

# Seminar: Multiples Testen in der Biostatistik

Dr. Franziska Kappenberg

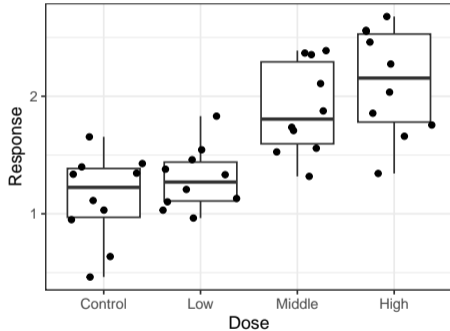
TU Dortmund, Fakultät Statistik

Sommersemester 2023

# Multiples Testen

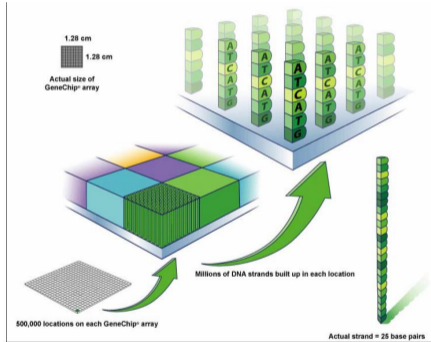
- ▶ Beim statistischen Testen wird üblicherweise in Fehler 1. und 2. Art unterschieden:
  - ▶ Fehler 1. Art: Lehne  $H_0$  ab, obwohl sie wahr ist (falsch positiv)
  - ▶ Fehler 2. Art: Lehne  $H_0$  nicht ab, obwohl sie falsch ist (falsch negativ)
- ▶ Für einen einzelnen Test wird die Wahrscheinlichkeit für einen Fehler 1. Art durch das Signifikanzniveau  $\alpha$  beschränkt
- ▶ Beim Durchführen mehrerer Tests tritt ein *Multiplizitätsproblem* auf, d.h. rein durch Zufall werden wahre Nullhypothesen abgelehnt, das Niveau  $\alpha$  wird nicht eingehalten
- ▶ Betrachtung verschiedener Fehlerraten, die auf starke / schwache Art kontrolliert werden können
- ▶ Anwendungsfälle speziell in der Biostatistik:

# Multiple Testen in der Biostatistik



- ▶ Multiple Vergleiche verschiedener Dosen gegen die Kontrolle (oder sogar verschiedener Dosen untereinander)
- ▶ Hohe Korrelation zwischen den Tests: Vernachlässigung dieser bei der Adjustierung führt zu konservativen Testverfahren
- ▶ Dunnett- oder Williams-Prozedur für jeweiligen Vergleich einer oder mehrerer gepoolter Konzentrationen gegen die Kontrolle unter Berücksichtigung der Korrelationen

# Multiples Testen in der Biostatistik



ID	Symbol	logFC	FoldChange	PValue	
1	223673_at	RFX4	4.992321	31.83012697	7.018472e-11
2	242054_s_at	SIX3	-4.939741	0.03258290	1.687460e-10
3	206345_s_at	PON1	-2.687722	0.15520833	6.269496e-10
4	200872_at	S100A10	3.092432	8.52932855	6.339838e-10
5	203789_s_at	SEMA3C	4.630311	24.76637262	9.475129e-10
6	214967_at	214967_at	3.223367	9.33963761	1.072445e-09
7	212097_at	CAV1	3.450423	10.93152428	1.449335e-09
8	218656_s_at	LHFPL6	3.619022	12.28666611	1.534997e-09
9	201116_s_at	CPE	3.229400	9.37877585	1.691989e-09
10	207173_x_at	CDH11	2.867834	7.29968183	1.878705e-09
11	209465_x_at	PTN	2.281272	4.86106277	1.944845e-09
12	201117_s_at	CPE	3.882014	14.74357402	1.951766e-09
13	219622_at	RAB20	2.327792	5.02036453	2.381772e-09
14	235756_at	235756_at	2.386481	5.22880473	2.667772e-09
15	207197_at	ZIC3	3.490329	11.23812002	2.944226e-09
16	211737_x_at	PTN	2.398581	5.27284391	2.974993e-09
17	201325_s_at	EMP1	3.217829	9.30385696	3.134636e-09
18	209466_x_at	PTN	2.340639	5.06526905	3.175761e-09
19	227198_at	AFF3	3.276533	9.69024527	3.538708e-09
20	202404_s_at	COL1A2	2.091125	4.26080203	3.545054e-09
21	207172_s_at	CDH11	2.415358	5.33451647	3.694033e-09
22	225681_at	CTHRC1	3.387362	10.46399434	3.886090e-09
23	1552520_at	TMEM74	-3.811204	0.07123825	4.098073e-09

Showing 1 to 23 of 54,675 entries, 10 total columns

↔ Bei mehreren zehntausend Genen soll nicht jeder Fehler 1. Art verhindert werden, sondern nur die Rate an falsch positiven Ergebnissen beschränkt werden

# Mögliche Themen

- ▶ Grundlagen (Signifikanzniveaus und ihre Kontrolle, Fehlerraten, ...)
- ▶ Einfache Prozeduren für FWER und FDR
- ▶ (Bayes-Statistik und FDR)
- ▶ Multiple Comparison Procedures: Dunnett-, Williams-Prozedur (und ähnliche)
- ▶ Resampling-basierte Verfahren
- ▶ Abschlusstestverfahren
- ▶ Grafische Methoden
- ▶ ...

# Organisatorisches

- ▶ Deutschsprachiges Seminar für Bachelor und Master
- ▶ Voraussetzungen: Schätzen und Testen, Interesse an biostatistischen Daten
- ▶ Prüfungsleistung:
  - ▶ Vortrag (BA: 30min, MA: 45min)
  - ▶ Nach Absprache enthält der Vortrag sowohl eine theoretische Vorstellung als auch eine kleine Simulationsstudie
  - ▶ Ausgearbeiteter Methodenteil (ca. 7 Seiten)
  - ▶ Aktive Teilnahme an der Diskussion
- ▶ Organisatorisches
  - ▶ Wöchentlicher Termin im Semester: Dienstags, 10:15 - 11:45 Uhr
  - ▶ Bei Interesse: Anmeldung per Mail (mit Name, Studiengang, Matrikelnummer) an [kappenberg@statistik.tu-dortmund.de](mailto:kappenberg@statistik.tu-dortmund.de) **bis zum 15.2.23**
  - ▶ Themenvergabe bis Ende Februar