
Contents

Preface	VII
List of Figures	XXVII
List of Tables	XXXI
1 Multiple Hypothesis Testing	1
1.1 Introduction	1
1.1.1 Motivation	1
1.1.2 Bibliography for proposed multiple testing methodology	2
1.1.3 Overview of applications to biomedical and genomic research	4
1.1.4 Road map	6
1.2 Multiple hypothesis testing framework	9
1.2.1 Overview	9
1.2.2 Data generating distribution	10
1.2.3 Parameters	11
1.2.4 Null and alternative hypotheses	12
1.2.5 Test statistics	13
1.2.6 Multiple testing procedures	15
1.2.7 Rejection regions	15
1.2.8 Errors in multiple hypothesis testing: Type I, Type II, and Type III errors	17
1.2.9 Type I error rates	18
1.2.10 Power	22
1.2.11 Type I error rates and power: Comparisons and examples	23
1.2.12 Unadjusted and adjusted p -values	27
1.2.13 Stepwise multiple testing procedures	34

2	Test Statistics Null Distribution	49
2.1	Introduction	49
2.1.1	Motivation	49
2.1.2	Outline	51
2.2	Type I error control and choice of a test statistics null distribution	52
2.2.1	Type I error control	52
2.2.2	Sketch of proposed approach to Type I error control	53
2.2.3	Characterization of test statistics null distribution in terms of null domination conditions	55
2.2.4	Contrast with other approaches	59
2.3	Null shift and scale-transformed test statistics null distribution	60
2.3.1	Explicit construction for the test statistics null distribution	60
2.3.2	Bootstrap estimation of the test statistics null distribution	65
2.4	Null quantile-transformed test statistics null distribution	69
2.4.1	Explicit construction for the test statistics null distribution	70
2.4.2	Bootstrap estimation of the test statistics null distribution	72
2.4.3	Comparison of null shift and scale-transformed and null quantile-transformed null distributions	73
2.5	Null distribution for transformations of the test statistics	75
2.5.1	Null distribution for transformed test statistics	75
2.5.2	Example: Absolute value transformation	77
2.5.3	Example: Null shift and scale and null quantile transformations	78
2.5.4	Bootstrap estimation of the null distribution for transformed test statistics	79
2.6	Testing single-parameter null hypotheses based on t -statistics	79
2.6.1	Set-up and assumptions	79
2.6.2	Test statistics null distribution	80
2.6.3	Estimation of the test statistics null distribution	82
2.6.4	Example: Tests for means	83
2.6.5	Example: Tests for correlation coefficients	83
2.6.6	Example: Tests for regression coefficients	84
2.7	Testing multiple-parameter null hypotheses based on F -statistics	87
2.7.1	Set-up and assumptions	87
2.7.2	Test statistics null distribution	88
2.7.3	Estimation of the test statistics null distribution	93
2.8	Weak and strong Type I error control and subset pivotality	94
2.8.1	Weak and strong control of a Type I error rate	95

2.8.2	Subset pivotality	97
2.9	Test statistics null distributions based on bootstrap and permutation data generating distributions	98
2.9.1	The two-sample test of means problem	99
2.9.2	Distribution of the test statistics under two different data generating distributions	100
2.9.3	Bootstrap and permutation test statistics null distributions	104
3	Overview of Multiple Testing Procedures	109
3.1	Introduction	109
3.1.1	Set-up	109
3.1.2	Type I error control and choice of a test statistics null distribution	110
3.1.3	Marginal multiple testing procedures	111
3.1.4	Joint multiple testing procedures	112
3.2	Multiple testing procedures for controlling the number of Type I errors: FWER	112
3.2.1	Controlling the number of Type I errors	112
3.2.2	FWER-controlling single-step procedures	113
3.2.3	FWER-controlling step-down procedures	121
3.2.4	FWER-controlling step-up procedures	127
3.3	Multiple testing procedures for controlling the number of Type I errors: gFWER	134
3.3.1	gFWER-controlling single-step and step-down Lehmann and Romano procedures	134
3.3.2	gFWER-controlling single-step common-cut-off and common-quantile procedures	137
3.3.3	gFWER-controlling augmentation multiple testing procedures	139
3.3.4	gFWER-controlling resampling-based empirical Bayes procedures	140
3.3.5	Other gFWER-controlling procedures	140
3.3.6	Comparison of gFWER-controlling procedures	140
3.4	Multiple testing procedures for controlling the proportion of Type I errors among the rejected hypotheses: FDR	145
3.4.1	Controlling the number vs. the proportion of Type I errors	145
3.4.2	FDR-controlling step-up Benjamini and Hochberg procedure	146
3.4.3	FDR-controlling step-up Benjamini and Yekutieli procedure	147
3.4.4	FDR-controlling resampling-based empirical Bayes procedures	148
3.4.5	Other FDR-controlling procedures	148

3.5	Multiple testing procedures for controlling the proportion of Type I errors among the rejected hypotheses: TPPFP	149
3.5.1	Controlling the expected value vs. tail probabilities for the proportion of Type I errors	149
3.5.2	TPPFP-controlling step-down Lehmann and Romano procedures	150
3.5.3	TPPFP-controlling augmentation multiple testing procedures	153
3.5.4	TPPFP-controlling resampling-based empirical Bayes procedures	154
3.5.5	Comparison of TPPFP-controlling procedures	155
4	Single-Step Multiple Testing Procedures for Controlling General Type I Error Rates, $\Theta(F_{V_n})$	161
4.1	Introduction	161
4.1.1	Motivation	161
4.1.2	Outline	163
4.2	$\Theta(F_{V_n})$ -controlling single-step procedures	163
4.2.1	Single-step common-quantile procedure	164
4.2.2	Single-step common-cut-off procedure	165
4.2.3	Asymptotic control of Type I error rate and test statistics null distribution	165
4.2.4	Common-cut-off vs. common-quantile procedures	168
4.3	Adjusted p -values for $\Theta(F_{V_n})$ -controlling single-step procedures	169
4.3.1	General Type I error rates, $\Theta(F_{V_n})$	169
4.3.2	Per-comparison error rate, PCER	171
4.3.3	Generalized family-wise error rate, gFWER	172
4.4	$\Theta(F_{V_n})$ -controlling bootstrap-based single-step procedures	174
4.4.1	Asymptotic control of Type I error rate for single-step procedures based on consistent estimator of test statistics null distribution	175
4.4.2	Bootstrap-based single-step procedures	183
4.5	$\Theta(F_{V_n})$ -controlling two-sided single-step procedures	187
4.5.1	Symmetric two-sided single-step common-quantile procedure	188
4.5.2	Symmetric two-sided single-step common-cut-off procedure	189
4.5.3	Asymptotic control of Type I error rate and test statistics null distribution	189
4.5.4	Bootstrap-based symmetric two-sided single-step procedures	190
4.6	Multiple hypothesis testing and confidence regions	191
4.6.1	Confidence regions for general Type I error rates, $\Theta(F_{V_n})$	191

4.6.2	Equivalence between Θ -specific single-step multiple testing procedures and confidence regions	194
4.6.3	Bootstrap-based confidence regions for general Type I error rates, $\Theta(F_{V_n})$	196
4.7	Optimal multiple testing procedures	197
5	Step-Down Multiple Testing Procedures for Controlling the Family-Wise Error Rate	199
5.1	Introduction	199
5.1.1	Motivation	199
5.1.2	Outline	201
5.2	FWER-controlling step-down common-cut-off procedure based on maxima of test statistics	202
5.2.1	Step-down maxT procedure	202
5.2.2	Asymptotic control of the FWER	203
5.2.3	Test statistics null distribution	208
5.2.4	Adjusted p -values	211
5.3	FWER-controlling step-down common-quantile procedure based on minima of unadjusted p -values	212
5.3.1	Step-down minP procedure	213
5.3.2	Asymptotic control of the FWER	215
5.3.3	Test statistics null distribution	218
5.3.4	Adjusted p -values	219
5.3.5	Comparison of joint step-down minP procedure to marginal step-down procedures	220
5.4	FWER-controlling step-up common-cut-off and common-quantile procedures	224
5.4.1	Candidate step-up maxT and minP procedures	224
5.4.2	Comparison of joint stepwise minP procedures to marginal stepwise Holm and Hochberg procedures	227
5.5	FWER-controlling bootstrap-based step-down procedures	227
5.5.1	Asymptotic control of FWER for step-down procedures based on consistent estimator of test statistics null distribution	228
5.5.2	Bootstrap-based step-down procedures	232
6	Augmentation Multiple Testing Procedures for Controlling Generalized Tail Probability Error Rates	235
6.1	Introduction	235
6.1.1	Motivation	235
6.1.2	Outline	237
6.1.3	Type I error rates	238
6.1.4	Augmentation multiple testing procedures	239
6.2	Augmentation multiple testing procedures for controlling the generalized family-wise error rate, $gFWER(k) = \Pr(V_n > k)$	242

6.2.1	gFWER-controlling augmentation multiple testing procedures	242
6.2.2	Finite sample and asymptotic control of the gFWER . .	243
6.2.3	Adjusted p -values for gFWER-controlling augmentation multiple testing procedures	244
6.3	Augmentation multiple testing procedures for controlling the tail probability for the proportion of false positives, $TPFP(q) = \Pr(V_n/R_n > q)$	245
6.3.1	TPFP-controlling augmentation multiple testing procedures	245
6.3.2	Finite sample and asymptotic control of the TPFP . .	247
6.3.3	Adjusted p -values for TPFP-controlling augmentation multiple testing procedures	250
6.4	TPFP-based multiple testing procedures for controlling the false discovery rate, $FDR = E[V_n/R_n]$	251
6.4.1	FDR-controlling TPFP-based multiple testing procedures	251
6.4.2	Adjusted p -values for FDR-controlling TPFP-based multiple testing procedures	255
6.5	General results on augmentation multiple testing procedures . .	256
6.5.1	Augmentation multiple testing procedures for controlling the generalized tail probability error rate, $gTP(q, g) = \Pr(g(V_n, R_n) > q)$	257
6.5.2	Adjusted p -values for general augmentation multiple testing procedures	262
6.5.3	gFWER-controlling augmentation multiple testing procedures	264
6.5.4	TPFP-controlling augmentation multiple testing procedures	265
6.5.5	gTPFP-controlling augmentation multiple testing procedures	267
6.6	gTP-based multiple testing procedures for controlling the generalized expected value, $gEV(g) = E[g(V_n, R_n)]$	269
6.6.1	gEV-controlling gTP-based multiple testing procedures	270
6.6.2	Adjusted p -values for gEV-controlling gTP-based multiple testing procedures	271
6.7	Initial FWER- and gFWER-controlling multiple testing procedures	272
6.8	Discussion	273
7	Resampling-Based Empirical Bayes Multiple Testing Procedures for Controlling Generalized Tail Probability Error Rates	289
7.1	Introduction	289

7.1.1	Motivation	289
7.1.2	Outline	290
7.2	gTP-controlling resampling-based empirical Bayes procedures .	291
7.2.1	Notation	291
7.2.2	gTP control and optimal test statistic cut-offs.....	292
7.2.3	Overview of gTP-controlling resampling-based empirical Bayes procedures	294
7.2.4	Working model for distributions of null test statistics and guessed sets of true null hypotheses	295
7.2.5	gTP-controlling resampling-based empirical Bayes procedures	298
7.3	Adjusted p -values for gTP-controlling resampling-based empirical Bayes procedures	300
7.3.1	Adjusted p -values for common-cut-off procedure	300
7.3.2	Adjusted p -values for common-quantile procedure	302
7.4	Finite sample rationale for gTP control by resampling-based empirical Bayes procedures	303
7.4.1	Procedures based on constant guessed set of true null hypotheses and observed test statistics	303
7.4.2	Procedures based on constant guessed set of true null hypotheses and null test statistics	305
7.4.3	Procedures based on random guessed sets of true null hypotheses and null test statistics	305
7.5	Formal asymptotic gTP control results for resampling-based empirical Bayes procedures	306
7.5.1	Asymptotic control of gTP by resampling-based empirical Bayes Procedure 7.1	306
7.5.2	Assumptions for Theorem 7.2	307
7.5.3	Proof of Theorem 7.2	310
7.6	gTP-controlling resampling-based weighted empirical Bayes procedures.....	312
7.7	FDR-controlling empirical Bayes procedures	313
7.7.1	FDR-controlling empirical Bayes q -value-based procedures	314
7.7.2	Equivalence between empirical Bayes q -value-based procedure and frequentist step-up Benjamini and Hochberg procedure	316
7.8	Discussion	318
	Color Plates	321
8	Simulation Studies: Assessment of Test Statistics Null Distributions	345
8.1	Introduction	345
8.1.1	Motivation	345

8.1.2	Outline	347
8.2	Bootstrap-based multiple testing procedures	348
8.2.1	Null shift and scale-transformed test statistics null distribution	348
8.2.2	Bootstrap estimation of the null shift and scale-transformed test statistics null distribution...	349
8.2.3	Bootstrap-based single-step maxT procedure	350
8.3	Simulation Study 1: Tests for regression coefficients in linear models with dependent covariates and error terms	351
8.3.1	Simulation model	351
8.3.2	Multiple testing procedures	352
8.3.3	Simulation study design	354
8.3.4	Simulation study results	356
8.4	Simulation Study 2: Tests for correlation coefficients	360
8.4.1	Simulation model	360
8.4.2	Multiple testing procedures	360
8.4.3	Simulation study design	363
8.4.4	Simulation study results	364
9	Identification of Differentially Expressed and Co-Expressed Genes in High-Throughput Gene Expression Experiments	367
9.1	Introduction	367
9.2	Apolipoprotein AI experiment of Callow et al. (2000)	368
9.2.1	Apo AI dataset	368
9.2.2	Multiple testing procedures	370
9.2.3	Software implementation using the Bioconductor R package multtest	372
9.2.4	Results	376
9.3	Cancer microRNA study of Lu et al. (2005)	402
9.3.1	Cancer miRNA dataset	403
9.3.2	Multiple testing procedures	403
9.3.3	Results	405
10	Multiple Tests of Association with Biological Annotation Metadata	413
10.1	Introduction	413
10.1.1	Motivation	413
10.1.2	Contrast with other approaches	414
10.1.3	Outline	416
10.2	Statistical framework for multiple tests of association with biological annotation metadata	417
10.2.1	Gene-annotation profiles	417
10.2.2	Gene-parameter profiles	418

10.2.3	Association measures for gene-annotation and gene-parameter profiles	419
10.2.4	Multiple hypothesis testing	422
10.3	The Gene Ontology	425
10.3.1	Overview of the Gene Ontology	425
10.3.2	Overview of R and Bioconductor software for GO annotation metadata analysis	428
10.3.3	The annotation metadata package GO	430
10.3.4	Affymetrix chip-specific annotation metadata packages: The hgu95av2 package	433
10.3.5	Assembling a GO gene-annotation matrix	437
10.4	Tests of association between GO annotation and differential gene expression in ALL	439
10.4.1	Acute lymphoblastic leukemia study of Chiaretti et al. (2004)	439
10.4.2	Multiple hypothesis testing framework	441
10.4.3	Results	448
10.5	Discussion	453
11	HIV-1 Sequence Variation and Viral Replication Capacity .	477
11.1	Introduction	477
11.2	HIV-1 dataset of Segal et al. (2004)	477
11.2.1	HIV-1 sequence variation and viral replication capacity	477
11.2.2	HIV-1 dataset	478
11.3	Multiple testing procedures	479
11.3.1	Multiple testing analysis, Part I	480
11.3.2	Multiple testing analysis, Part II	480
11.4	Software implementation in SAS	481
11.5	Results	482
11.5.1	Multiple testing analysis, Part I	482
11.5.2	Multiple testing analysis, Part II	483
11.5.3	Biological interpretation	483
11.6	Discussion	484
12	Genetic Mapping of Complex Human Traits Using Single Nucleotide Polymorphisms: The ObeLinks Project	489
12.1	Introduction	489
12.1.1	Motivation	489
12.1.2	Outline	490
12.2	The ObeLinks Project	491
12.2.1	ObeLinks dataset	491
12.2.2	Galois lattices	493
12.3	Multiple testing procedures	495
12.4	Results	497
12.4.1	Body mass index	497

XXVI Contents

12.4.2	Glucose metabolism	498
12.5	Discussion	501
13	Software Implementation	519
13.1	R package <code>multtest</code>	519
13.1.1	Introduction	519
13.1.2	Overview	520
13.1.3	MTP function for resampling-based multiple testing procedures	522
13.1.4	Numerical and graphical summaries of a multiple testing procedure	527
13.1.5	Software design	528
13.2	SAS macros	529
A	Summary of Multiple Testing Procedures	533
B	Miscellaneous Mathematical and Statistical Results	551
B.1	Probability inequalities	551
B.2	Convergence results	552
B.3	Properties of floor and ceiling functions	553
C	SAS Code	555
	References	561
	Author Index	575
	Subject Index	579

Multiple Testing Procedures with Applications to
Genomics

Dudoit, S.; van der Laan, M.J.

2008, XXXIII, 590 p., Hardcover

ISBN: 978-0-387-49316-9