

Biostatistics for Health Sciences

Review Sheet



<http://mathvault.ca>

June 1, 2017

Contents

1	Descriptive Statistics	2
1.1	Variables	2
1.1.1	Qualitative	2
1.1.2	Quantitative	2
1.2	Chart	2
1.3	Boxplot	3
1.3.1	5-Number Summary	3
1.3.2	Outlier Thresholds	3
1.4	Linear Transformation	3
1.4.1	Type	3
1.4.2	Effects On	3
1.4.3	Coefficient of Variation	3
2	Probability	3
2.1	Combinatorics	3
2.2	Laws	3
2.3	Real-Life Example: Diagnostic Tests	3
3	Distributions	4
3.1	Discrete	4
3.1.1	Binomial Distribution	4
3.1.2	Sample Proportion	4
3.1.3	Geometric Distribution	4
3.1.4	Negative Binomial Distribution	4
3.1.5	Hypergeometric Distribution	4
3.1.6	Poisson Distribution	4
3.2	Continuous	4

3.2.1	Uniform Distribution	4
3.2.2	Normal Distribution	4
4	Parametric Inferential Statistics	4
4.1	Theory	4
4.2	Hypotheses Testing	4
4.2.1	One-Sample Mean	4
4.2.2	Two-Independent-Sample Means	4
4.2.3	Paired Differences	4
4.2.4	One-Sample Proportion	4
4.2.5	Two-Sample Proportions	4
4.2.6	One-Sample χ^2 (Goodness-of-Fit)	4
4.2.7	$R \times C$ Contingency Table	4
4.2.8	Correlation	5
4.3	Confidence Interval	5
4.3.1	One-Sample Mean	5
4.3.2	Two-Independent-Sample Means	5
4.3.3	Paired Differences	5
4.3.4	One-Sample Proportion	5
4.3.5	Two-Sample Proportions	5
4.3.6	Correlation	5
5	Non-Parametric Methods	5
5.1	Sign Test (One-Sample Median)	5
5.2	Signed-Rank Test (Paired Differences)	5
5.3	Rank-Sum Test (Two Independent Samples)	5
5.4	Confidence Interval for One-Sample Median	5
5.5	Spearman's Rank Correlation	5
6	Simple Regression	5
6.1	Model	5
6.2	Theory	5

1 Descriptive Statistics

1.1.2 Quantitative

1.1 Variables

- Discrete
- Continuous

1.1.1 Qualitative

- Nominal
- Ordinal

1.2 Chart

- Pie Chart

- (Stacked) Bar Chart
- Histogram
- (Cumulative) Frequency Polygon
- Stemplot

1.3 Boxplot

1.3.1 5-Number Summary

- Central Box (Q1, Q2, Q3)
- Max and Min

1.3.2 Outlier Thresholds

- $1.5 \times \text{IQR}$ away from the central box

1.4 Linear Transformation

1.4.1 Type

- Shifting
- Rescaling

1.4.2 Effects On

- Mean
- Median
- Quartiles
- IQR
- Range
- Standard Deviation

1.4.3 Coefficient of Variation

- $\frac{s}{\bar{X}}$ in %
- Variability with respect to \bar{X}

- Invariant under change of units

2 Probability

2.1 Combinatorics

- Multiplication Rule
- Permutation Rule
- Combination Rule (Binomial Coefficient)
- Partition Rule (Multinomial Coefficient)

2.2 Laws

- Union
- Complement
- Intersection
 - General Formula
 - Formula for Independent Events
- Law of Total Probability
- Conditional Probability

2.3 Real-Life Example: Diagnostic Tests

- # of
 - True Positive
 - False Positive
 - True Negative
 - False Negative
- Sensitivity (i.e., $P(\text{Diagnosed}|\text{Disease})$)
- Specificity (i.e., $P(\text{Undiagnosed}|\text{No Disease})$)
- Gold Standard
- Prevalence

3 Distributions

3.1 Discrete

3.1.1 Binomial Distribution

- Distribution of the # of successes among n

3.1.2 Sample Proportion

3.1.3 Geometric Distribution

- Distribution of the first-success trial

3.1.4 Negative Binomial Distribution

- Distribution of the k^{th} -success trial

3.1.5 Hypergeometric Distribution

3.1.6 Poisson Distribution

3.2 Continuous

3.2.1 Uniform Distribution

3.2.2 Normal Distribution

- $X \sim \text{Normal}(\mu, \sigma) \implies \frac{X - \mu}{\sigma} = Z$
- Assessing Normality
 - Mound-Shape Histogram
 - 68-95-99 Rule
 - Normal Probability Plot (Z-Score vs. Variable)

4 Parametric Inferential Statistics

4.1 Theory

- Unbiased Estimator
- Central Limit Theorem
- Standard Error
- Relationship between α and β
- Power

4.2 Hypotheses Testing

4.2.1 One-Sample Mean

- Large Sample
- Small Sample

4.2.2 Two-Independent-Sample Means

- Large Sample
- Small Sample (Pooled Variance, Total Degree of Freedom)

4.2.3 Paired Differences

4.2.4 One-Sample Proportion

4.2.5 Two-Sample Proportions

- Pooled Proportion

4.2.6 One-Sample χ^2 (Goodness-of-Fit)

4.2.7 $R \times C$ Contingency Table

- Multi-Sample χ^2
 - Yate Correction (2×2)

- Fisher's Exact Test (2×2)
 - $X = \#$ of successes in Group 1

4.2.8 Correlation

- Exact Statistic
- Large Sample: $\frac{1}{2} \ln \frac{1+r}{1-r} \sim N\left(\frac{1}{2} \ln \frac{1+\rho}{1-\rho}, \sqrt{\frac{1}{n-3}}\right)$

4.3 Confidence Interval

4.3.1 One-Sample Mean

- Large Sample
- Small Sample

4.3.2 Two-Independent-Sample Means

- Large Sample
- Small Sample

4.3.3 Paired Differences

4.3.4 One-Sample Proportion

4.3.5 Two-Sample Proportions

4.3.6 Correlation

- Small Sample
- Large Sample

5 Non-Parametric Methods

5.1 Sign Test (One-Sample Median)

- Exact Statistic ($X \stackrel{df}{=} \#$ of data $> \Delta$)

- Large Sample (Normal Approximation)

5.2 Signed-Rank Test (Paired Differences)

- Exact Statistic ($R^+ \stackrel{df}{=} \text{Total Rank of Positive Data}$)
- Large Sample: $R^+ \sim N\left(\frac{n(n+1)}{4}, \sqrt{\frac{n(n+1)(2n+1)}{24}}\right)$

5.3 Rank-Sum Test (Two Independent Samples)

- Exact Statistic ($R_1 \stackrel{def}{=} \text{Total Rank of Sample 1}$)

5.4 Confidence Interval for One-Sample Median

5.5 Spearman's Rank Correlation

6 Simple Regression

6.1 Model

- $Y = \alpha + \beta x + \epsilon$
- β estimated by $b = \frac{SS_{xy}}{SS_x}$
- α estimated by $a = \bar{Y} - b\bar{X}$
- $\epsilon \sim N(0, \sigma)$, σ estimated by $\sqrt{\frac{SS_{residual}}{N-2}}$

6.2 Theory

- $SS_{Total} = SS_{Residual} + SS_{Regression}$
- $r^2 = \frac{SS_{Regression}}{SS_{Total}}$