

2.3 Das Prinzip des Abschlußtests

Die Pionierarbeit zum **Prinzip der Abschlußtests**, kurz **Abschluß-Prinzip** genannt (engl.: closed principle), ist eine Publikation von MARCUS, PERITZ und GABRIEL (1976). In SONNEMANN (1982) wird die Philosophie der Abschlußverfahren ausführlich dargestellt.

Das Testen nach diesem Prinzip ist komplizierter und erfordert mehr Rechenarbeit als nach dem Bonferroni-Prinzip. Der Vorteil von **Abschlußtests** (engl.: closed tests) liegt allgemein darin, daß sie unter allen multiplen Testverfahren, die bei einem bestimmten multiplen Testproblem ein vorgegebenes multiples Signifikanzniveau α einhalten, die höchste Güte aufweisen, also die meisten der falschen Nullhypothesen ablehnen.

Wir wollen das Abschluß-Prinzip am paarweisen Vergleich der erwarteten Hemmhofdurchmesser aus Beispiel 1.1 demonstrieren. Wir beschränken uns aber der Einfachheit halber auf den Vergleich der Mutanten 1 bis 4, d.h. auf die sechs Elementarhypothesen $H_{12}, H_{13}, H_{14}, H_{23}, H_{24}, H_{34}$, wobei $H_{ij}: \mu_i = \mu_j$, denn die Anzahl der notwendigen Rechenschritte ist bereits bei fünf Mutanten erheblich größer.

Aus diesen sechs Elementarhypothesen können durch 'und-Verknüpfung' neue Hypothesen gebildet werden, z.B. aus H_{12} und H_{34} die Hypothese

$$H_{12}H_{34}: \mu_1 = \mu_2 \text{ und } \mu_3 = \mu_4$$

oder aus H_{12} und H_{23} die Hypothese

$$H_{12}H_{23}: \mu_1 = \mu_2 \text{ und } \mu_2 = \mu_3,$$

die wir logischerweise auch schreiben können als

$$H_{12}H_{23}: \mu_1 = \mu_2 = \mu_3.$$

Solche Kombinationen von Hypothesen bezeichnet man als **Durchschnittshypothesen**.

Eine Durchschnittshypothese ist genau dann wahr, wenn all ihre Einzelhypothesen wahr sind. So ist z.B. $H_{12}H_{34}$ wahr, wenn sowohl $\mu_1 = \mu_2$ als auch $\mu_3 = \mu_4$ gilt.

Wir haben nun in Tabelle 2.1 alle möglichen Durchschnittshypothesen, die sich aus den sechs interessierenden Elementarhypothesen $H_{12}, H_{13}, H_{14}, H_{23}, H_{24}, H_{34}$ bilden lassen, schematisch und in einer Hierarchie dargestellt.

Die Entscheidungsprozedur nach dem Abschluß-Prinzip verlangt, sowohl die Elementarhypothesen als auch deren Durchschnittshypothesen zu prüfen. Das erfordert, daß für alle diese Hypothesen geeignete Tests zur Verfügung stehen.

Die Entscheidungsprozedur, die ein multiples Signifikanzniveau α gewährleistet, läßt sich dann folgendermaßen allgemein beschreiben.

Es ist zunächst jede Elementarhypothese sowie jede Durchschnittshypothese auf dem gleichen lokalen Signifikanzniveau α zu prüfen, ohne bereits eine Entscheidung bzgl. Annahme oder Ablehnung der Elementarhypothesen zu treffen. Die endgültige Entscheidung bezüglich der Elementarhypothesen erfolgt dann nach folgender Regel:

Horn und Vollandt (1995)

Lehne eine Elementarhypothese genau dann ab, wenn sich bei ihrer Prüfung und bei der Prüfung jeder Durchschnittshypothese, die diese Elementarhypothese enthält, Signifikanz auf dem lokalen Signifikanzniveau α ergab.

Nach dieser Regel ist beispielsweise die Elementarhypothese H_{12} : $\mu_1 = \mu_2$ genau dann abzulehnen, wenn H_{12} selbst sowie alle Durchschnittshypothesen, die H_{12} enthalten, mit für sie geeigneten Tests auf dem lokalen Niveau α abgelehnt wurden.

Tab. 2.1: Hierarchisches Schema der möglichen Durchschnittshypothesen, die sich bei den paarweisen Vergleichen von vier Mutanten ergeben

$H_{12}H_{13}H_{14}H_{23}H_{24}H_{34}$					
$H_{12}H_{13}H_{14}H_{23}H_{24}$ $H_{12}H_{13}H_{23}H_{24}H_{34}$		$H_{12}H_{13}H_{14}H_{23}H_{34}$ $H_{12}H_{14}H_{23}H_{24}H_{34}$		$H_{12}H_{13}H_{14}H_{24}H_{34}$ $H_{13}H_{14}H_{23}H_{24}H_{34}$	
$H_{12}H_{13}H_{14}H_{23}$ $H_{12}H_{13}H_{14}H_{34}$ $H_{12}H_{14}H_{24}H_{34}$	$H_{12}H_{13}H_{14}H_{24}$ $H_{12}H_{13}H_{23}H_{34}$ $H_{13}H_{14}H_{24}H_{34}$	$H_{12}H_{13}H_{23}H_{24}$ $H_{12}H_{14}H_{23}H_{34}$ $H_{12}H_{23}H_{24}H_{34}$	$H_{12}H_{14}H_{23}H_{24}$ $H_{13}H_{14}H_{23}H_{34}$ $H_{13}H_{23}H_{24}H_{34}$	$H_{13}H_{14}H_{23}H_{24}$ $H_{12}H_{13}H_{24}H_{34}$ $H_{14}H_{23}H_{24}H_{34}$	
$H_{12}H_{13}H_{14}$ $H_{12}H_{14}H_{24}$ $H_{12}H_{13}H_{34}$ $H_{14}H_{23}H_{34}$	$H_{12}H_{13}H_{23}$ $H_{13}H_{14}H_{24}$ $H_{12}H_{14}H_{34}$ $H_{12}H_{24}H_{34}$	$H_{12}H_{14}H_{23}$ $H_{12}H_{23}H_{24}$ $H_{13}H_{14}H_{34}$ $H_{13}H_{24}H_{34}$	$H_{13}H_{14}H_{23}$ $H_{13}H_{23}H_{24}$ $H_{12}H_{23}H_{34}$ $H_{14}H_{24}H_{34}$	$H_{12}H_{13}H_{24}$ $H_{14}H_{23}H_{24}$ $H_{13}H_{23}H_{34}$ $H_{23}H_{24}H_{34}$	
$H_{12}H_{13}$ $H_{13}H_{14}$ $H_{14}H_{24}$	$H_{12}H_{14}$ $H_{13}H_{23}$ $H_{14}H_{34}$	$H_{12}H_{23}$ $H_{13}H_{24}$ $H_{23}H_{24}$	$H_{12}H_{24}$ $H_{13}H_{34}$ $H_{23}H_{34}$	$H_{12}H_{34}$ $H_{14}H_{23}$ $H_{24}H_{34}$	
H_{12}	H_{13}	H_{14}	H_{23}	H_{24}	H_{34}

Nehmen wir an, daß die Hemmhofdurchmesser der vier zu vergleichenden Mutanten Normalverteilungen mit gleichen, unbekannten Varianzen besitzen. Dann läßt sich für jede Hypothese in Tabelle 2.1 ein geeigneter Test finden.

Beispielsweise kann die Elementarhypothese H_{12} : $\mu_1 = \mu_2$ mit dem multiplen Zweistichproben-t-Test auf dem (lokalen) Signifikanzniveau α geprüft werden.

Zur Prüfung der Hypothese $H_{12}H_{23}$: $\mu_1 = \mu_2 = \mu_3$ eignet sich die einfache Varianzanalyse auf dem Niveau α .

Für die Durchschnittshypothese $H_{12}H_{34}$: $\mu_1 = \mu_2$ und $\mu_3 = \mu_4$ existiert kein spezieller Test. Sie kann nach dem Bonferroni-Prinzip geprüft werden. Dazu wird jede der beiden Elementarhypothesen H_{12} und H_{34} mit dem multiplen Zweistichproben-t-Test auf dem Niveau $\alpha/2$ geprüft. Die Durchschnittshypothese wird genau dann abgelehnt, wenn mindestens eine der Elementarhypothesen abgelehnt wurde.

Man erkennt, welche enorme Rechenarbeit das Abschluß-Prinzip in der beschriebenen Form erfordert. Zu einer wesentlichen Verminderung des Rechenaufwandes führt folgende Überlegung:

Man kann leicht feststellen, daß mehrere Durchschnittshypothesen inhaltlich das gleiche bedeuten, d.h. **logisch äquivalent** sind. So sind sämtliche Hypothesen, die in den ersten sechs Zeilen von Tabelle 2.1 stehen, gleichbedeutend mit der Aussage ' $\mu_1 = \mu_2 = \mu_3 = \mu_4$ '. Beispielsweise besagt die Hypothese $H_{12}H_{13}H_{14}H_{23}H_{24}$, daß $\mu_1 = \mu_2$ und $\mu_1 = \mu_3$ und $\mu_1 = \mu_4$ und $\mu_2 = \mu_3$ und $\mu_2 = \mu_4$. Die Hypothese $H_{12}H_{13}H_{14}H_{23}$ besagt, daß $\mu_1 = \mu_2$ und $\mu_1 = \mu_3$ und $\mu_1 = \mu_4$ und $\mu_2 = \mu_3$. Beide Hypothesen bedeuten also, daß ' $\mu_1 = \mu_2 = \mu_3 = \mu_4$ ', und sind somit logisch äquivalent.

Dagegen bedeutet die in der siebenten Zeile stehende Durchschnittshypothese $H_{12}H_{13}H_{23}$ lediglich, daß ' $\mu_1 = \mu_2 = \mu_3$ '. Die in der achten Zeile stehende Durchschnittshypothese $H_{12}H_{14}H_{24}$ ist gleichbedeutend mit der Aussage ' $\mu_1 = \mu_2 = \mu_4$ ' usw.

Insgesamt lassen sich die 62 Hypothesen in Tabelle 2.1 auf nur 14 logisch unterschiedliche Aussagen reduzieren, die wir wieder als Hypothesen auffassen können. Diese sind in Tabelle 2.2 zusammengefaßt.

Tab. 2.2: Zusammenfassung der logisch unterschiedlichen Hypothesen

H_{1234} : $\mu_1 = \mu_2 = \mu_3 = \mu_4$					
H_{123} : $\mu_1 = \mu_2 = \mu_3$ H_{134} : $\mu_1 = \mu_3 = \mu_4$		H_{124} : $\mu_1 = \mu_2 = \mu_4$ H_{234} : $\mu_2 = \mu_3 = \mu_4$			
$H_{12}H_{34}$: $\mu_1 = \mu_2, \mu_3 = \mu_4$		$H_{13}H_{24}$: $\mu_1 = \mu_3, \mu_2 = \mu_4$		$H_{14}H_{23}$: $\mu_2 = \mu_4, \mu_3 = \mu_4$	
H_{12} : $\mu_1 = \mu_2$	H_{13} : $\mu_1 = \mu_3$	H_{14} : $\mu_1 = \mu_4$	H_{23} : $\mu_2 = \mu_3$	H_{24} : $\mu_2 = \mu_4$	H_{34} : $\mu_3 = \mu_4$

Es läßt sich beweisen und erscheint auch plausibel, daß es genügt, nur die 14 Hypothesen in Tabelle 2.2 zu prüfen. So wäre beispielsweise die Hypothese H_{12} : $\mu_1 = \mu_2$ endgültig abzulehnen, wenn alle Hypothesen, die die Aussage ' $\mu_1 = \mu_2$ ' enthalten, abzulehnen sind. Das sind die Hypothesen H_{1234} , H_{123} , H_{124} und $H_{12}H_{34}$ sowie H_{12} selbst.

Anmerkung 1. Wenn man aus zwei Hypothesen in Tabelle 2.1 eine neue Durchschnittshypothese bildet, dann entsteht dabei stets eine Durchschnittshypothese, die bereits in Tabelle 2.1 vorhanden ist. Das gleiche trifft auch auf die Hypothesen in Tabelle 2.2 zu.

Man sagt daher, daß die Hypothesen in Tabelle 2.1 und ebenso die in Tabelle 2.2 ein **durchschnittsabgeschlossenes System** bilden, s. HOMMEL (1986) und BERNHARD (1991). Daher kommt die Bezeichnung *Abschlußtest*.

Ein durchschnittsabgeschlossenes Hypothesensystem ist eine Voraussetzung für das Abschlußprinzip.

Anmerkung 2. Falls für irgendeine Durchschnittshypothese kein spezieller Test zur Verfügung steht, kann ihre Prüfung auf folgende universelle Weise erfolgen: Man prüft alle Elementarhypothesen, aus denen die Durchschnittshypothese gebildet ist, mit einem multiplen Test auf dem multiplen oder globalen Niveau α und lehnt die Durchschnittshypothese ab, sobald wenigstens eine der Elementarhypothesen abzulehnen ist. (Denn dann ist die Wahrscheinlichkeit, mindestens eine wahre, in der Durchschnittshypothese enthaltene Elementarhypothese irrtümlich abzulehnen, höchstens α , selbst wenn alle diese Elementarhypothesen wahr sind und damit die Durchschnittshypothese wahr ist. Folglich wird dann eine wahre Durchschnittshypothese nur höchstens mit Wahrscheinlichkeit α abgelehnt.)

Aus dem Bonferroni-Prinzip ergibt sich folgender einfache multiple Test zum multiplen Niveau α für k in einer Durchschnittshypothese enthaltene Elementarhypothesen: Man prüft jede Elementarhypothese auf dem Signifikanzniveau α/k und lehnt die Durchschnittshypothese genau dann ab, wenn mindestens eine der k Elementarhypothesen abzulehnen war.

Werden von vornherein alle Durchschnittshypothesen auf diese Weise nach dem Bonferroni-Prinzip getestet, führt das zu einer wesentlichen Einsparung von Testschritten und im Vergleich zum ursprünglichen Abschlußverfahren zu einer leicht durchführbaren Prozedur. Sie wird als Holm-Prozedur oder **Holm-Verfahren** bezeichnet und im nächsten Abschnitt beschrieben.

Die Anwendung des Bonferroni-Prinzips führt allerdings, wie man aus 2.1 schließen kann, zu einer relativ geringen Güte der Tests der Durchschnittshypothesen.

3 Paarweise Vergleiche

Inhalt

3.1	Verfahren bei Normalverteilungen	25
3.1.1	Ein Einschnitt-Verfahren mit Kontrolle des lokalen Signifikanzniveaus und gewöhnliche Konfidenzintervalle (multipler Zweistichproben-t-Test)	27
3.1.2	Ein Einschnitt-Verfahren mit Kontrolle des multiplen Signifikanzniveaus und simultane Konfidenzintervalle (Tukey-Kramer-Verfahren)	30
3.1.3	Ein Mehrschritt-Verfahren mit Kontrolle des multiplen Signifikanzniveaus (Holm-Prinzip)	32
3.2	Verfahren bei Binomialverteilungen	35
3.2.1	Ein Einschnitt-Verfahren mit Kontrolle des lokalen Signifikanzniveaus	37
3.2.2	Ein Einschnitt-Verfahren mit Kontrolle des multiplen Signifikanzniveaus	38
3.2.3	Ein Mehrschritt-Verfahren mit Kontrolle des multiplen Signifikanzniveaus (Holm-Prinzip)	40
3.3	Verteilungsfreie Verfahren	42
3.3.1	Ein Einschnitt-Verfahren mit Kontrolle des lokalen Signifikanzniveaus (Test von Wilcoxon-Mann-Whitney)	45
3.3.2	Ein Einschnitt-Verfahren mit Kontrolle des multiplen Signifikanzniveaus (Steel-Verfahren)	48
3.3.3	Ein Mehrschritt-Verfahren mit Kontrolle des multiplen Signifikanzniveaus (Holm-Prinzip)	49

In diesem Kapitel geht es um das Problem des paarweisen Vergleichs von a Populationen, d.h. des Vergleichs jeder Population mit jeder anderen.

3.1 Verfahren bei Normalverteilungen

In diesem Abschnitt 3.1 wird stets vorausgesetzt, daß die vorliegenden Daten normalverteilt sind. Die Methoden dienen ausschließlich dem Vergleich der Erwartungswerte μ_i von Normalverteilungen.

Problem

Gegeben seien a unabhängige Stichproben aus normalverteilten Populationen mit der gleichen, unbekannten Varianz σ^2 . Es sollen die unbekannten Populationsmittel μ_1, \dots, μ_a paarweise miteinander verglichen werden.

Dabei soll entweder das lokale oder das multiple Signifikanzniveau den Wert α haben. Außerdem sollen nach Möglichkeit für die unbekannten Erwartungswertdifferenzen $\mu_i - \mu_j$ gewöhnliche oder simultane Konfidenzintervalle zum Konfidenzniveau $1 - \alpha$ berechnet werden.

Symbole

Wir bezeichnen mit

$$x_{11}, \dots, x_{1n_1}$$

$$x_{21}, \dots, x_{2n_2}$$

⋮

⋮

$$x_{a1}, \dots, x_{an_a}$$

die Werte der a Stichproben

$$\bar{x}_1, \dots, \bar{x}_a$$

die a Stichprobenmittel

$$s_1^2, \dots, s_a^2$$

die a Stichprobenvarianzen

$$s^2 = \frac{\sum_{i=1}^a (n_i - 1) s_i^2}{n_1 + \dots + n_a - a}$$

die Schätzung der gemeinsamen Varianz σ^2

s

die Wurzel aus s^2

Hypothesen

Null- und Alternativhypothesen lauten

$$H_{ij}: \mu_i = \mu_j, H_{Aij}: \mu_i \neq \mu_j$$

bei zweiseitiger Fragestellung

bzw.

$$H_{ij}: \mu_i \geq \mu_j, H_{Aij}: \mu_i < \mu_j$$

bei einseitiger Fragestellung.

$$H_{ij}: \mu_i \leq \mu_j, H_{Aij}: \mu_i > \mu_j$$

Prüfgrößen

Für alle Tests in 3.1 berechnet man die Prüfgrößen

$$t_{ij} = \frac{\bar{x}_i - \bar{x}_j}{s \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \quad (i, j = 1, \dots, a; i \neq j).$$

3.1.1 Ein Einschritt-Verfahren mit Kontrolle des lokalen Signifikanzniveaus und gewöhnliche Konfidenzintervalle (multipler Zweistichproben-t-Test)

Das hier beschriebene Einschritt-Verfahren, das meistens als **multipler t-Test** und von uns als **multipler Zweistichproben-t-Test** bezeichnet wird, ist anzuwenden, wenn je zwei Erwartungswerte von a Normalverteilungen miteinander zu vergleichen sind und dabei ein bestimmtes *lokales Signifikanzniveau* eingehalten werden soll bzw., wenn *gewöhnliche Konfidenzintervalle* erwünscht sind.

Symbole

Wir bezeichnen mit

$$t_{f, 1-\alpha}$$

das $(1-\alpha)$ -Quantil der t-Verteilung zum Freiheitsgrad f (Tafel 2)

$$t_{f, 1-\alpha/2}$$

das $(1-\alpha/2)$ -Quantil der t-Verteilung zum Freiheitsgrad f (Tafel 2)

$$f = n_1 + \dots + n_a - a$$

den Freiheitsgrad

Gewöhnliche Konfidenzintervalle zum Konfidenzniveau $1-\alpha$

Gewöhnliche einseitige untere bzw. obere Konfidenzgrenzen für $\mu_i - \mu_j$ ($i, j = 1, \dots, a; i \neq j$) ermittelt man nach den Formeln

$$\bar{x}_i - \bar{x}_j - t_{f, 1-\alpha} s \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

bzw.

$$\bar{x}_i - \bar{x}_j + t_{f, 1-\alpha} s \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}.$$

Gewöhnliche zweiseitige Konfidenzgrenzen, d.h. die untere und die obere Grenze eines gewöhnlichen zweiseitig begrenzten Konfidenzintervalls für $\mu_i - \mu_j$ ($i, j = 1, \dots, a; i \neq j$) berechnet man nach der Formel

$$\bar{x}_i - \bar{x}_j \pm t_{f, 1-\alpha/2} s \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}.$$

Entscheidungsregel

Zweiseitige Fragestellung:

$H_{ij}: \mu_i = \mu_j$ ist zugunsten von $H_{Aij}: \mu_i \neq \mu_j$ abzulehnen, falls

$$|t_{ij}| > t_{f, 1-\alpha/2}$$

oder, was gleichbedeutend ist, falls das zweiseitige Konfidenzintervall für $\mu_i - \mu_j$ den Wert Null nicht enthält.

Einseitige Fragestellungen:

$H_{ij}: \mu_i \geq \mu_j$ ist zugunsten von $H_{Aij}: \mu_i < \mu_j$ abzulehnen, falls

$$t_{ij} < -t_{f,1-\alpha}$$

oder, was gleichbedeutend ist, falls die einseitige obere Konfidenzgrenze für $\mu_i - \mu_j$ kleiner als Null ist.

$H_{ij}: \mu_i \leq \mu_j$ ist zugunsten von $H_{Aij}: \mu_i > \mu_j$ abzulehnen, falls

$$t_{ij} > t_{f,1-\alpha}$$

oder, was gleichbedeutend ist, falls die einseitige untere Konfidenzgrenze für $\mu_i - \mu_j$ größer als Null ist.

Signifikanzniveau

Mit dem multiplen t-Test wird ein lokales Signifikanzniveau α eingehalten.

Anmerkung. Im Falle von zwei Stichproben ist der multiple Zweistichproben-t-Test mit dem gewöhnlichen Zweistichproben-t-Test identisch (deshalb bezeichnen wir ihn als multiplen Zweistichproben-t-Test und unterscheiden ihn von dem von uns so genannten multiplen Einstichproben-t-Test, vgl. 5.1.1). Beim Vergleich von mehr als zwei Stichproben ist der multiple t-Test dem gewöhnlichen t-Test überlegen. Denn σ^2 wird nicht nur aus den beiden jeweils zu vergleichenden Stichproben, sondern aus allen vorliegenden Stichproben geschätzt, wodurch die Schätzung genauer wird. Das wird dann bei dem Freiheitsgrad berücksichtigt. Der Freiheitsgrad der t-Verteilung, mit deren Quantil die t-Prüfgröße des gewöhnlichen Zweistichproben-t-Tests zu vergleichen ist, ist $n_1 + n_2 - 2$. Dagegen ist beim multiplen t-Test $n_1 + \dots + n_a - a$ der Freiheitsgrad. Signifikanz liegt vor, wenn die Prüfgröße des t-Tests das Quantil der t-Verteilung überschreitet. Das Quantil der t-Verteilung ist um so kleiner, je größer der Freiheitsgrad ist. Es ist also beim Freiheitsgrad $n_1 + \dots + n_a - a$ kleiner und wird daher von der Prüfgröße eher überschritten als beim Freiheitsgrad $n_1 + n_2 - 2$. (So ist in Beispiel 1.1 $n_1 = \dots = n_6 = 8$ und damit $n_1 + n_2 - 2 = 14$ und $n_1 + \dots + n_6 - 6 = 42$. Das 0.95-Quantil der t-Verteilung zum Freiheitsgrad 14 ist $t_{14,0.95} = 1.76$, das zum Freiheitsgrad 42 ist $t_{42,0.95} = 1.68$ und damit kleiner.) Daher ergibt sich beim Vergleich von zwei Stichproben mit dem multiplen t-Test eher Signifikanz als mit dem gewöhnlichen t-Test. Der multiple t-Test besitzt also eine höhere Güte.

Beispiel 3.1: Anhand der Daten aus Beispiel 1.1 soll jede Mutante mit jeder anderen Mutante bezüglich der erwarteten Hemmhofdurchmesser μ_i verglichen werden. Gleichzeitig sollen zweiseitige Konfidenzgrenzen, d.h. Konfidenzintervalle für die Differenzen $\mu_i - \mu_j$ der erwarteten Hemmhofdurchmesser berechnet werden. Das erfordert $6 \cdot 5/2 = 15$ zweiseitige Paarvergleiche bzw. Konfidenzintervalle. Dabei sei ausdrücklich gefordert, daß jeder der 15 Paarvergleiche auf dem Signifikanzniveau $\alpha = 0.05$ durchgeführt wird. Das ist gleichbedeutend damit, daß ein *lokales* Signifikanzniveau $\alpha = 0.05$ gefordert wird. Für die Konfidenzintervalle bedeutet diese Forderung, daß *gewöhnliche* Konfidenzintervalle zum Konfidenzniveau $1 - \alpha = 0.95$ zu berechnen sind.

Die Hemmhofdurchmesser werden als normalverteilt mit gleichen Varianzen vorausgesetzt. Daher läßt sich für die 15 Vergleiche und die 15 Konfidenzintervalle der multiple Zweistichproben-t-Test verwenden.

Für Beispiel 1.1 berechneten wir als gemeinsame Varianzschätzung $s^2 = 0.28$ und daraus $s = 0.53$. Der Freiheitsgrad ist $f = n_1 + \dots + n_a - a = 6 \cdot 8 - 6 = 42$. In Tafel 2 finden wir das entsprechende Quantil der t-Verteilung $t_{f,1-\alpha/2} = t_{42,0.975} = 2.02$. Die Differenzen $\bar{x}_i - \bar{x}_j$ sind bereits in Tabelle 1.2 und die Prüfgrößenwerte in Tabelle 1.3 und nochmals in Tabelle 3.1 zusammengefaßt (für $i < j$).

Die Absolutbeträge $|t_{ij}|$ der Prüfgrößenwerte sind mit 2.02 zu vergleichen.

Tab. 3.1: Werte der Prüfgrößen t_{ij} und Markierung '+' der Signifikanzen mit dem multiplen t-Test

j	i				
	1	2	3	4	5
2	0.45				
3	-2.87 ⁺	-3.32 ⁺			
4	0.15	-0.30	3.02 ⁺		
5	3.13 ⁺	2.68 ⁺	6.00 ⁺	2.98 ⁺	
6	2.38 ⁺	1.92	5.24 ⁺	2.23 ⁺	-0.75

In Tabelle 3.1 sind alle Prüfgrößenwerte mit einem '+' markiert, deren Absolutbetrag größer als 2.02 ist. In diesen Fällen ist die Hypothese $H_{ij}: \mu_i = \mu_j$ zugunsten der Alternative $H_{Aij}: \mu_i \neq \mu_j$ abzulehnen. D.h., es unterscheiden sich die Mutanten 1 und 3, 1 und 5, 1 und 6, 2 und 3, 2 und 5, 3 und 4, 3 und 5, 3 und 6, 4 und 5 sowie 4 und 6 signifikant in ihren mittleren Hemmhofdurchmessern. Für die 95%-Konfidenzintervalle berechnen wir

$$t_{f,1-\alpha/2} s \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} = 2.02 \cdot 0.53 \sqrt{\frac{1}{8} + \frac{1}{8}} = 0.54$$

Das ist die halbe Länge der Konfidenzintervalle.

Tab. 3.2: Gewöhnliche (dem multiplen Zweistichproben-t-Test entsprechende) 95%-Konfidenzintervalle für $\mu_i - \mu_j$

j	i				
	1	2	3	4	5
2	[-0.42, 0.66]				
3	[-1.30, -0.22]	[-1.42, -0.34]			
4	[-0.50, 0.58]	[-0.62, 0.46]	[0.26, 1.34]		
5	[0.29, 1.37]	[0.17, 1.25]	[1.05, 2.13]	[0.25, 1.33]	
6	[0.09, 1.17]	[-0.03, 1.05]	[0.85, 1.93]	[0.05, 1.13]	[-0.74, 0.34]

Tabelle 3.2 enthält die 95%-Konfidenzintervalle für alle 15 Differenzen $\mu_i - \mu_j$ für $i < j$. (Für $i > j$ erhält man die Intervalle durch Spiegelung an Null. Wenn z.B. das Intervall für $\mu_i - \mu_j$ die Grenzen -3 und +4 hat, dann hat das Intervall für $\mu_j - \mu_i$ die Grenzen -4 und +3.) Es ist die Hypothese $H_{ij}: \mu_i = \mu_j$ abzulehnen, wenn das zugehörige Konfidenzintervall die Null nicht überdeckt, d.h., wenn die Intervallgrenzen entweder beide positiv oder beide negativ sind. Das sind, wie

man sich leicht überzeugt, genau die Paare (i,j), bei denen in der vorangegangenen Tabelle die Markierung '+' steht.

3.1.2 Ein Einschnitt-Verfahren mit Kontrolle des multiplen Signifikanzniveaus und simultane Konfidenzintervalle (Tukey-Kramer-Verfahren)

Das sogenannte **Tukey-Kramer-Verfahren**, ein Einschnitt-Verfahren, ist zu verwenden, wenn die Erwartungswerte von a Normalverteilungen mit gleicher Varianz paarweise miteinander zu vergleichen sind und dabei ein bestimmtes *multipl. Signifikanzniveau* eingehalten werden soll bzw., wenn *simultane Konfidenzintervalle* erwünscht sind. Es hat eine höhere Güte als vergleichbare Einschnitt-Verfahren.

Das Verfahren wird (für gleich große Stichproben) auch als **Studentisierte Range-Prozedur** oder als **HSD-Verfahren** oder **WSD-Verfahren** bezeichnet (hsd=honestly significant difference, wsd=wholly significant difference).

Symbole

Wir bezeichnen mit

$q_{a,f,1-\alpha}$ das $(1-\alpha)$ -Quantil der Verteilung der studentisierten Spannweite mit dem Parameter a und dem Freiheitsgrad f (Tafel 4)

$f=n_1+\dots+n_a-a$ den Freiheitsgrad

Simultane Konfidenzintervalle zum Konfidenzniveau $1-\alpha$

Simultane zweiseitige Konfidenzgrenzen, d.h. die untere und die obere Grenze eines simultanen zweiseitig begrenzten Konfidenzintervalls für $\mu_i-\mu_j$ berechnet man nach der Formel

$$\bar{x}_i - \bar{x}_j \pm \frac{q_{a,f,1-\alpha}}{\sqrt{2}} s \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}.$$

(Einseitige Konfidenzintervalle werden mit dieser Methode nicht ermittelt.)

Entscheidungsregel

Zweiseitige Fragestellung:

$H_{ij}: \mu_i = \mu_j$ ist zugunsten von $H_{Aij}: \mu_i \neq \mu_j$ abzulehnen, falls

$$|t_{ij}| > \frac{q_{a,f,1-\alpha}}{\sqrt{2}}$$

oder, was gleichbedeutend ist, falls das simultane zweiseitige Konfidenzintervall für $\mu_i-\mu_j$ den Wert Null nicht enthält.

(Hier sind nur zweiseitige Vergleiche möglich.)

Signifikanzniveau

Mit dem Tukey-Kramer-Verfahren wird ein multiples Signifikanzniveau α eingehalten.

Anmerkung. Einseitige Fragestellungen werden mit diesem Verfahren nicht geprüft. In TUKEY (1953) ist die Methode nur für gleiche Stichprobenumfänge hergeleitet. Für ungleiche Stichprobenumfänge ist das Verfahren leicht konservativ. D.h., bei den Testentscheidungen kommt es mit etwas geringerer Wahrscheinlichkeit zu Signifikanzen, bzw. die Konfidenzintervalle werden etwas zu lang. Das Tukey-Kramer-Verfahren ist dennoch anderen Verfahren darin überlegen, daß es die kürzesten simultanen Konfidenzintervalle liefert bzw. mindestens zu so vielen Signifikanzen führt, wie andere Einschnitt-Verfahren, s. HOCHBERG und TAMHANE (1987).

Beispiel 3.2: Wie in Beispiel 3.1 soll jede Mutante mit jeder anderen Mutante bezüglich der erwarteten Hemmhofdurchmesser zweiseitig verglichen werden, jetzt jedoch auf dem *multiplen* Signifikanzniveau $\alpha=0.05$.

Gleichzeitig sollen zweiseitige *simultane* Konfidenzintervalle zum Konfidenzniveau $1-\alpha=0.95$ für die Differenzen $\mu_i-\mu_j$ der erwarteten Hemmhofdurchmesser berechnet werden.

Die Demonstration verschiedener Verfahren am gleichen Datenmaterial erspart Rechenarbeit und hat den weiteren Vorteil, daß die Verfahren an Hand der mit ihnen erzielten Ergebnisse verglichen werden können. Der Leser darf jedoch dabei nicht zu der Auffassung gelangen, daß sich jedes Verfahren für jede praktische Situation eignet und daß es z.B. gleichgültig sei, welchen Signifikanzniveau-Typ man in einem bestimmten Fall kontrolliert.

Die Hemmhofdurchmesser werden wie in Beispiel 3.1 als normalverteilt mit gleichen Varianzen vorausgesetzt. Deshalb kann für die 15 Vergleiche und die 15 simultanen Konfidenzintervalle das Tukey-Kramer-Verfahren angewandt werden.

Für Beispiel 1.1 berechneten wir als gemeinsame Varianzschätzung $s^2=0.28$ und daraus $s=0.53$. Der Freiheitsgrad ist $f=n_1+\dots+n_a-a=6\cdot 8-6=42$. In Tafel 4 finden wir das entsprechende Quantil der studentisierten Spannweite $q_{a,f,1-\alpha}=q_{6,42,0.95}=4.22$. Die Differenzen $\bar{x}_i-\bar{x}_j$ sind bereits in Tabelle 1.2 zusammengefaßt. Die Prüfgrößenwerte aus Tabelle 1.3 sind nochmals in Tabelle 3.3 angegeben. Ihre Absolutbeträge sind mit $4.22/\sqrt{2}=2.98$ zu vergleichen.

Tab. 3.3: Werte der Prüfgrößen t_{ij} und Markierung '+' der Signifikanzen mit dem Tukey-Kramer-Verfahren

j	i				
	1	2	3	4	5
2	0.45				
3	-2.87	-3.32 ⁺			
4	0.15	-0.30	3.02 ⁺		
5	3.13 ⁺	2.68	6.00 ⁺	2.98	
6	2.38	1.92	5.24 ⁺	2.23	-0.75

In Tabelle 3.3 sind alle Werte mit '+' markiert, deren Absolutbeträge größer als 2.98 sind. In diesen Fällen ist die Hypothese $H_{ij}: \mu_i = \mu_j$ zugunsten der Alternative $H_{Aij}: \mu_i \neq \mu_j$ abzulehnen. D.h., es unterscheiden sich die Mutanten 1 und 5, 2 und 3, 3 und 4, 3 und 5 sowie 3 und 6 signifikant in ihren mittleren Hemmhofdurchmessern.

der Alternative $H_{A16}: \mu_1 > \mu_6$ abzulehnen. D.h., die erwarteten Hemmhofdurchmesser μ_1 bis μ_4 sind gesichert größer als der erwartete Hemmhofdurchmesser μ_6 der Kontrollmutante. Die unteren 95%-Konfidenzgrenzen berechnen sich nach der Formel

$$\bar{x}_i - \bar{x}_6 - t_{\alpha,1-\alpha} \cdot s \sqrt{\frac{1}{n_i} + \frac{1}{n_6}} = \bar{x}_i - \bar{x}_6 - 1.68 \cdot 0.53 \sqrt{\frac{1}{8} + \frac{1}{8}} = \bar{x}_i - \bar{x}_6 - 0.44 \quad (i=1, \dots, 5).$$

Damit erhalten wir als untere 95%-Konfidenzgrenzen für $\mu_1 - \mu_6, \dots, \mu_5 - \mu_6$ die Werte

$$\begin{aligned} \bar{x}_1 - \bar{x}_6 - 0.44 &= 0.63 - 0.44 = 0.19 \\ \bar{x}_2 - \bar{x}_6 - 0.44 &= 0.51 - 0.44 = 0.07 \\ \bar{x}_3 - \bar{x}_6 - 0.44 &= 1.39 - 0.44 = 0.95 \\ \bar{x}_4 - \bar{x}_6 - 0.44 &= 0.59 - 0.44 = 0.15 \\ \bar{x}_5 - \bar{x}_6 - 0.44 &= -0.20 - 0.44 = -0.64. \end{aligned}$$

Da die unteren Konfidenzgrenzen für $\mu_1 - \mu_6$ bis $\mu_4 - \mu_6$ positiv ausfallen, gelten auch hier μ_1 bis μ_4 als gesichert größer als μ_6 .

4.1.2 Ein Einschritt-Verfahren mit Kontrolle des multiplen Signifikanzniveaus und simultane Konfidenzintervalle (Dunnett-Verfahren)

Wenn die Erwartungswerte von a Normalverteilungen gegen den Erwartungswert μ_0 einer ebenfalls normalverteilten Grundgesamtheit zu vergleichen sind und dabei ein bestimmtes *multipl. Signifikanzniveau* eingehalten werden soll bzw., wenn *simultane Konfidenzintervalle* erwünscht sind, ist das sogenannte **Dunnett-Verfahren** zu empfehlen. Es hat eine höhere Güte als vergleichbare Einschritt-Verfahren.

Diese Methode wurde in DUNNETT (1955) für einseitige und in DUNNETT (1964) für zweiseitige Vergleiche veröffentlicht.

Symbole

Wir bezeichnen mit

$t_{a,f,r,1-\alpha}$	das für den einseitigen Test erforderliche $(1-\alpha)$ -Quantil der a -dimensionalen t -Verteilung zum Freiheitsgrad f und zum Korrelationskoeffizienten r (Tafel 3a)
$ t _{a,f,r,1-\alpha}$	das für den zweiseitigen Test erforderliche $(1-\alpha)$ -Quantil der a -dimensionalen t -Verteilung zum Freiheitsgrad f und zum Korrelationskoeffizienten r (Tafel 3b)
$f = n_0 + \dots + n_a - a - 1$	den Freiheitsgrad
r	den Korrelationskoeffizienten

Dabei erhält man r folgendermaßen aus den Stichprobenumfängen: Zunächst berechnet man alle $m=a(a-1)/2$ Größen

$$r_{ij} = \frac{1}{\sqrt{(1 + \frac{n_0}{n_i})(1 + \frac{n_0}{n_j})}} \quad (i < j).$$

Danach berechnet man aus den r_{ij} das arithmetische Mittel

$$r = \frac{r_{12} + r_{13} + \dots + r_{1a} + r_{23} + r_{24} + \dots + r_{2a} + \dots + r_{(a-1)a}}{m},$$

d.h.

$$r = \frac{1}{m} \sum_{i=1}^{a-1} \sum_{j>i}^a r_{ij}.$$

Falls $n_1 = \dots = n_a = n$, haben alle r_{ij} und damit auch ihr arithmetisches Mittel den gleichen Wert $r = n/(n+n_0)$. Falls sogar $n_0 = n_1 = \dots = n_a$, ergibt sich für alle r_{ij} und somit für ihren Mittelwert $r = 0.5$.

Simultane Konfidenzintervalle zum Konfidenzniveau $1-\alpha$

Simultane einseitige untere bzw. obere Konfidenzgrenzen für $\mu_i - \mu_0$ berechnet man nach den Formeln

$$\bar{x}_i - \bar{x}_0 - t_{a,f,r,1-\alpha} s \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}$$

bzw.

$$\bar{x}_i - \bar{x}_0 + t_{a,f,r,1-\alpha} s \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}$$

Simultane zweiseitige Konfidenzgrenzen, d.h. die untere und die obere Grenze eines zweiseitig begrenzten simultanen Konfidenzintervalls für $\mu_i - \mu_0$ berechnet man nach der Formel

$$\bar{x}_i - \bar{x}_0 \pm |t|_{a,f,r,1-\alpha} s \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}$$

Entscheidungsregel

Zweiseitige Fragestellung:

$H_{i0}: \mu_i = \mu_0$ ist zugunsten von $H_{Ai0}: \mu_i \neq \mu_0$ abzulehnen, falls

$$|t_{i0}| > |t|_{a,f,r,1-\alpha}$$

oder, was gleichbedeutend ist, falls das zweiseitige Konfidenzintervall für $\mu_i - \mu_0$ den Wert Null nicht enthält.

Einseitige Fragestellungen:

$H_{i0}: \mu_i \geq \mu_0$ ist zugunsten von $H_{Ai0}: \mu_i < \mu_0$ abzulehnen, falls

$$t_{i0} < -t_{a,f,r,1-\alpha}$$

oder, was gleichbedeutend ist, falls die einseitige obere Konfidenzgrenze für $\mu_i - \mu_0$ kleiner als Null ist.

$H_{i0}: \mu_i \leq \mu_0$ ist zugunsten von $H_{Ai0}: \mu_i > \mu_0$ abzulehnen, falls

$$t_{i0} > t_{a,f,r,1-\alpha}$$

oder, was gleichbedeutend ist, falls die einseitige untere Konfidenzgrenze für $\mu_i - \mu_0$ größer als Null ist.

Signifikanzniveau

Mit diesem Verfahren wird ein multiples Signifikanzniveau α eingehalten.

Anmerkung 1. Falls die gemeinsame Varianz σ^2 bekannt ist (ein Fall, der eher von theoretischem Interesse ist, da er praktisch selten eintritt), sind lediglich die Quantile zum Freiheitsgrad $f = \infty$ zu verwenden, und in den allgemeinen Formeln ist s durch σ zu ersetzen.

Anmerkung 2. Interessanterweise sind beim Dunnett-Test für einseitige Vergleiche die Quantile einer anderen Prüfverteilung als für zweiseitige Vergleiche zu verwenden, während bei vielen anderen Tests, wie z.B. dem t-Test, beim einseitigen Vergleich das $(1-\alpha)$ -Quantil und beim zweiseitigen Vergleich das $(1-\alpha/2)$ -Quantil ein und derselben Prüfverteilung zu verwenden ist.

Den einseitigen Test findet man erstmals in DUNNETT (1955), den zweiseitigen in DUNNETT (1964).

Anmerkung 3. Die oben beschriebenen Entscheidungsregeln sind nur dann exakt, wenn $n_1 = \dots = n_a$ gilt, d.h., wenn alle r_{ij} gleich sind (n_0 kann beliebig sein). Falls jedoch die Stichprobenumfänge n_1, \dots, n_a und damit die r_{ij} nicht alle gleich sind, führt die Verwendung der Quantile $t_{a,f,r,1-\alpha}$ bzw. $|t|_{a,f,r,1-\alpha}$ mit dem Mittelwert r zu einer Näherung, die von DUNNETT (1985) vorgeschlagen und in einer ausführlichen numerischen Studie als gut nachgewiesen wurde.

Die Quantile für das exakte Verfahren, die von n_0, n_1, \dots, n_a und damit von allen r_{ij} abhängen, lassen sich nur mit einem speziellen Computerprogramm berechnen, z.B. dem von DUNNETT (1989).

Beispiel 4.2: Wie in Beispiel 4.1 ist unter der Annahme von Normalverteilungen und gleichen Varianzen zu prüfen, welche Mutanten größere erwartete Hemmhofdurchmesser als die Kontrollmutante haben (einseitige Fragestellung), jetzt jedoch mit dem Dunnett-Verfahren auf dem multiplen Signifikanzniveau $\alpha = 0.05$.

Gleichzeitig sollen simultane untere Konfidenzgrenzen zum Konfidenzniveau $1-\alpha = 0.95$ für die Differenzen $\mu_i - \mu_6$ ($i = 1, \dots, 5$) berechnet werden.

Der Freiheitsgrad ist $f = n_1 + \dots + n_6 - 6 = 6 \cdot 8 - 6 = 42$. Die Stichprobenumfänge haben alle den gleichen Wert 8, so daß $r = 0.5$. In Tafel 3a finden wir das für den einseitigen Dunnett-Test erforderliche Quantil

St.
lie
en
ler

Symbole

Wir bezeichnen mit

$$t_{k,f,r,1-\alpha}$$

das für den einseitigen Test erforderliche $(1-\alpha)$ -Quantil der k -dimensionalen t -Verteilung ($1 \leq k \leq a$) zum Freiheitsgrad f und zum Korrelationskoeffizienten r (Tafel 3a)

$$|t|_{k,f,r,1-\alpha}$$

das für den zweiseitigen Test erforderliche $(1-\alpha)$ -Quantil der k -dimensionalen t -Verteilung ($1 \leq k \leq a$) zum Freiheitsgrad f und zum Korrelationskoeffizienten r (Tafel 3b)

en

$$f = n_0 + an - a - 1$$

den Freiheitsgrad

$$r = \frac{n}{n + n_0}$$

den Korrelationskoeffizienten

Im Falle $n_0 = n_1 = \dots = n_a$ ergibt sich $r = 0.5$.

μ_6

Entscheidungsprozedur

Der Einfachheit halber beschreiben wir hier nur den zweiseitigen Test. Die einseitige Prüfung erfolgt analog. Im ersten Schritt vergleicht man die a absoluten Prüfgrößenwerte $|t_{i0}|$ mit dem Quantil des zweiseitigen Dunnett-Tests $|t|_{a,f,r,1-\alpha}$ wie in 4.1.2. Ist keiner der Werte $|t_{i0}|$ größer als dieses Quantil, wird keine Hypothese abgelehnt, und die Prozedur endet.

Sind l_1 Werte größer als das Quantil, werden die zugehörigen l_1 Hypothesen H_{i0} abgelehnt. Falls $l_1 = a$, endet die Prozedur.

Falls $l_1 < a$, vergleicht man die restlichen $a - l_1$ absoluten Prüfgrößenwerte mit dem Quantil $|t|_{a-l_1,f,r,1-\alpha}$.

Ist keiner von ihnen größer, werden keine weiteren Hypothesen abgelehnt, und die Prüfung ist beendet. Sind jedoch l_2 Werte größer, werden die zugehörigen l_2 Hypothesen zusätzlich zu den l_1 Hypothesen abgelehnt.

Falls $l_1 + l_2 < a$, werden die restlichen $a - l_1 - l_2$ absoluten Prüfgrößenwerte nun mit dem Quantil $|t|_{a-l_1-l_2,f,r,1-\alpha}$ verglichen usw.

Bleibt nach einigen Schritten nur eine einzige (noch nicht abgelehnte) Elementarhypothese H_{i_0} übrig, wird über diese durch Vergleich des zugehörigen absoluten Prüfgrößenwertes mit dem Quantil $|t|_{1,f,r,1-\alpha} = t_{f,1-\alpha/2}$ der gewöhnlichen t -Verteilung (Tafel 2) entschieden.

β ,

4.1.3 Ein Mehrschritt-Verfahren mit Kontrolle des multiplen Signifikanzniveaus (Dunnett-Abschluß-Verfahren)

Auf der Grundlage des in 2.3 beschriebenen Abschluß-Prinzips und des in 4.1.2 beschriebenen Dunnett-Verfahrens läßt sich bei Normalverteilungen mit gleichen Varianzen für den Vergleich von a Populationen gegen eine Kontrollpopulation ein Mehrschritt-Verfahren konstruieren, das eine höhere Güte als das Dunnett-Verfahren hat. Wir bezeichnen es als **Dunnett-Abschluß-Verfahren**. Eine allgemeine Beschreibung findet man in DUNNETT und TAMHANE (1991).

Wir beschreiben das Verfahren nur für den Fall $n_1 = \dots = n_a = n$ (n_0 kann von n verschieden sein), da es dann relativ einfach ist. Für diesen Fall wurde es bereits in MARCUS, PERITZ und GABRIEL (1976) beschrieben.

Signifikanzniveau

Mit diesem Verfahren wird ein multiples Signifikanzniveau α eingehalten.

Anmerkung. Nach dem Abschluß-Prinzip sind alle Durchschnittshypothesen, die sich aus den Elementarhypothesen H_{i0} zusammensetzen, auf dem Signifikanzniveau α zu prüfen. Wie bereits in 2.3 erläutert,

lehnen wir eine Durchschnittshypothese ab, wenn mindestens eine der Elementarhypothesen, aus der sie sich zusammensetzt, auf dem multiplen Niveau α abzulehnen ist. So prüfen wir beispielsweise eine aus k Elementarhypothesen zusammengesetzte Durchschnittshypothese, indem wir alle k Elementarhypothesen mit dem Dunnett-Verfahren prüfen, d.h., jede der k Prüfgrößen mit $t_{k,f,r,1-\alpha}$ bzw. $|t|_{k,f,r,1-\alpha}$ vergleichen.

Im Falle ungleicher n_i wird das Verfahren komplizierter, da sich für jede Durchschnittshypothese aus den entsprechenden Stichprobenumfängen ein anderer mittlerer r -Wert und damit ein anderes angenähertes (oder exaktes) Quantil $t_{k,f,r,1-\alpha}$ bzw. $|t|_{k,f,r,1-\alpha}$ ergeben kann.

Beispiel 4.3: Unter der gleichen Voraussetzung wie in Beispiel 4.2 ist auf dem multiplen Signifikanzniveau $\alpha=0.05$ zu prüfen, welche Mutanten größere erwartete Hemmhofdurchmesser als die Kontrollmutante haben (einseitige Fragestellung), jetzt jedoch mit dem *Dunnett-Abschluß-Verfahren*. Im ersten Schritt vergleichen wir die Prüfgrößenwerte

$$t_{i6} = \frac{\bar{x}_i - \bar{x}_6}{s \sqrt{\frac{1}{n_i} + \frac{1}{n_6}}} \quad (i=1, \dots, 5)$$

mit $t_{5,42,0.5,0.95}=2.31$. Da t_{16} und t_{36} größer als 2.31 sind, sind die $l_1=2$ Hypothesen H_{16} und H_{36} abzulehnen.

Im zweiten Schritt prüfen wir die restlichen $a-l_1=3$ Hypothesen.

Dazu vergleichen wir die zugehörigen Prüfgrößenwerte mit dem Quantil $t_{3,42,0.5,0.95}=2.12$. Da der zu H_{46} gehörige Prüfgrößenwert 2.23 größer ist als 2.12, ist auch H_{46} abzulehnen. Es ist $l_2=1$.

Im dritten Schritt prüfen wir die $a-l_1-l_2=5-2-1=2$ noch nicht abgelehnten Hypothesen H_{26} und H_{56} . Dazu vergleichen wir die beiden zugehörigen Prüfgrößenwerte mit dem Quantil $t_{2,42,0.5,0.95}=1.97$. Da beide kleiner als 1.97 sind, werden keine weiteren Hypothesen abgelehnt, und die Prozedur endet.

Tab. 4.3: Werte der Prüfgrößen t_{i6} und Markierung '+' der Signifikanzen nach dem Dunnett-Abschluß-Verfahren ($\alpha=0.05$)

i	1	2	3	4	5
t_{i6}	2.38 ⁺	1.92	5.24 ⁺	2.23 ⁺	-0.75

In Tabelle 4.3 haben wir die Prüfgrößenwerte der abgelehnten Hypothesen mit '+' markiert. Die erwarteten Hemmhofdurchmesser μ_1 , μ_3 und μ_4 sind also nach dem Dunnett-Abschluß-Verfahren gesichert größer als der erwartete Hemmhofdurchmesser μ_6 der Kontrollmutante.

6.4 Die Scheffé-Methode und die Duncan-Methode

Da die Verfahren von SCHEFFÉ (1953) und DUNCAN (1955) in der Literatur häufig erwähnt und des öfteren angewandt werden, wollen wir hier kurz auf sie eingehen.

Das **Verfahren von Scheffé**, das oft als **S-Prozedur** bezeichnet wird, ist wie das Tukey-Kramer-Verfahren ein Einschnitt-Verfahren. Es ist auch an die Voraussetzung von Normalverteilungen mit gleichen Varianzen gebunden.

Allgemein ist es anwendbar, um für sogenannte **lineare Kontraste** oder **Linearkombinationen** der Erwartungswerte μ_i simultane Konfidenzintervalle anzugeben bzw. Hypothesen zu prüfen. Eine Linearkombination beschreibt man allgemein durch $b_1\mu_1 + b_2\mu_2 + \dots + b_a\mu_a$, wobei die Nebenbedingung $b_1 + \dots + b_a = 0$ erfüllt sein muß.

Gilt speziell $b_1 = 1$, $b_2 = -1$ und $b_3 = \dots = b_a = 0$, so bleibt als Linearkombination die Differenz $\mu_1 - \mu_2$ übrig. Ebenso sind alle anderen Differenzen $\mu_i - \mu_j$ Spezialfälle. Man kann also speziell für alle Paardifferenzen $\mu_i - \mu_j$ simultane Konfidenzintervalle angeben bzw. alle Paarhypothesen $H_{ij}: \mu_i = \mu_j$ bei Kontrolle des multiplen Niveaus prüfen.

Allgemein gibt man simultane Konfidenzintervalle für mehrere interessierende Linearkombinationen $b_1\mu_1 + \dots + b_a\mu_a$ (mit $b_1 + \dots + b_a = 0$) an bzw. prüft alle entsprechenden Hypothesen ' $b_1\mu_1 + \dots + b_a\mu_a = 0$ ' unter Kontrolle des *multiplen* Signifikanzniveaus.

Speziell prüft man die Paarhypothesen $H_{ij}: \mu_i = \mu_j$ durch folgende Regel auf dem multiplen Niveau α :

Lehne H_{ij} ab, wenn

$$|t_{ij}| > \sqrt{(a-1)F_{a-1, f, 1-\alpha}}$$

Dabei ist t_{ij} die bereits beim multiplen t-Test, beim Tukey-Kramer-Verfahren und bei anderen Methoden zu verwendende Prüfgröße, $f = n_1 + \dots + n_a - a$, und $F_{a-1, f, 1-\alpha}$ ist das $(1-\alpha)$ -Quantil der F-Verteilung zu den Freiheitsgraden $a-1$ und f .

Das Scheffé-Verfahren hat bei der Prüfung aller Paarhypothesen eine geringere Güte als das Tukey-Kramer-Verfahren und alle vergleichbaren Methoden und ist deshalb nicht für die Prüfung aller Paarhypothesen H_{ij} zu empfehlen.

Das **Verfahren von Duncan**, das gelegentlich als **D-Prozedur** bezeichnet wird, ist (wie das Newman-Keuls-Verfahren und das Ryan-Verfahren) eine Step-down-Prozedur. Es ist nur bei Normalverteilungen mit gleichen Varianzen und gleichen Stichprobenumfängen anwendbar und prüft alle Paarhypothesen $H_{ij}: \mu_i = \mu_j$. Die Methode verlangt wie das Ryan-Verfahren für jeden Testschritt ein anderes Niveau α_i/m_i . Auf Grund seiner Niveaus führt das Duncan-Verfahren häufiger als das Ryan- oder Newman-Keuls-Verfahren zu Signifikanz. Jedoch kontrolliert die Duncan-Methode weder das multiple noch das globale Signifikanzniveau, sondern nur das *lokale* Niveau.

Die Interpretation der Testergebnisse ist schwierig, und das Verfahren kann nicht empfohlen werden.

Signifikanzniveau

Die Scheffé-Methode garantiert, daß das *multiple* Niveau α ist. Das Duncan-Verfahren garantiert, daß das *lokale* Niveau α ist.