universität
wien

MAGISTERARBEIT

Titel der Magisterarbeit

Gewichtete Multiple Testverfahren

Verfasser

Mag. rer. nat. Wolfgang Remmel

angestrebter akademischer Grad

Magister der Sozial- und Wirtschaftswissenschaften
(Mag. rer. soc. oec.)

Wien, im Jänner 2008

Studienkennzahl lt. Studienblatt:	A 066 951
Studienrichtung lt. Studienblatt:	Magisterstudium Statistik
Betreuer:	Ao. Univ. Prof. Mag. Dr. Martin Posch

Danksagungen

Zuerst möchte ich meinem Magisterarbeitsbetreuer Martin Posch danken, der mir über manche Klippe des multiplen Testens hinweggeholfen hat und auch bei den laienhaftesten Behauptungen meinerseits nie die Geduld verloren hat.
Weiters danke ich meinem Vater Leo Remmel, der sich getraut hat, eine umlautlose provisorische Fassung durchzulesen.
Die Mensa habe ich in den letzten Jahren kaum mehr besucht, weshalb eine Danksagung diesmal ausfällt.
Mein besonderer Dank gilt meiner lieben Frau Karin Remmel natürlich nicht nur für entscheidende logistische Hilfe am Anfang des Studiums, sondern auch wegen einer winterlichen Reise nach Leiston in der Grafschaft Suffolk.
Die Magisterbeit wurde, und das kommt wohl nicht sehr überraschend mit LaTeX und die Computerprogramme mit R verfaßt. Beides sind open source Produkte und den Autoren dieser Produkte wird herzlichst gedankt.

Inhaltsverzeichnis

Übersicht

Das simultane Testen mehrerer Hypothesen hat eine lange Geschichte ([18]). In der Varianzanalyse etwa ist ein Test über die Ablehnung der globalen Nullhypothese nicht sehr interessant, da sie meist sowieso nicht gilt. Das Interesse gilt vor allem der Fragestellung welche Mittelwerte sich wie voneinander unterscheiden.
In klinischen Versuchen gibt es oft mehrere Zeitpunkte zu denen eine Hypothese getestet werden soll und/oder mehrere Fragestellungen sollen gleichzeitig beantwortet werden.
In der Genetik, Hirnforschung oder Astronomie etwa sind dann oft tausend und mehr Hypothesen gleichzeitig zu testen.

Neben den in der Literatur bereits behandelten klassischen (ungewichteten) multiplen Testverfahren, den gewichteten Verfahren, bei denen die Gewichte a priori festgelegt werden, den gewichteten Verfahren bei denen die Gewichte aus den Daten bestimmt werden und der FWER kontrolliert wird behandeln wir auch den Fall bei dem die Gewichte aus den Daten bestimmt werden und die FDR kontrolliert wird.

Im Kapitel *Klassisches multiples Testen* behandeln wir die Bonferroni Korrektur (jede Hypothese wird zu einem von der Anzahl m der Hypothesen abhängenden kleineren Niveau α/m getestet). Weiters wird eine Modifikation der Bonferroni Korrektur, die Prozedur von Bonferroni-Holm vorgestellt. Kontrolliert wird jeweils die Wahrscheinlichkeit mindestens eine wahre Nullhypothese zu verwerfen (Familiy wide error rate FWER).

Im Kapitel *Anteil falscher Ablehnungen (FDR)* wird die FDR (der Erwartungswert des Quotienten aus der Anzahl der abgelehnten wahren Nullhypothesen V und der Anzahl aller abgelehnter Hypothesen R, bzw 0, wenn keine Hypothese abgelehnt wird) eingeführt.
Der Satz von Benjamini-Hochberg wird bewiesen. Er besagt, daß bei der Durchführung der Benjamini-Hochberg Prozedur im Fall unabhängiger P-Werte der wahren Nullhypothesen die FDR $\leq \alpha$ ist.

Die Benjamini-Hochberg Prozedur wird folgendermaßen durchgeführt: Sei m die Anzahl der Hypothesen. Die P-Werte P_i werden aufsteigende der Größe nach geordnet. Die geordneten P-Werte werden mit $P_{(i)}$ bezeichnet, die entsprechend geordneten Hypothesen mit $H_{(i)}$. Das maximale i mit $P_{(i)} \leq \frac{i}{m}\alpha$ wird bestimmt. Alle Hypothesen $H_{(j)}, j = 1, \ldots, i$ werden abgelehnt.
Es wird auch eine von Storey eingeführte der FDR verwandten Größe, die pFDR ($= E(V/R|R > 0)$) vorgestellt. Weiters stellen wir die Prozedur von Benjamini-Liu vor, die weniger Annahmen an die gemeinsame Verteilung der P-Werte stellt als die BH-Prozedur. Die Benjamini-Liu (BL) Prozedur verlangt nicht mehr wie die BH-Prozedur, daß die P-Werte der wahren Nullhypothesen unabhängig sind.

Im Kapitel *Gewichtete Prozeduren (a priori Gewichtung)* werden a priori bestimmte Gewichte eingeführt. Diese werden unabhängig von den P-Werten der einzelnen Hypothesen festgelegt. Es wird im wesentlichen dieselbe Prozedur wie bei der (ungewichteten) Benjamini-Hochberg Prozedur durchgeführt. Die Reihung der Hypothesen erfolgt nur nicht durch die P-Werte, sondern durch die $Q_i = P_i/W_i$ (W_i sind die Gewichte) Werte. Das Ziel ist durch (möglichst richtige) Priorinformationen wahre Alternativhypothesen (noch) eher abzulehnen als wahre Nullhypothesen als dies (nur) durch die P-Werte geschehen würde. Durch eine "richtige" Gewichtung einer wahren Alternative durch größere Gewichte als die einer wahren Nullhypothese wandern die wahren Alternativhypothesen bildlich gesprochen in der Reihung nach links. Das kann auch der Fall sein, wenn zufällig der P-Werte einer wahren Nullhypothese kleiner als ein P-Wert einer wahren Alternative ist. Daß in so einer Konstellation eine Umreihung stattfindet ist eines der Ziele der (a priori) Gewichtung.
Weiters werden gewichtete Analoga zu den Prozdeduren von Bonferroni und Bonferroni-Holm formuliert und bewiesen. Ein gewichtetes Analogon zum Satz von Benjamini-Hochberg unter der Einschränkung daß nicht nur die P-Werte der wahren Nullhypothesen sondern alle P-Werte voneinander unabhängig sind. Weiters werden asymptotische Resultate behandelt und auf die Macht der gewichteten Prozedur wird eingegangen.

Im Kapitel *Gewichtete Prozeduren (a posteriori Gewichtung)* werden die Gewichte nicht a priori bestimmt sondern aus den Daten gewonnen. Eine Prozedur im parametrischen normalverteilten Fall wird behandelt sowie nichtparametrische Prozeduren im Ein- und Zweistichprobenfall. Die Gewichte

werden im parametrischen Fall durch die Größe der Varianz und im nichtparametrischen Fall im Einstichprobenfall durch den Median der Absolutwerte und im Zweistichprobenfall durch den Interquantilabstand bestimmt. Diese so bestimmten Gewichte g_i werden noch durch Potenzierung mit $\eta \geq 0$ (die Gewichte sind also letztendlich g_i^η) parametrisiert.
Es wird sowohl die gewichtete Bonferroni-Holm (wir nennen sie auch Kropf Prozedur, wenn die Gewichte wie oben beschrieben durch die Daten bestimmt werden) als auch die gewichtete Benjamini-Hochberg Prozedur behandelt. Im Fall der gewichtete Bonferroni-Holm Prozedur wird bewiesen, daß diese den FWER zum Niveau α einhält. Im Falle der (gewichteten) Benjamini-Hochberg Prozedur werden Simulationen durchgeführt um Aussagen über die FDR zu erhalten. Bis auf eine Konstellation (Zweistichprobenfall, 9 bzw 14 Untersuchungseinheiten pro Gruppe, parametrischer Test und Gewichtsparameter $\eta = 2$) war die FDR jeweils ≤ 0.05. In der gerade erwähnten Konstellation war bei 7 Simulationen a 10000 Durchläufe (d.h. die FDR, die ja ein Erwartungswert ist, wurde als Mittelwert von 10000 unabhängigen Größen V/R bestimmt) die FDR ≥ 0.056, sowie bei einer Simulation a 100000 Durchläufe ebenso. Die Macht (die durchschnittliche Anzahl abgelehnter wahrer Alternativhypothesen) war bei bestmöglicher Wahl der Gewichte (bzw der η's) bei allen gewichteten Benjamini-Hochberg Prozeduren immer größer als bei der gewichteten Bonferroni-Holm Prozedur. Es gab allerdings stets (Gewichts) Konstellationen, bei denen die Macht der Bonferroni-Holm Prozedur größer war als die der Benjamini-Hochberg Prozedur. Außerdem war die Macht der Benjamini-Hochberg Prozedur sensibler gegenüber Veränderungen der Gewichte.
Weiters wurden die genannten Prozeduren auf den Golub-Datensatz der Genetik angewandt (der Golubdatensatz ist ein Microarraydatensatz der zwei verschiedene Typen von Leukämie untersucht).

Kapitel 1

Klassisches multiples Testen

Die Grundproblematik des multiplen Testens sowie die klassischen Methoden von Bonferroni und Holm sowie ein Beispiel, das diese Grundproblematik illustriert, werden vorgestellt.

1.1 Grundidee des multiplen Testens

1.1.1 Die grundlegende Anordnung

1. **Testproblem**: Es liegen m Nullhypothesen vor. Für jede der m Nullhypothesen ist eine Entscheidung zu treffen, ob diese angenommen oder abgelehnt werden soll.

2. **Fehler**: Es soll nun für jede Konstellation von wahren und falschen Nullhypothesen eine Aussage über die gemachten Fehler (ein Fehler wird begangen, wenn eine wahre Nullhypothese abgelehnt wird) getroffen werden.

3. **Definition 1 FWER** *Der Familienweite Fehler (Familiywise error rate: FWER) wird definiert als die Wahrscheinlichkeit mindestens eine wahre Nullhypothese abzulehnen.*

4. **Teststatistiken**: Es wird angenommen, daß die Entscheidung über die Annahme oder Ablehnung einer Hypothese mit Hilfe einer Teststatistik T vorgenommen wird mit welcher P-Werte P ausgerechnet werden können.

Wir können dann die P-Werte aufsteigend ordnen

$$P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(m)}.$$

Die entsprechende Ordnung der Hypothesen ist

$$H_{(1)} \leq H_{(2)} \leq \cdots \leq H_{(m)}.$$

Hypothesen, die links stehen werden eher den (wahren) Alternativen entsprechen, die Frage ist nun wo, bildlich gesprochen, in der obigen H-Reihe die Grenze gezogen werden soll (im Sinne, daß alle Hypothesen, die links dieser Grenze stehen abgelehnt werden).
Später werden wir sehen, daß die P-Werte nicht die einzige Möglichkeit sind, die Hypothesen zu ordnen.

1.1.2 Ein Beispiel

Das Grundproblem des multiplen Testens ist, dass, wenn man nur oft genug ein Zufallsexperiment ausführt mit Sicherheit seltene Ereignisse eintreten. Den Ereignissen im multiplen Testen entsprechen die Ergebnisse der einzelnen Tests. Bei einem einzelnen Test ist es selten, dass unter Gültigkeit der Nullhyphothese eine Ablehnung eben dieser eintrifft (so sind Tests ja konstruiert).
Bei oftmaliger Durchführung solcher Tests wird es wahrscheinlich, daß man ein falsches signifikantes Ereignis erhält.
Wir haben

$$\begin{aligned} P(\text{mindestens eine Hypothese wird abgelehnt}) &= \\ 1 - P(\text{Alle Hypothesen werden angenommen}). & \end{aligned}$$

Wenn die Tests unabhängig sind, haben wir (A_i sei die Annahmemenge der Hypothese i und die T_i die Teststatistik für das i-te Testproblem)

$$\begin{aligned} P(\text{Alle Hypothesen werden angenommen}) &= \\ P(\{T_1 \in A_1\} \cap \{T_2 \in A_2\} \cap \cdots \cap \{T_m \in A_m\}) &= \\ P(\{T_1 \in A_1\}) \cdot P(\{T_2 \in A_2\}) \cdots P(\{T_m \in A_m\}) &= \\ (1-\alpha)^n. & \end{aligned}$$

	$\frac{\#\text{Abg Null}}{\#\text{Tests}}$	$\frac{\#>1\text{Abl.}}{\#\text{Exp}}$
$(\alpha = 0.01,\ m = 40)$	0.0095	0.34
$(\alpha = 0.01,\ m = 20)$	0.0085	0.15
$(\alpha = 0.05,\ m = 40)$	0.0515	0.88
$(\alpha = 0.05,\ m = 20)$	0.0565	0.66

Tabelle 1.1:

und $(1-\alpha)^n$ geht gegen 0, für $n \to \infty$.
Wir haben also asymptotisch

$$P(\text{wenigstens ein Test wird abgelehnt}) \to 1.$$

Wir führen nun mit dem Computer folgendes Experiment durch.
Gegeben sind zwei Gruppen. In jeder Gruppe sind n Untersuchungseinheiten und m Merkmale wurden jeweils unabhängig voneinander erhoben. Jedes Merkmal ist $N(0,1)$ verteilt, die Gruppen unterscheiden sich also nicht voneinander.
Wir haben nun $m = 40$ gewählt, und einmal $n = 20$ mit Signifikanzniveau $\alpha = 0.01$ und $\alpha = 0.05$ gwählt und das Experiment auch mit $m = 20$ Untersuchungseinheiten durchgeführt.
Insgesamt wurden 100 Experimente durchgeführt.
Die Simulation wurde mit dem im letzten Kapitel angeführten R-Programm durchgeführt.
Der Programmaufruf war

```
erg <- Kapitel1(anzExp=100,m=40,n=40,alfa=0.01)
sum(erg)/(anzExp*m)
(length(erg[erg>0]))/anzExp
```

Die Ergebnisse sind in Tabelle 1.1, die Bezeichnungen in Tabelle 1.2 angegeben.

#Abg Null	Anzahl der abgelehnten Nullhypothesen insgesamt
#Tests	Anzahl der Tests insgesamt
#> 1Abl.	Anzahl der Experimente mit mindestens einer abgelehnten Nullhypothese
#Exp	Anzahl der durchgeführten Experimente ($= anzExp \cdot m$)

Tabelle 1.2:

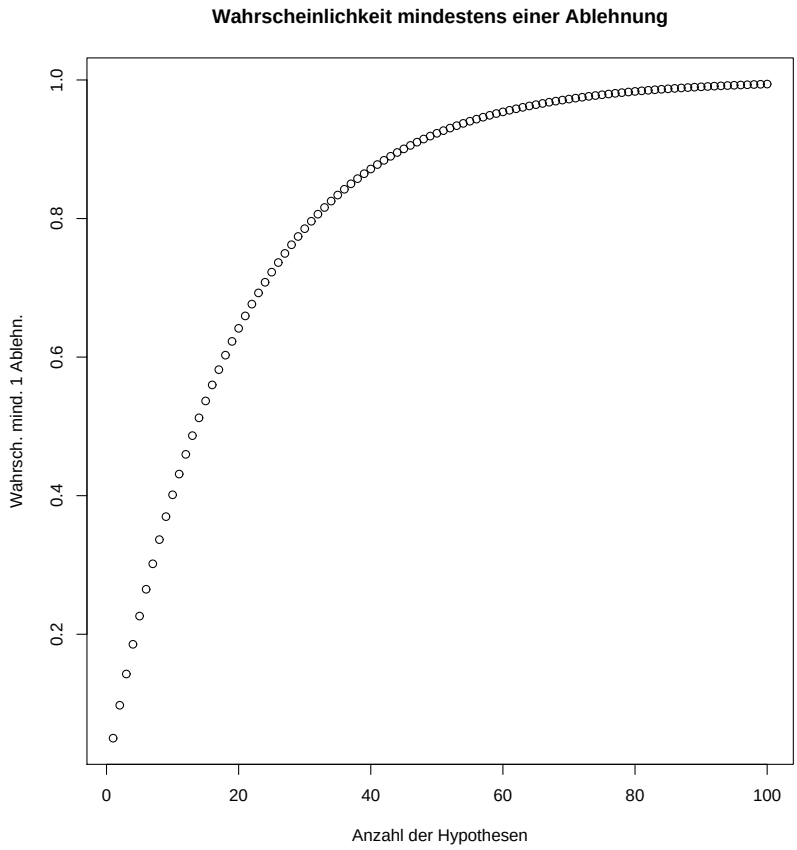

Abbildung 1.1: *Wahrscheinlichkeit der Anzahl mindestens eine wahre Nullhypothese abzulehnen bei einer Irrtumswahrscheinlichkeit von* $\alpha = 0.05$ *als Funktion der Anzahl der Hypothesen, wenn alle Nullhypothesen wahre Nullhypohtesen sind.*

In der ersten Spalte der Tabelle 1.1 sieht man das ungefähr erwartete Ergebnis. In dieser steht eine Annäherung an die Irrtumswahrscheinlichkeiten. In der zweiten Spalte der Tabelle 1.1 steht eine Approximation an die Wahrscheinlichkeit, mindestens eine wahre Nullhypthese irrtümlich abzulehnen.
In Abbildung 1.1 ist die Ablehnwahrscheinlickeit mindestens einer Hypothese exakt ausgerechnet.
Man sieht, dass dies bei einer Irrtumswahrscheinlichkeit von $\alpha = 0.05$ schon sehr hoch ist. Eine Möglichkeit diesem Problem zu begegnen (später werden wir noch andere kennenlernen) ist die Irrtumswahrscheinlichkeiten der einzelnen Tests zu verringern. Dies werden wir in den nächsten Abschnitten behandeln.

1.2 Die Bonferroni Methode

Definition 2 *Die* **Bonferroni Methode**
Verwerfe alle Hypothesen, für deren P-Werte P gilt:

$$P \leq \frac{\alpha}{m}.$$

Wir bringen jetzt den Beweis, dass die Bonferroni Methode den FWER zum Niveau α einhält, d.h. dass die Wahrscheinlichkeit, dass eine oder mehrere wahre Nullhypothesen abgelehnt werden kleiner oder gleich α ist.
Es gebe also m Hypothesen. m_0 von diesen seien wahre Nullhypothesen. Es gelte

$$A_i = \{\omega : H_i \text{ wird abgelehnt}\}$$

(ω ist das generische Element des Wahrscheinlichkeitsraumes). Weiters haben wir ($|M_0| = m_0$)

$$M_0 = \{j : H_j \text{ ist eine wahre Nullhypothese}\}.$$

Es sei dann

$$A = \bigcup_{i \in M_0} A_i.$$

A ist das Ereignis, mindestens eine wahre Nullhypothes abzulehnen. Es gilt dann

$$P(A) = P(\bigcup_{i \in M_0} A_i) \leq \sum_{i \in M_0} P(A_i) \leq \sum_{i=1}^{m_0} \frac{\alpha}{m} = \frac{\alpha}{m} m_0 \leq \alpha.$$

QED

1.3 Die Bonferroni-Holm Methode

Definition 3 *Die* **Holm Methode**
Man ordne die P-Werte:

$$P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(m)}.$$

Es sei R die Indexmenge der abgelehnten Hypothesen.
Die Prozedur wird nun folgendermaßen durchgeführt.

1. *Initialisiere $R = R_0 = \emptyset$.*
2. *Wenn $P_{(1)} \leq \frac{\alpha}{m}$ gilt, setze $R = R_1 = R_0 \cup \{(1)\}$, wobei (1) der Index des kleinsten P-Wertes ist*
   sonst stoppe.
3. *Wenn $P_{(2)} \leq \frac{\alpha}{m-1}$ dann setze $R = R_2 = R_1 \cup \{(2)\}$, wobei (2) der Index des zweit kleinsten P-Wertes ist*
   sonst stoppe.
4. *u.s.w*
5. *Wenn $P_{(m)} \leq \frac{\alpha}{m-(m-1)}$ dann setze $R = R_m = R_{m-1} \cup \{(m)\}$, wobei (m) der Index des m-kleinsten (also größten) P-Wertes ist*
   sonst stoppe.

Wir zeigen nun, daß die Bonferroni-Holm Methode den FWER zum Niveau α einhält [20].
R ist die Indexmenge der Hypothesen, die abgelehnt werden, I die Indexmenge der wahren Nullhypothesen und l die Anzahl der Elemente in I.
Im ersten Schritt zeigen wir (ω ist das generische Element des Ereignisraumes, die nächste Beziehung ist so zu lesen, daß das auf der linke Seite beschriebene Ereignis das auf der rechten Seite impliziert, die rechte Seite bedeutet, daß keine wahre Nullhypothese abgelehnt wird)

$$\{\omega : P_i \geq \alpha/l, \text{ für alle } i \in I\} \subseteq \{\omega : A \cap I = \emptyset\}.$$

Dies folgt aus der Konstruktion der Holm-Prozedur. Tritt nämlich das Ereignis auf der linken Seite ein, so kann eine wahre Nullhypothese in der Prozedur frühestens in dem Schritt abgelehnt werden in dem der Schwellenwert gleich

$\alpha/(l-1)$ ist. Dies ist der $(m-l+2)$. Schritt. Um in der Prozedur bis zu diesem Schritt zu gelangen, müßten aber $m-l+1$ wahre Alternativen schon abgelehnt werden, was nicht möglich ist, da es l wahre Nullhypothesen und damit $m-l$ wahre Alternativen gibt.
Wir bilden nun die Komplementärbeziehung zu obiger ($A \subseteq B \Longrightarrow \bar{B} \subseteq \bar{A}$). Diese ist

$$\{\omega : A \cap I \neq \emptyset\} \subseteq \{\omega : P_i < \alpha/l, \text{ für mindestens ein } i \in I\}.$$

Die Wahrscheinlichkeit des Ereignises auf der linken Seite ist der FWER, den wir durch Berechnen der Wahrscheinlichkeit der rechten Seite abschätzen können.

$$\begin{aligned} P(\{\omega : A \cap I \neq \emptyset\}) &\leq P(\bigcup_{i \in I} \{\omega : P_i < \alpha/l\}) \\ &\leq \sum_{i \in I} P(\omega : P_i < \alpha/l) \\ &\leq l \cdot \alpha/l = \alpha. \end{aligned}$$

QED

1.4 Literaturhinweise

Die Methoden von Bonferroni und Holm findet man an sehr vielen Orten, z.B in [19], oder in der Originalarbeit von Holm [11] oder in [20].
Viel über Multiples Testen findet man in den Standardwerken [9] und [18] oder auch im (elementarer gehaltenen) Werk [13].
Im Internet gibt es massenhaft Einführungen in das multiple Testen, es sei nur [8] angeführt, die mich zu dem Einführungsbeispiel inspiriert hat.

Kapitel 2

Anteil falscher Ablehnungen (FDR)

In diesem Kapitel wird die FalseDiscoveryRate (FDR), d.h. der Erwartungswert des Quotienten der *Anzahl abgelehnte wahre Nullhypothesen* durch die *Anzahl alle abgelehnten Hypothesen* vorgestellt.
Weiters wird ein Schätzwert von Storey sowie eine verteilungsfreie Prozedur vorgestellt.

2.1 Die Benjamini-Hochberg Prozedur

2.1.1 Das Modell

Wir haben das folgende Szenario (siehe die Tabelle 2.1). In der Tabelle 2.1 ist

	Anz. angenommen	Anz. abgelehnt	Insgesamt
Wahre Nullhypothese	U	V	m_0
Falsche Nullhypothese	T	S	$m - m_0$
Summe	$m - R$	R	m

Tabelle 2.1:

m, die Anzahl der getesteten Hypothesen und R, die Anzahl der abgelehnten Hypothesen beobachtbar. Die restlichen Grössen U, V, S, T und m_0 sind nicht beobachtbar.

2.1.2 FDR

Wir definieren nun die falsche Erkennungsrate , die FalseDiscoveryRate (FDR) als ($1_{\{.\}}$ bezeichnet die Indikatorfunktion)

Definition 4

$$FDR \ = E(V/R \cdot 1_{\{R>0\}}).$$

Wir bezeichnen die False Discovery Proportion (FDP) V/R mit Q.

$$Q = V/R \cdot 1_{\{R>0\}}.$$

2.1.3 Die Benjamini-Hochberg Prozedur

Wir ordnen wieder die P-Werte

$$P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(m)}.$$

und entsprechend die Hypothesen H_i.
Die Benjamini-Hochberg (BH) Prozedur ist nun folgendermaßen definiert

Definition 5 Benjamini-Hochberg *Prozedur*

$$\textit{Es sei } k \textit{ das grösste } i \textit{ für das } P_{(i)} \leq \frac{i}{m}\alpha \textit{ gilt.}$$

Folgende Hypothesen werden dann verworfen

$$H_{(i)} \textit{ für alle } (i) \leq (k).$$

Es werden also $H_{(i)}, H_{(2)}, \ldots, H_{(k)}$ *abgelehnt.*

Bemerkungen:

1. Sind alle Nullhypothesen wahr, so gilt

   $$E(V/R \cdot 1_{\{R>0\}}) = 1 \cdot P(V > 0) + 0 \cdot P(V = 0)$$

   und das ist der FWER (siehe Definition 1)

2. Es gilt definitionsgemäß immer $V/R \cdot 1_{\{R>0\}} \leq 1$ und damit

$$1_{\{V\geq 1\}} \geq V/R \cdot 1_{\{R>0\}}$$

und damit (Ungleichung umdrehen und Erwartungswert nehmen)

$$\text{FDR } = E(V/R \cdot 1_{\{R>0\}}) \leq E(1_{\{V\geq 1\}}) = \text{ FWER.}$$

Wir zitieren eine Proposition aus der Theorie der Ordnungsstatistiken:

Proposition 1 *Es seien $X_1 \ldots X_n$ unabhängig im Intervall $(0,1)$ gleichverteilt.*
Dann ist die Dichte $f(x)$ der n ten Ordnungsstatistik $X_{(n)}$ gleich (F sei die Verteilungsfunktion von f)

$$f(x) = n[F(x)]^{n-1}$$

für $0 \leq x \leq 1$.

Für den Beweis siehe [23].
Noch eine Proposition

Proposition 2 *Die P-Werte unter der Nullhypothese sind Gleichverteilt (unter gewissen Regularitätsvorraussetzungen an die Verteilungsfunktionen)*

Beweis:
Wir nehmen unter Einschränkung der Allgemeinheit an (der allgemeine Beweis geht aber sinngemäß), daß der P-Wert durch $P(T \geq u)$ gegeben ist. Sei nun $P(T \geq u) = \alpha$ (daß es zu jedem α ein u mit dieser Beziehung gibt, folgt aus der ¨genügenden¨ Glattheit der Verteilungsfunktionen), es gilt dann (beachte die Doppelbedeutung von P, siehe [21] und [22]) $P(P \leq \alpha) = P(T \geq u) = \alpha$ (wegen der Monotonie von $f(u) = P(T \geq u)$ in u). **QED**
Für den Beweis von Theorem 2.1.3 brauchen wir noch folgende Proposition

Proposition 3 *Es seien X_1 und X_2 auf $[0,1]$ gleichverteilt. Dann gilt $X_{(1)}/X_{(2)}$ ist ebenfalls auf $[0,1]$ gleichverteilt.*

Beweis:
Wir zitieren ein Ergebnis über die Dichte des Quotienten zweier Zufallsvariablen (siehe [10] Seite 22).
Wenn $X = (X_1, X_2)$ die Dichte f hat, so hat X_1/X_2 die Dichte

$$\int f(xy, y) \cdot |y| dy.$$

Auf [23] übernehmen wir wieder die Tatsache, daß die Dichte des kleinsten und größten Wertes durch

$$g(y_1, y_n) = 2 \cdot 1_{\{y_1 < y_2\}}$$

gegeben ist. Für die verlangte Dichte h von $X_{(1)}/X_{(2)}$ haben wir dann

$$h(x) = \int_0^1 g(xy, y) \cdot |y| dy = 2 \cdot \int_0^1 y dy = 2 \cdot \frac{y^2}{2}|_0^1 = 1.$$

QED

Proposition 4 *Alle m Hypothesen seien wahre Nullhypothesen und die P-Werte seien unabhängig.*
Dann gilt:

$$FDR \ = E(V/R \cdot 1_{\{R>0\}}) \leq \alpha.$$

Beweis: (der Beweis ist aus [**?**] adaptiert)
Wir könnten den Satz von Genovese im 3. Kapitel zitieren, siehe Abschnitt 3.4.2 (mit den Gewichten identisch 1, ziehen es aber vor keine Anleihen aus anderen Kapiteln zu nehmen, siehe auch [21])
Wir vereinbaren die Bezeichnungen

$$q_k = \frac{\alpha k}{m}$$

und

$$\bar{P}_i = \{P_j : j \neq i\}.$$

Weiters definieren wir die Menge $R_{k,i}$

$$R_{k,i} = \{k-1 \ P_j \in \bar{P}_i \text{ sind } \leq q_k \text{ und die anderen } P_{(u)} \in \bar{P}_i \text{ sind } > \frac{\alpha u}{m}\}.$$

Wir haben dann

$$\begin{aligned} E(V/R) \cdot 1_{\{R>0\}} &= E(1 \cdot 1_{\{R>0\}}) = \sum_{k=1}^{m} \frac{k}{k} P(R = k) \\ &= \sum_{k=1}^{m} \sum_{i=1}^{m} \frac{1}{k} P(P_i \leq \frac{k\alpha}{m} \cap R_{k,i}) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{m}\sum_{i=1}^{m}\frac{1}{k}P(P_i \le \frac{k\alpha}{m})\cdot P(R_{k,i}) \\
&= \sum_{k=1}^{m}\sum_{i=1}^{m}\frac{1}{k}\cdot\frac{k\alpha}{m}\cdot P(R_{k,i}) \\
&= \frac{\alpha}{m}\sum_{i=1}^{m}\sum_{k=1}^{m}P(R_{k,i}) \\
&\le \frac{\alpha}{m}\sum_{k=1}^{m}1 = \alpha.
\end{aligned}
$$

Den Schritt in die zweite Zeile sieht man so:
Wir halten $R = k$ fest. Es müssen dann k P-Werte kleiner oder gleich q_k sein und für die $m - k$ restlichen P-Werte $P_{(j)}$ muß jweils

$$P_{(j)} > \frac{\alpha j}{m}$$

gelten.
Ist ω ein generische Elementarereignis des Ereignisses $\{P_{i_1}, P_{i_2}, \ldots, P_{i_k} \le q_k,$ die restlichen $P_{(u)} > \frac{\alpha u}{m}\}$, so liegt ω genau in jeder der Mengen $\{P_{i_j} \le \frac{k\alpha}{m} \cap R_{k,i_j}\}, j = 1, \ldots, k$. Jedes ω kommt also genau k mal in einer der Mengen $\{P_i \le \frac{k\alpha}{m} \cap R_{k,i}\}$ vor.
Im vorletzten Schritt ist es für die Gültigkeit der Proposition nicht notwendig, sich Gedanken zu machen, ob $\sum_{k=1}^{m} P(R_{k,i}) = 1$ oder nur ≤ 1 gilt, da letzteres reicht. **QED**

Theorem 1 *Die P-Werte der wahren Nullhypothesen seien unabhängig.*
Es gilt dann:
Die BH-Prozedor hält die FDR zum Niveau α ein. [1]

Beweis (wir erinnern daran, daß die Verteilungsfunktionen dergestalt sind, daß die P-Werte gleichverteilt sind):
Zuerst beweisen wir ein Lemma:
m ist die Anzahl aller Hypothesen, m_0 die Anzahl der wahren Nullhypothesen und $m_1 = m - m_0$ die Anzahl der wahren Alternativhypothesen.
Wir denken uns die Hypothesen so numeriert, daß

- $P_1, P_2, \ldots, P_{m_0}$ die P-Werte der wahren Nullhypothesen sind

- $P_{m_0+1}, P_{m_0+2}, \ldots, P_m$ die P-Werte der wahren Alternativhypothesen sind.

Lemma 1 *Die P-Werte der m_0 Nullhypothesen werden wie in Theorem 2.1.3 als unabhängig angenommen. Die BH-Prozedur erfüllt dann folgende Beziehung (d.h. man führt die BH-Prozedur durch und errechnet $Q = V/R \cdot 1_{\{R>0\}}$)*

$$E(Q|P_{m_0+1} = p_1, \ldots, P_m = p_{m_1}) \leq \frac{m_0}{m}\alpha.$$

Beweis des Lemmas:
Für den Beweis nehmen wir an, daß es mindestens eine wahre Alternative gibt (den Fall, daß alle Hypothesen wahre Nullhypothesen sind, haben wir in Proposition 4 schon bewiesen).
Ist $m_0 = 0$ dann ist Q definitionsgemäß gleich 0 und damit die (bedingte) Verteilung auch gleich 0 und damit die Behaupung trivialerweise wahr.
Wir beweisen das Lemma durch Induktion nach m, für beliegiges $m_0 > 0$.
Induktionsanfang:
Es ist $m = 1$.

1. $m_0 = 0$. Diesen Fall haben wir schon erledigt.

2. $m_0 = 1$: $\Longrightarrow \frac{V}{R}$ ist Bernoulli 1 oder 0, je nachdem, ob H_1 abgelehnt wird oder nicht.
   $E(V/R \cdot 1_{\{R>0\}}|\cdot)$ ist dann der unbedingte Erwartungswert. Und laut Testkonstruktion wird H_1 abgelehnt, wenn $P_1 \leq \frac{1}{1}\alpha$ gilt.

Induktionsschritt: Die Behauptung aus dem Lemma gelte für m Hypothesen.
Wir bezeichnen nun die m_0 P-Werte der wahren Nullhypothesen durch $P_i^{'}$, wobei $i = 1, 2, \ldots, m_0$ gilt. Den größten Wert bezeichnen wir mit der Standardbezeichnung aus der Theorie der Ordnungsstatistiken mit $P_{(m_0)}^{'}$. Die P-Werte der wahren Nullhypothesen seien gleichverteilt auf $[(0, 1)]$.
Wir bezeichnen weiters die P-Werte der wahren Alternativen der Reihe nach mit

$$p_1 \leq p_2 \leq \ldots \leq p_{m_1}.$$

Sei j_0 das größte $j \leq m_1$ sodaß

$$p_j \leq \frac{m_0 + j}{m + 1}\alpha$$

und sei

$$p^{''} = \frac{m_0 + j_0}{m+1}\alpha.$$

Gibt es kein so ein j_0, so sei $j_0 = 0$ und $p^{''} = 0$ (eines der gleich folgenden Integrale fällt dann weg).
Wir schreiben nun den (bedingten) Erwartungswertes sinngemäß (die stetige Form) nach der Formel $(P(S) = P(S|A)P(A) + P(S|\bar{A})P(\bar{A}))$, bedingt wird nach dem größten P-Wert der wahren Nullhypothesen $P^{'}_{(m_0)} = p$. Wir erhalten dann

$$\begin{aligned} E(Q|P_{m_0+1} = p_1, \ldots, P_m = p_{p_{m_1}}) &= \int_0^{p^{''}} E(Q|P^{'}_{m_0} = p, P_{m_0+1} = p_1, . & (2.1)\\ & \quad P_m = p_{m_1}) f_{P^{'}_{m_0}}(p)dp & (2.2)\\ &+ \int_{p^{''}}^1 E(Q|P^{'}_{m_0} = p, P_{m_0+1} = p_1, . & (2.3)\\ & \quad P_m = p_{m_1}) f_{P^{'}_{m_0}}(p)dp & (2.4) \end{aligned}$$

wobei $f_{P^{'}_{m_0}} = m_0 p^{m_0-1}$ die Dichte der Verteilung des größten P-Wertes der wahren Nullhypothesen ist.
Wir behandeln nun die beiden Integrale.

1. Integral

   Es gilt $P^{'}_{m_0} \leq p^{''}$ und daher werden alle wahren Nullhypothesen abgelehnt. Es gilt dann

   $$Q = \frac{m_0}{m_0 + j_0}.$$

   Diese Beziehung gilt, weil wegen der Konstruktion von j_0 zumindest die Hypothesen abgelehnt werden, die den P-Werten der wahren Alternativen $p_1, p_2, \ldots, p_{j_0}$ entsprechen.
   Es kann aber auch keine weitere Hypothese abgelehnt werden. Alle Nullhypothesen werden schon abgelehnt und für p_{j_0+1} gilt wegen der Konstruktion von j_0

   $$p_{j_0+1} > \frac{m_0 + j_0 + 1}{m+1} \cdot \alpha.$$

   Der bedingte Erwartungswert ist also konstant und kann vor das Integral gebracht werden, es bleibt nur mehr eine einfache Integration eines Potenzterms. Wir erhalten für das Integral also

   $$\frac{m_0}{m_0 + j_0}(p^{''})^{m_0} = \frac{m_0}{m_0 + j_0}\frac{m_0 + j_0}{m+1}\alpha(p^{''})^{m_0-1} = \frac{m_0}{m+1}\alpha(p^{''})^{m_0-1}.$$

2. Integral

Wir betrachten hier jedes Intervall der P-Werte der wahren Alternativen, in dem der größte der P-Werte der wahren Nullhpyothesen, wir haben ihn mit p bezeichnet vorkommt extra.
Das erste ist ein abgeschnittenes Intervall, nämlich

$$p_{j_0} \leq p'' < p < p_{j_0+1}$$

die "allgemeinen" Intervalle sind dann

$$p_{j_0} < p_j \leq p < p_{j+1}.$$

Wegen der Definition von j_0 kann kein p_j die BH-Bedingung erfüllen (j_0 ist ja schon der maximale Index), p kann aber auch kein Schwellenwert sein, da ja $p > p_j$ gilt, und dann auch die Hypothese mit P-Wert p_j abgelehnt werden müsste.
Eine Hypothese $H_{(i)}$ (das sind die gemäß den nach der Größe geordneten P-Werten mitgeordneten Hypothesen) kann nur abgelehnt werden, wenn es einen Index k gibt, der die BH-Bedingung erfüllt, also

$$p_{(k)} \leq \frac{k}{m+1}\alpha$$

und $(i) \leq (k)$ gilt. Da die p_j der Alternativen von vorher nach der Größe geordnet Werte waren, kann man die $m_1 - j + 1$ größten dieser Indizi ausschließen (wegen der Konstruktion von j_0), also muss für k gelten

$$i \leq k \leq m_0 + m_1 - (m_1 - j + 1) = m_0 + j - 1.$$

In der vorletzten Beziehung dividieren wir noch durch p und erweitern den Bruch mit $m_0 + j - 1$ und erhalten somit

$$\frac{p_{(k)}}{p} \leq \frac{k}{m_0 + j - 1}\frac{m_0 + j - 1}{(m+1)p}\alpha.$$

Da ja auf p bedingt wurde, sind nun die $m_0 - 1$ Werte $\frac{P_i'}{p}$ wieder unabhängig gleichverteil auf dem Intervall $(0, 1)$ (Siehe Proposition 3). Wir fassen nun diese Brüche wieder als P-Werte von wahren Nullhypothesen auf (sozusagen jede Größe, die gleichverteilt auf $(0, 1)$ ist, kann man als P-Wert auffassen).

Wir verwenden nun die Induktionsannahme, indem wir bemerken, daß $m_0 + j - 1 \leq m$ gilt (der 1er wars) und den Ausdruck

$$\frac{m_0 + j - 1}{(m+1)p}\alpha$$

die Rolle des (neuen) α spielen lassen (außerdem haben wir schon gezeigt, daß die $p_{(k)}$ und die $\frac{p_{(k)}}{p}$ die gleichen Hypothesen ablehnen). Wir haben dann

$$\begin{aligned} E(Q|P'_{(m_0)} = p, P_{m_0+1} = p_1, \ldots, P_m = p_{m_1}) &\leq \frac{m_0 - 1}{m_0 + j - 1}\frac{m_0 + j - 1}{(m+1)p}\alpha \\ &= \frac{m_0 - 1}{(m+1)p}\alpha. \end{aligned}$$

Jetzt integrieren wir das zweite Integral als Summe der Integrale jeweils über die Grenezn (p_j, p_{j+1}) sowie das erste RumpfÏntervall und setzen die eben gewonnen Abschätzung ein und erhalten (die einzelnen Integrale nicht hinschreibend, diese Aufschlitterung braucht man nur um die Abschätzung einzusetzen und einen konstanten Term herauszuheben)

$$\int_{p''}^{1} E(Q|P'_{(m_0)} = p, P_{m_0+1} = p_1, \ldots, P_m = p_{m_1}) f_{P'}(p)dp \leq \int_{p''}^{1} \frac{m_0 - 1}{(m+1)p}\alpha m_0 p^{m_0-1}dp$$

nach Herausheben bleibt dann

$$= \frac{m_0}{m+1}\alpha \int_{p''}^{1}(m_0 - 1)p^{m_0-2} = \frac{m_0}{m+1}\alpha(1 - p''^{m_0-1}).$$

Nun addieren wir die Abschätzungen der beiden Integrale, der Term p''^{m_0-1} hebt sich freundlicherweise weg und es bleibt nur mehr die im Lemma geforderte Abschätzung

$$\frac{m_0}{m+1}\alpha.$$

QED

Jetzt zeigen wir aus dem Lemma, daß die BH-Prozedur den FDR-Fehler α

einhält. Wir verwenden dazu die Formel von der totalen Wahscheinlichkeit ($\bigcup_{i=1}^{n} A_i = \Omega$)

$$E(T) = \sum_{i=1}^{n} P(A_i)E(T|A_i)$$

und im stetigen Fall sinngemäß als Integral. Durch das Lemma wissen wir, daß

$$E(Q|P_{m_0+1} = p_1, \ldots, P_m = p_{m_1}) \leq \frac{m_0}{m}\alpha$$

gilt. Wir summieren nun über alle möglichen Tupel ($p_1, \ldots, p_m$, bzw im stetigen Fall integrieren wir und erhalten dann , es gilt ja $\sum P(A_i) = 1$.

$$E(Q) \leq \frac{m_0}{m}\alpha \leq \alpha.$$

QED

2.2 Die Benjamini-Liu Prozedur

Wir erwaehnen noch kurz eine Prozedur, die weniger Annahmen an die gemeinsame Verteilung der P-Werte stellt als die BH-Prozedur. Die Benjamini-Liu (BL) Prozedur verlangt nicht mehr wie die BH-Prozedur, daß die P-Werte unabhängig sind. [3]
Wir ordnen die P-Werte wieder nach der Groesse:

$$P_{(1)} \leq P_{(2)} \leq \ldots P_{(m)}$$

und bezeichnen die dadurch umgeordneten Hypothesen mit $H_{(i)}$. Wir definieren nun m kritische Werte durch

$$d_i = \min(1, \frac{m}{(m-i+1)^2}\alpha) \text{ fuer } 1 \leq i \leq m.$$

Da der Nenner mit wachsenden i kleiner wird, wird der Ausdruck im Minimum groesser, bleibt aber insgesamt immer kleiner oder gleich 1, also gilt

$$0 \leq d_1 \leq d_2 \ldots \leq d_m \leq 1.$$

Die BL-Prozedur geht nun folgendermassen:

- Gilt $P_{(1)} \leq d_1$ dann lehne die Hypothese H_1 ab, sonst höre auf.
- Gilt $P_{(2)} \leq d_2$ dann lehne die Hypothese H_2 ab, sonst höre auf.
- Und so weiter, das letzte Mal wo $P_{(i)} \leq d_i$ gilt, ist das letzte Mal wo eine Hypothese abgelehnt wird.

Wir beweisen nun, dass diese Prozedur die FDR zum Niveau α einhält.
Zuerst betrachten wir die Grenzfälle.
Sei $m_0 = 0$ Klarerweise ist dann Q gleich 0.
Der andere Grenzfall ist $m_0 = m$. Es gilt dann $V = R$ und damit haben wir.

$$\begin{aligned} E(Q) = E(I_{V>0}) = P(V > 0) &= P(P_{(1)} \leq d_1) \\ &\leq \sum_{i=1}^{m} P(P_i \leq d_1) \\ &= m d_1 = m \frac{m}{m^2} \alpha = \alpha. \end{aligned}$$

Die erste Gleichung oben folgt daraus, dass $V = R$ und damit $\frac{V}{R} = 1$, und $E(Q) = 0P(V = 0) + 1P(V > 0)$ gilt. Die dritte Gleichung folgt daraus, dass mindestens eine Hypothese genau dann abgelehnt wird, wenn zumindest die erste Hypothese abgelehnt wird. Die Ungleichung folgt aus $\{P_{(1)} \leq d_1\} \subseteq \bigcup_i \{P_i \leq d_1\}$. Die drittletzte Gleichung folgt daraus, dass die P-Werte unter der Nullhypothese gleichverteilt sind und in der vorletzten Gleichung wird die Definition von d_1 eingesetzt.
Wir kommen nun zu den allgemeinen Fällen, d.h. $1 \leq m_0 \leq m - 1$.
Zuerst ein paar Bezeichnungen:

- $m_1 = m - m_0$, und das ist > 0, wir sind ja im allgemeinen Fall.
- $P_1^{'}, \ldots, P_{m_0}^{'}$ bezeichnen die P-Werte der m_0 wahren Nullhypothesen
- $p_1, \ldots, p_{m_1}$ seien die P-Werte der m_1 wahren Alternativen
- $Q_e(P_1^{'}, \ldots, P_{m_1}^{'}) = E(Q|P_1^{'}, \ldots, P_{m_1}^{'})$.

Wir führen noch eine Größe G ein. Wir ordnen dazu die P-Werte der wahren Alternativen

$$p_1 \leq p_2 \leq \ldots \leq p_{(m_1)}.$$

Wir vergleichen nun nur die geordneten p_i Werte der wahren Alternativen der Reihe nach mit den d_i (Wir lassen also die wahren Nullhypothesen einmal

aus). Wir vergleichen nun die p_i so lange der Reihe nach in Aufsteigender Form mit den d_i

- Gilt $p_i \leq d_i$ für alle $1 \leq i \leq m_1$, dann sei $G = m_1$
- Es gilt $p_1 > d_1$, dann sei $G = 0$
- (der allgemeine Fall) wenn das erste Mal $p_{i+1} > d_{i+1}$ gilt, sei $G = i$

Wir haben nun folgende Beziehungen, unten erklären wir dann die einzelnen Schritte

$$
\begin{aligned}
Q_e(p_1, \ldots, p_{m_1}) &= E(\frac{V}{R} I_{V>0} | p_1, \ldots, p_{m_1}) && (2.5)\\
&\leq E(\frac{V}{V+G} I_{V>0} | p_1, \ldots, p_{m_1}) && (2.6)\\
&\leq \frac{m_0}{G+m_0} E(I_{V>0} | p_1, \ldots, p_{m_1}) && (2.7)\\
&\leq \frac{m_0}{G+m_0} P(\min(P_1^{'}, \ldots, P_{m_0}^{'}) \leq d_{G+1}) && (2.8)\\
&\leq \frac{m_0}{G+m_0} \sum_{i=1}^{m_0} P(P_i^{'} \leq d_{G+1}) && (2.9)\\
&\leq \frac{m_0}{G+m_0} (m-G) d_{G+1} && (2.10)\\
&= \frac{m_0(m-G)}{G+m_0} \min(1, \frac{m}{(m-G)^2}) \cdot \alpha && (2.11)\\
&\leq \frac{m_0 m}{(G+m_0)(m-G)} \cdot \alpha && (2.12)\\
&\leq \alpha. && (2.13)
\end{aligned}
$$

Die erste Ungleichung folgt wegen $R \geq G + V$, dies folgt daraus, daß wenn auch die P-Werte der wahren Nullhypothesen betrachtet werden die P-Werte bildlich gesprochen nach rechts verschoben werden und eher noch weitere wahre Alternativen abgelehnt werden. Die nächste Ungleichung folgt aus $\frac{V}{G+V} \leq \frac{m_0}{G+m_0}$ (d.h. die Zufallsvariable $\frac{V}{G+V}$ wird durch eine Konstante (bedingt auf $p_1, p_2, \ldots, p_{m_1}$) nach oben abgeschätzt und diese kann dann aus dem (bedingten) Erwartungswert herausgehoben werden), dies sieht man indem man einfach ausmultipliziert, es bleibt zu zeigen $GV \leq Gm_0$ und $V \leq m_0$ gilt klarerweise (V ist ja die Anzahl der abgelehnten wahren Nullhypothesen).

Die nächste Ungleichung sieht man so: Der betrachtete bedingte Erwartungswerte ist ja gleich der (bedingten) Wahrscheinlichkeit der Ereignisses $\{I_{V>0}\}$. Die behauptete Beziehung sieht man nun so: Damit wenigstens eine wahre Nullhypothese abgelehnt werden kann (also $V > 0$ gilt) kann der kleinste P-Wert der wahren Nullhypothesen P^* entweder kleiner als eines der $P'_{(i)}$ mit $i \leq G$, und damit sicher kleiner als d_{G+1} sein, oder aber größer als alle $P'_{(i)}$ mit $i \leq G$ sein, und damit nach Definition der Prozedur muss der Index des kleinsten der P^* gleich $G+1$ und damit für dieses $P^* \leq d_{G+1}$ gelten.
Die nächste Ungleichung folgt wieder aus '(die kleinste Wahrscheinlichkeit kleiner oder gleich) ist enthalten in (ist in irgendeiner) '. Die darauffolgende Ungleichung folgt wieder aus der Tatsache dass für die P-Werte der wahren Nullhypthesen $P(P \leq a) \leq a$ gilt.
Für die letzte Ungleichung muss nun gelten

$$\frac{m_0 m}{(G+m_0)(m-G)} \leq 1.$$

Multiplizieren und auf beiden Seiten $m_0 m$ subtrahieren ergibt

$$\begin{aligned} 0 &\leq G(m-m_0) - G^2 && (2.14)\\ G^2 &\leq G(m-m_0) && (2.15)\\ G &\leq m - m_0 && (2.16)\\ m_0 + G &\leq m. && (2.17) \end{aligned}$$

und die letzte Gleichung gilt ja definitionsgemäß, da ja $G \leq m_1$ sein muss und $m_0 + m_1 = m$ gilt.
Jetzt sind alle bedingten Erwartungswerte kleiner oder gleich q und (sinngemäß) wegen der Formel der totalen Wahrscheinlichkeit auch der unbedingte Erwartungswert kleiner oder gleich q. **QED**

2.3 Die Methode von Storey

Storey (siehe [24]) definiert eine der FalseDiscoveryRate sehr ähnliche Größe.

Definition 6 $pFDR = E(\frac{V}{R}|R > 0)$.

Gilt $m_0 = m$, so ist diese Größe gleich 1. Deshalb wird die pFDR nicht gegenüber irgendeines α's kontrolliert, sondern geschätzt.

Definition 7 Schwellenwert γ *einer Prozedur.*
Wenn alle Hypothesen H_i*, deren entsprechender P-Wert* $P_i \leq \gamma$ *ist, abgelehnt werden, nennt man* γ *den* **Schwellenwert** *der Prozedur.*

Wir vereinbaren noch die Bezeichnung

$$\pi_0 = \frac{m_0}{m}.$$

Wir bringen ein Resultat von Storey ohne Beweis

Theorem 2 *Es gilt*

$$pFDR(\gamma) = \frac{\pi_0 P(P \leq \gamma | H = 0)}{P(P \leq \gamma)} = \frac{\pi_0 \gamma}{P(P \leq \gamma)}.$$

Die zweite Gleichung gilt wieder wegen der Gleichverteiltheit der P-Werte unter der Nullhypothese.

2.3.1 Heuristik der Schätzung von π_0

Definition 8 *Es sei*

$$W(\lambda) = 1 - R(\lambda)$$

und

$$R(\lambda) = \{\#p | p \leq \lambda\}.$$

Wir geben nun einen Schätzer von π_0 an

$$\hat{\pi}_0(\lambda) = \frac{W(\lambda)}{(1-\lambda)m}$$

wobei λ ein geeignet zu wählender Parameter ist.
$\hat{\pi}_0(\lambda)$ ist ein konservativer (d.h. überschätzt π_0) Schätzer, siehe Storey [24].
Die Heuristik hinter obiger Formel von $\hat{\pi}_0(\lambda)$ ist folgende:
Wenn ein Test eine (genügend) große Macht hat (man denke dabei etwa an einen Test über zwei Mittelwerte, die weit genug auseinander liegen und/oder viele Beobachtungen vorliegen), dann werden die meisten beobachteten P-Werte, die nahe bei 1 liegen von den wahren Nullhypothesen stammen (ist der P-Wert nahe bei 1, so wird der ihn erzeugende Beobachtungswert nahe bei dem Mittelwert der Nullhypothese liegen, ein Wert, der sehr wahrscheinlich bei Vorliegen der Nullhypothese und unwahrscheinlich bei Vorliegen der

Alternativhypothese ist).

Da die P-Werte unter Vorliegen der Nullhypothese gleichverteilt sind werden ungefähr $m_0(1-\lambda)$ P-Werte im Intervall $[\lambda, 1]$ liegen (dort erwartet man ja kaum P-Werte der Alternativen) und es gilt

$$W(\lambda) \approx m_0(1-\lambda) = m\pi_0(1-\lambda).$$

Der rechts stehende Ausdruck ist der erwartete Wert an P-Werten in $[\lambda, 1]$ und damit, da wie gesagt in diesem Intervall vor allem P-Werte der wahren Nullhypothesen liegen, praktisch alle betrachteten P-Werte.

2.3.2 Formel pFDR

Wir geben nun mit Hilfe des Schätzers von π_0 einen Schätzer für die pFDR an.
Wir erinnern zuerst an die Formel aus dem Theorem

$$\text{pFDR}(\gamma) = \frac{\pi_0\gamma}{P(P \leq \gamma)}.$$

Für π_0 haben wir im vorigen Abschnitt einen heuristischen Schätzwert abgeleitet. Wir wenden uns jetzt der Wahrscheinlichkeit

$$P(P \leq \gamma)$$

zu. Wir verwenden einfach die relative Häufigkeit

$$\widehat{P(P \leq \gamma)} = \frac{\#\{p_i \leq \gamma\}}{m} = \frac{R(\gamma)}{m}$$

wobei $R(\gamma)$ durch den Ausdruck im Zähler im mittleren Term definiert ist (siehe Definition 8).
Wir haben also insgesamt (λ ist ein Parameter für die heuristische Bestimmung von π_0 und γ ist der Schwellenwert), das m hebt sich weg

$$\widehat{pFDR}_\lambda(\gamma) = \frac{\hat{\pi}_0(\lambda)\gamma}{\hat{P}(P \leq \gamma)} = \frac{W(\lambda)\gamma}{(1-\lambda)R(\gamma)}.$$

Storey [24] schlägt nun vor, da der Schätzer für $P(P \leq \gamma)$ nur sinnvoll ist, wenn $R(\lambda) > 0$ gilt, $R(\gamma)$ durch $R(\gamma) \vee 1$ (also wenn $R(\gamma) = 0$, $R(\gamma)$ durch

1 zu ersetzen) zu ersetzen und die Formel noch durch eine untere Schranke von $P(R(\gamma) > 0)$ zu dividieren.
Wir erhalten dann schlußendlich

$$p\widehat{FD}R_\lambda(\gamma) = \frac{W(\lambda)\gamma}{(1-\lambda)(R(\gamma) \vee 1)(1-(1-\gamma)^m)}.$$

Nebenbei bemerkt: Die FDR ist ein unbedingter Erwartungswert, sodaß wir folgende Formel erhalten (der Bedingungsfaktor fällt weg und wir ersetzen wieder $R(\gamma)$ durch das Größere von $R(\gamma)$ und 1)

$$\widehat{FDR}_\lambda(\gamma) = \frac{W(\lambda)\gamma}{(1-\lambda)(R(\gamma) \vee 1)}.$$

Wir geben nun noch die Prozedur von Storey an:

1. Rechne füe alle m Hypothesen die P-Werte $p_1, \ldots, p_m$ aus.
2. Schätze π_0 und $P(P \leq \gamma)$ durch

   $$\hat{\pi_0}(\lambda) = \frac{W(\lambda)}{(1-\lambda)m}$$

   und

   $$\hat{P}(P \leq \gamma) = \frac{\max(R(\gamma), 1)}{m}$$

   wobei $R(\gamma) = \#\{p_i \leq \gamma\}$ und $W(\lambda) = \#\{p_i > \lambda\}$ bedeutet.
3. Für einen Ablehnbereich $[0, \gamma]$ schätze man pFDR(γ) durch

   $$p\hat{FD}R_\lambda(\gamma) = \frac{\hat{\pi_0}(\lambda)\gamma}{\hat{P}(P \leq \gamma)(1-(1-\gamma)^m)}$$

   für ein gewisses λ (wie man λ optimal wählt behandeln wir hier nicht).
4. Bür B bootstrap Stichproben von $p_1 \ldots, p_m$ berechne man eine Schätzung der pFDR wie in den vorigen Schritten beschrieben.
5. Man nehme nun ein $1-\alpha$ Quantil dieser Stichprobe als obere Konfidenzschranke.

2.4 Literaturhinweise

Die Originalarbeit über die FDR stammt von Benjamini und Hochberg [1]. Das Kapitel über den Zugang von Storey fußt auf [24] und [25]. In letzterer ist manches besser dargestellt, etwa die Heuristik der Schätzung von π_0. Die nichtparametrische Prozedur von Benjamini-Liu ist in [3] zu finden. Das Material über Ordnungsstatistiken habe ich dem 10. Kapitel von [23] entnommen.

Kapitel 3

Gewichtete Prozeduren (a priori Gewichtung)

Diese Kapitel handelt von der Verwendung von Priorinformationen über die Gültigkeit von Nullhypothesen und Alternativen. Mit Hilfe dieser Informationen werden die Gewichte bestimmt, große falls eher an die Alternative geglaubt wird, kleine, falls eher an die Gültigkeit der Nullhypothese geglaubt wird.

3.1 Idee der gewichteten P-Werte

Wir erinnern an die Grundmethode des multiplen Testens. Es werden die m beobachteten Tests in eine aufsteigende Ordnung gebracht (m ist die Anzahl der Tests)

$$H_{(1)} \leq H_{(2)} \leq \cdots \leq H_{(m)}$$

wobei die links stehenden Tests eher ablehnungswürdig sind als die rechts stehenden.
Diese aufsteigende Ordnung wird durch eine Funktion $g(H)$ erzeugt (d.h. $H_i \leq H_j$, falls $g(H_i) \leq g(H_j)$ gilt).
Mögliche Funktionen sind etwa

$$g(H_i) = P_i$$

wobei die P_i die P-Werte bezüglich einer Teststatistik sind.
Eine andere Möglichkeit wäre

$$g(H_i) = \frac{P_i}{W_i}$$

wobei die W_i geeignet zu wählende Gewichte sind.
Dann wird ein Schwellenwert T (Threshholdvalue) festgelegt, sodass alle Hypothesen H_i mit $g(H_i) \leq T$ abgelehnt werden.
Für die Gewichte kann man jetzt entweder versuchen, die Information der erhobenen Daten zu verwenden (siehe das nächste Kapitel) oder aber etwaige Priorinformationen über die Daten ausnützen. Man wird also versuchen diese Information zu benützen um die H_i-Werte in der obigen List so umzustellen, daß Hypothesen, die wahren Alternavtiven entsprechen in der Reihung nach links und solche, die wahren Nullhypothesen entsprechen nach rechts rücken.
In diesem Kapitel untersuchen wir Priorgewichtungen. Wir multiplizieren die P-Werte mit einer Zahl, den Gewichten, sodass die Hypothesen von vermuteten Alternativen bildlich gesprochen nach Links rücken, bzw die $g(H_i)$ Werte kleiner werden.

3.2 Die gewichtete BH-Prozedur

Die gewichtete BH-Prozedur ist analog der BH-Prozedur durchzuführen

1. Gewichte W_i bestimmen, für die

   $$\frac{1}{m}\sum_{i=1}^{m} W_i = 1$$

   gilt.

2.

   $$Q_i = \frac{P_i}{W_i}$$

   ausrechnen.

3. Q_i der Größe nach ordnen, parallel dazu die H_i. Die geordneten Q_i, H_i werden mit $Q_{(i)}, H_{(i)}$ bezeichnet.

4. Suche maximales (j) mit

$$Q_{(j)} \leq \frac{j}{m}\alpha.$$

5. Lehne alle $H_{(i)}$ mit $(i) \leq (j)$ ab.

3.3 Ein allgemeines Modell der Gewichtung

Wir haben wieder folgende Tabelle In der ersten Zeile steht, welche Hypothese

Natur	$H_1, H_2, \ldots, H_m$
Experte	$E_1, E_2, \ldots, E_m$

Tabelle 3.1:

in Wirklichkeit gilt, in der zweiten, was der Experte glaubt.
Dabei ist $H_i = 1$, falls die Alternative gilt, sonst $H_i = 0$ und analog $E_i = 1$ wenn der Experte an die Gültigkeit der Alternative glaubt und $E_i = 0$, wenn der Experte an die Gültigkeit der Nullhypothese glaubt.
Der Experte konstruiert nun Gewichte W_i, sodass

$$\frac{1}{m}\sum_{i=1}^{m} W_i = 1$$

gilt.
Weiters wird der Experte W_i einen Wert > 1 angeben, falls er glaubt, daß die Alternative (wir schreiben dafür immer $H = 1$) zutrifft.
Der Sinn dabei ist, dass wenn P_i/W_i gebildet wird, die Hypothese H_i nach ¨links¨ wandert. Es bedeutet das eine gewisse Absicherung gegen nicht signifikante Ergebnisse von wahren Alternativen. Genauso wandert bei $W_i < 1$ die Hypothese eher nach rechts, was bei einer richtigen Nullhypothese ebenfalls günstig ist. Weiters berechnen wir die Randverteilung der P-Werte (wir verwenden hier den Buchstaben P in doppelter Bedeutung).

$$\begin{aligned} P(P \leq t) &= P(P \leq t|H = 1)P(H = 1) + P(P \leq t|H = 0)P(H = 0) \\ &= F(t)P(H = 1) + tP(H = 0) \\ &= aF(t) + (1 - a)t \end{aligned}$$

wobei F die Verteilungsfunktion der P-Werte unter Gültigkeit der Alternativhypothese ist und wieder $a = P(H = 1)$ gesetzt wurde, siehe die Tabelle unten.
Wir bezeichnen noch

$$P(P \leq t) = G(t)$$

und wie im vorigen Kapitel die Häufigkeiten des Vorkommens der Hypothesen mit den Bezeichnungen aus Tabelle 3.2 Weiters haben wir, daß bedingt

$P(H = 0)$	$1 - a$
$P(H = 1)$	a

Tabelle 3.2:

auf den Zustand der Natur, also des Vektors $(H_1, \ldots, H_m)$ die P-Werte und die Gewichte unabhängig sind, daß die Gewichte vor den Beobachtungen bestimmt werden müssen, auf Grund von Raten oder besser a priori Wissen.
Eine Möglichkeit die Gewichte zu bestimmen besteht darin, zuerst m Zufallsveränderliche U_i zu bestimmen und aus diesen die Gewichte W_i berechnet. Zuerst wird der Mittelwert der U_i gebildet

$$\bar{U}_m = \frac{1}{m} \sum_{i=1}^{m} U_i$$

Die Gewichte werden dann durch

$$W_i = \frac{U_i}{\bar{U}_m}$$

bestimmt.
Weiters bezeichnen wir die (bedingte) Verteilung von $(U_i|H_1)$ mit V_1 und die analoge nach H_0 mit V_0. Es gilt nun

$$P(W_i \leq t|H = 0) = P(\frac{U_i}{\bar{U}_m} \leq t|H = 0)$$

Wir haben nun weiters genau wie im vorigen Abschnitt (wir lassen der Einfachheit halber die Bedingung auf $H = 0$ weg, weiters bezeichnen wir mit $U_{\bar{i}}$ den Mittelwert aller der U_j mit $j \neq i$)

$$\begin{aligned} P(U_i \leq t(U_i + U_{\bar{i}})) &= P(U_i - tU_i \leq tU_{\bar{i}}) \\ &= P(U_i(1-t) \leq tU_{\bar{i}}) \\ &= P(U_i \leq \frac{t}{1-t} U_{\bar{i}}) \\ &= M_0(t) \end{aligned}$$

wobei die letzte Gleichung rechts (die mit M_0) eine Definition ist (das geht, da die U_i als gleichverteilt und unabhängig vorrausgesetzt werden). Analog erhält man

$$P(W_i \leq t|H = 1) = M_1(t).$$

Die Randverteilung von W_i ist dann

$$\begin{aligned} P(W_i \leq t) &= P(W_i \leq t|H = 0)P(H = 0) + P(W_i \leq t|H = 1)P(H = 1) \\ &= M_0(t)(1 - a) + M_1(t)a \end{aligned}$$

Man hat dann [1]

$$W_i(t) = M_0(t)(1 - a) + M_1(t)a$$

und wenn man auf beiden Seiten den Erwartungswert bildet (mit den Bezeichnungen $E(M_0) = \mu_0$ und $E(M_1) = \mu_1$; strenggenommen bilden wir das Integral $\int t dW(t)$, usw. Nimmt man die W_i als diskret an, hat man Summen und das ist dann etwas anschaulicher)

$$E(W) = (1 - a)\mu_0 + a\mu_1 = \mu = 1$$

wegen der Konstruktion der W_i. Wir haben ja

$$\frac{1}{m}\sum_i W_i = \frac{1}{m}\sum_i \frac{U_i}{\bar{U}_i} = 1$$

und da die W_i die selben Verteilung haben, gilt

$$E(W_i) = 1.$$

3.4 FDR

3.4.1 Vorbemerkung

Wir wollen den Schwellenwert t_0 der Benjamini-Hochberg (BH) Prozedur in einer anderen Form darstellen. [2]
Es sei t_0 definiert durch

$$t_0 = \sup\{t : \hat{B}(t) \leq \alpha\}$$

[1] $W_i(t)$ bezeichnet in einer leichten Doppelbedeutung, siehe auch [22] die Verteilungsfunktion von W_i

[2] Siehe auch [6], wo die von t abhängige FDP als Stochastischer Prozess aufgefaßt wird und asymptotische Resultate erziehlt werden.

dabei gilt

$$\hat{B}(t) = \frac{t}{\hat{G}(t)}$$

und $\hat{G_m}(t)$ ist die empirische Verteilungsfunktion der P-Werte

$$G(t) = P(P \leq t)$$

ist. Die ursprüngliche Definition des Schwellenwertes der BH-Prozedur ist folgendermaßen.
In der BH-Prozedur werden die P-Werte geordnet

$$P_{(1)} \leq P_{(2)} \leq \ldots P_{(m)}.$$

Man sucht dann das maximale j, sodaß

$$P_{(j)} \leq \frac{j}{m}\alpha$$

gilt, das heißt daß

$$\frac{P_{(j)}}{\frac{j}{m}} \leq \alpha$$

gilt.
Der Schwellenwert wäre dann $P_{(j)}$.
Die beiden soeben definierten Schwellenwerte $t_0, P_{(j)}$ müssen nicht unbedingt gleich sein, d.h. es kann $t_0 \neq P_{(j)}$ gelten, aber es werden immer die gleichen Hypothesen abgelehnt.
Gilt nämlich für ein bestimmtes t_0

$$t_0 = \sup\left\{t : \frac{t}{\hat{G}(t)} \leq \alpha\right\}$$

so muss

$$P_{(j)} \leq t_0 < P_{(j+1)}$$

für ein bestimmtes (j) gelten.
Da im Intervall

$$P_{(j)} \leq t < P_{(j+1)}$$

$\hat{G}(t) = j/m$ konstant ist (es werden keine weiteren Hypothesen abgelehnt), und $P_{(j)} \leq t$ gilt, muß

$$\frac{P_{(j)}}{j/m} = \frac{P_{(j)}}{\hat{G}(t_0)} \leq \frac{t_0}{\hat{G}(t_0)} \leq \alpha$$

gelten.
Und

$$\frac{P_{(j+1)}}{(j+1)/m} \leq \alpha$$

kann nicht gelten, da ja

$$t_0 < P_{(j+1)}$$

gilt.

3.4.2 Gewichtete Prozedur hält FDR ein

Die False Discovery Proportion bezüglich eines Grenzwertes wird folgendermassen erklärt

$$\mathrm{FDP}(T) = \frac{\sum_{i=1}^{m} 1_{\{P_1 \leq T\}}(1 - H_i)}{\sum_{i=1}^{m} 1_{\{P_i \leq T\}}}$$

Der Schwellenwert wird nun erklärt durch ($Q = \frac{P}{W}$)

$$\sup\left\{ Q_{(i)} : Q_{(i)} \leq \frac{i\alpha}{m} \right\}$$

Wir betrachten nun eine Versuchsreihe ($H = (H_1, \ldots, H_m)$). Es gilt

$$E\left(\frac{V}{R}1_{\{R>0\}}\right) = E\left(E\left(\frac{V}{R}1_{\{R>0\}}\right)|H\right)$$

Wir betrachten zuerst

$$E\left(\frac{V}{R}1_{\{R>0\}}|H\right)$$

Dieser Erwartungswert ¨lebt¨ auf dem Produktraum $P_1 \times \cdots \times P_m \times W_1 \cdots \times W_m$.
Wir rechnen den Erwartungswert zuerst für fixe $w_1 \cdots \times w_m$ aus. Wir betrachten jeweils einzeln alle Terme, für die $R = k$ gilt.
Wir definieren Mengen $R_{k,i}$.

$$R_{k,i} = \{k{-}1Q \text{ Werte sind } \leq q_k \text{ und die anderen } Q_{(u)} \text{ jeweils } > \frac{\alpha u}{m}, j, u \neq i\}$$

Wir können wegen der Unabhängigkeit der P_i von den W_i die gesamte Dichte als Produkt der Dichten der P_i und W_i schreiben. Folgende Wahrscheinlichkeiten werden bei der Bereichnung des Erwartungswertes vorkommen: (es sei $q_k = \frac{k\alpha}{m}$)

$$A_i = \{Q_i \leq q_k \cap R_{k,i}\}$$

Das Ereignis A_i tritt ein, wenn genau k Hypothesen abgelehnt werden und eine davon i ist.
Es gilt (Ω bezeichnet den ganzen Wahrscheinlichkeitsraum)

- $\bigcup_{k=1}^{m} R_{k,i} = \Omega$
- $R_{k,i} \cap R_{j,i} = \{\}$, für $k \neq j$

Entscheidend ist die zweite Beziehung, sie folgt sofort aus der Konstruktion der $R_{k,i}$.
N_0 bedeute die Menge der wahren Nullhypothesen und B_i sei folgendermaßen definiert

$$B_{k,i} = \{j : j = i, \text{ oder } j = \text{ eines der } Q_{(j)} \leq q_k\}$$

Es gilt nun:

1. $M_{i_0} = \{\omega$:Nur ein $i_0 \in B_{i_0}$ in $N_0\}$
2. $M_{i_0,i} = \{\omega$:Ein i_0 und genau ein anderes $i \in B_{i_0}$ in $N_0\}$
3. $M_{i_0,i,j} = \{\omega$:Ein i_0 und genau zwei andere $i, j \in B_{i_0}$ in $N_0\}$
4. usw

Diese Mengen $M_{i_o,\cdot}$ sind alle disjunkt.
Nehmen wir etwa den zweiten Fall her: Diese Menge kommt noch genau ein anderes Mal vor, nämlich wenn das zweite i die Rolle des bestimmten i_0 annimmt und i_0 die des zweiten i.
Die Summe ergibt dann: 2 mal eine Wahrscheinlichkeit.
Macht man dies mit allen 2er Konstellationen, so erhält man den Beitrag von $\frac{2}{k}$ zu $E(\frac{V}{R}|H)$ (man partitioniert sozusagen die Menge $Q_i \leq q_k \cap R_{i,k}$ nach dem Auftreten wahrer Nullhypothesen).
Der Erwartungswert von $\frac{V}{R}$ (die Bedingung auf H und die fixen $w_1, \ldots, w_m$ lassen wir weg) ist dann also

$$\begin{aligned} E(V/R|H) &= \sum_{k=1}^{m} \sum_{i \in (H=0)} \frac{1}{k} P(Q_i \leq q_k \cap R_{k,i}) \\ &= \sum_{i \in (H=0)} \sum_{k=1}^{m} \frac{1}{k} P(P_i \leq q_k W_i \cap R_{k,i}) \\ &= \sum_{i \in (H=0)} \sum_{k=1}^{m} \frac{1}{k} P(P_i \leq q_k W_i) P(R_{k,i}) \end{aligned}$$

Der Schritt von der zweiten in die dritte Zeile ist möglich, weil in die Berechnung der Wahrscheinlichkeit von $R_{k,i}$ nur P_j mit $j \neq i$ eingehen und die P_i als unabhängig angenommen wurden.
Jetzt integrieren wir über die w und bezeichnen den Integraloperator mit E. Unten, wenn die W_i noch fest sind, rechnen wir die Wahrscheinlichkeit $P(P_i \leq q_k w_i)$ aus.

$$\begin{aligned} E\sum_i \sum_k \frac{1}{k} P(P_i \leq q_k W_i) P(R_{k,i}) &= \sum_i E \sum_k \frac{1}{k} q_k w_i P(R_{k,i}) \\ &= \sum_i E \sum_k \frac{1}{k} \frac{\alpha k}{m} w_i P(R_{k,i}) \\ &= \sum_i E \sum_k \frac{\alpha}{m} w_i P(R_{k,i}) \\ &= \sum_i E \frac{\alpha}{m} w_i \sum_k P(R_{k,i}) \end{aligned}$$

i ist ja von k unabhängig, kann also vor das Summenzeichen gezogen werden. Da $\sum R_{k,i} = 1$ gilt, m bleibt also von der letzten Zeile über

$$\begin{aligned} E(V/R|H) &= \sum_i \frac{\alpha}{m} E w_i \\ &= \frac{\alpha}{m} \sum_i \mu_0 \\ &= \frac{\alpha}{m} \mu_0 \sum_{i=1}^{m} (1 - H_i) \end{aligned}$$

Die letzte Summe ist genau die Anzahl der wahren Nullhypothesen.
Integriert man über die H_i, so erhält man

$$E(V/R) = \alpha \mu_0 E\left(\sum \frac{1}{m} - \frac{H_i}{m}\right) = \alpha \mu_0 \sum \left(\frac{1}{m} - \frac{a}{m}\right) = \alpha \mu_0 (1 - a)$$

Da nun $1 = (1-a)\mu_0 + a\mu_1$ gilt, ist $(1-a)\mu_0 \leq 1$ und damit $\mu_0 \leq \frac{1}{1-a}$ und damit insgesamt

$$E(V/R) = \mathrm{FDR} \leq \alpha \frac{1}{1-a}(1-a) = \alpha$$

3.5 Verteilungsfunktion von $\frac{P}{W}$

Wir tun der Einfachheit halber so, als ob W diskret verteilt wäre (dies vereinfacht die Schreibweise und erleichtert das intuitive Verständnis). $\frac{P}{W}$ lebt sozusagen auf dem Kreuzprodukt von H und W, d.h. wir waehlen zufaellig einen Punkt (h, w) aus und beobachten dann P/W. Die Verteilungen von H und W sind natuerlich idealerweise stark abhaengig.
Wir schicken noch kurz folgende elementare Beziehung voraus (hier begegnet uns wieder die Schwierigkeit des Bezeichnungswesens: P steht hier fuer probability nicht fuer den P-Wert Siehe [21] aber auch [22] ¨An mehrfachbedeutungen von Symbolen muß man sich frühzeitig gewöhnen¨, sinngemäß aus dem Gedächtnis des Autors entnommen).

$$P(W = w, H = h) = P(W = w|H = h)\cdot P(H = h) = P(H = h|W = w)\cdot P(W = w)$$

Wir bemerken noch, daß es für die Ableitung unten nicht wichtig ist, daß für die W_i

$$\frac{1}{m}\sum_i W_i = 1$$

gilt (dies wird für die asypmtotischen Ergebnisse wichtig sein).
Und jetzt kommen wir zur Verteilungsfunktion von $\frac{P}{W}$. (Wir verwenden P in zwei Bedeutungen. Es bezeichnet sowohl den P-Wert als auch P wie Probability; außerdem schreiben wir auch etwa $P(w|h)$ für $P(W = w|H = h)$ und $P(h)$ für $P(H = h)$)

$$\begin{aligned}
D(t) &= P\left(\left\{\frac{P}{W} \le t\right\}\right)\sum_{w,k} P\left(\left\{\frac{P}{W} \le t\right\} \cap \left\{W = w \cap H = h\right\}\right) \\
&= \sum_{(w,h)} \frac{P(\{\frac{P}{W} \le t\} \cap \{W = w, H = h\})}{P(W = w \cap H = h)} \cdot P(W = w \cap H = h) \\
&= \sum_{(w,h)} P(\{\frac{P}{W} \le t\}|\{W = w, H = h\}) \cdot P(h)P(w|h) \\
&= \sum_{w}\sum_{h=0}^{h=1} P\left(\left\{\frac{P}{W} \le t\right\}|\left\{W = w, H = h\right\}\right)P(h)P(w|h) \\
&= \sum_{h=0}^{h=1}(\sum_{w} P\left(\left\{\frac{P}{W} \le t\right\}|\left\{W = w, H = h\right\})P(h)P(w|h))
\end{aligned}$$

$$
\begin{aligned}
&= \sum_w P\left(\left\{\frac{P}{W} \le t\right\} | \left\{W = w, H = 0\right\}\right) P(H = 0) P(w|H = 0)) \\
&+ \sum_w P\left(\left\{\frac{P}{W} \le t\right\} | \left\{W = w, H = 1\right\}\right) P(H = 1) P(w|H = 1)) \\
&= \sum_w P(\{P \le wt\} | \{W = w, H = 0\}) P(H = 0) P(w|H = 0)) \\
&+ \sum_w P(\{P \le wt\} | \{W = w, H = 1\}) P(H = 1) P(w|H = 1)) \\
&= \sum_w wt(1-a) P(w|H = 0) + \sum_w F(wt) a P(w|H = 1) \\
&= t(1-a) \sum_w w P(w|H = 0) + a \sum_w F(wt) P(w|H = 1)
\end{aligned}
$$

3.6 Gewichteter Bonferroni

Die gewichtete Bonferroni Prozedur hat folgenden Ablehnungsbereich:

$$R_0 = \{j : Q_j \le \frac{\alpha}{m}\}$$

Weiters bezeichnen wir die Indici aller wahren Nullhypothesen mit

$$S_0 = \{i : H_i = 0\}$$

Theorem 3 *Die gewichtete Bonferroni-Prozedur hält den familienweiten Fehler ein.*

Wir betrachten wieder zuerst die Ablehnungswahrscheinlichkeit für fixe $w_1, \ldots, w_m$ und integrieren dann.
Die Ablehnungswahrscheinlichkeit bedingt auf $w_1, \ldots, w_m$ ist also

$$
\begin{aligned}
P(\#(R_0 \cap S_0) > 0) &= P(P_j \le \frac{\alpha w_j}{m} \text{ für mindestens ein } j \in S_0) \\
&\le \sum_{j=1}^{m} P(P_j \le \frac{\alpha w_j}{m} \text{ und } H_j = 0) \\
&= \sum_{j=1}^{m} P(P_j \le \frac{\alpha w_j}{m} | H_j) P(H_j = 0)
\end{aligned}
$$

Jetzt integrieren wir über die $w_1, \ldots, w_m$ und beachten, daß wir Summe und Integral vertauschen dürfen (Wahrscheinlichkeiten sind zwischen 0 und 1 und

die Summer ist endlich).
Wir haben also unbedingt

$$\begin{aligned} P(\#(R_\cap S_0) > 0) &= \sum_{j=1}^{m} \int P(P_j \leq \frac{\alpha w_j}{m} | H_j)(1-a) \\ &= (1-a) \sum_{j=1}^{m} \int \frac{\alpha w_j}{m} \\ &= (1-a) \frac{\alpha}{m} \sum_{j=1}^{m} \int w_j \\ &= (1-a) \frac{\alpha}{m} \mu_0 m \\ &= (1-a) \alpha \mu_0 \\ &\leq \alpha \end{aligned}$$

Für die letzte Ungleichung erinnern wir an die Konstruktion der w_i von früher, nämlich daß

$$(1-a)\mu_0 \leq 1$$

gilt. **QED**

3.7 Gewichteter Holm

S_0 hat dieselbe Bedeutung wie im vorigen Abschnitt, S_0 ist die Menge der Indizi aller wahren Nullhypothesen.
Die Holm Prozedur:
Es seien wieder $Q_i = \frac{P_i}{W_i}$ und

$$Q_{(1)} \leq \ldots \leq Q_{(m)}$$

Ist schon $Q_1 > \frac{\alpha}{m}$ so ist die Ablehnmenge $R_0 = \{\}$.
Sonst ist i_0 das größte i_0, sodaß

$$Q_{(j)} \leq \frac{\alpha}{\sum_{k=j}^{m} W_{(k)}}$$

für alle $1 \leq j \leq i_0$ gilt.
R_0 ist dann gleich $\{(1), \ldots, (i_0)\}$. Wir schreiben noch

$$N_0 = \sum_{i=1}^{m} W_i (1 - H_i)$$

also die Summe aller W_i, die den wahren Nullhypothesen entsprechen.
Wir nehmen nun eine fixe Konstellation der Hypothesen $H = (H_1, \ldots, H_m)$ und betrachten die Menge

$$A = \{Q_i > \frac{\alpha}{N_0} \text{ für alle } i_0 \in S_0\}$$

Wir zeigen nun, daß $P(A) \geq 1 - \alpha$ gilt.

$$\begin{aligned}
P(A) &= 1 - P(Q_i \leq \frac{\alpha}{N_0} \text{ für mindestens ein } i_0 \in S_0) \\
&= 1 - \int P(Q_i \leq \frac{\alpha}{N_0} \text{ für mindestens ein } i_0 \in S_0 | w) dM_0 \\
&\geq 1 - \int \sum_{i \in S_0} P(Q_i \leq \frac{\alpha}{N_0} | w) dM_0 \\
&= 1 - \int \sum_{i \in S_0} P(P_i \leq \frac{w_i \alpha}{N_0} | w) dM_0 \\
&= 1 - \int \sum_{i \in S_0} \frac{w_i \alpha}{N_0} dM_0 \\
&= 1 - \int \alpha dM_0 \\
&= 1 - \alpha
\end{aligned}$$

Wir zeigen nun, daß wenn A eintritt, und das geschieht ja mit mindestens einer Wahrscheinlichkeit von $1-\alpha$ keine Nullhypothese abgelehnt wird, dann folgt daß höchstens mit einer Wahrscheinlichkeit von α eine wahre Nullhypothese abgelehnt wird.
Wir bezeichnen mit ν den Index der Nullhypothese, die am weitesten ¨links¨ in der nach der Größe der Q_i geordneten Hypothesen liegt. Wird diese nicht abgelehnt, dann wird überhaupt keine Nullhypothese abgelehnt, da die Prozedur dann ja abbricht.
Wir haben (die erste Ungleichung gilt, weil wir annehmen, daß A eingetreten ist)

$$Q_{(\nu)} > \frac{\alpha}{N_0} = \frac{\alpha}{\sum_{i \in S_0} N_0} \geq \frac{\alpha}{\sum_{i=\nu}^{m} W_{(i)}}$$

Die letzte Ungleichung gilt, weil ν der Index der ersten wahren Nullhypothese ist, alle Indizi von S_0 also in der Summe vorkommen müssen.
Die eben hergeleitete Ungleichung sagt aber, daß wegen der Konstruktion der (gewichteten) Holm-Prozedur $H_{(\nu)}$ nicht abgelehnt werden kann, was wir zeigen wollten. **QED**

3.8 Asymptotische Resultate

Wir definieren nun die Schwellenwerte basierend auf der theoretischen und der empirischen Verteilung der Q_i.
Wir definieren $C(t)$.

$$C(t) = \frac{D(t)}{t} \ , \ \hat{C}_m(t) = \frac{\hat{D}_m(t)}{t}$$

Der Schwellenwert der theoretischen Verteilung

$$t_* = \ \sup\{t : C(t) \geq \frac{1}{\alpha}\}$$

Der Schwellenwert der empirischen Verteilung

$$T_m = \ \sup\{t : \hat{C}_m(t) \geq \frac{1}{\alpha}\}$$

Lemma 2 *Es sei F (die Verteilungsfunktion der P-Werte unter Gültigkeit der Alternative) konkav auf $[0, 1]$. Dann folgt*

1. *G ist konkav auf $[0, 1]$*
2. *D ist konkav auf $[0, 1]$*
3. *C ist monoton fallend auf $[0, 1]$*

Ist F strikt konkav, so gilt für die ersten beiden Punkte auch strikt konkav.

Beweis:
Wir haben die Beziehung

$$G = (1 - a)U + aF$$

Da U (das ist in diesem Fall die Verteilungsfunktion der Gleichverteilung) wenigsten konkav (natürlich nicht strikt konkav) ist, folgt die erste Behauptung.
Für den Beweis der zweiten Behauptung schreiben wir noch einmal die Verteilungsfunktion von $P/Q = D$ an

$$D(t) = (1 - a)\mu_0 t + a \int F(wt) dM_1(w)$$

Wir erinnern daran, daß es für die Ableitung der obigen Gleichung nicht wichtig ist, ob der Mittelwert der W gleich 1 ist, oder nicht. Wir haben also

$$\begin{aligned} D((1-\lambda)t_0 + \lambda t_1)) &= (1-a)\mu_0((1-\lambda)t_0 + \lambda t_1) + a\int F(w(1-\lambda)t_0 + \lambda t_1))dM_1(w) \\ &> (1-\lambda)D(t_0) + \lambda D(t_1) \end{aligned}$$

Für die dritte Behauptung bezeichnen wir $1 > t_1 > t_0 > 0$ und bemerken, dass ja $F(0) = 0$ gilt, da die P-Werte positiv sind, daraus folgt $D(0) = 0$, wie man etwa durch Einsetzen in obige Formel sieht.
Wir haben nun folgende Beziehung

$$D(t_1) = \frac{D(t_1)}{t_1} = \frac{(1-t_0/t_1)D(0) + (t_0/t_1)D(t_1)}{t_0} \leq \frac{D(t_0)}{t_0} = C(t_0)$$

was ja die dritte Aussage ist (die beiden äußeren Gleichungen sind Definitionen, die zweite Gleichung trivial wegen $D(0) = 0$ und die Ungleichung folgt aus der konkavität). **QED**

Jetzt kommen wir zum angekündigten Theorem

Theorem 4 *Ist $C(t)$ (Definition siehe Anfang dieses Kapitels) monoton fallend, so gilt*

1. *$T_m \to t_*$ für $m \to \infty$*
2. *$E|FDP(T_m) - FDP(t_*)| \to 0$*

Bemerkung: Ist F konkav, so ist, siehe das Lemma C monoton fallend.
Beweis:
Wir führen den Beweis in zwei Schritten

1. Wir zeigen die asymptotischen Beziehungen für den Fall, daß die W_i voneinander unabhängig sind
2. Dann zeigen wir, daß für einen Spezialfall unabhängiger W_i ($\tilde{W}_{=\frac{U_i}{E(U_i)}}$) und den obig definierten $W_i = \frac{U_i}{U_i}$ die Funktionen D asymptotisch gleich sind.

Zum ersten Fall:
Die W_i sind jetzt also unabhängig, nicht notwendigerweise mit Mittelwert

gleich 1.
Wir halten $b, \epsilon > 0$ fest. Definitionsgemäß gilt dann

$$C(t_* + b) < \frac{1}{\alpha}$$

wegen der Definition von C. Wir schreiben obige Ungleichung als Gleichung mit einem positiven δ.

$$C(t_* + b) = \frac{1}{\alpha} - \delta$$

Wir haben jetzt folgende Beziehung für alle $t > t_* + b$

$$\begin{aligned} \hat{C}_m(t) &\leq \frac{\hat{D}_m(t)}{t} \leq \frac{D(t) + \sup_u |\hat{D}_m(u) - D(u)|}{t} \leq \frac{D(t) + \epsilon}{t} \\ &= C(t) + \frac{\epsilon}{t} \leq C(t_* + b) + \frac{\epsilon}{t} = \frac{1}{\alpha} - \delta + \frac{\epsilon}{t} \leq \frac{1}{\alpha} \end{aligned}$$

Die letzte Ungleichung gilt für m groß genug, da dann ja ϵ beliebig klein wird. Die obigen Beziehungen sieht man so:

1. Die erste Ungleichung ist definitionsgemäß richtig.

2. Die zweite Ungleichung gilt, weil das Supremum ja über alle u genommen wird, der Zähler also sicher größer ist, wenn für u t eingesetzt wird.

3. Die letzte Ungleichung der ersten Zeile gilt, weil nach dem sogenannten Zentralsatz der Statistik (von Gliwenko-Cantelli) die empirische Verteilungsfunktion die theoretische Verteilungsfunktion beliebig genau approximiert.
   Eigentlich müßte in der obigen Beziehung statt $\hat{C}_m(t)$ korrekterweise $\hat{C}_m(t)(\omega)$ stehen. ω ist hier das generische Element im Satz von Gliwenko-Cantelli, das ist eine unendliche Folge von Zufallsvariablen mit denen die empirische Verteilungsfunktion gebildet wird. Für fast alle ω gilt dann die Aussage des Satzes von Gliwenko-Cantelli und damit obige Beziehung.

4. In der zweiten Beziehung ist nur bemerkenswert, daß die erste Ungleichung gilt, weil ja $t > t_* + b$ vorausgesetzt wurde, und

$$-\delta + \frac{1}{\alpha} < 0$$

gilt, weil ja mit $m \to \infty$ ϵ beliebig klein wird.

Aus der Definition des Schwellwertes T_m folgt nun, dass

$$T_m \leq t_* + b$$

gelten muß (sonst könnte ja t_* nicht das Supremum von $C(t) \geq \frac{1}{\alpha}$ sein).
Ebenso kann man zeigen, daß

$$T_m \geq t_* - b$$

gilt und damit

$$|T_m - t_*| \leq b$$

für ein beliebiges positives b.
Für die false discovery proportion haben wir

$$\mathrm{FDP}(T_m) = \frac{m^{-1} \sum_i (1 - H_i) 1_{\{P_i/W_i \leq T_m\}}}{m^{-1} \sum_i 1_{\{P_i/W_i \leq T_m\}}} = \frac{\hat{V}_m(T_m)}{\hat{D}_m(T_m)}$$

Wenn nun m groß genug ist, hat man

$$\begin{aligned} |\hat{D}_m(T_m) - D(t_*)| &= |\hat{D}_m(T_m) - D(T_m) + D(T_m) - D(t_*)| \\ &\leq \sup_u |\hat{D}_m(u) - D(u)| + |D(T_m) - D(t_*)| \to 0 \end{aligned}$$

wieder wegen Gliwenko-Cantelli und unter der Vorraussetzung, daß D stetig ist.
Ebenso gilt, daß

$$\hat{V}_m(T_m) \to V(t_*)$$

für $m \to \infty$ gilt. Dabei ist

$$V(t) = E(1 - H_i) 1_{\{P_i/W_i \leq t\}} \leq t.$$

Jetzt betrachten wir den Fall, daß der Mittelwert der W_i gleich 1 sein muß.
Wir haben wieder $U_1, \ldots, U_m$ unabhängig und gleichverteilt und setzen

$$W_i = \frac{U_i}{\bar{U}_m}$$

Weiters setzen wir noch

$$\tilde{W}_i = \frac{U_i}{EU_1}$$

d.h. die $\tilde{W}_i$ sind unabhängig.
Wir zeigen jetzt, daß P_i/W_i und $P_i/\tilde{W}_i$ asymptotisch gleich sind. Daraus folgen dann die Behauptungen für die W_i.

Wir stellen jetzt eine Beziehung zwischen der empirischen Verteilungsfunktion von $P_/W_i$ und $P_i/\tilde{W}_i$ her.

$$\begin{aligned}\hat{D}_m &= \frac{1}{m}\sum_i 1_{\frac{P_i}{\tilde{W}_i}\leq t} = \frac{1}{m}\sum_i 1_{\frac{P_i}{\tilde{W}_i}\leq t\frac{EU_1}{EU_1+(\bar{U}_m)-EU_1}} \\ &\leq \frac{1}{m}\sum_i 1_{\frac{P_i}{\tilde{W}_i}\leq t\frac{EU_1}{EU_1-\epsilon}} \leq P\left(\frac{P_i}{\tilde{W}_i}\leq\frac{EU_1}{EU_1-\epsilon}\right)+\epsilon\end{aligned}$$

falls nur m groß genug ist. Die Gleichung in der ersten Zeile findet man durch einfaches Umformen

$$\begin{aligned}\frac{P_i}{W_i} &\leq t \\ \frac{P_i}{\frac{U_i}{\bar{U}_m}} &\leq t \\ \frac{P_i\bar{U}_m}{\frac{U_i}{EU_i}} &\leq tEU_1 \\ \frac{P_i}{\tilde{W}_i} &\leq t\frac{EU_1}{\bar{U}_m}\end{aligned}$$

Die erste Gleichung ist Definition, die zweite nur einen Bruch umformen.
Die erste Ungleichung in der zweite Zeile oben folgt aus dem Gesetz der großen Zahlen, die zweite wieder aus der Tatsache, daß die empirische Verteilungsfunktion die theoretische Verteilungsfunktion beliebig genau annähert.
Umgekehrt kann man auch zeigen, dass

$$\hat{D}_m \geq \sum_i 1_{\frac{P_i}{\tilde{W}_i}\leq t\frac{EU_1}{EU_1+\epsilon}} \geq P(\frac{P_i}{\tilde{W}_i}\leq\frac{EU_1}{EU_1-\epsilon})-\epsilon$$

gilt.
Da ϵ beliebig war, nähert sich der Ausdruck ind der rechten Seite der Ungleichung in der Wahrscheinlichkeit P praktisch t an, womit man die behauptete Asymptotische Beziehung erhält. **QED**

3.9 Macht

Die Macht H als Funktion des Ablehnungsschwellenwertes t ist folgendermassen definiert.

$$\begin{aligned} H(t) &= P(\frac{P}{W} \leq t | H = 1) \\ &= P(P \leq Wt | H = 1) \\ &= \int F(wt) dw_1 \end{aligned}$$

In der zweiten Gleichung haben wir wieder ein Integral im Produktraum $p \times w$ und ¨integrieren¨ bei festem w zuerst über p, das ergibt wt. dw_1 bezeichnet das Wahrscheinlichkeitsmaß $P(W|H = 1)$.
Die Werte des Ablehnungsschwellenwertes der gewichteten Prozedur sei mit t^w und die der ungewichteten mit t^0 (also alle $w_i = 1$) bezeichnet. Wir bilden nun den Quotienten der Macht des gewichteten und der ungewichteten Falles, also

$$\frac{H(t^w)}{H(t^0)}$$

Ist dieser Quotien größer als 1 bedeutet dies, daß die gewichtete Prozedur eine größere Macht als die Ungewichtete hat.
Wir leiten jetzt sowohl Ausdrücke für $H(t^w)$ also auch für $H(t^0)$ her.
Zuerst bemerken wir, dass definitionsgemäß für die Ablehnungsschwellenwerte folgende Beziehungen gelten

$$\frac{t^w}{D(t^w)} = \alpha = \frac{t^0}{G(t^0)}$$

Wir verwenden nun den linken Teil der obigen Beziehung und die Formel für $D(t^w)$

$$D(t) = (1 - a)\mu_0 t + a \int F(wt) dw_1$$

und erhalten unter Beachtung von $(1 - a)\mu_0 + a\mu_1 = 1$

$$\begin{aligned} D(t^w) &= (1 - a)\mu_0 t + aH(t^w) \\ \frac{t^w}{\alpha} &= (1 - a)\mu_0 t^w + aH()t^w \\ \frac{1}{\alpha} &= 1 - a\mu_1 + a\frac{H(t^w)}{t^w} \end{aligned}$$

und schlußendlich

$$\frac{H(t^w)}{t^w} = \frac{1}{a\alpha} - \frac{1}{a} + \mu_1$$

Wir schreiben noch einmal die Formel für $D(t)$ auf:

$$D(t) = (1-a)\mu_0 t + a \int F(wt) dw_1$$

Wenn jetzt W gleich 1 ist, also der ungewichtete Fall vorliegt, ist im Integral $F(wt)$ konstant gleich $F(t)$ und man erhält, wenn die Formel an der Stelle t^0 betrachtet wird und statt D G geschrieben wird, μ_0 ist jetzt gleich 1)

$$\begin{aligned} G(t^0) &= (1-a)t^0 + aF(t^0) = \frac{t^0}{\alpha} \\ F(t^0)a &= \frac{t^0}{\alpha} - (1-a)t^0 \\ \frac{F(t^0)}{t^0} &= \frac{1}{a\alpha} - \frac{1}{a} + 1 \end{aligned}$$

Nun bilden wir den Quotienten der eben berechneten Ausdrücke und erhalten

$$\frac{\frac{H(t^w)}{t^w}}{\frac{F(t^0)}{t^0}} = \frac{K+\mu_1}{1+K}$$

wobei

$$K = \frac{1}{a\alpha} - \frac{1}{a}$$

gilt. Wir bringen nun den Ausdruck auf die Form

$$\frac{K+\mu_1}{1+K} = 1 + L$$

und bestimmen L.

$$L = \frac{K+\mu_1}{1+K} - \frac{1+K}{1+K}$$

und das ist

$$\frac{\mu_1 - 1}{1+K}$$

und weiters berechnen wir noch

$$1 + K = 1 + \frac{1}{a\alpha} - \frac{1}{a} = \frac{a\alpha + 1 - \alpha}{a\alpha} = \frac{1 - (\alpha(1-a))}{a\alpha}$$

und erhalten schlußendlich

$$\frac{H(t^w)}{F(t^0)} = \frac{t^w}{t^0}(1 + (\mu_1 - 1)\frac{a\alpha}{1 - (\alpha(1-a))})$$

Leider hängen t^w und t^0 von H und F ab. Deshalb sind Aussagen über den Quontienten H/F durch obige Gleichung unmittelbar schwer zu machen.
Wir bilden nun eine Funktion von λ, t^λ, wobei bei $\lambda = 0$ $t^0 = t^0$ und bei $\lambda = 1$ $t^1 = t^w$ gilt.
Dann bestimmen wir die Ableitung der Funktion t^λ nach λ an der Stelle $\lambda = 0$ und entwickeln t^λ in eine Taylorreihe mit linearem Term. Wir schreiben $t^\lambda = f(\lambda)$.

$$t^\lambda = f(\lambda) = f(0) + \frac{df}{d\lambda}_{(0)} \cdot \lambda$$

und erhalten dann

$$\begin{aligned} t^w = f^1 = f(1) &= f(0) + \frac{df}{d\lambda}_{(0)} \cdot 1 \\ &= t^0 + \frac{df}{d\lambda}_{(0)} \cdot 1 \end{aligned}$$

und schlussendlich

$$\frac{t^w}{t^0} = 1 + \frac{1}{t^0}\frac{df}{d\lambda}_{(0)}$$

Dazu betrachten wir im Folgenden eine einparametrige Familie von Gewichten

$$W^\lambda = \lambda W + 1 - \lambda$$

Es gilt dann

$$W^1 = 1w + 1 - 1 = w$$

und

$$W^0 = 0w + 1 - 0 = 1$$

Der Erwartungswert von W^λ ist dann

$$E(W^\lambda | H = j) = \lambda E(w) + 1 - \lambda = \lambda\mu_j + 1 - \lambda = 1 + \lambda(\mu_j - 1)$$

für $j = 0, 1$

3.9.1 Die Funktion $R(t, \lambda)$

Wir definieren folgende Funktion (der Sinn wird sein mit Hilfe des Satzes vom impliziten Differenzieren einen Zusammenhang von t^λ und λ herzustellen).

$$R(t, \lambda) = \int \frac{F((\lambda w + 1 - \lambda)t)}{t} dw_1 - \lambda(\mu_1 - 1) - \frac{F(t^0)}{t^0)}$$

Wir wollen jetzt zeigen, dass

$$R(t^\lambda, \lambda) = 0$$

für alle $0 \leq \lambda \leq 1$ gilt.
Es gilt einmal folgende Beziehung wegen der Definition (t^λ ist der sblehn-Schwellenwert zu den Gewichten W^λ) der Ablehnschwellenwerte.

$$\frac{D(t^\lambda)}{t^\lambda} = \alpha = \frac{G(t^0)}{t^0}$$

Zuerst leiten wir noch eine Beziehung zwischen $G(t)$ und $F(t)$ her. Es gilt einmal definitionsgemäß folgendes:

$$G(t) = P(P \leq t)$$

und

$$F(t) = P(P \leq t | H = 1) = \frac{P(P \leq t \cap H = 1)}{P(H = 1))}$$

Wir haben dann

$$\begin{aligned}
\frac{G(t)}{a} &= \frac{G(t)}{P(H = 1)} \\
&= \frac{P(P \leq t \cap H = 1) + P(P \leq t \cap H = 0)}{P(H = 1)} \\
&= F(t) + \frac{P(P \leq t \cap H = 0)}{P(H = 1)} \\
&= F(t) + \frac{P(P \leq t | H = 0) \cdot P(H = 0)}{P(H = 1)} \\
&= F(t) + \frac{t(1 - a)}{a}
\end{aligned}$$

und darauf folgt dann

$$G(t) = aF(t) + t(1-a)$$

Es gilt dann weiters

$$D^{\lambda}(t) = (1-a)\mu_0 t\lambda + (1-a)(1-\lambda)t + a\int F((\lambda w + 1 - \lambda)t)dw_1$$

was man folgendermaßen sieht.
Man beachte zuerst einmal

$$1 = (1-a)\mu_0 + a\mu_1$$

Man nimmt in der vorletzen Zeile der Ableitung zu $D(t)$ weiter oben in Abschnitt 3.5, statt w einfach $\lambda w + 1 - \lambda$ und erhaelt dann fuer den ersten Term

$$\begin{aligned}(1-a)t\int(\lambda w + 1 - \lambda)dw_0 &= (1-a)t\lambda\mu_0 + (1-a)t(1-\lambda) \\ &= (1-a\mu_1)t\lambda + (1-a)t(11-\lambda) \\ &= t(\lambda - a\mu_1\lambda) + (1-a)(1-\lambda)t\end{aligned}$$

Weiters haben wir

$$\frac{D^{\lambda}}{t^{\lambda}} = \alpha = \frac{G(t^0)}{t^0}$$

und damit

$$\frac{D^{\lambda}}{t^{\lambda}} - \frac{G(t^0)}{t^0} = 0$$

und daraus folgt folgendes (wir schreiben der Einfachheit halber statt $F((\lambda w + 1 - \lambda)t)$ einfach F)

$$\begin{aligned}0 &= (1-a)\mu_0\lambda + (1-a)(1-\lambda) + a\int\frac{F}{t^{\lambda}}dm_1 - \frac{aF(t^0) + t^0(1-a)}{t^0} \\ &= (1-a\mu_1)\lambda + 1 - a - \lambda + a\lambda - 1 + a + a\int\frac{F}{t^{\lambda}}dm_1 - a\frac{F(t^0)}{t^0} - (1-a) \\ &= (1-a\mu_1)\lambda - \lambda(1-a) + a\int\frac{F}{t^{\lambda}}dm_1 - a\frac{F(t^0)}{t^0} \\ &= \lambda(1 - a\mu_1 - 1 + a) + a\int\frac{F}{t^{\lambda}}dm_1 - a\frac{F(t^0)}{t^0} \\ &= a(1-\mu_1)\lambda + a\int\frac{F}{t^{\lambda}}dm_1 - a\frac{F(t^0)}{t^0}\end{aligned}$$

Unter der Vorraussetzung, daß $a \neq 0$ gilt, also Alternativen vorhanden sind, kann man in der letzen Gleichung durch a dividieren und was übrigbleibt ist genau $R(t^\lambda, \lambda)$.

3.9.2 Implizites Differenzieren

Wir erinnern an die Definition von $R(t, \lambda)$

$$R(t,\lambda) = \int \frac{F((\lambda w + 1 - \lambda) \cdot t)}{t} dw_1 - \lambda(\mu_1 - 1) - \frac{F(t^0)}{t^0}$$

Wir differenzieren nun unter der Annahme, dass unter dem Integralzeichen differenziert werden darf und die Verteilungsfunktion F differenzierbar mit Ableitung also Dichte f ist.
Zuerst die partielle Ableitung nach t (Bemerkung: dw_1 bedeutet, dass w die Integrationsvariable und das Mass M_1 ist, also die Verteilung der Gewichte unter der Alternative).
Der zweite und dritte Summand ist bezüglich t eine Konstante.

$$\frac{\delta R(t,\lambda)}{\delta t} = \int \frac{f((\lambda w + 1 - \lambda)t) \cdot (\lambda w + 1 - \lambda) \cdot t - F((\lambda w + 1 - \lambda)t))}{t^2} dw_1$$

Und nun die partielle Ableitung nach λ

$$\frac{\delta R(t,\lambda)}{\delta \lambda} = \int \frac{f((\lambda w + 1 - \lambda)t) \cdot t \cdot (w - 1)}{t} dw_1 - (\mu_1 - 1)$$

Wir berechnen nun die partielle Ableitung nach t an der Stelle $(t^0, 0)$.

$$\frac{\delta R(t,\lambda)}{\delta t}_{(t^0,0)} = \int \frac{f(t^0) \cdot 1 \cdot t^0 - F(t^0)}{t^{0^2}} dw_1$$

und da ja alles was im Integral steht eine Konstante ist, ist das gleich

$$\frac{t^0 f(t^0) - F(t^0)}{t^{0^2}}$$

Jetzt berechnen wir noch den Wert der partiellen Ableitung nach λ an der Stelle $(t^0, 0)$.

$$\frac{\delta R(t,\lambda)}{\delta \lambda}_{(t^0,0)} = \int \frac{f(t^0) \cdot t^0 \cdot (w - 1)}{t^0} dw_1 - (\mu_1 - 1)$$

und das ist gleich, wenn man $f(t^0)$ als Konstante heraushebt und beachtet, dass

$$\int (w-1)dw_1 = \int w dw_1 - \int 1 \cdot dw_1 = \mu_1 - 1$$

gilt

$$f(t^0)(\mu_1 - 1) - (\mu_1 - 1) = (\mu_1 - 1) \cdot (f(t^0) - 1).$$

Es ist ja t^λ eine Funktion von λ. Wir wollen nun die Ableitung von t^λ and der Stelle $\lambda = 0$ bstimmen und verwenden dazu den Satz ueber implizite Funktionen. Wir erhalten dann:

$$\frac{dt^\lambda}{d\lambda}_0 = \frac{-1}{\frac{\delta R(t,\lambda)}{\delta t}_{(t^0,0)}} \cdot \frac{\delta R(t,\lambda)}{\delta \lambda}_{(t^0,0)}$$

Wir setzen die obig errechneten Werte fuer die Ableitungen ein und erhalten dann, sogleich den Kehrwert bildend und auch den so erhaltenen Nenner mit -1 multiplizierend

$$\frac{dt^\lambda}{d\lambda}_0 = \frac{{t^0}^2}{F(t^0) - t^0 f(t^0)} \cdot (\mu_1 - 1)(f(t^0) - 1)$$

Wir haben also dann

$$\begin{aligned} \frac{t^w}{t^0} &= 1 + \frac{1}{t^0}\frac{df}{d\lambda}_{(0)} \\ &= 1 + \frac{t^0}{F(t^0) - t^0 f(t^0)}(\mu_1 - 1)(f(t^0) - 1) \end{aligned}$$

3.9.3 Ein Beispiel

Wir nehmen als Nullhypothese ein Normalvertilung mit Varianz 1 und Erwartungswert 1. Als Alternativen nehmen wir Normalverteilungen mit Erwartungswert θ und Varianz 1.
Wir rechnen jetzt die Dichte der P-Werte unter der Alternative aus. Als Ablehnungsbereich nehmen wir

$$T(x) = \bar{x} \geq g$$

Unter der Nullhypothese gilt nun

$$\begin{aligned} P(\bar{x} \geq g) &\leq t \\ 1 - P(\bar{x} \leq g) &\leq t \\ 1 - t &\leq P(\bar{x} \leq g) = \Phi(g) \end{aligned}$$

und erhält dann weiter

$$1 - t = \Phi(g)$$

und dann den Ablehnungwert

$$g = \Phi^{-1}(1 - t)$$

Jetzt berechnen wir die Verteilungsfunktion F der P-Werte unter der Alternative.
Wir berechnen dazu

$$P_A(\bar{x} \geq g)$$

und erhalten

$$\begin{aligned} P(\bar{x} - \theta \geq g - \theta) &= 1 - \Phi(g - \theta) \\ &= 1 - \Phi(\Phi^{-1}(1 - t) - \theta) = F(t) \end{aligned}$$

Jetzt leiten wir $F(t)$ ab, und erhalten (wir schreiben ab der zweiten Zeile g für $\Phi^{-1}(1 - t)$)
Das Minus der zweiten inneren Ableitung wird gleich zum Plus vorne im Ausdruck.

$$F^{'}(t) = \frac{1}{\pi} e^{-\frac{(\Phi^{-1}(1-t)-\theta)^2}{2}} \frac{e^{(\Phi^{-1}(\dot{1-t}))^2/2}}{\frac{1}{\pi}}$$

der zweite e-Faktor kommt von der inversen Ableitung, wir fahren also fort

$$\begin{aligned} F^{'}(t) &= e^{-\frac{(g-\theta)^2}{2}} \dot{e}^{\frac{g^2}{2}} \\ &= e^{-\frac{g^2}{2} + g\theta - \frac{\theta^2}{2}} e^{\frac{g^2}{2}} \\ &= e^{g\theta - \frac{\theta^2}{2}} \end{aligned}$$

also bleibt ausgeschrieben über

$$e^{-\frac{\theta^2}{2} + \theta\Phi^{-1}(1-t)}$$

In der obigen linearen Taylorentwicklung untersuchen wir, wann diese Dichte > 1 ist.

$$\begin{aligned} e^{-\frac{\theta^2}{2} + \theta\Phi^{-1}(1-t)} &\geq 1 \\ -\frac{\theta^2}{2} + \theta\Phi^{-1}(1 - t) &\geq 0 \\ -\frac{\theta}{2} + \Phi^{-1}(1 - t) &\geq 0 \end{aligned}$$

was zu der Bedingung

$$\Phi^{-1}(1-t) \geq \theta$$

führt. Der interessierende Wert t^0 ist nicht größer als α, also bei $\alpha = 0.05$ ist die Linke Seite so bei 2, also gilt obige Beziehung sicher bis $\theta = 4$.
Die Taylorbeziehung haben wir ja nur bis zum linearen Term verwendet. Obige Rechnung sind also nur ein Hinweis, wenn auch ein halbwegs plausibler für die Machtbeziehungen.

3.10 Binäre Gewichtung

Wir denken uns ein Quadrat in dem gitternetzartig viele kleine Quadrate eingebettet sind. Die Natur kennzeichnet jedes Quadrat mit einer H_0-Farbe, falls die Nullhypothese wahr ist und mit einer H_1-Farbe falls die Alternativhypothese wahr ist. Diese Färbung ist exakt nur der Natur bekannt.
Nun kommt ein Experte und färbt ebenfalls die Quadrate an, mit einer E_0-Farbe, falls seiner Meinung nach die Nullhypthese wahr ist und mit einer E_1-Farbe, falls die Alternative wahr ist.
Die Wahrscheinlichkeit eines der vielen kleinen Quadrate auszuwählen sei durch die diskrete Gleichverteilung gegeben.
Wir vereinbaren folgende Bezeichnung

$$|E_i H_j| = \text{ Anzahl der } H_j \text{ Farbkästchen, die auch } E_i \text{ Farbkästchen sind}$$

analog bedeutet auch $|H_j|$ Anzahl der H_j-Farbkästchen.
Folgende relative Häufigkeiten sind zwar nicht beobachtbar, aber der Natur bekannt (wir schreiben sie dann auch gleich als Wahrscheinlichkeit):

$$\frac{|E_0 H_0|}{|H_0|} = P(E_0|H_0)$$

analog haben wir auch $P(E_1|H_0), P(E_0|H_1)$ und $P(E_1|H_1)$.
Weiters schreiben wir auch

$$a = (\text{ relativer Anteil der } H_1 \text{ Farkästchen}) = P(H = 1)$$

Diese Größe ist auch nicht direkt beobachtbar.
Die Randwahrscheinlichkeiten $P(E_i)$ sind nun sowohl beobachtbar als auch durch die nichtbeobachtbaren Größen darstellbar.

$$\begin{aligned} P(E_0) &= P(H_0)P(E_0|H_0) + P(H_1)P(E_0|H_1) \\ &= (1-a)P(E_0|H_0) + aP(E_0|H_1) \end{aligned}$$

und natürlich

$$P(E_1) = 1 - P(E_0)$$

Nun führen wir noch Indikatorvariablen für die Färbungen des Experten ein Die i laufen über alle kleine Kästchen und die U_i sind Bernoulli verteilte

$U_i = 1$	Wenn Experte Kästchen mit Nummer i auf E_1 färbt
$U_i = 0$	Wenn Experte Kästchen mit Nummer i auf E_0 färbt

Tabelle 3.3:

Variablen und die Wahrscheinlichkeiten

$$P(U_i = k), P(U_i = k|H_j) \quad k = 0, 1$$

sind die gleichen, wenn man $U_i = k$ durch E_k ersetzt (wir schreiben auch $\gamma = P(U_i)$).
Für eine zufällige Auswahl von m Hypothesen kann der Experte für jedes der ausgewählten Kästchen eine Gewichtung angeben

$$W_i = \frac{S_i}{\bar{S}_m} = \frac{1 + (r-1)U_i}{1 + (r-1)\bar{U}_m}$$

wobei

$$\bar{U}_m = \frac{1}{m} \sum_i U_i$$

bedeuten soll.
Die Größe $1 + (r-1)U_i$ kann für alle Kästchen vom Experten a priori angegeben werden [3]. Die Normierung erfolgt dann zur Zeit des Experiments [4].
Analog $\bar{S}_m$ (es gilt $S_i = 1 + (r-1)U_i$ und $\bar{S}_m = 1 + (r-1)\bar{U}_m$).
Es gilt natürlich

$$\frac{1}{m} \sum (1 + (r-1)U_i) = (1 + (r-1)) \frac{1}{m} \sum U_i = (1 + (r-1))\bar{U}_m$$

und damit

$$\frac{1}{m} \sum_{i=1}^{m} \frac{1 + (r-1)U_i}{1 + (r-1)\bar{U}_m} = \frac{1 + (r-1)\frac{1}{m}\sum_{i=1}^{m} U_i}{1 + (r-1)\bar{U}_m} = 1$$

[3] r ist vorläufig ein weiterer Parameter, seine Interpretation wird weiter unten gegeben
[4] unter ¨Experiment¨ verstehen wir hier die Durchführung der m Hypothesentests

Wir bestimmen nun die bedingte Wahrscheinlichkeitsverteilung der W_I (Wir bezeichnen noch mit $S_{\bar{i}}$ die Summe über alle S_j mit $j \neq i$).
Weiters bemerken wir noch, daß die $S_{\bar{i}}$ wegen der gleichen Verteilung der S_j für alle i dieselbe Verteilung haben.

$$\begin{aligned} P(W_i \leq t|H_1) &= P(\frac{S_i}{\bar{S}_m} \leq t|H_1) \\ &= P(S_i \leq t(S_i + S_{\bar{i}})|H_1)) \\ &= P(S_i(1-t) \leq tS_{\bar{i}}|H_1) \\ &= P(S_i \leq \frac{t}{1-t}S_{\bar{i}}|H_1) \\ &= M_1(t) \end{aligned}$$

analog für die Bedingung auf H_0. Wir betrachten nun die Größe

$$\eta = \frac{P(U=1|H=1)}{P(U=1|H=0)}$$

Ist der Experte gut, so wird η gross > 1 sein, ist er schlecht so wird $\eta < 1$ sein. Ist $\eta = 1$, so könnte man auch würfeln.
Wir haben nun folgende beide Formeln für die bedingte Wahrscheinlichkeit der U (wir lassen wenn möglich den Index weg).

$$P(U=1|H=1) = \frac{\eta\gamma}{a\eta + 1 - a}$$

sowie

$$P(U=1|H=1) = \frac{\gamma}{a\eta + 1 - a}$$

Hierbei gilt $P(U=1) = \gamma$.
Wir haben nun folgendes:

1. Zuerst zeigen wir, dass wenn man von obigen beiden Formeln ausgeht, γ wirklich gleich $P(U=1)$ ist.
2. Dann leiten wir noch obige Formeln ab unter der Annahme, daß $P(U=1) = \gamma$ gilt.

Punkt 1):
Man muss dazu nur die Definition der bedingten Wahrscheinlichkeit verwenden und beide obigen Gleichungen

$$P(U=1|H=1) = \frac{P(U=1 \cap H=1)}{a} = \frac{\eta\gamma}{a\eta + 1 - a}$$

ebenso die andere Gleichung und damit gilt

$$P(U=1) = P((U=1 \cap H=1) \cup (U=1 \cap H=0)) = P(U=1 \cap H=1) + P(U=1 \cap H=0)$$

$$\begin{aligned} P(U=1) &= \frac{\eta\gamma}{a\eta+1-a}a + \frac{\gamma}{a\eta+1-a}(1-a) \\ &= \frac{\gamma\eta a + \gamma(1-a)}{a\eta+1-a} = \frac{\gamma(a\eta+1-a)}{a\eta+1-a} = \gamma \end{aligned}$$

Punkt 2):
Wir verwenden die Gleichung von η und erhalten dann

$$P(U=1|H=1) = \eta P(U=1|H=0) = \eta\frac{P(U=1 \cap H=0)}{1-a} = \frac{P(H=0 \cap U=1)}{1-a}$$

$$= \eta\frac{P(H=0|U=1)}{1-a}P(U=1)$$

Es bleibt nun zu zeigen, daß der mittlere Term gleich $\frac{1}{a\eta+1-a}$ ist.
Wir betrachten also, die obige Gleichung fortführend, wir wollen zeigen ,dass

$$P(H=0|U=1)(a\eta+1-a) = 1-a$$

gilt. Wir haben also

$$\begin{aligned} & P(H=0|U=1)(a\eta+1-a) \\ = \;& a\eta P(H=0|U=1) + P(H=0|U=1) - aP(H=0|U=1) \\ = \;& a\frac{P(U=1|H=1)}{P(U=1|H=0)}P(H=0|U=1) + P(H=0|U=1) - aP(H=0|U=1) \\ = \;& a\frac{P(U=1|H=1)}{P(U=1 \cap H=0)}P(H=0)\frac{P(H=0 \cap U=1)}{P(U=1)} \\ & + P(H=0|U=1) - aP(H=0|U=1) \\ = \;& a(1-a)\frac{P(U=1|H=1)}{P(U=1)} + P(H=0|U=1) - aP(H=0|U=1) \\ = \;& (1-a)(a\frac{P(U=1|H=1)}{P(U=1)} + P(H=0|U=1)) \\ = \;& (1-a)(\frac{P(U=1 \cap H=1)}{P(U=1)} + \frac{P(H=0 \cap U=1)}{P(U=1)}) \\ = \;& (1-a)\frac{P(U=1)}{P(U=1)} = 1-a \end{aligned}$$

und damit stimmt die obige Gleichung.
Die Gewichte sind

$$W_i = \frac{1 + (r-1)U_i}{1 + (r-1)\bar{U}_m}$$

Klarerweise folgt aus

$$w_0 = \frac{1}{\text{Nenner}} \ , \ w_1 = \frac{1 + r - 1}{\text{Nenner}}$$

daß

$$r = \frac{w_1}{w_0}$$

gilt [5].
Die Güte von U_i, d.h. vom obigen η liefert der Experte, r unser Glauben an den Experten. Ein großes r bedeutet, daß w_1 groß gegenüber w_0 ist und damit P_i/W_i für $H_i = 1$ sehr klein und P_i/W_i für $H_i = 0$ sehr groß wird.

3.11 Literaturhinweise

Dieses Kapitel beruht auf der Arbeit von [**?**]. Die Berechnung der Funktion $R(t, \lambda)$ sowie die Ableitung von $f(\lambda) = t^\lambda$ an der Stell $\lambda = 0$ wird in der zitierten Arbeit nicht angeführt. Meiner Meinung nach haben sich die Autoren leicht geirrt, was die Anwendbarkeit dieser Methode ein wenig einschränkt. Sie ist dennoch hier angeführt, da die Idee interessant ist und vielleicht in anderem Zusammenhang nützlich sein kann.

Weitere Informationen über gewichtete Prozeduren findet man in [2], wo unter anderem etwa eine gewichtete FDR (WFDR) definiert wird ($w = (w_1, w_2, \ldots, w_m), \sum_{i=1}^{i=m} w_i = m$)

$$Q(w) = \frac{\sum_{i=1}^{i=m} w_i V_i}{\sum_{i=1}^{i=m} w_i R_i} \cdot 1_{\{\sum_{i=1}^{i=m} w_i R_i > 0\}}$$

und eine Aussage analog der zur FDR in Kapitel 2 (siehe Abschnitt 2.1.3) bewiesen wird.
Viel Interessantes über die FDR wird sicherlich im von mir noch nicht eingesehenen Buch [4] zu finden sein (das Buch ist noch nicht erschienen).

[5] w_1 ist die Größe des Gewichts, wenn der Experte die Alternative für wahr hält ($U_i = 1$). w_0 ist die Größe des Gewichts, wenn der Experte die Nullhypothese für wahr hält($U_i = 0$).

Kapitel 4

Gewichtete Prozeduren (a posteriori Gewichtung)

Prozeduren bei denen aus den Daten mehr als nur der P-Wert verwendet wird, werden vorgestellt. Die ¨Umreihung¨ der P-Werte erfolgt entweder durch Kenngrößen der empirischen Verteilung (Median, Interquantilwert und Quadratsummen) beziehungsweise durch daraus abgeleitete Gewichte. Sowohl parametrische als auch nichtparametrische Methoden werden vorgestellt.

4.1 Einleitung

Wenn es keine guten Priorinformationen über die Richtigkeit und Nichtrichtigkeit der Hypothesen gibt, welche dazu dienen, siehe das vorige Kapitel, Gewichte zu bestimmen, wäre es gut, wenn man aus den Daten gewisse Informationen für die Bestimmung der Gewichte ziehen könnte.
Man denke etwa an folgende Prozedur.

1. Ordne die Variablen nach einem gewissen Kriterium (z.B abfallend nach dem größten Stichprobenwert jeder Variablen)
2. Wähle die extremste (also im obigen Fall die mit dem grössten Stichprobenwert) X_{i_0} aus
3. Führe für die Variable X_{i_0} einen Test mit der Teststatistik T aus.
4. Falls X_{i_0} auf Niveau α nicht signifikant ist, höre auf.

5. Andernfalls mache weiter in beliebiger Reihung bis ein nichtsignifikantes Ereignis eintritt (die Reihung ist für die Macht wichtig, für die Einhaltung des α Fehlers ist sie egal.)

Kurz gesagt bringen wir die Variablen zuerst in eine Reihung und führen dann einen hierarchischen Test durch.
Das Problem hier ist, dass wir auch bei Unabhängigkeit der Zufallsgrössen nicht wissen ob die Teststatistik bei dieser Datengesteuerten Auswahl wirklich die erfordeliche Verteilung hat.

4.2 Ungewichtete Bonferroni und Bonferrroni-Holm Prozeduren (Die Reihung der Hypothesen wird aus den Daten bestimmt)

Wir behandeln in diesem Abschnitt ungewichtete Prozeduren, wobei die Reihung der Hypothesen nicht wie in anderen Fällen nur durch die P-Werte oder durch a priori bestimmte Gewichte W und die damit ausgerechneten Q-Werte ($Q = P/W$) erfolgt sondern durch $Q = P/W$-Werte, wobei die Gewichte W aus den Daten berechneten werden, bestimmt wird.

4.2.1 Prozeduren normalverteilter Größen

Wir betrachten jetzt den Fall normalverteilter Grössen.
Wir haben n unabhängige p-dimensionale normalverteilte Vektoren x_j. Es gilt also

$$x_j \sim N_p(\mu, \Sigma)$$

wobei wie üblich μ der Mittelwertsvektor und Σ die Varianzkovarianz-Matrix der Zufallsgröße x_j ist und p die Anzahl der Variablen angibt.
Wir betrachten nun die globale Nullhypothese

$$H : \mu = 0$$

Weiters fassen wir die n Beobachtungsvektoren zu einer $n \times p$ Matrix X zusammen, wobei in den Zeilen also jeweils die p beobachteten Variablen

und in den Spalten die n Beobachtungen der jeweiligen Variablen stehen.

$$X = \begin{pmatrix} x_1^{'} \\ \cdot \\ \cdot \\ \cdot \\ x_n^{'} \end{pmatrix}$$

Man kann nun die Matrix X als einen $n \cdot p$-dimensionale normalverteilte Grösse auffassen.

$$X \sim N_{n \cdot p}(1_n \cdot \mu^{'}, I_n \otimes \Sigma)$$

- 1_n ist der (Spalten) n-Vektor aus lauter Einsen.
- $\otimes$ bezeichnet das sogenannte Kroneckerprodukt

$$A_{m \times n} \otimes B_{r \times s} = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \cdot & \cdot & \cdot \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix}$$

- I_n ist die n-dimensionale Einheitsmatrix

$1_n \cdot \mu^{'}$ hat jeweils μ als Beobachtungsvektor.
$I_n \otimes \Sigma$ hat also wegen der Unabhängigkeit der Beobachtungsvektoren die erwartete Blockmatrixstruktur.
Es werden nun zur Dimensionsreduktion Scorevektoren erzeugt, d.h. wir bilden die eindimensionalen Grössen

$$z_j = d^{'} \cdot x_j \quad (j = 1, \ldots, n)$$

Diese Scorevektoren dürfen von den Daten abhängen.
Wir zitieren nun ohne Beweis den Satz von Läuter [16]

Theorem 5 *X sei wie oben. Es wird weiters die Gültigkeit der Nullhypothese $H : \mu = 0$ angenommen.*
Es sei weiters d ein p-dimensionaler Gewichtsvektor, der nur von der Grösse (einer Matrix)

$$W = X^{'}X$$

abhängt.
Ausserdem sei Xd mit Wahrscheinlichkeit 1 nicht gleich dem Nullvektor.

Es gilt nun:
Es sei

$$\bar{z} = \frac{1}{n} z \cdot 1_n \text{ also der Mittelwert der } z_j \text{ Grössen}$$

und

$$s^2 = \frac{1}{n-1}(z^{'} z - n\bar{z}^2) \text{ also die empirische Varianz der } z_j \text{ Grössen}$$

Es gilt nun:
Die Grösse

$$t = \frac{\sqrt{n}\bar{z}}{s}$$

ist dann tatsächlich t-Verteilt (mit $n-1$ Freiheitsgraden) unter Annahme der globalen Nullhypothese.

Bemerkung:
Die t-Größe wird mit Hilfe des datenabhängigen Gewichtsvektors d berechnet. Es kann also nicht erwartet werden, daß sie für beliebige Gewichtsvektoren t-verteilt ist.
Beispiel:
Wir verwenden nur die Diagonalelemente von $W = X^{'}X$, also

$$w_{ii} = \sum_{j=1}^{n} x_{ji}^2 (j = 1, \ldots, n)$$

Wir ordnen dann die Hypothesen nach abfallenden w_{ii}.
Sei nun i_0 das i mit dem größten w_{ii}. Wir bilden dann den Gewichtsvektor $d = (0, \cdots, i_0, \cdots, 0)$ mit 1 an i_0 ter Stelle. Das Bildungsgesetzt dieses Gewichtsvektors hängt nur von W ab. Die t-Größe der $i_0$ten Variablen ist dann gemäß dem Satz von Läuter wirklich t-verteilt (obwohl i_0 nicht konstant ist). Diese Konstruktion kann man auch mit der Variablen mit dem 2. größten, der 3. größten, etc. w_{ii} Wert vornehmen.
Eine Prozedur analog wie in der Einleitung (Abschnitt 4.1) beschrieben ist also möglich.

4.2.2 Prozeduren für beliebig verteilte Größen

Wir betrachten den Einstichprobenfall, p Variablen werden jeweils n mal beobachtet, wobei die p Variablen eine p-dimensionale symmetrische stetige

Dichte haben.
Hier die Bezeichnungen:

$$x_j = \begin{pmatrix} x_{j1} \\ \cdot \\ \cdot \\ \cdot \\ x_{jp} \end{pmatrix}$$

also die einmalige Beobachtung der p Variablen.
Der Vektor, um den die Verteilung symmetrisch ist, werde mit

$$\mu = \begin{pmatrix} \mu_1 \\ \cdot \\ \cdot \\ \cdot \\ \mu_p \end{pmatrix}$$

bezeichnet.
Zu Testen sind nun die Hypothesen

$$H_i : \mu_i = 0$$

für $i = 1, \ldots, p$.
Wir versuchen nun eine Reihung der Variablen sowie eine Testgröße zu finden, deren Verteilung wir berechnen können.
Wir betrachten nun jede Variable $x_{\cdot i}$ einzeln. Wir haben dann n Beobachtungswerte

$$x_{1i}, x_{2i}, \ldots, x_{ni}$$

Wir ordnen nun die Absolutwerte dieser n Werte aufsteigend

$$|x_{(1)i}| \leq |x_{(2)i}| \leq \ldots \leq |x_{(n)i}|$$

Es ist dadurch eine Ordnungsfunktion r_{ji} durch

$$x_{ji} = x_{(r_{ji})i}$$

definiert (z.B: i halten wir fest und lassen es gleich in der Notation weg; x_7 sei etwa das drittgrößte also

$$x_7 = x_{(3)} = x_{(r_7)},$$

also gilt $r_7 = 3$).
Der Median der absoluten Beobachtungswerte ist nun durch

$$m_i = \begin{cases} |x_{((n+1)/2)i}| & \text{für ungerade } n \\ (|x_{(n/2)i}| - |x_{(n/2+1)i}|)/2 & \text{für gerade } n \end{cases}$$

gegeben.
Der Median der absoluten Beobachtungswerte dient zum Ordnen der Variablen.
Als Teststatistik nehmen wir dieselbe Größe wie beim einstichproben Wilcoxontest, also die Teststatistik

$$W_i = \sum_{i=1}^{n} r_{ji} \text{pos}(x_{ij})$$

wobei

$$\text{pos}(x) = \begin{cases} 1 & \text{wenn } x > 0 \\ 0 & \text{sonst} \end{cases}$$

gilt (d.h. es werden die Rangnummern der positiven Werte addiert). Da die Dichten als stetig vorrausgesetzt wurden, machen wir uns keine Gedanken über Bindungen.
Die (multiple) Testprozedur ist nun folgendermaßen definiert:
Prozedur II:

1. Sortiere die Variablen x_j nach absteigenden Medianwerten der Absolutwerte.

2. Führe gemäß der durchgeführten Ordnung in Punkt (1) den Einstichproben Wilcoxontest zum Niveau α durch, solange das Testergebnis ein signifikantes Resultat liefert.
   Stoppe die Prozedur bei dem ersten nichtsignifikanten Ergebnis. Genau die bisher abgelehnten Tests bilden die Ablehnmenge

Theorem 6 *Die eben beschriebene Prozedur hält den familienweiten α Fehler ein*

Beweis
Wir führen den Beweis in zwei Teilen [15]

1. Wir betrachten den Vektor y aller Variablen wo die Nullhypothese wahr ist.
   Wir zeigen in diesem ersten Schritt dann, daß die Wilcoxon-Teststatistik der Variablen mit dem größten Median der Absolutwerte den familienweiten α Fehler (FWER) einhält.

2. In diesem Schritt folgern wir, daß die Prozedur insgesamt den familienweiten α Fehler (FWER) einhält (genauer: Das Experiment wird durchgeführt, die Variable mit dem größten Median der Absolutwerte eruiert, die Summe der Ränge der positiven Werte gebildet. Ist dieser Wert dieser Teststatistik größer als der α-Grenzwert der Wilcoxonstatistik, so wird die Hypothese abgelehnt. Diese Prozedur hält dann den α-Fehler ein).

Die Beweisidee (sowie auch in späteren Abschnitten) ist, den Stichprobenraum so zu partitionieren, daß die Aussage auf jeder Partition gilt und dann (sinngemäß) den Satz von der totalen Wahrscheinlichkeit anzuwenden.
1) Wir teilen den Beobachtungsraum in Invarianzklassen ein.
Dabei heißen zwei Stichproben (die y und z sind jeweils p-Vektoren, p ist die Anzahl der Variablen)

$$y_1, \ldots, y_n$$

und

$$z_1, \ldots, z_n$$

äquivalent, wenn

$$y_1 = v_1 z_1, \ldots, y_n = v_n z_n$$

gilt, wobei die v_j gleich -1 oder 1 sein können.
Es gibt genau 2^n solcher Vektoren $v = (v_1, \ldots, v_n)$.
Da wir die Stichproben unabhängig ziehen und die Dichten unter der Nullhypothese als symmetrisch bezüglich des Nullpunktes vorrausgesetzt sind, haben wir für die Dichte der gemeinsamen Verteilung der y bzw der z (wir befinden uns jetzt im $n \cdot p$ dimensionalem Raum)

$$\begin{aligned} f_{n \cdot p_0}(y_1, \ldots, y_n) &= f_{p_0}(y_1) \ldots f_{p_0}(y_n) \\ &= f_{p_0}(v_1 z_1) \ldots f_{p_0}(v_n z_n) \\ &= f_{p_0}(z_1) \ldots f_{p_0}(z_n) \\ &= f_{n \cdot p_0}(z_1, \ldots, z_n) \end{aligned}$$

Die bedingte Verteilung einer Stichprobe, bedingt auf seine Invarianzklasse ist dann die diskrete Gleichverteilung mit

$$P(\cdot) = \frac{1}{2^n}$$

Man denke etwa an den diskreten Fall. Jedes der y hat die gleiche absolute Wahrscheinlichkeit w, also ist

$$P(y|A) = \frac{P(y \cap A)}{P(A)} = \frac{w}{\sum_{|A|} w} = \frac{1}{2^n}$$

da ja die Äquivalenzklassen aus 2^n Elementen besteht.
Wir haben also: Für eine fixierte Invarianzklasse sind die Ränge und der Median der Absolutwerte der Variablen auch fixiert, da sich die Beobachtungswerte nur jeweils im Vorzeichen unterscheiden.
Es sei nun i der Index der Variablen mit dem größten Median. Wenn man dann die Wilcoxon Statistik für alle Beobachtungspunkte der Invarianzklassen bildet erhält man genau die Nullverteilung der Wilcoxon Statistik (Die Absolutwerte sind ja gleich, nur im Vorzeichen unterscheiden sie sich. In der Äquivalenzklasse kommen alle Vorzeichenkombinationen vor. Da die bedingte Verteilung die Gleichverteilung ist, haben alle diese 2^n Möglichkeiten dieselbe Wahrscheinlichkeit und diese Konstruktion entspricht genau der Konstruktion der Nullverteilung der Wilcoxonteststatistik).
Da diese Verteilung dieselbe für alle Invarianzklassen ist, gilt dies auch für die unbedingte Verteilung.
(dies ist der Satz von der totalen Wahrscheinlichkeit $P(A) = \sum_n P(A|B_n)P(B_n)$ (B_n eine Partition des Grundraumes)
In diesem Fall ist also die bedingte Verteilung $P(A|B_n)$ gleich der unbedingten Verteilung $P(A)$.
Diese Konstruktion kann aber genauso für die Variable mit dem 2. größten, 3. größten, etc Median der Absolutwerte machen.
2) Bringt man nun im allgemeinen Fall eine Ordnung in die Variablen gemäß den fallend geordneten Medians der Absolutwerte der Beobachtungspunkte, so lehnt man höchstens zusätzlich richtige Alternativen ab.

4.3 Gewichtete Bonferroni und Bonferrroni-Holm Prozeduren

4.3.1 Parametrische Prozeduren

Prozedur III
Diese Prozedur wird wie im nächsten Kapitel durchgeführt, nur als Test wird der t-Test genommen und als Gewichte wieder die w_{ii}, bzw die ${w_{ii}}^h$ mit geeignetem h.

4.3.2 Nichtparametrische Prozeduren (Einstichprobenfall)

Wir beginnen gleich mit der Definition der gewichteten Prozedur im Einstichprobenfall.
Prozedur IV[1]:

1. Für jede der p Variablen führe den Einstichproben Wilcoxontest zum Niveau α durch.

2. Für jede der p Variablen x_i bestimme man den Median m_i der Absolutwerte und bestimme dann die Gewichte

   $$g_i = {m_i}^{\eta}$$

   für ein fixes $\eta \geq 0$

3. Man berechne die gewichteten P-Werte

   $$Q_i = \frac{P_i}{g_i}$$

   und ordne diese aufsteigend

   $$Q_{i_1} \leq Q_{i_2} \leq \cdots \leq Q_{i_p}$$

   Weiters definiere man die Indexmengen

   $$S_j = \{i_j, i_{j+1}, \ldots, i_p\}$$

   $(j = 1, \ldots, p)$

[1]Es handelt sich wieder um die gewichtete Bonferroni-Holm Prozedur, im Gegensatz zu der Prozedur in Abschnitt 3.7 werden hier die Gewichte nicht a priori bestimmt, sondern aus den Daten gewonnen

4. Die entsprechenden Hypothesen H_{i_j} werden dann abgelehnt, solange

$$Q_{i_j} \leq \frac{\alpha}{\sum_{h \in S_j} g_h}$$

gilt. Man höhre auf, wenn das erste Mal diese Ungleichung nicht gilt.

Wir beweisen nun, daß diese Prozedur den Familienweiten Fehler α einhält, wieder mit der Idee einer Partition wie im vorigen Kapitel.
Zuerst definieren wir noch eine Menge, und zwar

$$S_0 = \{ \text{ Subskripte aller Variablen, deren Erwartungswerte} = 0 \text{ ist} \}$$

Wir bezeichnen nun das Subskript der ersten wahren Nullhypothese der nach gewichteten P-Werten geordneten Hypothesen mit i_0.
Wir müssen nun zeigen, daß die Variable x_{i_0} mit höchstens der Wahrscheinlichkeit α abgelehnt wird.
Dazu bemerken wir, daß

$$S_{i_0} \supseteq S_0$$

gilt (i_0 ist ja die ¨erste¨ abgelehnte Nullhypothese, S_{i_0} enthält alle Indizi der wahren Nullhypothesen und allenfalls solche wahrer Alternativen).
Wir haben nun folgende Ungleichungen (die Wahrscheinlichkeiten werden wieder auf die Äquivalenzklassen wie im vorigen Abschnitt bedingt [2])

$$\begin{aligned}
P(Q_{i_0} \leq \frac{\alpha}{\sum_{h \in S_{i_0}} g_h}) &\leq P(Q_{i_0} \leq \frac{\alpha}{\sum_{h \in S_0} g_h}) \\
&= P(\min_{i \in S_0} Q_i \leq \frac{\alpha}{\sum_{h \in S_0} g_h}) \\
&= P(\cup_{i \in S_0} (Q_i \leq \frac{\alpha}{\sum_{h \in S_0} g_h})) \\
&= P(\cup_{i \in S_0} (P_i \leq \frac{g_i \alpha}{\sum_{h \in S_0} g_h})) \\
&\leq \sum_{i \in S_0} P(P_i \leq \frac{g_i \alpha}{\sum_{h \in S_0} g_h}))
\end{aligned}$$

Obige Ableitungen verwenden eigentliche nur, daß die Q_i der Größe nach geordnet sind.

[2] Wir erinnern: $(y_1, \ldots, y_n)$ und $(z_1, \ldots, z_n)$ heißen äquivalent, wenn $y_1 = v_1 z_1, \ldots, y_n = v_n z_n$ für $v_i = 1$ oder $v_i = -1$ gilt

1. Die erste Ungleichung ist klar, da die Summe im Nenner rechts weniger Summanden als die Linke hat, der Nenner damit kleiner und der Ausdruck rechts damit größer wird.

2. Die erste Gleichung folgt aus der Tatsache, daß

$$Q_{i_0} \leq Q_i(\text{ für alle } i \in S_0)$$

gilt.

3. Die zweite Gleichung folgt ebenso aus der Monotonie der Q_i

4. In der dritten Gleichung wird einfach in jeder Ungleichung mit g_i multipliziert.

5. Die letzte Ungleichung folgt aus der bekannten allgemeinen Formel

$$P(\cup A_i) = \sum P(A_i)$$

Es gilt nun

$$P(P_i \leq c) \leq c$$

denn P_i ist ein P-Wert, der aber nicht unbedingt stetig verteilt ist. und damit haben wir weiters

$$\begin{aligned} P(Q_{i_0} \leq \frac{\alpha}{\sum_{h \in S_{i_0}} g_h}) &\leq \sum_{i \in S_0} P(P_i \leq \frac{g_i \alpha}{\sum_{h \in S_0 g_h})} \\ &\leq \sum_{i \in S_0} \frac{\alpha g_i}{\sum_{h \in S_0} g_h} = \alpha \end{aligned}$$

Dies gilt für die bedingte Wahrscheinlichkeit auf allen Äquivalenzklassen, also auch unbedingt.
Schlussendlich kann wieder passieren, daß zusätzlich noch wahre Alternativen abgelehnt werden, was ja für die Power sowieso gut ist, und nichts an der Einhaltung des FWER ändert. **QED**

4.3.3 Nichtparametrische Prozeduren (Zweistichprobenfall)

Wir müssen wieder untersuchen, wie die P-Werte und Gewichte bestimmt werden und ob die Beweise vom Einstichprobenfall (problemlos) adaptiert

werden können.
Wir haben folgendes Szenario:
Wir haben zwei Stichproben

$$x_j^{(k)} = \begin{pmatrix} x_{j1}^k \\ \cdot \\ \cdot \\ \cdot \\ x_{jp}^k \end{pmatrix} \quad (k = 1, 2; j = 1, \cdots, n_k)$$

d.h. k gibt die Stichprobe an und n_1 die Anzahl der Probanden in der ersten und n_2 die Anzahl der Probanden in der zweiten Stichprobe.
Die Annahme ist nun, daß die Stichprobe aus Grundgesamtheiten kommen, mit

$$f_p^2(x) = f_p^1(x + \mu)$$

mit $\mu = (\mu_1, \cdots, \mu_p)'$.
Die Nullhypothesen sind nun

$$H_i : \mu_i = 0$$

Wir fassen für jede Variable i die Werte aus beiden Beobachtungsgruppen zusammen

$$x_{1i}^{(1)}, \cdots, x_{n_1 i}^{(1)}, x_{12}^{(2)}, \cdots, x_{n_2 i}^{(2)}$$

Es sei nun $n = n_1 + n_2$. Wir ordnen die obigen Variablen nach der Größe

$$x_{(1)i} \leq x_{(2)i} \leq \ldots \leq x_{(n)i}$$

Dann haben wir nunmehr

1. P-Werte: Wilcoxon 2-stichproben Test
2. Gewichte

$$g_i = x_{(0.75(n+1))i} - x_{(0.25(n+1))i}$$

Ähnlich wie in den vorherigen Beweisen werden wir versuchen geeignete Invarianzklassen zu finden.
Dieses Mal betrachten wir die n Stichprobenelemente (jedes dieser Stichprobenelemente ist ein p-Vektor, wobei p die Anzahl der Variablen ist).
Wir haben also $n_1 + n_2 = n$ Stichprobenelemente

$$y_1^{(1)}, \ldots, y_{n_1}^{(1)}, y_1^{(2)}, \ldots, y_{n_2}^{(2)}$$

Vorrausgesetzt wird die Gültigkeit der Nullhypothese (beide Gruppen unterscheiden sich nicht im Mittelwert).
Wir halten nun so eine n Stichprobe fest.
Wir sagen dann, daß alle n-Stichproben, deren Elemente eine Permutation der festgehaltenen sind, eine Äquivalenzklasse bilden. Wegen der Stetigkeit der Verteilungen schließen wir identische Elemente in der Stichprobe aus.
Wegen der Unabhängigkeit der Stichprobenelemente hat jeder dieser $n!$ Elemente dieselbe bedingte Wahrscheinlichkeit (bedingt auf die zugehörige Äquivalenzklasse).
Die Variable mit dem größten Interquantilabstand ist für alle Elemente der Äquivalenzklasse dieselbe (die n Werte für jeder der p Variablen werden ja ine aufsteigende Reihung gebracht und diese ist für alle Elemente der Äquivalenzklasse dieselbe).
Der Rest folgt dann wieder wie vorher. Auf der Äquivalenzklasse sind die Elemente wieder bedingt gleichverteilt, sodaß die bedingte Testverteilung gleich der unbedingten Wilcoxonteststatistik ist. **QED**

4.4 Gewichtete FDR (Gewichte aus den Daten bestimmt)

Wir behandeln nun die Benjamini-Hochberg (BH) Prozedur bei datengesteuerten Gewichten. Die Gewichte und natürlich die P-Werte werden wie in den vorigen Abschnitten bestimmt.
Mit den

$$Q_i = \frac{P_i}{G_i}$$

Werten wird dann die BH-Prozedur durchgeführt. Der einfacheren Referenz wegen schreiben wir die Prozedur noch einmal explizit hin:

4.4.1 Parametrische Prozedur

Parametrische gewichtete BH-Prozedur

1. Bestimme P-Werte P_i (mit Hilfe des t-Tests für jede Variable)

2. Bestimme aus den Daten jeweils die w-Werte

$$w_i = \sum_{j=1}^{n} x_{ji}^2$$

3. Bestimme für ein festes η

$$G_i = w_i^{\eta}$$

4. Normieren (Mittelwert der Gewichte = 1)

$$\begin{aligned} mG &= \frac{\sum_i G_i}{n} \\ g_i &= \frac{G_i}{mG} \end{aligned}$$

5. Bilde

$$Q_i = \frac{P_i}{g_i}$$

6. Ordne die Q_i

$$Q_{i_1} \leq Q_{i_2} \leq \cdots Q_{i_n}$$

7. Bestimme das maximale j, sodass

$$Q_{i_j} \leq \frac{j}{n}\alpha$$

8. Lehne alle

$$Q_{i_1} \leq Q_{i_2} \leq \cdots Q_{i_j}$$

ab.

9. Gilt für alle j

$$Q_{i_j} > \frac{j}{n}\alpha$$

lehne keine Hypothese ab.

4.4.2 Nichparametrische Prozedur

Diese wird genause wie die parametrische DBH-Prozedur durchgeführt, bis auf die ersten beiden Punkte

1. Führe für jede Variable den Wilcoxon Test durch
2. Bestimme die w durch

$$w_i = \text{median}(|x_{1i}|, |x_{2i}|, \ldots, |x_{ni}|)$$

4.5 Simulationen

Bemerkungen zu den verwendeten Prozeduren

- *Kropf Prozedur*: Damit bezeichnen wir die gewichtete Bonferroni-Holm Prozedur, wenn die Gewichte wie in den vorigen Abschnitten dargelegt aus den Daten berechnet werden (Median der Absolutwerte im Einstichprobenfall, Interquantilabstand im Zweistichprobenfall) und der FWER kontrolliert wird.
- *(Gewichtete) Benjamini-Hochberg Prozedur*: Damit bezeichnen wir die im vorigen Kapitel (siehe Seite 36) definierte gewichtete Benjamini-Hochberg Prozedur. Die Berechnung der Gewichte erfolgt wie bei der Kropf Prozedur.
- *Parametrisch, nichtparametrisch*: Wenn wir sagen, daß eine parametrische Prozedur durchgeführt wurde meinen wir, daß zur Bestimmung der P-Werte ein parametrischer Test (in unserem Fall der t-Test) verwendet wurde. Mit nichtparametrischer Prozedur meinen wir analog, daß zur Bestimmung der P-Werte eine nichtparametrische Prozedur (in unserem Fall der Wilcoxon-Test) verwendet wurde.
- Für die Fallzahlen der Simulationen haben wir uns zur besseren Vergleichbarkeit an die Arbeit von [14] angelehnt (in runden Klammern die Kurzbezeichnungen).

  - Anzahl der Nullhypothesen (m_0): 12525.
  - Anzahl der Alternativhypothesen (m_1): 100

- Mittelwert der Nullhypothese (μ_0): 0.
- Mittelwert der Alternativhypothese (m_1): 1.15
- Varianzen gleich (σ_0 für die Varianz der Nullhypothese, σ_1 für die der Alternative, bzw nur σ falls diese gleich sind)1.
- Normalverteilte Größen
- Gruppengröße: 15

Wo die Fallzahlen abweichen wird explizit daraufhingewiesen.

4.5.1 Globale Nullhypothese gewichtete FDR

Im nichtparametrischen Fall haben wir eine Simulation durchgeführt, ob bei Durchführung der gewichteten FDR-Prozedur im Fall, daß alle Hypothesen wahre Nullhypothesen sind, also die sogenannte globale Nullhypothese vorliegt, die FDR kleiner oder gleich 0.05 ist, also den FDR-α Fehler einhält. Dazu wurde ein Punktschäter für den Erwartungswert der FalseDiscovery-Proportion (FDP) und das ist ja die FDR errechnet. Wir erinnern an die Definitionen (V = Anzahl der abgelehnten wahren Nullhypohtesen, R = Anzahl aller abgelehnter Hypothesen)

$$\text{FDR} = E(\frac{V}{R}1_{\{R>0\}})$$

Es wurden eine Million FDP simuliert, siehe das Programm *GlobalNullFDR* [3]. Als Gewichte wurde die Gewichte mit den Hochzahlen $0, 4, 16, 40$ verwendet. errechnet.

η	0	4	16	30
FDR	0.049734	0.049576	0.049562	0.0496

Tabelle 4.1: η gibt die Hochzahl in den Gewichten g^η an. FDR ist jeweils der Durchschnitt von einer Million unabhängig simulierter FDP (V/R). Alle Hypothesen sind wahre Nullhypothesen

[3] Aus Speicherplatzgründen wurden 1000*1000 also eine Million FDP berechnet. Mit der Hilfsfunktion *Extrahieren* wurde die Schätzwerte errechnet

4.5.2 Adaptive FDR

In der Simulation gab es 12525 wahre Nullhypothesen (N(0,1))und 100 wahre Alternativen (N(1.15,1)). Bei dieser Prozedur wurde die gewichtete Benjamini-Hochberg Prozedur in jedem Schritt jeweils mit den Hochzahlen $(0, 1, 2, 3, 4, 6, 8, 10, 12, 14, 16, 20, 25, 30, 60)$ durchgeführt, wobei jeweils diejenigen Hypothesen abgelehnt wurden , wo bei der entsprechenden Hochzahl die meisten Alternativen abgelehnt wurden.
Es wurden zehntausend Simulationsläufe durchgeführt.
Als Schätzwert für die FDR ergab sich

$$\hat{\text{FDR}} = 0.0405495$$

4.5.3 Einstichprobenfall

hh	Hochzahl bei der Bestimmung der Gewichte
FDR	False Discovery Rate
fdrA	Durchschnittschliche Anzahl abgelehnter Alternativhypothesen. Gewichtete Benjamini-Hochberg Prozedur
fwerA	Durchschnittschliche Anzahl abgelehnter Alternativhypothesen. Kropf Prozedur

Tabelle 4.2: *fdrA ist also die durchschnittliche Anzahl der abgelehnten Alternativen, wenn die gewichtete BH-Prozedur durchgeführt wurde und fwerA analog wenn die gewichtete Kropf-Prozedur aus Kapitel 4 durchgeführt wurde*

Nichtparametrisch 10000 Durchläufe						
hh	0	2	8	16	30	100
FDR	0.0204	0.0374	0.0296	0.00669	0.0015	0.0011
fdr	6.7348	40.5173	61.2834	38.4376	12.5472	2.3254
fwer	0.0000	0.0000	16.2311	24.0996	20.2771	11.6885

Tabelle 4.3: $m_0 = 12525$, $m_1 = 100$, $\mu_0 = 0$, $\mu_1 = 1.15$, $\sigma = 1$, *Anzahl der Probanden* $= 15$, $\alpha = 0.05$. *Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

Nichtparametrisch 10000 Durchläufe (verschiedene hh)			
hh	FDR	**fdr**	**fwer**
0	0.0205	6.9	0.0
2	0.037	40.72	0.0
4	0.037	56.2	3.0
8	0.029	61.3	16.2
10	0.022	57.7	20.2
12	0.015	51.9	22.5
14	0.010	45.2	23.8
16	0.007	38.5	24.2
18	0.005	32.4	24.1
20	0.0037	27.2	23.6
25	0.002	18.1	22.0
30	0.0015	12.7	20.3
50	0.00078	5.2	15.7
100	0.0005	2.3	11.7

Tabelle 4.4: $m_0 = 12525$, $m_1 = 100$, $\mu_0 = 0$, $\mu_1 = 1.15$, $\sigma = 1$, *Anzahl der Probanden* $= 15$, $\alpha = 0.05$. *Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

Parametrisch				
hh	0	4	16	30
FDR	0.049	0.042	0.0008	0.00019
fdr	48.0	81.8	36.7	11.6
fwer	7.8	31.7	44.0	33.4

Tabelle 4.5: $m_0 = 12525$, $m_1 = 100$, $\mu_0 = 0$, $\mu_1 = 1.15$, $\sigma = 1$, *Anzahl der Probanden* $= 15$, $\alpha = 0.05$. *Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

Parametrisch (AnzAltern:500)				
hh	0	4	16	30
FDR	0.047	0.02	0.00007	0.000008
fdr	397.1	452.7	161.9	33.6
fwer	39.2	144.8	198.0	160.4

Tabelle 4.6: $m_0 = 12125$, $m_1 = 500$, $\mu_0 = 0$, $\mu_1 = 1.15$, $\sigma = 1$, *Anzahl der Probanden* $= 15$, $\alpha = 0.05$. *Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

Parametrisch (AnzAltern:1000)				
hh	0	4	16	30
FDR	0.04	0.01	0.00002	0.0000008
fdr	881.8	921.8	311.0	54.3
fwer	78.7	260.6	331.1	296.4

Tabelle 4.7: $m_0 = 11625$, $m_1 = 1000$, $\mu_0 = 0$, $\mu_1 = 1.15$, $\sigma = 1$, *Anzahl der Probanden* $= 15$, $\alpha = 0.05$. *Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

Parametrisch (Varianzen:1.5)				
hh	0	4	16	30
FDR	0.049	0.045	0.0083	0.0044
fdr	13.8	48.8	21.2	6.8
fwer	2.5	11.2	16.9	11.4

Tabelle 4.8: $m_0 = 12525$, $m_1 = 100$, $\mu_0 = 0$, $\mu_1 = 1.15$, $\sigma = 1.5$, *Anzahl der Probanden* $= 15$, $\alpha = 0.05$. *Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

Parametrisch (Varianzen:1.5,mA=2)				
hh	0	4	16	30
fdr	95.1152	99.5697	72.9505	29.6753
fwer	40.8146	89.6045	95.6768	91.3665

Tabelle 4.9: $m_0 = 12525$, $m_1 = 100$, $\mu = 0$, $\mu_1 = 2$, *Varianzen* 1 *Anzahl der Probanden* $= 15$, $\alpha = 0.05$. *Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

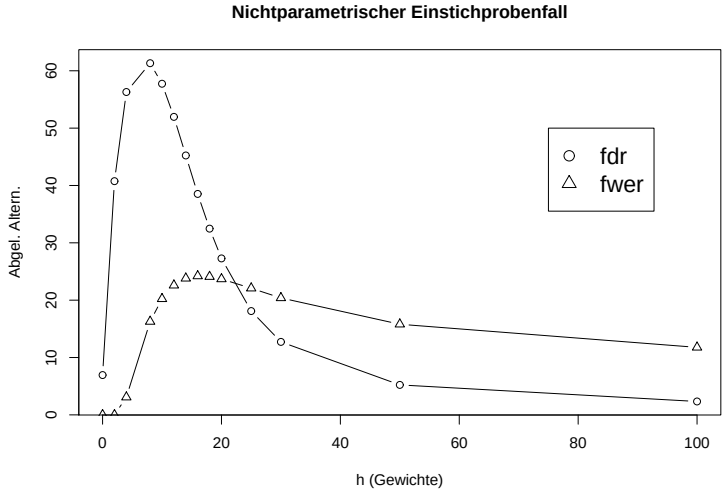

Abbildung 4.1: $m_0 = 12525$, $m_1 = 100$, $\mu_0 = 0$, $\mu_1 = 1.15$, $\sigma = 1$, *normalverteilt. Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

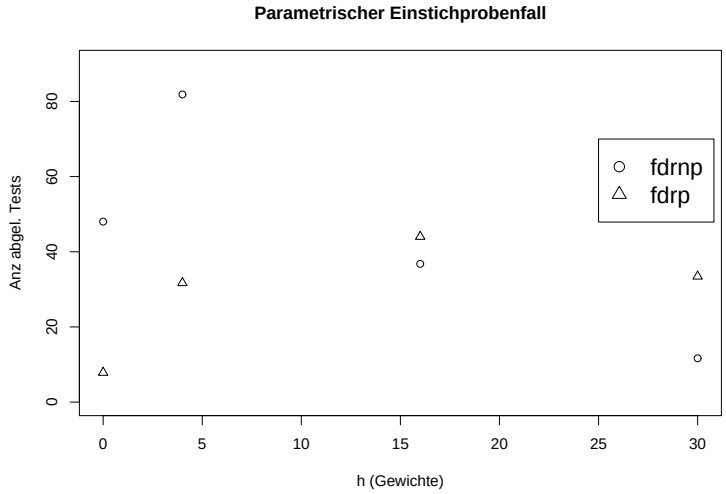

Abbildung 4.2: $m_0 = 12525$, $m_1 = 100$, $\mu_0 = 0$, $\mu_1 = 1.15$, $\sigma = 1$, *normalverteilt. Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

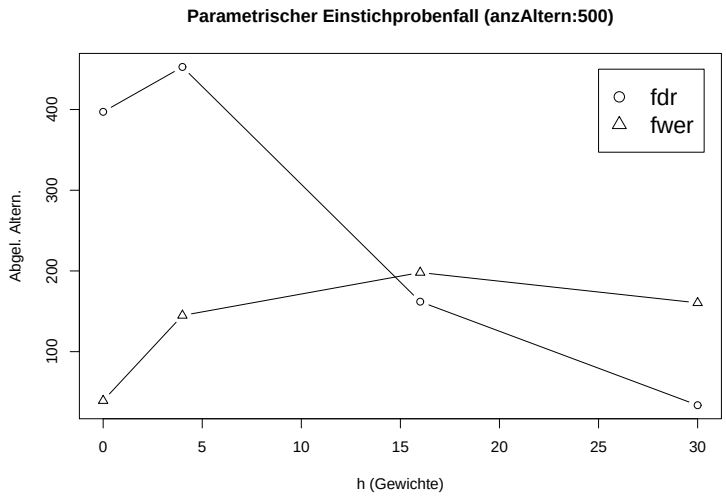

Abbildung 4.3: $m_0 = 12125$, $m_1 = 500$, $\mu_0 = 0$, $\mu_1 = 1.15$, $\sigma = 1$, *normalverteilt. Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

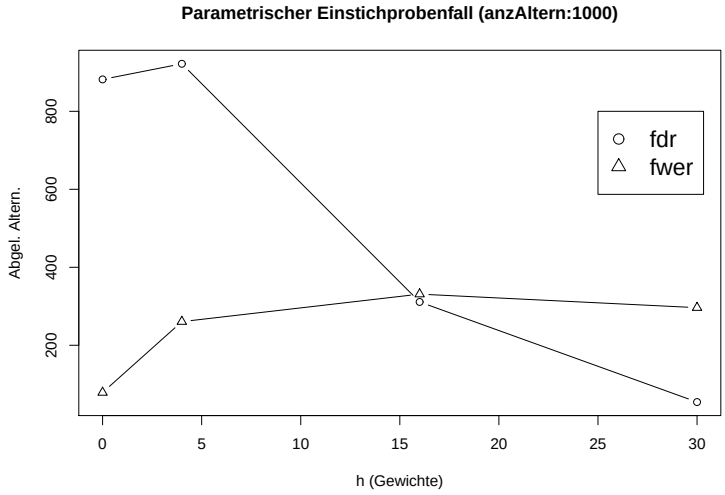

Abbildung 4.4: $m_0 = 11625$, $m_1 = 1000$, $\mu_0 = 0$, $\mu_1 = 1.15$, $sigam = 1$, *normalverteilt. Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

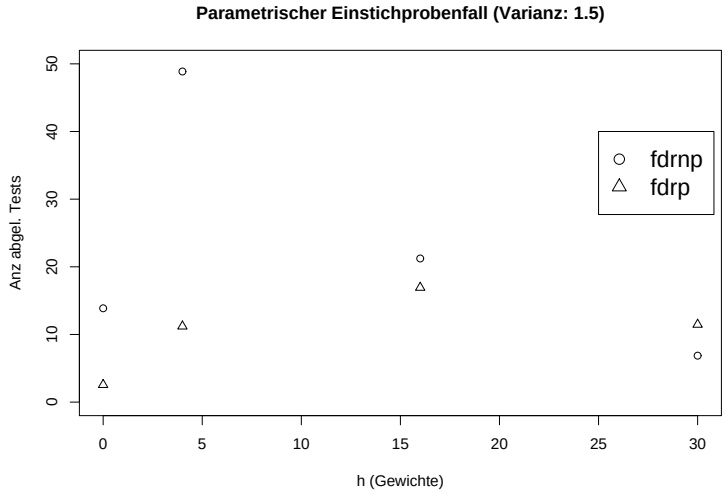

Abbildung 4.5: $m_0 = 12525$, $m_1 = 100$, $\mu_0 = 0$, $\mu_1 = 1.15$, $\sigma = 1.5$, *normalverteilt. Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

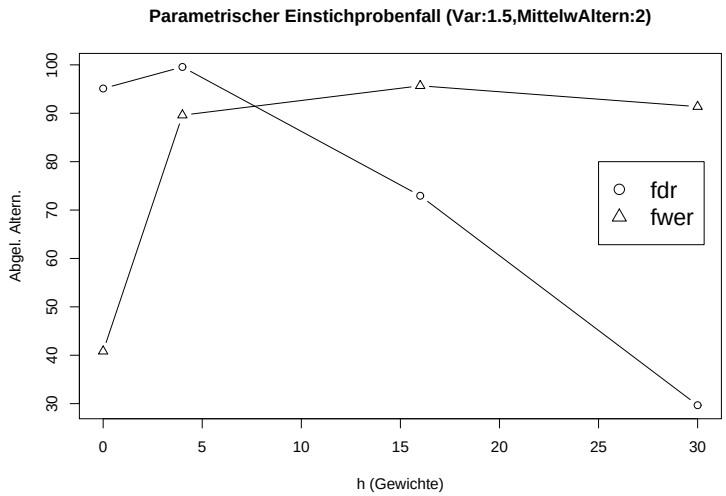

Abbildung 4.6: $m_0 = 12525$, $m_1 = 100$, $\mu_0 = 0$, $\mu_1 = 2$, $\sigma = 1.5$, *normalverteilt. Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

4.5.4 Zweistichprobenfall

Nichtparametrisch 10000 Durchläufe						
hh	0	2	8	16	30	100
FDR	0.046	0.04	0.042	0.028	0.012	0.007
fdr	51.7	59.7	65.7	53.2	24.8	3.3
fwer	12.0	16.2	25.8	27.3	19.4	6.4

Tabelle 4.10: $m_0 = 12525$, $m_1 = 100$, $\mu_0 = 0$, $\mu_1 = 1.15$, $\sigma = 1$, *Anzahl der Probanden pro Stichprobe* $= 15$, $\alpha = 0.05$.*Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

Parametrisch 10000 Durchläufe						
hh	0	2	8	16	30	100
FDR	0.044	0.044	0.034	0.009	0.003	0.0016
fdrA	55.5	68.4	74.5	50.4	18.8	2.8
fwerA	13.9	21.4	36.8	36.2	24.84	10.4

Tabelle 4.11: $m_0 = 12525$, $m_1 = 100$, $\mu_0 = 0$, $\mu_1 = 1.15$,$\sigma = 1$, *Anzahl der Probanden pro Stichprobe* $= 15$, $\alpha = 0.05$.*Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

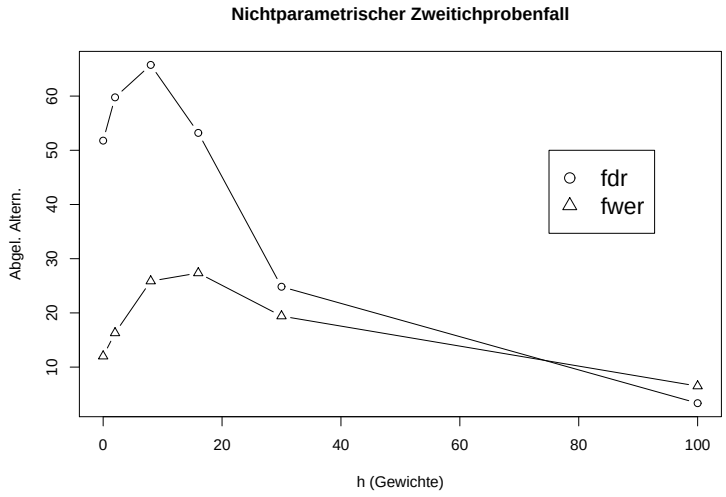

Abbildung 4.7: $m_0 = 12525$, $m_1 = 100$, $\mu_0 = 0$, $\mu_1 = 1.15$, $\sigma = 1$, *Anzahl der Probanden pro Stichprobe* $= 15$, $\alpha = 0.05$. *Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

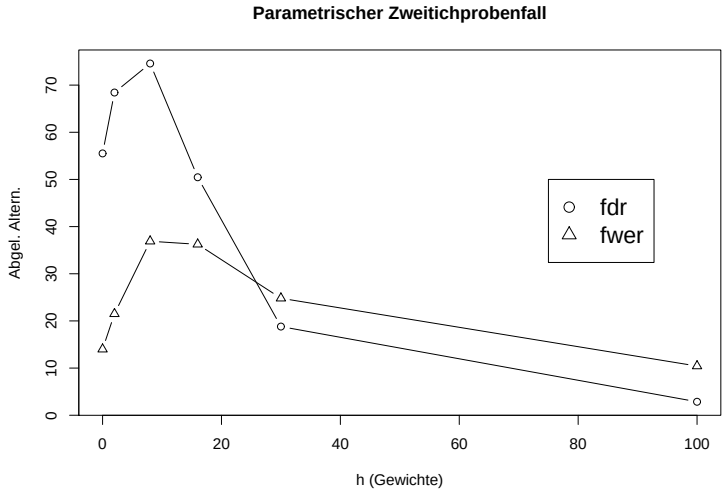

Abbildung 4.8: $m_0 = 12525$, $m_1 = 100$, $\mu_0 = 0$, $\mu_1 = 1.15$, $\sigma = 1$, *Anzahl der Probanden pro Stichprobe* $= 15$, $\alpha = 0.05$. *Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario.*

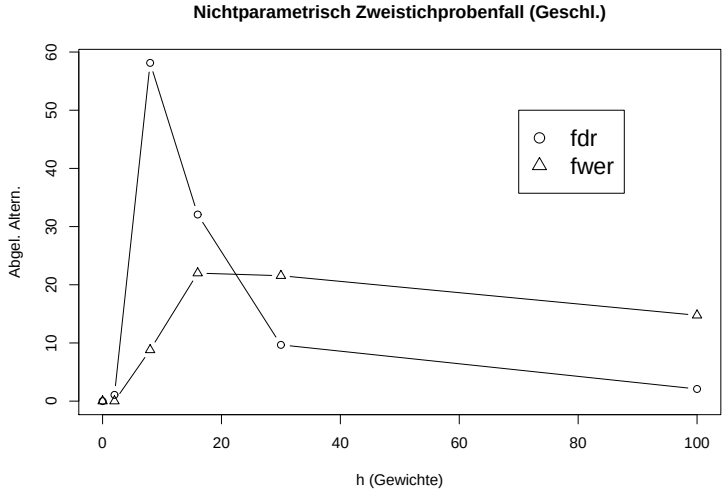

Abbildung 4.9: $m_0 = 6$, $m_1 = 9$. *Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario (nicht-parametrischer Zweistichprobenfall).*

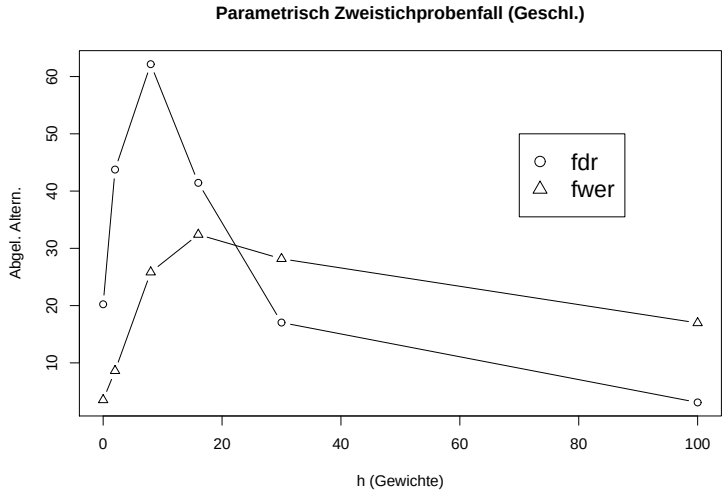

Abbildung 4.10: $m_0 = 6$, $m_1 = 9$. *Durchschnittliche Anzahl der abgelehnten Alternativhypothesen für die gewichtete BH (***fdr***) und die gewichtete Kropf (***fwer***) Prozedur für das Szenario (parametrischer Zweistichprobenfall).*

h	2	8
FDR	0.06184395	0.05606500
FDR	0.0620654068	0.0561913503

Tabelle 4.12: *Die Fälle der Hochzahlen (zur Bestimmung der Gewichte in g^h) im Kleinstichprobenbereich (***6** *und* **9***), bei denen die FDR größer als* 0.05 *war.* $m_0 = 12525$*,* $m_1 = 100$*,* $\mu_0 = 0$*,* $\mu_1 = 1.15$*,* $\sigma = 1$*. Durchgeführt wurde die parametrische Benjamin-Hochberg Prozedur*

FDR	0.056394	0.056668	0.056691	0.056470	0.056529	0.056390	0.056803

Abbildung 4.11: *Die FDR bei Hochzahl* 2*,* $m_0 = 9$ *Probanden in einer Gruppe,* **14** *in der anderen, parametrische Benjamini-Hochberg Prozedur Prozedur,* $m_0 = 12525$*,* $m_1 = 100$*,* $\mu_0 = 0$*,* $\mu_1 = 2.4$*,* $\sigma = 1$

Es kann Zufall sein, aber, siehe Tabelle 4.12 die FDR ist bei 7 Durchläufen a 10000 jeweils über 0.05 (bei den anderen Hochzahlen, insbesondere auch 0 war das nicht der Fall). Wir erinnern, als Test wurde der t-Test gewählt und die Größe der beiden Gruppen war nicht sehr groß (6 und 9).
Dieses Phänomen ist bei Wahl des Wilcoxon-Tests nicht aufgetreten.

4.6 Ein Beispiel

Wir machen nun plausibel, warum bei gewissen Konstellationen (wenn die Hochzahl ν groß wird und damit die Gewichte extremere Werte annehmen) die gewichtete Bonferroni-Holm Prozedur mehr wahre Alternativen ablehnt als die gewichtete BH Prozedur.
Es seien zwei Hypothesen gegeben, beides wahre Alternativen. Die P-Werte seien gleich $(0.04, 0.04)$, die Gewichte $W = (1 - \epsilon, \epsilon)$, wobei ϵ wie auch sonst in der Mathematik oft der Fallist, sehr klein sei. Die normierten Gewichte sind dann $(2(1 - \epsilon), 2\epsilon)$, also das erste praktisch 2 und das andere fast 0. Bildet man dann P/W, so ist ein Wert moderat der andere aber wegen der Division durch eine sehr kleine Zahl sehr groß. Dadurch lehnt die (gewichtete) Bonferroni-Holm Prozedur beide Hypothesen ab, die (gewichtete) Benjamini-Hochberg Prozedur aber nur die erste.

4.7 Der Golub Datensatz

Der *golub* Datensatz ist im package *multtest* auf der Internetseite *http://www.bioconductor.org/* zu finden.
Maßzahlen von Genen werden in zwei Gruppen verschiedener Leukämieerkrankungen untersucht (weitergehende Informationen sind auch im erwähnten package zu finden).
3051 Gene sind im Datensatz zu finden, eine Gruppe umfaßte 27 Probanden, die andere 11.
Wir haben untersucht wieviel Nullhypothesen (die Gruppen unterscheiden sich nicht in der Genmaßzahl) bei der Kropfprozedur und der gewichteten BH-Prozedur abgelehnt werden.

	Kropf Proz.		gew. BH Proz	
	nicht parametrisch	parametrisch	nicht parametrisch	parametrisch
0	93	104	680	696
1	96	117	592	614
2	100	116	524	535
4	88	103	446	459
8	33	55	397	407
12	12	30	413	423
16	9	17	415	428
20	11	14	421	431
15	9	19	416	428
30	8	12	436	449
50	7	8	451	462

Tabelle 4.13: *Anzahl der abgelehnten Tests im Golub Datensatz*

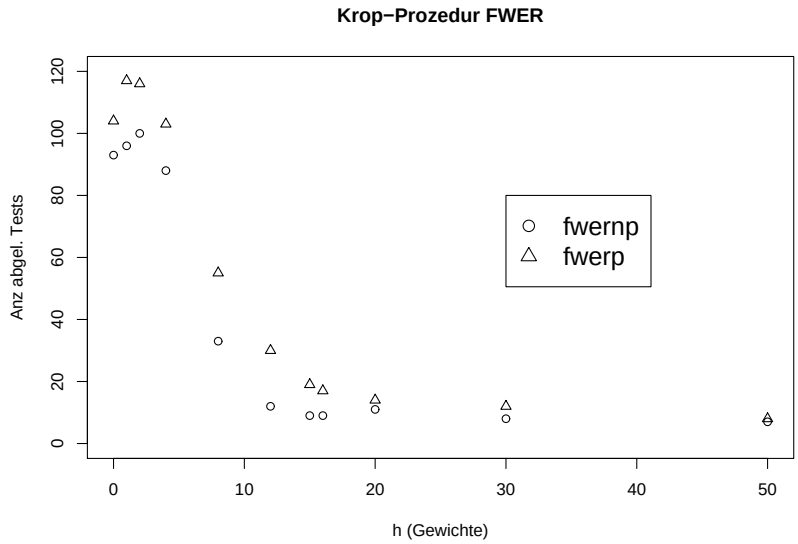

Abbildung 4.12: *Anzahl der abgelehnten Tests des Golub Datensatzes bei der gewichteten Kropf Prozedur im parametrischen Fall (fwerp) und nichtparametrischen Fall (fwernp)*

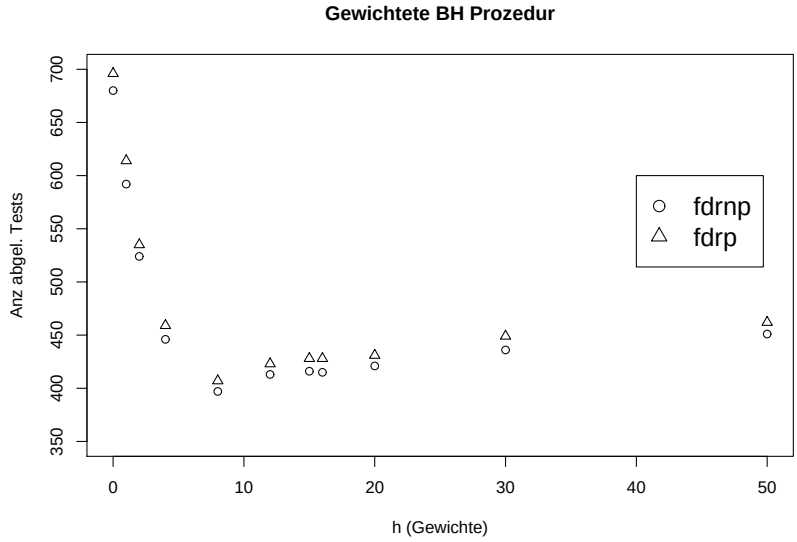

Abbildung 4.13: *Anzahl der abgelehnten Tests des Golub Datensatzes bei der gewichteten BH Prozedur im parametrischen Fall (fdrp) und nichtparametrischen Fall (fdrnp)*

4.8 Bemerkungen

1. Die Macht (durchschnittliche Anzahl abgelehnter wahrer Alternativen) wird im nichtparametrischen Fall für den Einstichprobenfall in Abbildung 1 und für den Zweistichprobenfall in Abbildungen 7, sowie 9 dargestellt.

2. Die Macht im parametrischen Fall für verschiedene Szenarien (unterschiedlich viele wahre Alternativen und/oder Varianzen und Mittelwerte der Alternativen) wird für den Einstichprobenfall in den Abbildungen 2 bis 6 dargestellt, sowie für den Zweistichprobenfall in den Abbildungen 8 und 10.

3. Die Simulationen deuten darauf hin, daß die FDR bei (fast, siehe unten) allen Gewichtungen den α Fehler einhält.

4. Die Macht ist zwar für große Exponenten h der Gewichte bei der Kropf-Prozedur besser, jedoch ist die bestmögliche Macht bei der gewichteten Benjamini-Hochberg Prozedur weit besser als die der Kropfprozedur (siehe auch das Beispiel in Kapitel 4.6).

5. Diese Aussagen gelten sowohl für den parametrischen als auch den nichtparametrischen Fall

6. Sie gelten auch für andere Szenarien, etwa andere Anzahl von Alternativhypothesen, andere Streuung oder Mittelwert der Alternativen.

7. Die Macht der Kropfprozedur ist stabiler gegenüber Änderungen des Gewichtsparameters η als die gewichtete Benjamini-Hochberg Prozedur.

8. Bei kleinen Stichprobengrößen (6 und 9, bzw 9 und 14) ergab sich bei zwei Hochzahlen (2 und 8) eine leichte Übertretung der α-Schranke, was natürlich auch eine Zufallserscheinung sein kann. Diese Phänomen ist allerdings bei besagten kleinen Stichproben bei allen Simulationen immer aufgetreten.

4.9 Literarturhinweise

Der nichtparametrische nichtgewichtete Teil dieses Kapitels beruht auf [15], bzw [12]. Die Methodik für die parametrischen nicht gewichteten Methoden sind in [14] zu finden. Beweise für den Satz von Läuter findet man in [16] und [17].

Eine Anwendung der Ideen in diesem Kapitel - die Betrachtung von Äquivalenzklassen, in dem folgenden Hinweis orbits genannt - auf Permutationstest findet man in [5].

Kapitel 5

R Programme

5.1 Beispiel 1

```
# Zufallszahlen
ZufallsZahlen1 <- function(m,n) {
   zZ <- rnorm(2*m*n)
   zM <- matrix(zZ,nrow=2*m,byrow=T)
   return(zM)
}

Pwerte1 <- function(zM) {
  av <- nrow(zM)/2
  e <- numeric(av)
  for (i in 1:av) {
     e[i] <- t.test(zM[i,],zM[(i+av),])[3]
  }
  return(e)
}

Kapitel1 <- function(anzExp=100,m=50,n=100,alfa=0.05) {
  erg <- numeric(anzExp)
  for (j in 1:anzExp) {
    zM <- ZufallsZahlen1(m,n)
    e  <- Pwerte1(zM)
    aS <- length(e[e<=alfa])
```

```
    erg[j] <- aS
  }
  return(erg)
}
```

5.2 Vergleiche der Macht

5.2.1 Nichtparametrisch Einstichprobenfall

```
# Diese Prozedur rechnet im Einstichprobenfall
# Nullhypothese:        N(0,1)
# Alternativhypothese:  N(1.15,1)
# die
# 1) Durchschnittliche Anzahl abgelehnter Hypothesen
#    a) Gewichtete Prozedur nach Kropf (fwera)
#    b) Gewichtete FDR nach Genovese (fdr)
# 2) FDR
# aus.


VglPowerGFDRuKFWEResnp <- function() {

# Werte aus Arbeit von Kropf

anzExp <- 50         #  Anzahl Experimente
alfa   <- 0.05
ap <- 15             #  Anzahl (Versuchs)Personen
av <- 12625          #  Anzahl Variablen
an <- 12525          #  Anzahl Nullhypothesen
aa <- 100            #  Anzahl Alternativhypothesen

vh  <- c(0,2,8,16,30,100)  # Vektor Hochzahlen
lvh <- length(vh)          # Laenge Vektor Hochzahlen

fdra  <- numeric(lvh)      # FalseDiscoveryRate Alternativ (abgelehnte)
fdp   <- numeric(lvh)      # FalseDisvoceryProportion
fwera <- numeric(lvh)      # FamilyWideErrorRate Alternavitv (abgelehnte)

for (i in 1:anzExp) {
    # Die erstene an Hypothesen sind die wahren Nullhypothesen
    # Die restlichen aa Hypothesen sind die wahren Alternativhypothesen
    # zM: Zeilen      = Variablen (bzw Hypothesen)
    # zM: Spalten     = Anzahl der Versuch(spersonen)
```

```
    # pW: PWerte      = TestFkt := Wilcoxon Einstichprobentest
    # rG: Rohgewichte = Median der Absolutwerte der Variablen

    zM <- ZufallsZahlen(anzProb=ap,anzNull=an,anzAltern=aa)
    pW <- Pwerte(zM,TestFkt)
    rG <- VektorAbsolutMedian(zM)
    z <- 1
    for (h in vh) {
       vG  <- rG^h                              # g Hoch h

       dF <- KFWERG(pw=pW,g=vG,alfa=alfa)       # Prozedur nach Kropf
                                                # dF := Nummern der
                                                # abgelehnten Hypothesen
       aFA <- length(dF[dF > an])               # Anzahl der abgel. Alternath.
       fwera[z] <- fwera[z] + aFA               # Summer der Anzahl der
                                                #   abgel. Alternath.
                                                #   bei Hochzahl h
                                                #   wird ausgerechnet

       dT <- KFDRG(pw=pW,g=vG,alfa=alfa)        # Gewichtete FDR nach Genovese
       aN <- length(dT[dT <= an]) - 1           # Anzahl abgel. Nullhypothesen
       aA <- length(dT[dT > an])                # Anzahl abgel. Alternatien
       fdra[z] <- fdra[z] + aA                  # Sum Anzahl abgel.
                                                #   Alternativen (Hochz. h)
       vr <-  aN + aA                           # Anzahl aller abgel. Hypothese
       vr[vr==0] <-  1                          # Gleich 1, falls 0 (vr ist Ska
       fdp[z] <- fdp[z] + aN/vr                 # FalseDiscoveryProportion
                                                #   bei Hochz. h
       z <- z + 1
    }
}

mfdra  <- fdra/anzExp    # Durchschnitt abgel. Alternativh. bei gew. FDR Pr
fdr    <- fdp/anzExp     # FalseDiscoveryRate (:= Durchschnitt der FDPropor
mfwera <- fwera/anzExp   # Durchschnitt abgel. Alternativh. bei Kropf Proz.

print(fdr)
print(mfdra)
```

```
print(mfwera)

}
```

5.2.2 Nichtparametrisch Zweistichprobenfall

```
# Diese Prozedur rechnet im Zweistichprobenfall
# Nullhypothese:        N(0,1)
# Alternativhypothese:  N(1.15,1)
# die
# 1) Durchschnittliche Anzahl abgelehnter Hypothesen
#    a) Gewichtete Prozedur nach Kropf (fwera)
#    b) Gewichtete FDR nach Genovese (fdr)
# 2) FDR
# aus.
VglPowerGFDRuKFWERzsnp <- function() {

# Werte aus Arbeit von Kropf

anzExp <- 100         #  Anzahl Experimente
alfa   <- 0.05
ap <- 15             #  Anzahl (Versuchs)Personen
av <- 12625          #  Anzahl Variablen
an <- 12525          #  Anzahl Nullhypothesen
aa <- 100            #  Anzahl Alternarivhypothesen

vh  <- c(0,2,8,16,30,100)  # Vektor Hochzahlen
lvh <- length(vh)          # Laenge Vektor Hochzahlen

fdra  <- numeric(lvh)      # FalseDiscoveryRate Alternativ (abgelehnte)
fdp   <- numeric(lvh)      # FalseDisvoceryProportion
fwera <- numeric(lvh)      # FamilyWideErrorRate Alternavitv (abgelehnte)

for (i in 1:anzExp) {
    # Die erstene an Hypothesen sind die wahren Nullhypothesen
    # Die restlichen aa Hypothesen sind die wahren Alternativhypothesen
    # zM: Zeilen      = Variablen (bzw Hypothesen)
    # zM: Spalten     = Anzahl der Versuch(spersonen)
    # pW: PWerte      = TestFkt := Wilcoxon Zweistichprobentest
    # rG: Rohgewichte = Interquantilbereich

    zM <- ZufallsZahlen2(anzProb=ap,anzNull=an,anzAltern=aa)
```

```
    pW <- Pwerte2(zM)
    rG <- VektorGewichte2(zM)
    z <- 1
    for (h in vh) {
       vG  <- rG^h                              # g Hoch h

       aFA <- MacheFWERa(pw=pW,g=vG,alfa=alfa,an)   # Prozedur nach Kropf
                                                    # Anzahl der abgelehnten
                                                    #   Alternativhypothesen
       fwera[z] <- fwera[z] + aFA    # Summe der Anzahl der abgelehnten
                                     #  Alternativhypothesen
                                     #  bei Hochzahl h (Nummer:z) wird
                                     #  ausgerechnet
       dT <- MacheFDR(pw=pW,g=vG,alfa=alfa,an) # Gewichtete FDR nach Genovese

       fdra[z] <- fdra[z] + dT[2]  # Summe Anzahl abgel. Alternativhypothesen
                                   #   bei Hochz. h

       fdp[z] <- fdp[z] + dT[1]  # FalseDiscoveryProportion bei Hochz. h
       z <- z + 1
    }
}

mfdra  <- fdra/anzExp    # Durchschnitt abgel. Alternativh. bei gew. FDR Proz.
fdr    <- fdp/anzExp     # FalseDiscoveryRate (:= Durchschnitt der FDProportion)
mfwera <- fwera/anzExp   # Durchschnitt abgel. Alternativh. bei Kropf Proz.

print(fdr)
print(mfdra)
print(mfwera)

}
```

5.2.3 Parametrisch Einstichprobenfall

```
PowerGFDRuKFWERExperiment <- function(anzExp=10,anzNull=12525,
           anzAltern=100,anzProb=15,alfa=0.05,vh=c(0,4,16,30)){
    MittelN <- 0
    MittelA <- 1.15

    VarianzN <- 1
    VarianzA <- 1

    # Mittelwerte ausrechnen
    zn <- rnorm(anzNull*anzExp,mean=MittelN,sd=sqrt(VarianzN*1/anzProb))
    za <- rnorm(anzAltern*anzExp,mean=MittelA,sd=sqrt(VarianzA*1/anzProb))
    mM <- matrix(c(zn,za),nrow=anzNull+anzAltern,byrow=T)

    # Varianzen   ausrechnen
    zn <- (VarianzN/(anzProb-1))*rchisq(anzNull*anzExp,anzProb-1)
    za <- (VarianzA/(anzProb-1))*rchisq(anzAltern*anzExp,anzProb-1)
    mS <- matrix(c(zn,za),nrow=anzNull+anzAltern,byrow=T)
      #print(mS)
    # Gewichte    ausrechnen
    gg <-((mM^2)*anzProb+ mS*(anzProb-1))
    saP <- sqrt(anzProb)
    tt <- saP*mM/sqrt(mS)

    # P-Werte     ausrechnen
    pw  <- apply(tt,c(1,2),function(x) {1-pt(x,anzProb-1)})

    ## Eventuell zur Kontrolle. Ausgegebner Wert sollte ungefaehr 0.05 sein
    #pwN <- pw[(1:anzNull),]
    #print((length(pwN[pwN<=0.05]))/(anzNull*anzExp))
    #cat("\n")

    fdr    <- numeric(anzExp)
    fwera  <- numeric(anzExp)
    fdra   <- numeric(anzExp)

    j <- 1
```

```
    lvh <- length(vh)
    mm <- matrix(0,nrow=3,ncol=lvh)
    for (h in vh) {
      g <- gg^h

      for (i in 1:anzExp) {
        fdrR    <- MacheFDR(pw[,i],g[,i],alfa,anzNull)
        fwera[i]  <- MacheFWERa(pw[,i],g[,i],alfa,anzNull)
        fdr[i]  <- fdrR[1]
        fdra[i]   <- fdrR[2]
      }

      mm[1,j] <- mean(fdr)
      mm[2,j] <- mean(fdra)
      mm[3,j] <- mean(fwera)
      j <- j + 1
    }
    return(mm)
}


PowerGFDRuKFWER <- function(anzExp=10,anzNull=12525,anzAltern=100,
                      anzProb=15,alfa=0.05,anzIt=10,vh=c(0,4,16,30)){
  rmm <- c()
  for (i in 1:anzIt){
     mm <- PowerGFDRuKFWERExperiment(anzExp=anzExp,anzNull=anzNull,
     anzAltern=anzAltern,anzProb=anzProb,alfa=alfa,vh=vh)
     rmm <- rbind(rmm,mm)
  }
  return(rbind(vh,rmm))
}

#erg <- PowerGFDRuKFWER(anzExp=10,anzNull=12525,anzAltern=100,
 anzProb=15,alfa=0.05,anzIt=10)
```

5.3 Adaptive gewichtete FDR

```
# Es wird folgende Prozedur verwendet
# 1) Ein Vektor vh von Hochzahlen ist gegeben
# 2) P-Werte des Experiments werden ausgerechnet
# 3) Gewichte werden ausgerechnet
# 4) Fuer alle h in vh werden die g_i^h gebildet
# 5) Es werden diejenigen Hypothesen (zu h) abgelehnt, bei denen im Schritt
#    (4) die meisten Alternativhypothesen abgelehnt werden.
# 6) Die FDP wird zugehoerig zum h von (5) gebildet
# 7) Die FDR wird mit diesen FDP's gebildet

FDRadaptivExperiment <- function(anzExp=10,anzNull=12525,
          anzAltern=100,anzProb=15,alfa=0.05,vh=c(0,4,16,30)){
    MittelN <- 0
    MittelA <- 1.15

    VarianzN <- 1
    VarianzA <- 1

    # Mittelwerte ausrechnen
    zn <- rnorm(anzNull*anzExp,mean=MittelN,sd=sqrt(VarianzN*1/anzProb))
    za <- rnorm(anzAltern*anzExp,mean=MittelA,sd=sqrt(VarianzA*1/anzProb))
    mM <- matrix(c(zn,za),nrow=anzNull+anzAltern,byrow=T)

    # Varianzen   ausrechnen
    zn <- (VarianzN/(anzProb-1))*rchisq(anzNull*anzExp,anzProb-1)
    za <- (VarianzA/(anzProb-1))*rchisq(anzAltern*anzExp,anzProb-1)
    mS <- matrix(c(zn,za),nrow=anzNull+anzAltern,byrow=T)

    # (Roh) Gewichte    ausrechnen
    gg <-((mM^2)*anzProb+ mS*(anzProb-1))
    saP <- sqrt(anzProb)
    tt <- saP*mM/sqrt(mS)    # t-Wert

    # P-Werte     ausrechnen
    pw  <- apply(tt,c(1,2),function(x) {1-pt(x,anzProb-1)})
```

```
    ## Eventuell zur Kontrolle.
    ## Ausgegebner Wert sollte ungefaehr 0.05 sein.
    #pwN <- pw[(1:anzNull),]
    #print((length(pwN[pwN<=0.05]))/(anzNull*anzExp))
    #cat("\n")

    fdr    <- numeric(anzExp)
    fwera  <- numeric(anzExp)
    fdra   <- numeric(anzExp)

    j <- 1
    lvh <- length(vh)
    mm <- matrix(0,nrow=1,ncol=lvh)
    fdp  <- numeric(lvh)
    fdpa <- numeric(lvh)
    fdr4 <- numeric(anzExp)
    for (i in 1:anzExp) {
       j <- 1
       for (h in vh) {
          g <- gg^h      # Gewichte
          fdpR <- MacheFDR(pw[,i],g[,i],alfa,anzNull)
            # Return: Vekt. mit 2 Kompnenten
            # 1) FalseDiscoveryProportion
            # 2) Anzahl der abgel. Alternativhyp.
          fdp[j]  <- fdpR[1]
          fdpa[j] <- fdpR[2]
          j <- j + 1
       }
       fdphm <- which.max(fdpa)    # Index (h) mit meisten abgelehnten
                                   #   Alternativhypothesen
       fdr4[i] <- fdp[fdphm]       # Zugehoerige FalseDiscoveryProportion
    }
    return(mean(fdr4))    # Return: FDR
}


FDRadaptiv <- function(anzExp=10,anzNull=12525,anzAltern=100,
                  anzProb=15,alfa=0.05,anzIt=10,vh=c(0,4,16,30)){
```

```
  rfdr <- numeric(anzIt)
  for (i in 1:anzIt){
     mm <- FDRadaptivExperiment(anzExp=anzExp,anzNull=anzNull,
             anzAltern=anzAltern,anzProb=anzProb,alfa=alfa,vh=vh)
     rfdr[i] <- mm
  }
  return(mean(rfdr))
}

#erg <- FDRadaptiv(anzExp=10,anzNull=12525,anzAltern=100,
  anzProb=15,alfa=0.05,anzIt=10,vh=c(0,1,2,3,4,6,8,10,12,14,16,20,25,30,60)
```

5.4 Gewichtete FDR-Globale Nullhypothese

```
# Es wird untersucht ob die gewichtete FDR Prozedur bei der
# globalen Nullhypothese, also alle Variable aus N(0,1)
# zum Niveau 0.05 eingehalten wird.
GlobalNullFDRExperiment <- function(anzExp=10,anzNull=12625,
         anzProb=15,alfa=0.05,vh=c(0,4,16,30)){
    MittelN <- 0
    MittelA <- 1.15

    VarianzN <- 1
    VarianzA <- 1

    # Mittelwerte ausrechnen
    zn <- rnorm(anzNull*anzExp,mean=MittelN,sd=sqrt(VarianzN*1/anzProb))
    mM <- matrix(zn,nrow=anzNull,byrow=T)

    # Varianzen   ausrechnen
    zn <- (VarianzN/(anzProb-1))*rchisq(anzNull*anzExp,anzProb-1)
    mS <- matrix(zn,nrow=anzNull,byrow=T)

    # Gewichte    ausrechnen
    gg <-((mM^2)*anzProb+ mS*(anzProb-1))
    saP <- sqrt(anzProb)
    tt <- saP*mM/sqrt(mS)

    # P-Werte     ausrechnen
    pw  <- apply(tt,c(1,2),function(x) {1-pt(x,anzProb-1)})

    ## Eventuell zur Kontrolle.
    ## Ausgegebner Wert sollte ungefaehr 0.05 sein.
    #pwN <- pw[(1:anzNull),]
    #print((length(pwN[pwN<=0.05]))/(anzNull*anzExp))
    #cat("\n")

    fdr    <- numeric(anzExp)
    fwera  <- numeric(anzExp)
    fdra   <- numeric(anzExp)
```

```
    j <- 1
    lvh <- length(vh)
    mm <- matrix(0,nrow=1,ncol=lvh)
    for (h in vh) {
      g <- gg^h
      #print(g)
      for (i in 1:anzExp) {
        fdrR    <- MacheFDR(pw[,i],g[,i],alfa,anzNull)
        fdr[i]  <- fdrR[1]
        fdra[i]   <- fdrR[2]
      }

      mm[1,j] <- mean(fdr)
      j <- j + 1
    }
    return(mm)
}


GlobalNullFDR <- function(anzExp=10,anzNull=12625,
       anzProb=15,alfa=0.05,anzIt=10,vh=c(0,4,16,30)){
  rmm <- c()
  for (i in 1:anzIt){
     mm <- GlobalNullFDRExperiment(anzExp=anzExp,
       anzNull=anzNull,anzProb=anzProb,alfa=alfa,vh=vh)
     rmm <- rbind(rmm,mm)
  }
  return(rbind(vh,rmm))
}

#GlobalNullerg <- GlobalNullFDR(anzExp=10,anzNull=12625,
         anzProb=15,alfa=0.05,anzIt=10)
```

5.5 Golub Test

```
GolubTest <- function(GolubDaten) {
  GD   <- GolubDaten
  aM   <- 27
  aF   <- 11
  alfa <- 0.05

  pW   <- PwerteMFt(zM=GD,aM=aM,aF=aF)
  vG   <- VektorGewichteMF(zM=GD,aM=aM,aF=aF)

  ergFWER <- KFWERG(pW,vG,alfa)
  ergFDR  <- KFDRG(pW,vG,alfa)

  erg <- list()
  erg[[1]] <- ergFWER
  erg[[2]] <- ergFDR
  return(erg)
}
```

5.6 Prozeduren und Utilities

```
# Gewichtete FDR-Prozedur nach Genovese et. al.

KFDRG  <- function(pw,g,alfa) {
     pw  <- c(0,pw)
     g   <- c(1,g)
       # PWerte (Zusatz von 0 aus programmtechnischen Gruenden)
       # Gewichte (Zusatz von 1 aus programmtechnischen Gruenden)
     nn  <- length(pw)
     pg  <- pw/g          # Pwerte/Gewichte := Q-Werte
     m   <- nn - 1        # "Wirkliche" Anzahl der P-Werte

     q   <- sort(pg,index.return=TRUE)    # Sortiere die  Q-Werte

     v   <- 0:m
     cc  <- v/m*alfa       # Vektor der Werte von Q_i <= i/m*alfa

     qq <- unlist(q["x"])
     bh  <- v[(qq <= cc) == TRUE]
       # Wo gilt ueberall Q_i <= i/m*alfa  (
       # bh := Benjamini-Hochberg besteht nur aus abg. Hyp.
     ind <- bh[length(bh)]    # Index des Groessten, wo (*) gilt
     erg <- (unlist(q["ix"])-1)[1:(ind+1)]
       # Alle Indizi der Hypothesen, die abgelehnt werden
       # -1, da Index wegen erstem Vektor jeweils zu hoch.
       # Alle nach den Q-Werten geordnete Hypothesen
       # bis einschliesslich zum Index ind werden
       # abgelehnt. Da die "dummy"-Hypothese an Position
       # 1 steht, steht die letzte abzulehnende "echte"
       # Hypothese an Stelle ind + 1
     return (erg)
}

KFWERG <- function(pw,g,alfa) {
   pw   <- c(0,pw)
   g    <- c(1,g)
     # PWerte (Zusatz von 0 aus programmtechnischen Gruenden)
```

```
     # Gewichte (Zusatz von 1 aus programmtechnischen Gruenden)
   nn   <- length(pw)
   pg   <- pw/g
   pgo  <- order(pg)
   dsum <- sum(g)
   wsum <- 0

   dieTests <- numeric(nn)

   # Die erste Hypothese wird auf jeden Fall abgelehnt,
   # dann wird pgo[1]=1 und damit wsum = 1.
   # Bei der ersten "echten" Iteration hat dann dsum den richtigen Wert

   for (i in 1:nn) {
        dsum <- dsum - wsum
        if (pg[pgo[i]]<= alfa/dsum) {
          dieTests[i] <- pgo[i] - 1
          wsum <-  g[pgo[i]]
        }
        else {
          return(dieTests[1:(i-1)])
        }
   }
   return(dieTests[1:nn])
}

MacheFDR <- function(pw,g,alfa,an) {
   dT <- KFDRG(pw,g/mean(g),alfa)
   aN <- length(dT[dT <= an]) - 1
   aA <- length(dT[dT > an])

   vr <-  aN + aA
   vr[vr==0] <-  1

   return (c(aN/vr,aA))
}

MacheFWERa <- function(pw,g,alfa,an) {
```

```
   dT <- KFWERG(pw,g,alfa)
   aA <- length(dT[dT > an])

   return(aA)
}

ZufallsZahlen <- function(anzProb,anzNull,anzAltern,mN=0,mA=1.15) {
   aN <- anzNull*anzProb
   aA <- anzAltern*anzProb
   aM <- anzNull + anzAltern

   zN <- rnorm(aN,mean=mN)
   zA <- rnorm(aA,mean=mA)

   alleN <- c(zN,zA)
   alleM <- matrix(alleN,nrow=aM,byrow=T)

   return(alleM)
   # die ersten anzNull Zeilen enthalten die
   #  Beobachtungswerte der wahren Nullhypothesen
   # die restlichen anzAltern Zeilen enhalten die
   #  Beobachtungswerte der wahren Alternativen
}

Pwerte <- function(zM,TF) {
  return(apply(zM,1,TF))
  # Teststatistik wird auf die Zeile der Matrix (Stichprobe)
  # angewendet
}

TestFkt <- function(v) {
  return(wilcox.test(v)[[3]])   # P-Wert des Wilcoxontest.
}

VektorAbsolutMedian <- function(zM) {
   abszM <- abs(zM)
   vaM   <- apply(abszM,1,median)
```

```
   return(vaM)
}

VektorGewichte <- function(zM,h){
  vaM <- VektorAbsolutMedian(zM)
  vG  <- vaM^h

  return(vG)
}

ZufallsZahlen2 <- function(anzProb,anzNull,anzAltern,mN=0,mA=1.15) {
   av <- anzNull + anzAltern
   aN <- (anzNull+av)*anzProb
   aA <- anzAltern*anzProb

   zN <- rnorm(aN,mean=mN)
   zA <- rnorm(aA,mean=mA)

   alleN <- c(zN,zA)
   alleM <- matrix(alleN,nrow=2*av,byrow=T)

   return(alleM)
   ## 1) an Zeilen Normal mean Null (0)
   ## 2) aa Zeilen Normal mean Null (0)
   ## 3) an Zeilen Normal mean Null (0)
   ## 4) aa Zeilen Normal mean Alternativ
   ####  (Mittelwert der Alternativhypothese)
   ### 1) + 3) --> Nullhypothese
   ### 2) + 4) --> Alternativhypothese
}

Pwerte2 <- function(zM) {
  av <- length(zM[,1])/2
  pW <- numeric(av)
  for (i in 1:av) {
     m <- matrix(c(zM[i,],zM[(i+av),]),ncol=2)
     pW[i] <- as.numeric(wilcox.test(m)[3])
  }
```

```
  return(pW)
}


VektorGewichte2 <- function(zM) {
  av <- length(zM[,1])/2
  vG <- numeric(av)
  for (i in 1:av) {
    a <- c(zM[i,],zM[(i+av),])
    qa <- quantile(a)
    vG[i] <- qa[4]-qa[2]
  }
  return(vG)
}
Extrahieren <- function(erg) {
  az <- nrow(erg)
  erg <- erg[2:az,]
  return(mean(erg))
}

Extrahieren3 <- function(erg,i) {
  az <- nrow(erg)
  erg <- erg[2:az,]
  azn <- nrow(erg)
  welche <- ((1:azn)%%3) == i
  welche <- (1:azn)[welche]
  alle <- erg[welche,]
  asp <- ncol(alle)
  rerg <- numeric(asp)
  for (i in 1:asp) {
    rerg[i] <- mean(alle[,i])
  }
  return(rerg)
}
###############
KFDRGalt <- function(pw,g,alfa) {
   # Alternativprozedur zu KFDR. Diese läuft wegen der
   # Schleife wohl etwas langsamer, da in Szenarien
```

```
   # mit nur wenig Alternativhypothesen die Schleife
   # sehr lange bis zum Abbruch läuft.
   pw   <- c(0,pw) #damit mindestens eine Hypothese abgelehnt wird,
                   # die dummy Hypothese
   g    <- c(1,g)  #muss nacher abgezogen werden.
   nn   <- length(pw)
   pg   <- pw/g
   pgo  <- order(pg)

   m <- nn - 1
   aDm <- alfa/m
   for (i in nn:1) {
 if (pg[pgo[i]] <= (i-1)*aDm) {
    return(pgo[1:i]-1)
 }
   }
   return(pgo[1])
}
```

Diese Funktionen werden von der Funktion *GolubTest* sowie von den Funktionen bei der Simulation kleiner, vor allem ungleich großer Stichproben verwendet.

```
ZufallsZahlenMF <- function(anzM=6,anzF=9,anzNull=12525,
       anzAltern=100,mN=0,mA=2.4) {
  av <- anzNull + anzAltern
  amM <- av*anzM
  amFnull <- anzNull*anzF
  amFaltern <- anzAltern*anzF

  zmM <- rnorm(amM)
  zmFnull <- rnorm(amFnull)
  zmFaltern <- rnorm(amFaltern,mean=mA)

  mM <- matrix(zmM,nrow=av)
  mFnull <- matrix(zmFnull,nrow=anzNull)
```

```
   mFaltern <- matrix(zmFaltern,nrow=anzAltern)

   MM <- matrix(0,nrow=av,ncol=anzM+anzF)

   MM[1:av,1:anzM] <- mM
   MM[1:anzNull,(anzM+1):(anzM+anzF)] <- mFnull
   MM[(anzNull+1):av,(anzM+1):(anzM+anzF)] <- mFaltern

   return(MM)

}

VektorGewichteMF <- function(zM,aM=6,aF=9) {
    av <- nrow(zM)  # AnzahlVariablen
    vG <- numeric(av)       # VektorGewichte
    for (i in 1:av) {
      a <- c(zM[i,1:aM],zM[i,(aM+1):(aM+aF)])
      qa <- quantile(a)
      vG[i] <- qa[4]-qa[2]
  }
  return(vG)
}

PwerteMFw <- function(zM,aM=6,aF=9) {
  av <- nrow(zM)  # AnzahlVariablen
  pW <- numeric(av)
  for (i in 1:av) {
    pW[i] <- as.numeric(wilcox.test(x=zM[i,1:aM],y=zM[i,(aM+1):(aM+aF)])[3]
  }
  return(pW)
}

PwerteMFt <- function(zM,aM=6,aF=9) {
  av <- nrow(zM)  # AnzahlVariablen
  pW <- numeric(av)
  for (i in 1:av) {
    pW[i] <- as.numeric(t.test(x=zM[i,1:aM],y=zM[i,(aM+1):(aM+aF)])[3])
  }
```

```
  return(pW)
}
```

Symbolverzeichnis

Es ist schon seit längerem bekannt, daß man bei der Auswahl des richtigen Variablennamen oft sehr jonglieren muß [21]. Das gilt natürlich auch für Bezeichner aller Art. Man muß sich auch daran gewöhnen, daß das gleiche Symbol in verschiedener Bedeutung verwendet wird [22].
For the convenience of the reader, wie man es auf Englisch sagen würde, listen wir hier die wichtigsten Bezeichner auf.
Wenn ein Bezeichner sehr häufig auftritt, etwa m für die Anzahl der zu testenden Hypothesen wird ∞ in der Seitenspalte angegeben, sonst mit ein paar Ausnahmen das erste Auftreten.

Symbol	Seite	Kurzbeschreibung
$B(t)$	40	$1/G(t)$
$\hat{B}_m(t)$	40	$1/\hat{G}_m(t)$
BH[a]	18	Benjamini-Hochberg Prozedur
$C(t)$	48	$1/D(t)$
$\hat{C}_m(t)$	48	$1/\hat{D}_m(t)$
$D(t)$	44	Verteilungsfunktion von P/W
$\hat{D}_m(t)$	44	Empirische Verteilungsfunktion von P/W
$F(t), F$	38	Verteilungsfunktion der P-Werte bei Gültigkeit der Alternative
FDP	18	$V/R \cdot 1_{\{R>0\}}$
FDR	18	$E(V/R \cdot 1_{\{R>0\}})$
FWER	9	Wahrscheinlichkeit, mindestens eine wahre Nullhypothese abzulehnen
g_i	∞	Gewichte
$G(t)$	40	Verteilungsfunktion der P-Werte
$\hat{G}_m(t)$	40	Empirische Verteilungsfunktion der P-Werte
H_i	∞	i-te Hypothese, aber auch $H_i = 1$, falls H_i Alternative, $= 0$ sonst.
$H(t)$	53	$P(P/W \leq t \mid H = 1)$
m	∞	Anzahl der betrachteten Hypothesen
m_0	∞	Anzahl der wahren Nullhypothesen
m_1	∞	Anzahl der wahren Alternativhypothesen
pFDR	29	$E(V/R \mid R > 0)$
P_i	∞	P-Wert der i-ten Hypothese
$P_{(i)}$	∞	i-t größter P-Wert
Q	18	Andere Bezeichnung für FDP
Q_i	∞	P_i/W_i
R	17	Anzahl abgelehnter Hypothesen
$R_{k,i}$	20,41	Siehe die Definition im Text
S	17	Anzahl abgelehnter wahrer Alternativen
T	17	Anzahl angenommener wahrer Alternativen
U	17	Anzahl angenommener wahrer Nullhypothesen
V	17	Anzahl abgelehnter wahrer Nullhypothesen
$V_m(t)$	51	$m^{-1} \sum_i (1 - H_i) 1_{\{P_i/W_i \leq t\}}$
$V(t)$	51	$V(t) = E(1 - H_i) 1_{\{P_i/W_i \leq t\}}$
W_i	∞	Gewicht

[a] BH kann auch für Bonferroni-Holm stehen. Der Bezeichnungsteufel hat zugeschlagen [21].

Literaturverzeichnis

[1] Benjamini Yoav, Hochberg Yosef
Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing
J.R. Statist. Soc. B (1995)
57,No 1, pp289-300

[2] Benjamini Yoav, Hochberg Yosef
Multiple Hypotheses Testing with Weights
Scandinavian Journal of Statistics, Vol. 24, 407-418, 1997

[3] Benjamini Yoav, Liu Wei
A distribution-free multiple-test procedure that controls the false discovery rate
Auf der homepage von Benhamini publiziert:
http://www.math.tau.ac.il/ ybenja/MyPapers.html

[4] Dudoit, Sandrine, van der Laan, Mark J.
Multiple Testing Procedures with Applications to Genomics
Springer

[5] Finos L, Salmaso L
Weighted Methods controlling the multiplicity when the number of variables is much higher than the number of observations
Journal of Nonparametric Statistics, Vol. 18, No.2, February 2006, 245-261

[6] Genovese Christopher, Wasserman Larry
A Large Sample Approach to False Discovery Control
Im Internet unter http://lib.stat.cmu.edu/ genovese/papers/

[7] Genovese Christopher, Roeder Kathryn, Wasserman Larry
False discovery control with p-value weighting
Biometrika(2006),**93**,3,pp 509-524

[8] Herve Abdi
The Bonferroni and Sidak Corrections for Multiple Comparisons
in: Neil Salkind (Ed.). Encyclopedia of Measurement and Statistics, Thousand Oaks (CA):Sage
(auch im Internet erhältlich)

[9] Hochberg Y, Tamhane A.
Multiple Comparison Procedures
New York, Wiley, 1987

[10] Hofbauer Franz
Wahrscheinlichkeitstheorie 1
Skriptum, erhältlich unter http://www.mat.univie.ac.at/

[11] Holm S
A simple sequentially rejective multiple test procedure
Scandinavian Journal of Statistics, **6**, 65-70

[12] Hommel Gerhard, Kropf Siegfried
Tests for Differentiation in Gene Expression Using a Data-Driven Order or Weights for Hypotheses
Biometrical Journal **47** (2005) 4,554-562

[13] Hsu, Jason C.
Multiple Comparisons: Theory and Methods
Chapmann Hall 1996

[14] Kropf Siegfried
Hochdimensionale multivariate Verfahren in der medizinischen Statistik
Shaker Verlag 2000

[15] Kropf Siegfried , Jürgen Läuter, Markus Eszlingerb, Knut Krohnb, Ralf Paschke
Nonparametric multiple test procedures with data-driven order of hypotheses and with weighted hypotheses

Journal of Statistical Planning and Inference
Volume 125, Issues 1-2, October 2004, Pages 31-47

[16] Läuter J.
Exact t and F Tests for Analyzing Repeated Measurements
Biometrics **52**,pp 964-970

[17] Läuter J, Glimm E., Kropf S.
New Multivariate Tests for Data with an Inherent Structure
Biometrical Journal **38**, 5-22.
Erratum: Biometrical Journal **38**,1015

[18] Miller R.G
Simultaneous Statistical Inference
Springer, 1981

[19] Moye Lemuel A.
Multiple Analyses in Clinical Trials: Fundamentals for Investigators
Springer 2003

[20] Pflug Georg
Mathematische Statistik: Skriptum zur Vorlesung von Prof. Pflug
nur persönlich bei Prof. Pflug, bzw im Sekratariat, Universitätsstrasse 5/9, 1090 Wien erhältlich.

[21] Reiter, Hans
Die Marmelade in der Sachertorte
Ordinarius zu Wien. Zu Lebzeiten in der hohen Abstraktion beheimatet, derzeitiger Aufenthaltsort unbekannt
persönliche Mitteilung

[22] Rindler, Harald
Bemerkungen über dies und jenes
Mathematiker, Musiker, Philosoph und Mensabesucher
persönliche Mitteilung

[23] Roussas, George G
A Course in Mathematical Statistics
Academic Press, 1997

[24] Storey John D.
A direct approach to false discovery rates
J.R.Statist.Soc **B** (2002)
64,Part 3, pp. 479-498

[25] Storey John D., Taylor Jonathan, Siegmung David
Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach
J.R.Statist.Soc **B** (2004)
66,Part 1, pp. 187-205

Lebenslauf

- 02/2007 Beginn der Magisterarbeit bei Univ Doz Dr. Martin Posch
- 11/2006 Alle Prüfungen abgelegt
- 10/2005 Beginn des Magisterstudiums der Statistik
- 12/2005 Winterreise nach Leiston, Suffolk
- 05/08/05 **Heirat** mit DI DDr. Karin Datzberger (ab ca 13:20 immer noch DI DDR Karin, aber nunmehr Remmel)
- 26/04/05 Erlangung der Lenkerberechtigung für PKW
- ab 01/2003 MA 14 (Elektronische Datenverarbeitung) Mainframe, Webservices
- 07/2002-01/2003 Firma Quester (Operating)
- 02/2000 - 06/2002 ICS (Webprogrammierung: ASP,PHP,TEX)
- 09/1990 - 12/1998 Post/Telekom (Auslandsauskunft)
- 12/1995 Abschluß des Diplomstudiums der Mathematik mit Auszeichnung. Diplomarbeit bei Prof. Krattenthaler über *extremal set theory*
- Ausgedehnte Italienreisen, auf den Spuren von Eissalons und Bahnhöfen nicht Goethes.
- 1969-1981 Besuch verschiedener Schulen, unter anderem des BRG XVIII, an welchem auch die Matura abgelegt wurde
- 23/12/1962 Geburt in Wien Favoriten

Abstract

In der vorliegenden Arbeit werden die klassischen multiplen Testverfahren von Bonferroni und Bonferroni-Holm sowohl im gewichteten als auch im ungewichteten Fall behandelt. Die False Discovery Rate FDR und die pFDR werden behandelt. Es wird die ungewichtete und gewichtete Benjamini-Hochberg Prozedur untersucht. Gewichte werden sowohl a priori als auch a posteriori aus den Daten bestimmt. Im Fall der aus den Daten bestimmten Gewichten wird auch die noch nicht in der Litertur behandelte FDR untersucht. Simulationen werden für alle Prozeduren durchgeführt.