

## Anhang 1

### *Friedman-Test*

Beim Friedman-Test handelt es sich um eine zweifache Rangvarianzanalyse, die zum Vergleich mehrerer verbundener Stichproben untereinander geeignet ist. Der Friedman-Test gehört zu den sog nichtparametrischen Verfahren, die verteilungsunabhängig sind [81]. Der größte Vorteil solcher Methoden liegt darin, daß sie auch Stichproben aus nicht-normalverteilten Grundgesamtheiten vergleichen können [82]. Obwohl es Tests gibt, die auf Abweichungen von der Normalverteilung reagieren, ist es unmöglich zu beweisen, dass eine Stichprobe einer normalverteilten Grundgesamtheit entstammt. Viele häufig benutzte Verfahren, wie der t-Test oder die einfaktorielle ANOVA, nehmen normalverteilte Grundgesamtheiten an [83]. Wenn diese Annahme zutrifft sind diese Tests trennschärfer als die entsprechenden nichtparametrischen Verfahren. Wenn aber keine normalverteilte Grundgesamtheit vorliegt, sollten diese Tests nicht benutzt werden. Ihre Trennschärfe (oder Power), d.h. ihre Fähigkeit, die Nullhypothese abzulehnen, wenn die Alternativhypothese tatsächlich zutrifft, ist dann gering.

In solchen Situationen sollte man entweder „normalisierende“ Transformationen durchführen (z. B. logarithmische Transformation) oder nichtparametrische Verfahren benutzen. Bei Abweichungen von der Normalverteilung sind nichtparametrische Tests trennschärfer. Der häufig benutzte Wilcoxon-Test für Paardifferenzen kann zwei verbundene Stichproben miteinander vergleichen. Liegen mehrere verbundene Stichproben vor, sollte der Friedman-Test benutzt werden. Es ist nicht zulässig, mehrere verbundene Stichproben durch einzelne Wilcoxon-Tests miteinander zu vergleichen, da die Aussagekraft der Teststatistik dadurch verfälscht wird.

Der Friedman-Test basiert auf Rängen und prüft die Nullhypothese, dass es keine signifikante Unterschiede zwischen den einzelnen Stichproben gibt. Bei der hier benutzten zweiseitigen Fragestellung besagt die Alternativhypothese, dass mindestens zwei der Stichproben unterschiedlichen Grundgesamtheiten entstammen. Das Signifikanzniveau ist 5% ( $\alpha=0,05$ ). Wenn die Nullhypothese abgelehnt wird, wissen wir zwar, dass es signifikante Unterschiede zwischen den Stichproben gibt, aber wir wissen nicht, um welche Unterschiede es sich handelt.

Da in der vorliegenden Arbeit die Varianz von Messwertpaaren als Maß für die Reproduzierbarkeit der Meßmethoden dient, wurden diese Varianzen auf signifikante Unterschiede geprüft. Wie im Methodenteil bereits erklärt wurde, gibt es für jede Drüse zwei Messwerte mit der jeweiligen Meßmethode (1mm-, 2mm-, 4mm- oder 8mm-Messung). Für jedes Wertepaar wird dann die Varianz berechnet. So ergibt sich für jede Meßmethode eine Stichprobe aus 22 Varianzen. Diese Stichproben werden dann mit dem Friedman-Test verglichen.

Wenn wir  $k$  Stichproben (in unserem Falle  $k=4$ ) und  $n$  Fälle (hier  $n=22$ ) haben, so wird für jeden Fall  $n$  den  $k$  Messwerten ein Rang zugeordnet (hier 1 bis 4). So läßt sich für jede der vier Meßmethoden aus den 22 Fällen eine Rangsumme bilden. Der Friedman-Test prüft nun, ob es signifikante Unterschiede zwischen diesen Rangsummen gibt. Die Prüfgröße ist  $\chi^2$ -verteilt mit  $k-1$  Freiheitsgraden. Eine ausführliche Erklärung der Teststatistik findet sich in [65].

## Anhang 2

### *Multipler Vergleich nach Wilcoxon und Wilcox*

Mit Hilfe des in Anhang 1 beschriebenen Friedman-Tests kann man prüfen, ob mehrere Stichproben aus identischen Grundgesamtheiten stammen. Sobald festgestellt wurde, daß es signifikante Unterschiede zwischen einzelnen Stichproben gibt, kann man untersuchen, welche Stichproben sich voneinander unterscheiden. Zu diesem Zweck wurde der multiple Vergleich nach Wilcoxon und Wilcox verwendet. Auch hier handelt es sich um eine nichtparametrische Methode, die auf Rangzahlen basiert.

Verglichen werden wieder  $k$  (hier  $k=4$ ) Stichproben (=Meßmethoden) mit  $n$  (hier  $n=22$ ) Fällen. Jeder Methode wird eine Rangzahl von 1 bis  $k$  zugeordnet, so daß 22 Rangordnungen resultieren. Für jede Methode werden die Ränge der einzelnen Meßwerte addiert. Dann werden die Differenzen aus den einzelnen Rangsummen gebildet. Diese Differenzen werden mit den tabellierten kritischen Differenzen aus der Originalarbeit von Wilcoxon und Wilcox verglichen. Bei der hier benutzten zweiseitigen Fragestellung ergibt sich für  $k=4$  und  $n=22$  die kritische Differenz 22 ( $\alpha=0,05$ ) [65]. Wenn die gemessenen Differenzen der Rangsummen diesen Wert überschreiten, muß die Nullhypothese (die zwei entsprechenden Stichproben entstammen identischen Grundgesamtheiten) abgelehnt werden.

Der multiple Vergleich nach Wilcoxon und Wilcox wurde so konzipiert, daß die Effekte, die sonst beim multiplen Testen auftreten würden, korrigiert werden.

## Anhang 3

### *Pearson-Korrelationskoeffizient*

In vielen Fällen ist es wichtig zu wissen, ob zwei Merkmale eines Objektes voneinander abhängig sind. Um einen eventuell bestehenden Zusammenhang anschaulich darzustellen, kann man die Werte des einen Merkmals (Variable  $y$ )

gegen die Werte des anderen Merkmals (Variable x) in einem Diagramm auftragen. Solche Diagramme können oft sehr nützlich sein, haben aber prinzipiell nur orientierenden Charakter. Es gibt aber mathematische Verfahren, die genauere Aussagen über die Abhängigkeit zweier Variablen liefern können.

Um die Stärke des Zusammenhangs zwischen zwei Variablen zu bestimmen, kann man den Korrelationskoeffizienten berechnen. Hier wurde der sog. Korrelationskoeffizient nach Pearson benutzt ( $r$ ) [84]. Er kann Werte zwischen  $-1$  und  $1$  annehmen, wobei der Absolutwert die Stärke der Korrelation und das Vorzeichen ihre Richtung angibt. Wenn der errechnete Korrelationskoeffizient ein positives Vorzeichen hat, sprechen wir von einer positiven Korrelation. In diesem Falle steigt der Wert der Variablen  $y$  an, wenn der Wert der Variablen  $x$  auch ansteigt. Wenn  $x$  ansteigt und  $y$  gleichzeitig kleiner wird, handelt es sich um eine negative Korrelation. So ergibt sich ein Korrelationskoeffizient mit einem negativen Vorzeichen. Ein Wert von  $1$  bzw.  $-1$  bedeutet maximale positive bzw. negative Korrelation. Wenn  $r$  den Wert Null hat, dann besteht überhaupt kein linearer Zusammenhang [85].

$0$  und  $1$  sind die Extrema für den Absolutwert des Korrelationskoeffizienten. Normalerweise liegt der Absolutwert von  $r$  irgendwo dazwischen. Aus den experimentell ermittelten Werten für  $x$  und  $y$  läßt sich der Korrelationskoeffizient mit Hilfe des Programms SPSS berechnen. Der ebenfalls automatisch erfolgende Vergleich mit einem tabellierten „kritischen Wert“ zeigt, ob es sich um einen statistisch signifikanten Zusammenhang handelt. Hier wurde eine zweiseitige Fragestellung benutzt ( $\alpha=0,05$ ). Der Pearson-Korrelationskoeffizient setzt angenähert normalverteilte Grundgesamtheiten und metrisch skalierte Variablen voraus.

