidence Interval	**W3- 10 min**
6.7. Confidence Intervals for the Population Proportion	**W3- 15 min**
7. **Hypothesis Testing**	
7.1. General Concepts	**W3- 20 min**
7.2. Hypothesis Testing for the Population Mean	**W3- 15 min**
7.3. Statistical Significance	
7.4. z -Tests of the Population Mean	
7.5. Interpretation of p –value	**W3- 15 min**
7.6. One-Sided Hypothesis Testing	
7.7. Two-Sided Hypothesis Testing	
7.8. Hypothesis Testing Using t –tests	**W3- 10 min**
7.9. Hypothesis Testing for Population Proportion	**W3- 10 min**
7.10. Test of Normality	**W3- 10 min**
7.11. nonparametric methods	**W3 -10 min**
8. **Statistical Inference for the Relationship Between Two Variables**	
8.1. Relationship Between a Numerical Variable and a Binary Variable	**W3- 15 min**
8.1.1.Two-Sample t -tests for Comparing the Means	
8.1.2.Paired t –test	
8.1.3. Nonparametric methods	
8.2. Inference about the Relationship Between Two Binary Variables	**W3- 15 min**
8.3. Inference Regarding the Linear Relationship Between Two Numerical Variables	
9. **Analysis of Variance (ANOVA)**	
9.1. One-Way ANOVA	**W4- 15 min**
9.2. Blocked ANOVA	**W4 – 15 min**
9.3. Two-Way ANOVA	**W4- 15 min**
9.4. The Assumptions of ANOVA	**W4- 15 min**
9.5. Repeated Measures ANOVA	**W4- 15 min**
9.5.1.The Assumptions of Repeated Measures	
9.6. nonparametric methods (independent and related samples)	**W4- 20 min**
10. **Analysis of Categorical Variables**	**W4- 30 min**
10.1. Test of Independence	
10.1.1. Pearson's χ^2 Test of Independence	

Introduction to Biostatistics and R Software Workshop

02.02.1400 – 06.03.1400
(Thursdays 9:00-12:00)

10.1.2. Fisher's Exact Test	
10.1.3. Clustered (Colerrated) binary data	
11. **Regression Analysis**	
11.1. Linear Regression Models with One Binary Explanatory Variable	**W4- 15 min**
11.2. Statistical Inference Using Simple Linear Regression Models	
11.3. Linear Regression Models with One Numerical Explanatory Variable	
11.4. Goodness of Fit	**W4-15 min**
11.5. Model Assumptions and Diagnostics	
11.6. Multiple Linear Regression	**W4- 25 min**
11.6.1. Interaction	
11.6.2. Model Assumptions and Diagnostics and Remedy (collinearity)	**W5- 30 min**
12. **Analysis of Covariance (ANCOVA)**	
12.1. Introduction and implementation in R	**W5- 10 min**
12.2. The Assumptions of ANCOVA	**W5- 20 min**
13. **Power Analysis and Sample Size Estimation**	
13.1. Statistical and clinical effect size	**W5- 30 min**
13.2. Type I and II Errors	
13.3. Power of a test	
14. **Multiple comparison**	
14.1. Error rates and general concepts	**W5- 10 min**
14.1.1. Family-wise error rate (FWER)	**W5- 30 min**
14.1.2. Per-comparison error rate (PCER)	
14.1.3. False discovery rate (FDR)	
15. **Practical examples for neuroimaging**	
15.1. research planning	**W6- 45 min**
15.2. data analysis and interpretation	**W6- 45 min**
15.3. statistical considerations	**W6- 45 min**
15.4. practical examples of applications	**W6- 45 min**