Statistische Bewertung und Analyse der Klausurergebnisse Statistik (Grundstudium)

Diplomarbeit

zur Erlangung des Grades
einer Diplom-Kauffrau

an der Wirtschaftswissenschaftlichen Fakultät
der Humboldt Universität zu Berlin

C.A.S.E.
CENTER FOR APPLIED STATISTICS AND ECONOMICS

eingereicht von

Ulrike Brandes

Matrikel-Nr. 139321

Betreuer : Dr. S. Klinke
Prüfer : Prof. Dr. Rönz

09. Januar 2004

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

Das Fach Statistik gehört zur obligatorischen Grundausbildung der Studenten an der Fakultät für Wirtschaftswissenschaften, da dieses Gebiet von großer Wichtigkeit für das weitere Verständnis in der Wirtschaftswelt ist. Erfahrungsgemäß stellt dieser Bereich die Studenten vor große Probleme. Ziel dieser Arbeit soll es daher zum einem sein, einen Überblick über die Leistungen der Studenten im Fach Statistik im Grundstudium zu erhalten.
Aufgrund der aktuellen finanziellen Lage stehen die Universitäten des Landes im starken Wettbewerb miteinander. Aus diesem Grunde ist es wichtig für die Fakultäten, die Stellung in der Universitätswelt zu kennen. Mit Hilfe dieser Arbeit soll es zum anderen möglich sein, zum Beispiel mit Hilfe der Durchfallquoten, sich mit anderen Universitäten im Bereich Statistik (Grundstudium) zu vergleichen.
Im ersten Kapitel wird daher ein deskriptiver und induktiver Überblick über die erfassten Klausuren im Grundstudium, mit einer anschließenden Aufstellung der möglichen Einflussfaktoren auf die Klausurergebnisse, gegeben. Im nächsten Kapitel werden die dazu benötigten statistischen Grundlagen erläutert. Das dritte Kapitel widmet sich der Analyse und Auswertung der Klausurergebnisse. Am Ende soll eine kleine Zusammenfassung die Ergebnisse des dritten Kapitels Revue passieren. Dann sollen mögliche Schlussfolgerungen für künftige Statistikveranstaltungen gezogen werden. Im Anhang der Arbeit befinden sich alle für das dritte Kapitel benötigten Tabellen, die mit Hilfe von SPSS 11.0 erstellt wurden.

2 Allgemeiner Überblick

2.1 Erhebung der Daten

Für die vorliegende Arbeit wurden 15 Klausuren mit insgesamt 1.747 Studenten erfasst. Die Zeitspanne reicht vom Wintersemester 2000/2001 bis zum Wintersemester 2002/2003. Die Klausuren können aus datenschutzrechtlichen Gründen nicht im Anhang dieser Arbeit aufgeführt werden, es ist aber möglich, sie am Institut für Statistik und Ökonometrie der Wirtschaftswissenschaftlichen Fakultät an der Humboldt-Universität einzusehen.
Unter den 15 Klausuren befinden sich drei Klausuren, die sich auf die alte Prüfungsordnung (PO) und je sechs Klausuren (Statistik I und der Statistik II), die sich auf die neue Prüfungsordnung beziehen. Der Unterschied zwischen der alten und neuen Prüfungsordnung besteht darin, dass nach alter Prüfungsordnung die Statistik I und Statistik II in einer Blockklausur über vier Stunden geschrieben wurden. Nach neuer Prüfungsordnung wird das Fach Statistik in zwei voneinander unabhängigen Prüfungen abgelegt.
In den Klausuren nach alter Prüfungsordnung waren maximal 100 Punkte zu erreichen. In den Teilklausuren nach neuer Prüfungsordnung wurden maximal je 50 Punkte vergeben. In der Klausur vom 09. Oktober 2002 konnten die Studenten ausnahmsweise 51 Punkte erlangen. Damit diese Klausuren mit den unterschiedlichen Gesamtpunkten miteinander verglichen werden können, werden die Gesamtpunkte auf 100 normiert.
Damit können die alte und die neue Prüfungsordnung miteinander verglichen werden. Bei der Erfassung der Daten war zu beachten, dass die Klausuren, um Betrugsversuche auszuschließen, in zwei Versionen geschrieben wurden. Dies wirkt sich jedoch nicht auf die Untersuchungen aus, da nur die Zahlen und nicht die Aufgabenstellungen des jeweiligen Themas geändert wurden.

2.2 Erhobene Daten

Für jede Klausur wurden die Matrikelnummer des Studenten, seine Klausurnote, seine Gesamtpunkte und die Punktezahl der jeweiligen Aufgabe erfasst. Dabei ist bei der Vergabe der Punkte zu beachten, dass der Student entweder die volle Punktezahl oder null Punkte für eine Aufgabe erhält. Ab dem Sommersemester 2002 wurde jedoch die Punktevergabe verändert, so dass der Student statt der vollen Punktezahl auch Teilpunkte erhalten konnte. Zusätzlich wurden im Sommersemester 2002 noch das Geschlecht, das Studienfach, die besuchte Übung und die Anzahl der Versuche, die Klausur zu bestehen, von den Studenten erfragt. Für das Wintersemester 2002/2003 wurden ebenfalls das Geschlecht und die Benutzung von MM-Stat von den teilnehmenden Studenten an den Klausuren erfasst. Zum Schluss wurde noch jede Aufgabenstellung der einzelnen Klausuren aufgenommen und in ihr Themengebiet, gemäß dem Aufbau der Statistik I - und der Statistik II – Vorlesung, eingegliedert. Die Veranstaltungen wurden im Wintersemester 2000/2001, im Sommersemester 2001 und im Wintersemester 2001/2002 von Prof. Dr. Rönz durchgeführt, im Sommersemester 2002 und im Wintersemester 2002/2003 dagegen von Dr. Klinke.

2.3 Deskriptiver Überblick über die Klausuren

In diesem Abschnitt soll ein kleiner deskriptiver Überblick über die einzelnen Klausuren gegeben werden. Dabei ist vielleicht interessant zu sehen, ob sich die Klausurnoten über die Zeit geändert haben. Es werden hier und im weiteren die normierten Gesamtpunkte verwendet, da dieses Merkmal mehr Ausprägungen hat als das Merkmal Klausurnote und somit feinere Nuancen erkannt werden können. Aus den normierten Gesamtpunkten können dann Rückschlüsse auf die Klausurnote gezogen werden (Tabelle A-1).
Tabelle A-2 zeigt, dass es beobachtbare Mittelwertunterschiede in den normierten Gesamtpunkten nach Klausurtyp gibt.

2.3.1 Die Klausur Statistik I

Die herangezogenen sechs Klausuren erzielen einen Median von 50 Punkten (Tabelle A-2). In der Abbildung 2-1 ist dabei folgendes zu erkennen:

1. Die Klausur vom 22. Februar 2001 hat mit 96 Punkten einen sehr hohen Median.
2. Die Streuung ist bei der Klausur vom 26. Juli 2002 am größten.
3. Die Klausur vom 09. Oktober ist im Mittel am schlechtesten ausgefallen. (Diese Klausur war sogar so schlecht, dass die Notenvergabe verändert werden musste, damit die Durchfallquote nicht über 50 % lag.)
4. An der Klausur vom 22. Februar 2001 haben nur 11 Studenten teilgenommen, so dass diese aus der Betrachtung herausgenommen wird.

Es ist eine abnehmende Tendenz, mit Ausnahme der Klausur vom 30. Juli 2002, in den Klausurergebnissen ersichtlich.

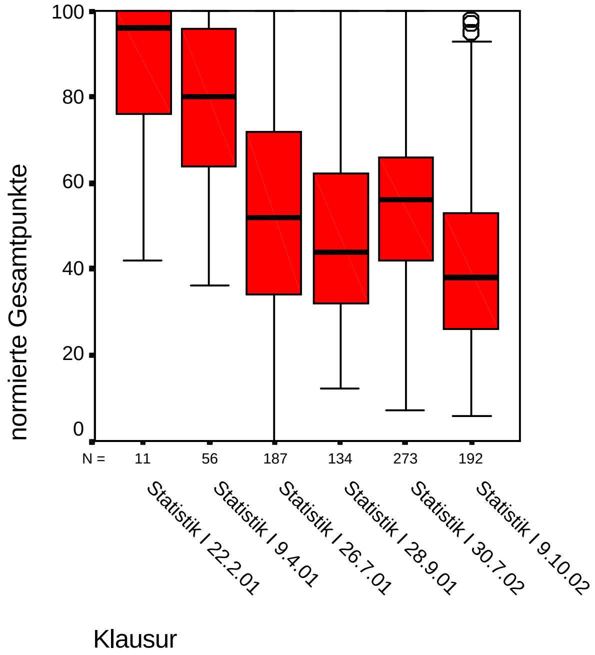


Abbildung 2-1 Boxplot der normierten Gesamtpunkte nach Klausur Statistik I

2.3.2 Die Klausur Statistik II

Die Statistik II – Klausuren erzielen insgesamt einen Median von 60 Punkten (Tabelle A-2). Damit sind die Statistik II - Klausuren im Durchschnitt besser ausgefallen als die Statistik I - Klausuren. Dies wird im Kapitel 2.4 noch mit Hilfe eines statistischen Tests nachgewiesen.

Der Abbildung 2-2 sind folgende Aussagen zu entnehmen :

1. Die Klausur vom 08. April 2002 ist mit einem Median von 74 Punkten (Tabelle A-2) besonders gut ausgefallen.
2. Die Streuung ist bei der Klausur vom 21. Februar 2002 am größten.
3. Der Median der Klausuren hat sich über die Zeit nicht stark verändert.
4. An der Klausur vom 22. Februar 2001 haben nur 9 Studenten teilgenommen, so dass diese ebenfalls aus der Betrachtung h erausgenommen wird.

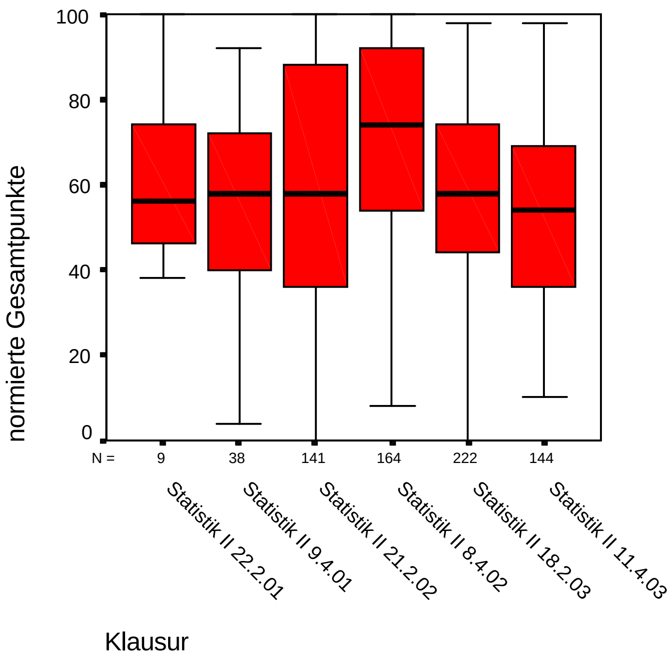


Abbildung 2-2 Boxplot der normierten Gesamtpunkte nach Klausur Statistik II

2.3.3 Statistik nach alter Prüfungsordnung

In den drei Klausuren, die nach alter Prüfungsordnung geschrieben wurden, ist ein abnehmender Trend in den Gesamtpunkten zu erkennen (Abbildung 2-3). Dabei ist zu beachten, dass die Klausur vom 08. April 2002 nur von 13 Studenten geschrieben wurde.

Der Grund für diese Tendenz ist, dass ein Wechsel von der alten zur neuen Prüfungsordnung vollzogen wurde, auch erkennbar an der abnehmenden Teilnehmerzahl. Entweder haben die Studenten schon das Fach Statistik in vorherigen Semestern abgeschlossen oder sind zur neuen Prüfungsordnung gewechselt. Somit sind nach alter Prüfungsordnung nur noch Studenten „übrig geblieben“, die möglicherweise die Klausur mehrfach wiederholen mussten oder die Ablegung des Fachs herauszögerten.

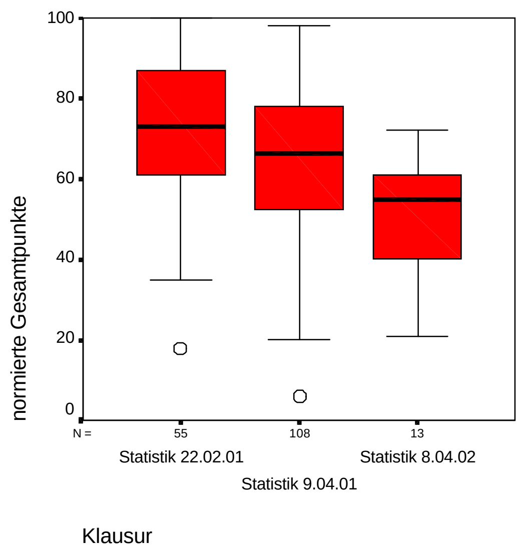


Abbildung 2-3 Boxplot der normierten Gesamtpunkte nach Klausur alte PO

2.4 Induktiver Überblick über die Klausuren

In diesem Abschnitt sollen die deskriptiven Beobachtungen des Kapitels 2.2. induktiv vertieft werden. Die hier verwendeten Tests werden im Kapitel 3 näher erläutert. Es wird ein Signifikanzniveau von $\alpha = 0{,}05$ angenommen.

Das Fehlerbalkendiagramm (Abbildung 2-4) zeigt, dass es signifikante Unterschiede in den Mittelwerten der Gesamtpunkte gibt. Damit kann die in Kapitel 2.3.2 gemachte These, dass die Statistik II – Klausur im Mittel bessere Ergebnisse erzielt hat, aufrechterhalten werden.

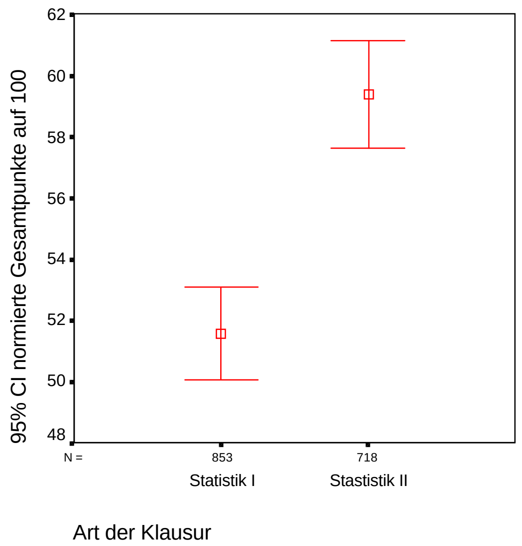


Abbildung 2-4 Fehlerbalkendiagramm der normierten Gesamtpunkte nach Klausurtyp

Da aufgrund der Ablehnung der Normalverteilung (Tabelle A-3) kein parametrischer Test durchgeführt werden kann, wird der Mann-Whitney-Test verwendet. Ein Signifikanzniveauvergleich (Tabelle A-4) beweist statistisch, dass die Homogenitätshypothese abgelehnt wird. Damit ist die oben gestellte These bekräftigt.

Als nächstes wird gezeigt, dass die Unterschiede der beobachteten Mittelwerte der Gesamtpunkte innerhalb der Klausurtypen (Tabelle A-2) signifikant sind. Denn es soll die These aufgestellt werden, dass es Einflüsse gibt die sich positiv auf das Ergebnis einer Klausur (wie z.B. die Statistik II - Klausur vom 08 April 2002) auswirken. Dazu muss jeder Klausurtyp einzeln untersucht werden.

Nach Ablehnung der Normalverteilung (Tabelle A-3) wird der Kruskal-Wallis-Test durchgeführt. Dieser zeigt, dass für jede Klausurart (Tabelle A-5 bis Tabelle A-7) die Homogenitätshypothese abgelehnt wird. Somit sind die Unterschiede der beobachteten Mittelwerte signifikant.

2.5 Einflüsse auf die Klausur

In den Kapiteln 2.3. und 2.4. wurden die vorgestellten Thesen bestätigt. Folglich müssen Faktoren existieren, die die Klausuren in einer bestimmten Richtung beeinflussen. Solche Einflüsse können zum Beispiel sein :

1. das Geschlecht des Studenten,
2. die Studienrichtung des Studenten,
3. die besuchte Übung des Studenten,
4. die Anzahl der Versuche zum bestehen der Klausur,
5. das benutzte Lehrmaterial zur Vorbereitung auf die Klausur,
6. die Vorbereitungszeit auf die Klausur,
7. die Aufgaben der jeweiligen Klausur,
8. die Mathematikleistungen des Studenten,

usw..

In dieser Arbeit kann nicht auf alle möglichen Faktoren eingegangen werden. Es werden nur Einflüsse untersucht, die mit Hilfe der Klausur erfasst wurden. Mit den Ergebnissen können dann auch die „Ausreißer“ der Statistik I und Statistik II Klausur erklärt werden.

3 Statistische Grundlagen

Bevor die Klausurergebnisse näher betrachtet werden soll jedoch erst ein kurzer Einblick in die benutzten statistischen Methoden der Vollständigkeit halber zum besseren Verständnis gegebenen werden.

3.1 Deskriptive Statistik

Die deskriptive Statistik wird angewendet, um eine tabellarische und grafische Einsicht in das Datenmaterial zu gewinnen. Bei großen Datenmengen bietet es sich daher an, diese durch Lageparameter, Streuungsmaße und entsprechende Diagramme näher zu charakterisieren.

3.1.1 Lagemaße

Lagemaße dienen dazu, das Zentrum einer Häufigkeitsverteilung näher zu beschreiben. Ein Lagemaß für metrisch skalierte Merkmale ist das arithmetische Mittel $\bar{x}$, wobei die Summe aller Beobachtungen durch die Anzahl der Beobachtungen dividiert wird, also:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \,,$$

wobei $x_1,\ldots, x_n$ die beobachteten Merkmalsausprägungen sind mit $i = 1,\ldots,n$. Der Mittelwert ist gegenüber Ausreißern (Ausreißer sind extrem niedrige oder hohe Werte innerhalb einer Reihe von Beobachtungen, die sich nicht sehr stark unterscheiden) sehr empfindlich, da auch diese in der Berechnung zu 100 % Berücksichtigung finden.

Der Median oder auch Zentralwert ist derjenige Wert $\tilde{x}$, für den gilt, dass 50 % der Beobachtungen größer bzw. kleiner sind als er. Er kann im Gegensatz zum arithmetischen Mittel auch für ordinal skalierte Merkmale angewendet werden, da die Beobachtungen ihrer Größe nach geordnet werden. Der Median ergibt sich also als:

$$\tilde{x} = \begin{cases} x_{\frac{(n+1)}{2}} & \text{, für gerade n} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{(n+2)}{2}}\right) & \text{, für ungerade n.} \end{cases}$$

Der Vorteil des Zentralwerts im Gegensatz zum Mittelwert liegt darin, dass dieser robust gegenüber Ausreißern ist.

Liegen die beobachteten Merkmale in einer geordneten Beobachtungsreihe vor, so wird das α-Quartil $\tilde{x}_\alpha$ definiert als:

$$\tilde{x}_\alpha = \begin{cases} x_{(k)} & \text{, wenn } n \cdot \alpha \notin \mathbb{Z} \text{ ist (k ist dann die auf } n \cdot \alpha \text{ folgende ganze Zahl)} \\ \frac{1}{2}\left(x_{(k)} + x_{(k+1)}\right) & \text{, wenn } n \cdot \alpha \in \mathbb{Z} \text{ ist} \end{cases}$$

Damit liegen n·a Beobachtungswerte unter und n·(1-α) Werte über dem α-Quartil $\tilde{x}_\alpha$. Der Median ist ein Spezialfall des α-Quartil für α = 0,5. Weitere Spezialfälle sind das 0,25-Quartil und 0,75-Quartil, die als unteres und oberes Quartil bezeichnet werden.

In dieser Arbeit soll der Median für die Klausurnote, da diese ordinal skaliert sind, und das arithmetische Mittel für die normierten Gesamtpunkte, da diese mit den vielen Merkmalsausprägungen von 0 bis 100 als metrisch angesehen werden können, verwendet werden. Das obere Quartil und das untere Quartil werden für die Bildung des Boxplotes herangezogen.

3.1.2 Streuungsmaße

Das Lagemaß allein ist nicht ausreichend aussagefähig über die Lage der Häufigkeitsverteilung. Daher werden Streuungsmaße bzw. Dispersionsmaße verwendet. Damit kann angeben werden, wie weit ein konkreter Merkmalswert von dem Zentrum entfernt ist. Dadurch können genauere Angaben zur Häufigkeitsverteilung gemacht werden.

Es können auch Größen als Dispersionmaße verwendet werden, die eine durchschnittliche quadratische Abweichung MQ der Merkmalswerte von einem Beobachtungswert c angeben. Grundidee für dieses Maß ist das Prinzip der Summe aus gewogenen Einzelwerten, berücksichtigt werden dadurch im Gegensatz zur Spannweite und Quartilsabstand alle Werte, also :

$$MQ(c) = \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2 .$$

Ist dieser Bezugspunkt c das arithmetische Mittel, so bezeichnet man diese mittlere quadratische Abweichung als Varianz s^2. Wird daraus die Wurzel gezogen erhält man die Standardabweichung s. Da die Varianz im Quadrat der ursprünglichen Einheit vorliegt und daraus schwer Rückschlüsse gemacht werden können, wird die Standardabweichung verwendet, die dann in der gleichen Einheit wie die Beobachtungen vorliegt.

Nur die Interpretation der Standardabweichung für den Grenzfall s = 0 ist offenbar. Die Standardabweichung nimmt den Wert null an, wenn keine Abstände zwischen den Merkmalsausprägungen existieren, wenn alle Merkmalsausprägungen also gleich sind. Ein kleiner Wert der Standardabweichung deutet auf eine kleine Streuung der Merkmalsausprägungen hin und ein großer Wert auf eine große Streuung der Merkmalsausprägungen. Die Größe der Standardabweichung hängt dabei nicht von der absoluten Größe von n ab, sondern nur von den relativen Häufigkeiten.

3.1.3 Graphische Darstellung

Das Balken- oder Säulendiagramm liefert einen visuellen Eindruck über die Häufigkeit für alle nominal, ordinal und metrisch skalierten Merkmale mit weniger als zehn Merkmalsausprägungen. Das Liniendiagramm eignet sich besonders gut für die Betrachtung von Merkmalen über die Zeit. Der Box-and-Whisker-Plot und das Fehlerbalkendiagramm zeigen die Lage- und Streuungsmaße metrisch skalierter Merkmale graphisch auf. Dabei nutzt der Boxplot Kenngrößen wie den größten Beobachtungswert $x_{(max)}$ und kleinsten Beobachtungswert $x_{(min)}$, sowie untere Quartil $x_{0,25}$ und obere Quartil $x_{0,75}$ und den Median $x_{0,5}$ zur Veranschaulichung. Die Box wird von dem 0,25- und 0,75-Quartil eingegrenzt, die damit auch den Interquartilsabstand angibt. In ihr befindet sich der Median, der je nach Lage Auskunft über die Schiefe der Verteilung gibt. Die Whisker (senkrechte Linien von der unteren und oberen Boxgrenze) haben eine Länge von 1,5 des Interquartilabstandes jeweils vom Ende der Box gemessen und enden in einem beobachteten Punkt. Zwischen dem 3- und 1,5-fachen Interquartilsabstand werden die Ausreißer mit einem Punkt markiert. Durch einen Stern sind Extremwerte markiert, die sich mehr als den dreifachen Interquartilsabstand von der Box entfernen.

Das Fehlerbalkendiagramm hilft grafisch Unterschiede zwischen den Mittelwerten zweier oder mehrere Stichproben aufzudecken. Als Fehlerbalken wird häufig das 95 %- Konfidenzintervall für den Mittelwert benutzt. Ein Konfidenzintervall ist ein Bereich, im dem sich ein alle möglichen Lageparameter mit einer bestimmten Wahrscheinlichkeit (hier 95 %) befinden können. Sobald sich die Konfidenzintervalle der Stichproben nicht überlappen, besteht ein wesentlicher Unterschiede zwischen den Mittelwerten.

3.2 Korrelationsanalyse

Wenn die lineare Abhängigkeit zwischen zwei Merkmalen X und Y gemessen werden soll, wird auf die Korrelationsanalyse zurückgegriffen. Mit ihr können die Art und der Grad des Zusammenhanges bestimmt werden. Für die grafische Darstellung, den so genannten Scatterplot, werden die Merkmale in ein Koordinatensystem eingetragen und anhand der so entstehenden Punktwolke Rückschlüsse auf den Zusammenhang gezogen. Die Maßzahl für die Korrelation ist der Korrelationskoeffizient r mit $-1 \leq r \leq 1$. Wenn $r = \pm 1$ besteht ein funktion aler Zusammenhang, d.h. alle Punkte liegen auf einer Geraden. Ist $r = 0$ so liegen die Merkmale X und Y unkorreliert vor. Im Folgenden soll nur auf den Zusammenhang ordinal skalierten Merkmale eingegangen werden, da nur dieser Merkmalstyp in der Arbeit behandelt wird.

3.2.1 Spearmansche Rangkorrelationskoeffizient

In solchen Fällen benutzt man zum einen den Rangkorrelationskoeffizienten von Spearman r_S mit

$$r_S = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n \cdot (n^2 - 1)}.$$

Dieser hat den Vorteil, dass er bei nicht normalverteilten Stichproben (Normalverteilung in den Grundgesamtheiten existiert nur sehr selten in der Realität) und bei sehr kleinem Stichprobenumfang angewendet werden kann. Voraussetzung für die Anwendung ist, dass die Merkmale mindestens ordinal skaliert und unabhängig sind. Zur Berechnung werden beide Stichproben der Merkmale X und Y in Rangreihen geordnet, indem man die Ausprägungen des Merkmals X und Y der Größe nach ordnet und durchnummeriert. Tritt die Ausprägung eines Merkmales x_i bzw. y_i mehrmals auf, so werden diesen gleichen Werten mittlere Rangplätze zugeordnet. Dann bildet man die quadrierte Differenz D^2, aus den n Rangpaaren ($r(x_i)$, $r(y_i)$) und summiert diese. Wenn die zwei Rangreihen gleich sind ergibt sich ein Korrelationskoeffizient r_s von eins, da die Differenzen null sind. Verhalten sich die Rangreihen invers zueinander, so ergibt sich ein r_s von minus eins. Damit kann die Frage beantwortet werden, ob ein positiver oder negativer Zusammenhang zwischen den Merkmalen X und Y besteht.

3.2.2 Kendallsche Rangkorrelationskoeffizient

Hier werden wieder den Ausprägungen der Merkmale X und Y Rangzahlen $r(x_i)$ bzw. $r(y_i)$ zugeordnet. Dann werden die geordneten Paare ($r(x_i)$, $r(y_i)$) gebildet und bezüglich $r(x_i)$ der Größe nach geordnet. Man erhält dann eine Folge von Paaren ($r(x_{i1})$, $r(y_{i1})$),…, ($r(x_{in})$, $r(y_{in})$), mit $r(x_{ik}) \leq r(x_{ik+1})$, die anschließend in einer 2 x n Matrix M der Form

$$M := \begin{pmatrix} r(x_{i1}) \ldots r(x_{in}) \\ r(y_{i1}) \ldots r(y_{in}) \end{pmatrix}$$

angeordnet werden. Anschließend zählt man für jedes Element $r(y_i)$ die Anzahl der Fehlstellungen, also die Anzahl aller Zahlen $r(y_j)$, die sich rechts von $r(y_i)$ in der Matrix M befinden und kleiner oder gleich $r(y_i)$ sind. Diese Zahl wird Q_i genannt. Damit ergibt sich folgende Formel für den Kendallschen Rangkorrelationskoeffizienten :

$$r_K = 1 - \frac{4\sum_{i=1}^{n} Q_i}{n(n-1)}.$$

Er hat den Vorteil, dass Bindungen (mehrere gleiche Elemente) berücksichtigt werden. Daher wird in dieser Arbeit auch der Kendallsche Rangkoeffizient benutzt, da davon ausgegangen werden muss, dass verstärkt Bindungen (gerade bei den normierten Gesamtpunkten) auftreten.

3.2.3 Cohens Kappa Koeffizient

Der Cohens Kappa Koeffizient misst den Grad der Übereinstimmung von Beurteilungen eines Gegenstandes durch zwei verschiedene Personen. Beide Personen beurteilen den Bebachtungswert des Merkmals X nach den gleichen Kriterien $Y_1,\ldots Y_l$, wobei l die Anzahl der Kriterien ist. In einer Kontingenztabelle steht in der Zelle jk die Kombination bezüglich des Kriteriums Y_j und des Kriteriums Y_k. Damit stehen in der Hauptdiagonalen die Anzahl der Übereinstimmungen bezüglich des Kriteriums Y_j.

Der Cohens Kappa Koeffizient ist folgendermaßen :

$$\kappa = \frac{n\sum_{j=1}^{J} h_{jj} - \sum_{j=1}^{J} h_{j+}h_{+j}}{n^2 - \sum_{j=1}^{J} h_{j+}h_{+j}}.$$

Er setzt sich zusammen aus der Summe der relativen Häufigkeit der übereinstimmenden Beurteilungen (sie stehen in der Diagonale der Kontingenztabelle), die um die übereinstimmenden Beurteilungen bereinigt wird, die auftreten können, wenn die Person das Objekt zufällig bewertet (sie sind die Summe des Produkts der Randhäufigkeiten). Diese Differenz ist der tatsächliche Anteil der über den Zufall hinausgehenden übereinstimmenden Beurteilungen. Zum Zwecke der Normierung wird sie durch den theoretischen Anteil dividiert, der über den Zufall hinausgehenden übereinstimmenden Bewertungen geht. Der Wertebereich des Cohens Kappa Koeffizienten liegt zwischen -1 und +1.

3.3 Vergleich zweier unabhängiger Stichprobenmittelwerte

In dieser Arbeit werden sehr oft die Mittelwerte zweier Grundgesamtheiten miteinander verglichen, um auf signifikante Unterschiede in den Grundgesamtheiten bezüglich eines Merkmales X schließen zu können. Je nachdem welche Voraussetzungen gegeben sind können der t-Test oder der Mann-Whitney-Test verwendet werden. Dabei gehört der t-Test zu den parametrischen Tests, die an die Verteilungen der Grundgesamtheiten bestimmte Voraussetzungen stellen. Nichtparametrische Tests, wie der Mann-Whithney-Test, stellen keine Voraussetzungen an die Verteilung der Grundgesamtheit.

Allgemeine Voraussetzungen für den Mittelwertvergleich zweier Grundgesamtheiten sind :

1. Die aus den Grundgesamtheiten gezogenen Zufallsstichproben $X_{1,1},\dots,X_{1,n_1}$ und $X_{2,1},\dots,X_{2,n_2}$ sind unabhängig voneinander.
2. Die Stichprobenumfänge sind n_1 und n_2.
3. Die Merkmal X muss mindestens ordinal skaliert sein.

In der Arbeit wird die Nullhypothese, dass die Mittelwerte der normierten Gesamtpunkte bezüglich eines Merkmals (z.B. Geschlecht) gleich sind, gegen die Hypothese, dass die Mittelwerte verschieden sind, getestet.

3.3.1 Der t-Test

Weitere Voraussetzungen für die Anwendung des t-Tests sind :

1. Die Zufallsvariablen in den Grundgesamtheiten folgen einer Normalverteilung.
2. Die Varianzen der Grundgesamtheiten σ_1^2 und σ_2^2 sind unbekannt aber gleich.

Dann kann mit dem Zweistichproben t-Test getestet werden, ob die Mittelwerte der beiden Stichproben gleich sind. Es wird getestet, ob

$$H_0 : \mu_1 = \mu_2 \text{ oder } H_1 : \mu_1 \neq \mu_2$$

zutrifft. Aufgrund der Voraussetzung der Normalverteilung der Grundgesamtheiten und der Schätzung der unbekannten Varianzen erhält man folgende Teststatistik :

$$\hat{t} = \frac{|\bar{x}_1 - \bar{x}_2|}{\hat{\sigma}_{(\bar{x}_1-\bar{x}_2)}} \quad ,$$

mit $\hat{\sigma}_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\frac{(n_1-1)\cdot\sigma_1^2 + (n_2-1)\cdot\sigma_2^2}{(n_1-1)+(n_2-1)}}\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$.

Die Nullhypothese wird verworfen, wenn $\hat{t}$ größer gleich $t_{n_1+n_2-2,1-\frac{\alpha}{2}}$ ist.

3.3.2 Der Mann-Whitney-Test

Der Mann-Whitney-Test ist das Gegenstück zum t-Test, denn er prüft ebenfalls die Nullhypothese $H_0 : \mu_1 = \mu_2$ gegen die Hypothese $H_0 : \mu_1 \neq \mu_2$, jedoch ohne Voraussetzungen an die Verteilungen der zwei Grundgesamtheiten zu stellen.

Die Beobachtungen der zwei Stichproben werden der Größe nach geordnet und anschließend durchnummeriert. Somit erhält jede Beobachtung eine Rangzahl r_{ij} (es wird sich gemerkt aus welcher Stichprobe die Rangzahl der Beobachtung stammt). Dann werden die Rangsummen der Stichproben mit $R_i = \sum_{i=1}^{m} r_{ij}$, m = 2, gebildet. Jetzt kann die Prüfgröße berechnet werden :

$$U = \min\left\{U_1 = n_1 n_2 + \frac{n_1(n_2+1)}{2} - R_1, U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2\right\}.$$

Diese wird für m, n ≥ 8 nach Mann-Whitney standardisiert :

$$z = \frac{\left|U - \frac{n_1 n_2}{2}\right|}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{2}}}$$

Treten in den Stichprobenwerten Bindungen auf, erhalten diese eine mittlere Rangzahl. Jedoch beeinflussen nur Bindungen die zwischen den Stichproben auftreten den Test, dieser muss dann in der folgenden Weise korrigiert werden :

$$z^{korr} = \frac{\left|U - \frac{n_1 n_2}{2}\right|}{\sqrt{\left(\frac{n_1 n_2}{(n_1+n_2)(n_1+n_2-1)}\right) \cdot \left(\frac{(n_1+n_2)^3 - n_1 + n_2}{12} - \sum_{i=1}^{i=r} \frac{t_i^3 - t_i}{12}\right)}}.$$

Ist $z^{korr} \geq z_\alpha$ kann die Nullhypothese statistisch nicht aufrechterhalten werden.

3.4 Vergleich mehrerer unabhängiger Stichprobenmittelwerte

Im Laufe dieser Arbeit werden nicht nur Mittelwerte miteinander verglichen die aus zwei Grundgesamtheiten stammen, sondern aus mehreren. Voraussetzungen für die Anwendung dieser speziellen Tests sind :

1. Die Variable X ist metrisch skaliert.
2. Es sind m Grundgesamtheiten gegeben.
3. Die aus den Grundgesamtheiten gezogenen m Zufallsstichproben $X_{j,1},\ldots,X_{j,n_j}$, mit j = 1,…, m, sind unabhängig voneinander.

4. Für den Stichprobenumfang gilt : $n = \sum_{j=1}^{m} n_j$.

3.4.1 ANOVA

Die Anova (Analysis of Variance) vergleicht die Erwartungswerte mehrerer Stichproben auf Gleichheit. Dabei bedient sich der Test (wie der Name besagt) der empirischen gemessenen Varianzen. Die Nullhypothese lautet dann :

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_m .$$

Weitere Voraussetzungen für die Anwendung der ANOVA sind :

1. Die m Grundgesamtheiten sind normalverteilt, dies kann mit dem Kolmogorov-Smirnov-Test (siehe Kapitel 3.6) geprüft werden.
2. Die Varianzen der Grundgesamtheiten $\sigma_1^2,\ldots,\sigma_m^2$ sind gleich, jedoch unbekannt, dies kann mit dem Levené-Test (siehe Kapitel 3.6) geprüft werden.

Werden die Ergebnisse in einer Datenmatrix dargestellt, so steht in der Zeile i, Spalte j die Variable x_{ij}, die bei der i-ten Beobachtung in der j-ten Stichprobe gemessen wurde. Aus dieser Datenmatrix läßt sich folgende Beziehung aufstellen :

$$SQT = SQZ + SQI.$$

Wobei $SQT = \sum_{j=1}^{m} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}\right)^2$ mit $\bar{x} = \frac{\sum_{j=1}^{m} \bar{x}_j n_j}{n}$ die Summe der Abstandsquadrate aller Beobachtungen xij vom Gesamt – arithmetischen Mittel,

$SQZ = \sum_{j=1}^{m} n_j \left(\bar{x}_j - \bar{x}\right)^2$ die Summe der Abstandsquadrate des Gruppen – arithmetischen Mittels vom Gesamt - arithmetischen Mittel und

$SQI = \sum_{j=1}^{m} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}_j\right)^2$ die Summe der Abstandsquadrate aller Beobachtungen von ihrem jeweiligen Gruppen –arithmetischen Mittel sind.

Für die Teststatistik werden die Quadratsummen SQZ und SQI in empirischen Varianzen überführt, in dem sie durch eine Zahl dividiert werden, die aus ihnen die mittlerer Abstandsquadrate MQZ und MQI macht. Die Ergebnisse sind erwartungstreue Schätzer für die Varianzen der Grundgesamtheiten, also :

$$MQZ = \frac{SQZ}{m-1} \text{ und } MQI = \frac{SQI}{n-m}.$$

Entstammen alle Zufallsstichproben aus derselben Grundgesamtheit, so sollten die Werte von MQZ und MQI ungefähr gleich groß sein. Somit ergibt sich die Teststatistik mit :

$$F = \frac{SQZ}{SQI} \cdot \frac{(n-m)}{(m-1)}.$$

Ist F größer als der kritische Wert $f_{\alpha, m-1, n-m}$ wird die Nullhypothese abgelehnt.

3.4.2 Der H-Test von Kruskal und Wallis

Mit diesem H-Test kann geprüft werden, ob die Mittelwerte der m Stichproben als signifikant verschieden voneinander angesehen werden können. Es werden also folgende Hypothesen gegeneinander getestet :

$$H_0 : \mu_1 = \mu_2 = ... = \mu_m \text{ gegen } H_1 : \mu_1 \neq \mu_m.$$

Für die Anwendung des H-Test von Kruskal und Wallis müssen folgende Voraussetzungen gegeben sein :

1. Die Verteilungsfunktionen $F_n(x)$ der m Grundgesamtheiten sind unbekannt (es muss keine Normalverteilung vorliegen).
2. Die m Zufallsstichproben $X_{j,1},...,X_{j,n_j}$, mit j = 1,…, m stammen aus stetigen Verteilungsfunktionen $F_j(x)$.

Um die Teststatistik H zu erhalten, werden zuerst alle m Stichproben kombiniert und ihre Werte der Größe nach geordnet. Dabei wird jedem Element x_{ij} eine Rangzahl r_{ij} zugeordnet. Treten Bindungen auf, so wird ihnen der mittlere in Frage stehende Rang zugeordnet. Die Rangsumme R_i ergibt sich als Summe der Ränge der i-ten Stichprobe. Die Teststatistik H ist dann definiert als

$$H = 1 - \frac{1}{n^3 - n} \sum_{i=1}^{g} (t_i^3 - t_i) \left[\frac{12}{n(n+1)} \sum_{i=1}^{m} \frac{1}{n_i} \left(R_i - \frac{n_i(n+1)}{2} \right)^2 \right].$$

Es wird also das gewichtete Abweichungsquadrat zwischen der Rangsumme mit dem Erwartungswert der Rangsumme der i-ten Stichprobe gebildet, korrigiert um das Auftreten von Bindungen, wobei t_i die Anzahl der jeweils gleichen Rangplätze in der Bindung i angibt. Der H-Test folgt approximativ der χ^2-Quadrat Verteilung. Ist H größer gleich dem kritischen Wert $\chi^2_{m-1,\alpha}$, so muss die Nullhypothese abgelehnt werden.

3.5 Wilcoxon-Test

Bisher wurden nur Tests betrachtet, die unabhängige Zufallstichproben voraussetzen. An dieser Stelle soll ein Test erläutert werden, der auf abhängige Zufallsstichproben eingeht. Dieser

Test stellt die Nullhypothese auf, dass zwei verbundene Zufallsstichproben gleichen Umfangs n denselben Mittelwert besitzen. Also, ob

$$H_0: \mu_1 = \mu_2 \text{ oder } H_1: \mu \neq \mu_2$$

zutrifft. Für das Aufstellen der Teststatistik wird zwischen den Beobachtungspaaren (X_i, Y_i) die Differenz d_i gebildet. Treten Paare auf, so dass $d_i = y_i - x_i = 0$, werden diese aus der Stichprobe entfernt. Den so entstehenden Differenzen werden ohne Beachtung der Vorzeichen Rangzahlen $r(d_i)$ zu geordnet. Diesen Rangzahlen $r(d_i)$ werden nun erst die entsprechenden Vorzeichen zugeordnet. R der folgenden Teststatistik ist dann die kleiner Rangsumme $R_{+/-}$ aus den positiven bzw. negativen Rangzahlen mit i = 1,…n der positiven bzw. negativen Wertdifferenzen. Somit ergibt sich die Teststatistik z, die bei $n \geq 25$ approximativ einer Standardnormalverteilung folgt :

$$z = \frac{R - \frac{n \cdot (n+1)}{4}}{\sqrt{\frac{n \cdot (n+1) \cdot (2n+1)}{24}}}.$$

Die Nullhypothese wird abgelehnt, wenn z größer ist als der kritische Wert z_α.

3.6 Exkurs : Kolmogorov-Smirnov-Test und Levené-Test

Für die Anwendung von Tests die Normalverteilung und Varianzenhomogenität voraussetzen (z.B. für den t-Test) wird der Kolmogorov-Smirnov-Test (für die Normalverteilung) und der Levené-Test (für die Varianzhomogenität) herangezogen.
Im Allgemeinen überprüft der Kolmogorov-Smirnov-Test, ob eine theoretische Verteilung mit der empirischen Verteilung übereinstimmt. Die Nullhypothese lautet in diesem Fall :

$$H_0: F_n(x) = F_0(x)$$

In unserem Fall interessiert jedoch nur die Normalverteilung. In diesem Fall wird überprüft, ob die unbekannte Verteilungsfunktion $F_n(x)$ der Grundgesamtheit eine Normalverteilung mit $\Phi = \left(\frac{x-\mu}{\sigma}\right)$ ist.

Dabei müssen folgende Voraussetzungen erfüllt sein :

1. Die Variable X ist metrisch skaliert und nicht klassiert.
2. Die theoretische Verteilung $F_n(x)$ ist stetig.
3. Die Parameter der empirischen Verteilungsfunktion $F_0(x)$ sind bekannt.

Die Teststatistik überprüft die maximale Differenz zwischen der theoretischen und empirischen Verteilungsfunktion. Sie lautet :

$$D_n^{NV} = \max_x \left| F_n(x) - \Phi\left(\frac{x - \bar{x}}{s}\right) \right|,$$

wobei μ durch $\bar{x}$ und σ durch s ersetzt wurde. Dies erfolgt, weil μ und σ unbekannt sind und aus der Stichprobe geschätzt werden müssen. Die Bestimmung des kritischen Wertes $d_{n,1-\alpha}$ erfolgt mit der Korrektur von Liffofors. Ist $D_n^{NV} \geq d_{n,1-\alpha}$ so wird die Nullhypothese abgelehnt.

Der Levené-Test prüft die Hypothesen $H_0 : \sigma_1^2 = ... = \sigma_m^2$ gegen $H_1 : \sigma_j^2 \neq \sigma_k^2, j \neq k$, also ob die Varianzen aus verschiedenen Grundgesamtheiten gleich sind, im speziellen Fall für die Anwendung des t-Testes, ob $H_0 : \sigma_1^2 = \sigma_2^2$ ist.

Dazu müssen jedoch folgende Voraussetzungen gegeben sein :

1. Es liegen m Grundgesamtheiten vor mit $m \geq 2$.
2. Die unabhängigen Zufallsvariablen $X_{j,1},\ldots,X_{j,n_j}$, mit j = 1,…, m, in den Grundgesamtheiten folgen einer stetigen Verteilung.
3. Die aus jeder Grundgesamtheit gezogenene Stichproben sind unabhängig voneinander.
4. Für den Stichprobenumfang gilt : $n = \sum_{j=1}^{m} n_j$.

Die Teststatistik lautet,

$$L = \frac{n-m}{m-1} \cdot \frac{\sum_{j=1}^{m} n_j \left(\bar{Y}_j - \bar{Y}\right)^2}{\sum_{j=1}^{m} \sum_{i=1}^{n_j} \left(Y_{ij} - \bar{Y}\right)^2},$$

wobei Y_{ij} die absolute Abweichung der Stichprobenvariablen X_{ij} vom Mittelwert der i-ten Stichprobe, $\bar{Y}_j$ der Mittelwert von Y_{ij} der i-ten Stichprobe, $\bar{Y}$ der Mittelwert aller Stichproben ist. Somit wird die Gesamtheit der Abweichung zwischen den Stichproben dividiert durch die Gesamtheit der Abweichung innerhalb der Stichprobe. Damit wird geprüft, ob die Stichproben aus den m Grundgesamtheiten mit gleichen mittleren Abweichungen stammen. Ist dies der Fall, so wird die Nullhypothese angenommen. Die Nullhypothese wird dagegen verworfen, wenn L den kritischen Wert von $F_{m-1,\, n-m,\, 1-\alpha}$ überscheitet.

3.7 Die hierarchische Clusteranalyse

Die Clusteranalyse wird benutzt, um Objekte derart in Gruppen, sog. Cluster, zu unterteilen, dass in einer Gruppe Objekte sind, die sich sehr ähneln und Objekte unterschiedlicher Gruppen möglichst verschieden voneinander sind. Die Ähnlichkeit der betrachteten Objekte wird

dabei anhand von mehrerer Merkmale gemessen. Voraussetzung für die Anwendung ist, dass die Merkmale mindestens ordinal skaliert sind.
Mittels eines sog. Distanzmaßes werden die Objekte bezügliche ihrer Unähnlichkeit auf bestimmte Merkmale hin in Gruppen unterteilt. Dabei soll an dieser Stelle auf das Quadrierte Euklidische Distanzmaß D^2 eingegangen werden, dass auch in der Arbeit verwendet wird. Es errechnet die Summe der quadrierten Differenzen zwischen den Variablen der beiden betrachteten Objekte Y und X, also :

$$D^2 = \sum_{i=1}^{v} (Y_i - X_i)^2 .$$

Wobei v die Anzahl der berücksichtigten Variablen zur Bewertung der Ähnlichkeit ist. Um nun die Clusteranalyse für alle Objekte durchzuführen, muss der Distanzwert für alle Paare berechnet werden (diese Werte bilden die Grundlage der Clusteranalyse). Diese Werte werden mit Hilfe einer Distanzmatrix grafisch dargestellt.
Bei der hierarchischen Clusteranalyse wird zuerst jedem Objekt ein Cluster zugeordnet. Die beiden Cluster zwischen denen die geringste Distanz besteht, werden dann zu einem Cluster zusammengefasst. Damit verringert sich die Anzahl der Cluster um eins. Nun werden für die Cluster erneut die Distanzwerte berechnet und wieder zwei Cluster zusammengefasst, die die geringste Distanz aufweisen. Dieses Verfahren wird solange fortgesetzt bis alle Objekte in einem Cluster zusammengefasst sind. Dieses Ergebnis ist natürlich ein Extrem, welches nicht gewünscht ist. Das tatsächliche Ergebnis liegt in den Zwischenschritten der Zusammenfassung, welches von Anwender an dem gewünschten Zwischenschritt ausgewählt werden kann. Diese Stufen werden mit einer sog. Agglomerationstabelle dargestellt.
Ab der zweiten Stufe der Zusammenfassung der Objekte zu Clustern muss der Distanzwert zwischen Clustern ermittelt werden. Da sich in Clustern mehrere Objekte befinden können, sind für diese Ermittlung mehrere Methoden denkbar. Für die Arbeit soll die Methode „Linkage zwischen den Gruppen“ angewendet werden. Die Methode konstruiert alle Paare aus in den beiden Clustern enthaltenen Objekten (die Paare enthalten somit Objekte aus beiden Clustern). Dann wird für jedes Paar die Distanz berechnet (wie in der oben gezeigten Weise). Das arithmetische Mittel dieser Distanzen bildet die Distanz zwischen den beiden Clustern.

4 Auswertung und Analyse der Klausurergebnisse

In diesem Kapitel sollen die möglichen Einflussfaktoren auf eine Klausurnote untersucht werden. Im Kapitel 2.4. wurden bereits mögliche Einflussfaktoren, wie beispielsweise das Geschlecht des Studenten, die vom Studenten besuchte Übung, die Benutzung von Lehrmaterial, die Vorbereitungszeit usw. aufgeführt. Dazu sind jedoch vorerst einige Anmerkungen auszuführen.

Zur besseren Anschaulichkeit basieren die Balkendiagramme nicht auf den normierten Gesamtpunkten, sondern auf der Klausurnote. Die Tests beruhen weiterhin auf den normierten Gesamtpunkten. Den Zusammenhang zwischen den Klausurnoten und den normierten Gesamtpunkten findet man in Tabelle A-1. Die durchgeführten Tests der nächsten Unterkapitel wurden im Kapitel 3 erläutert. Es wird dabei ein Signifikanzniveau von $\alpha = 0,05$ angenommen, sofern keine anderen Annahmen getroffen werden.

Die erhobenen Daten ermöglichen es, zu untersuchen, ob das Geschlecht des Studenten, sein gewähltes Studienfach, die von ihm besuchte Übung und die Benutzung des Lehrmaterials MM-Stat das Ergebnis einer Klausur beeinflussen kann. Weiterhin soll gezeigt werden, dass auch die Aufgabenstellungen der Klausuraufgaben wichtig für den Ausgang einer Klausur sind.

Und zum Schluss wird noch auf Daten, wie zum Beispiel die Durchfallquote oder auch der Zusammenhang zwischen der Lehrveranstaltung Statistik I und Statistik II eingegangen.

4.1 Einflussfaktoren

4.1.1 Einfluss des Merkmals Geschlecht

Im Folgenden soll die These untersucht werden, ob das Geschlecht einen Einfluss auf die Klausurnote hat. Für diese Untersuchung wurde in vier Klausuren (zwei aus Statistik I und zwei aus Statistik II) zusätzlich das Merkmal Geschlecht erhoben.

Für die Betrachtung werden die Statistik I - und die Statistik II – Klausur getrennt voneinander ausgewertet, da sich beide Klausurtypen, wie im Kapitel 2.4 gezeigt, unterscheiden.

4.1.1.1 Die Klausur Statistik I

Die erhobenen Daten stammen aus dem Sommersemester 2002. Die Klausuren wurden am 30. Juli und 09. Oktober geschrieben. Es haben 457 Studenten, 218 Männer und 239 Frauen, teilgenommen.

Die Abbildung 4-1 und die Abbildung 4-2 zeigen folgendes :

1. 75 Frauen und 49 Männer sind durchgefallen, das sind 26 Frauen mehr.
2. Mit 4.0 haben 22.2 % der Frauen und 19.7 % der Männer bestanden.
3. Die männlichen Studenten sind im Notenbereich von 2.3 bis 3.3 stärker vertreten.
4. Der Mittelwert der Männer beträgt 51.04 Punkte und der der Frauen 46.08 Punkte (Tabelle A-10).

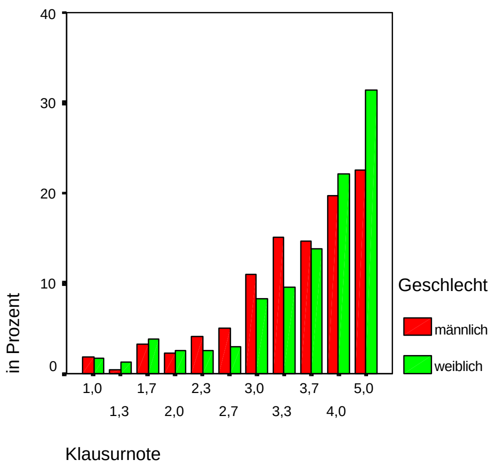


Abbildung 4-1 Balkendiagramm der Klausurnoten nach Geschlecht für Statistik I

Note * Geschlecht Crosstabulation

			Geschlecht		
			männlich	weiblich	Total
Note	1,0	% within Geschlecht	1,8%	1,7%	1,8%
	1,3	% within Geschlecht	,5%	1,3%	,9%
	1,7	% within Geschlecht	3,2%	3,8%	3,5%
	2,0	% within Geschlecht	2,3%	2,5%	2,4%
	2,3	% within Geschlecht	4,1%	2,5%	3,3%
	2,7	% within Geschlecht	5,0%	2,9%	3,9%
	3,0	% within Geschlecht	11,0%	8,4%	9,6%
	3,3	% within Geschlecht	15,1%	9,6%	12,3%
	3,7	% within Geschlecht	14,7%	13,8%	14,2%
	4,0	% within Geschlecht	19,7%	22,2%	21,0%
	5,0	% within Geschlecht	22,5%	31,4%	27,1%
Total		% within Geschlecht	100,0%	100,0%	100,0%
		% of Total	47,7%	52,3%	100,0%

Abbildung 4-2 Übersicht der Klausurnoten nach Geschlecht für Statistik I

Man kommt zu der These, dass die Männer im Durchschnitt die Klausur der Statistik I besser bestehen als die Frauen. Der Boxplot (Abbildung 4-3) erhärtet die These zusätzlich. Nun soll anhand eines Testes gezeigt werden, dass diese Mittelwertunterschiede signifikant sind. Die Annahme der unabhängigen Zufallsstichproben ist erfüllt, denn die Beobachtungen der Stichprobe der Männer hängen nicht von der Beobachtung der Stichprobe der Frauen ab. Die Prüfung der Zufallsvariablen in den Grundgesamtheiten hat ergeben, dass diese (Tabelle A-8) nicht normalverteilt sind. Daher kann der t-Test nicht durchgeführt werden und es wird auf den Mann-Whitney-Test zurückgegriffen.

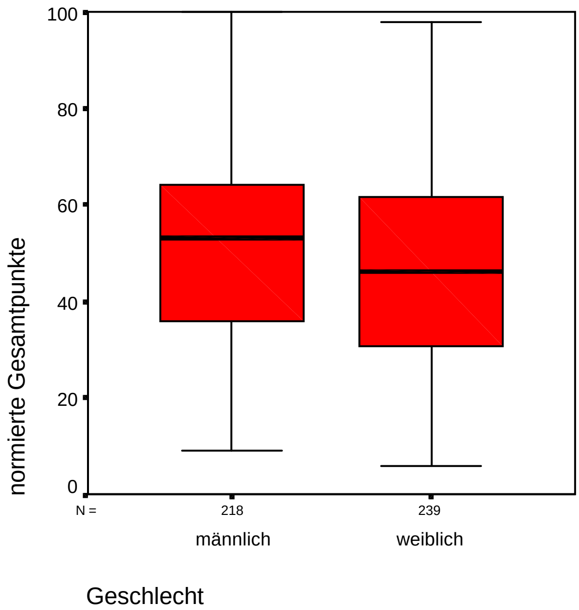


Abbildung 4-3 Boxplot der normierten Gesamtpunkte nach Geschlecht für Statistik I

Da das empirische Signifikanzniveau der Tabelle A-9 kleiner als α = 0,05 ist, kann die Nullhypothese nicht aufrechterhalten werden. Es wird daher davon ausgegangen, dass die mittleren normierten Gesamtpunkte der Frauen und Männer signifikant voneinander verschieden sind. Da der beobachtete Mittelwert der Männer größer ist als der der Frauen (Tabelle A-10) kann die Behauptung, dass die Männer im Mittel mit einer besseren Klausurnote bestehen als die Frauen nicht widerlegt werden.

4.1.1.2 Die Klausur Statistik II

Das Merkmal Geschlecht für die Lehrveranstaltung Statistik II wurde aus den geschriebenen Klausuren des Wintersemesters 2002/2003 entnommen. An den Klausuren vom 18. Februar und 11. April haben 184 männliche und 175 weibliche Studenten teilgenommen.
Das Balkendiagramm (Abbildung 4-4) und die Abbildung 4-5 lassen folgende Aussagen zu:

1. Es sind 39 Männer und 39 Frauen durchgefallen sind.
2. Mit 4.0 haben 34 Frauen und 32 Männer die Klausur Statistik II bestanden.
3. Mit 3.7 haben 20 % der Frauen und 9.8 % der Männer die Klausur abgeschlossen.
4. Mit Ausnahme der Klausurnote 2.3 haben mehr männliche Studenten im Notenbereich von 1.0 bis 3.3 die jeweilige Klausurnote erhalten.
5. Der Mittelwert der Männer beträgt 58.27 Punkte und der der Frauen 54.17 Punkte (Tabelle A-11).

Daher lässt sich wieder die These aufstellen, dass die männlichen Studenten im Mittel die Klausur besser bestehen als die weiblichen Studenten.

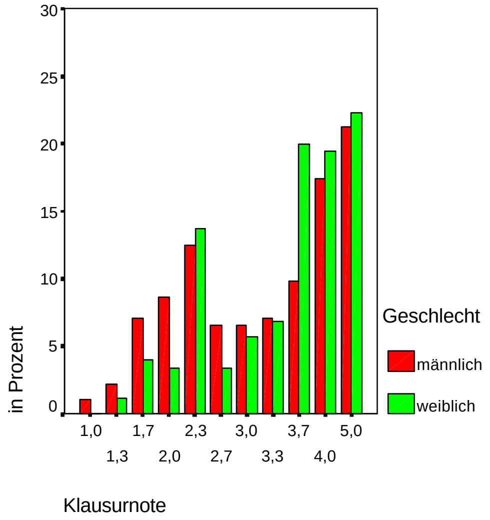


Abbildung 4-4 Balkendiagramm der Klausurnoten nach Geschlecht für Statistik II

Note * Geschlecht Crosstabulation

			Geschlecht		
			männlich	weiblich	Total
Note	1,0	% within Geschlecht	1,1%		,6%
	1,3	% within Geschlecht	2,2%	1,1%	1,7%
	1,7	% within Geschlecht	7,1%	4,0%	5,6%
	2,0	% within Geschlecht	8,7%	3,4%	6,1%
	2,3	% within Geschlecht	12,5%	13,7%	13,1%
	2,7	% within Geschlecht	6,5%	3,4%	5,0%
	3,0	% within Geschlecht	6,5%	5,7%	6,1%
	3,3	% within Geschlecht	7,1%	6,9%	7,0%
	3,7	% within Geschlecht	9,8%	20,0%	14,8%
	4,0	% within Geschlecht	17,4%	19,4%	18,4%
	5,0	% within Geschlecht	21,2%	22,3%	21,7%
Total		% within Geschlecht	100,0%	100,0%	100,0%
		% of Total	51,3%	48,7%	100,0%

Abbildung 4-5 Übersicht der Klausurnoten nach Geschlecht für Statistik II

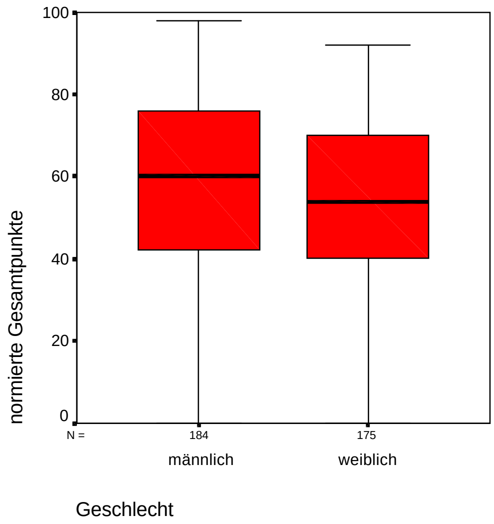


Abbildung 4-6 Boxplot der normierten Gesamtpunkte nach Geschlecht für Statistik II

Der Boxplot (Abbildung 4-6) zeigt, dass die Unterschiede zwischen den Geschlechtern nicht so stark ausgeprägt ist wie in der Statistik I Klausur. Gibt es daher noch signifikante Unterschiede im Geschlecht in Bezug auf das Abschneiden in der Klausur?

Wiederum soll ein Test zur Erhärtung der These angewendet werden. Die Unabhängigkeit der beiden Zufallsstichproben voneinander ist hier aus dem gleichen Grund wie bei der Statistik I Klausur gegeben. Der Kolmogorov-Smirnov-Test (Tabelle A-12) zeigt, dass keine Normal-

verteilung vorliegt. Daher wird wieder auf den Mann-Whitney-Test zurückgegriffen. Aufgrund des Signifikanzniveauvergleichs (Tabelle A-13) wird die Nullhypothese verworfen. Damit sind die Mittelwerte zwischen den Geschlechtern wesentlich verschieden voneinander und man gelangt zu dem gleichen Ergebnis wie in der Lehrveranstaltung Statistik I.
Somit sind die beobachteten Mittelwertunterschiede der normierten Gesamtpunkte nach Geschlecht signifikant und kein Zufall. Also kann die oben gestellte These erhärtet werden.
Eine mögliche Ursache in diesem Ergebnis liegt darin, dass Männer im Mathematikunterricht der Schule weniger Probleme haben, weil sie durch verstärktes Basteln und Werken in der Kindheit das räumliche Vorstellungsvermögen besser geschult haben als Mädchen. Auch mathematische Sachaufgaben liegen ihnen besser, da hier oft Themen angesprochen werden, die Jungen bzw. Männer mehr interessieren. Lässt man den Frauen jedoch mehr Zeit mit dem Lösen der Aufgaben werden diese Differenzen geringer. Dies hängt wohl damit zusammen, dass Mädchen bzw. Frauen mit weniger Zuversicht an die Lösung der Aufgaben herangehen. Ursache dafür ist, dass schon in der Schule von Mädchen geringere Mathematikleistungen erwartet werden als bei Jungen. Zusätzlich gelten in unserer Gesellschaft gute Mathematikleistungen als unweiblich, so dass Mädchen in der Pubertät ihre Ressourcen nicht voll ausschöpfen. Somit wird schon in der Kindheit der Grundstein für schlechtere Mathematikleistungen bei Frauen gelegt, die sich im Studium natürlich fortsetzen. (Zech, 1996)

4.1.2 Einfluss des Merkmals Übung

Im Grundstudium wird zur Statistik I und Statistik II vierzehntägig eine Übung für die Studenten angeboten. Die Übungen sind für die Studenten obligatorisch. Es wird nicht überprüft, ob ein Student tatsächlich an der Übung teilgenommen hat. In diesem Kapitel soll daher untersucht werden, ob der Besuch der Übungen sich positiv auf die Klausurnote auswirkt. Dazu wurde im Sommersemester 2002 die Frage nach der besuchten Übung gestellt.
Als erstes sollen die normierten Gesamtpunkte von Studenten, die an einer Übung teilgenommen haben und von den Studenten, die keine Angaben in der Erhebung zu diesem Thema gemacht haben, miteinander verglichen werden.
Das Balkendiagramm (Abbildung 4-7) und die Abbildung 4-8 zeigen folgendes:

1. Die sehr guten bis guten Noten wurden fast ausschließlich nur von Studenten erzielt, die an einer Übung teilgenommen haben.
2. Prozentual sind jedoch auch mehr Studenten durchgefallen, die eine Übung besucht haben.
3. Die Übung wird von 89.7 % der Studenten besucht

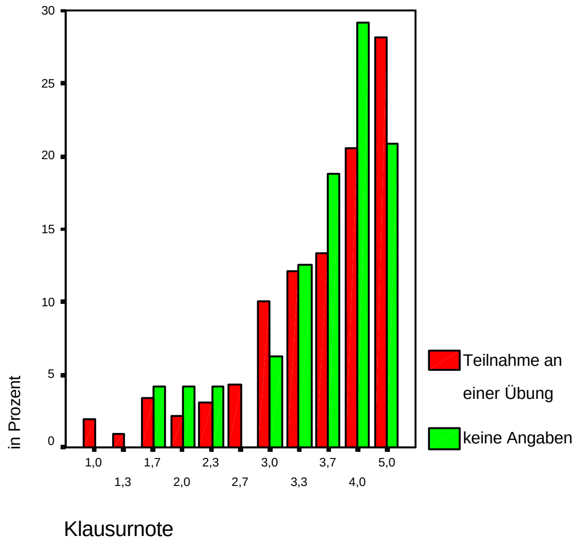


Abbildung 4-7 Balkendiagramm der Klausurnoten nach Teilnahme an Übung

Note * Vergleich Teilnahme an Übung und keine Teilnahme Crosstabulation

		Vergleich Teilnahme an Übung und keine Teilnahme		
		Teilnahme an einer Übung	k.A.	Total
Note	1,0	1,9%		1,7%
	1,3	1,0%		,9%
	1,7	3,3%	4,2%	3,4%
	2,0	2,1%	4,2%	2,4%
	2,3	3,1%	4,2%	3,2%
	2,7	4,3%		3,9%
	3,0	10,0%	6,3%	9,6%
	3,3	12,2%	12,5%	12,2%
	3,7	13,4%	18,8%	13,9%
	4,0	20,5%	29,2%	21,4%
	5,0	28,2%	20,8%	27,4%
Total		100,0% 89,7%	100% 10,3%	100% 100%

Abbildung 4-8 Übersicht der Klausurnoten nach Teilnahme an Übung

Kann daher die These aufgestellt werden, dass Studenten, die an einer Übung teilgenommen haben im Mittel mit einer besseren Klausurnote bestehen?

Auch hier liegt keine Normalverteilung vor (Tabelle A-14). Bei Anwendung des Mann-Whitney-Tests (Tabelle A-15) zeigt sich, dass der Besuch einer Übung keinen Einfluss auf die Mittelwerte der normierten Gesamtpunkte hat. Es ist also nicht bewiesen, dass ein Student, der keine Übung besucht die Klausur schlechter schreibt als ein anderer Student, die eine Übung besucht hat.

Das Ergebnis lässt sich in der Weise erklären, dass man aus Ausprägung „keine Angabe“ nicht unbedingt schließen kann, dass keine Übung besucht wurde. Der jeweilige Student kann auch zwischen den Übungen sehr oft gewechselt haben und konnte sich somit für keine „Hauptübung“ entscheiden. Aber es ist auch möglich, dass er sich mit anderen Lehrmaterialen gut auf die Klausur vorbereitet hat.

Im Folgenden soll die Behauptung untersucht werden, dass sich der Besuch einer bestimmten Übung positiv auf das Klausurergebnis auswirkt.

Die Abbildung 4-9 bis Abbildung 4-11 erlauben folgende Aussagen:

1. Die Übung 2 ist mit einem Mittelwert von 50.82 Punkten gut ausgefallen.
2. An dieser Übung haben auch die meisten Studenten teilgenommen.
3. In der Übung 3 gibt es die größte Durchfallquote.

Studenten, die keine Angaben zu dem Besuch der Übung gemacht haben, werden nicht weiter berücksichtigt.

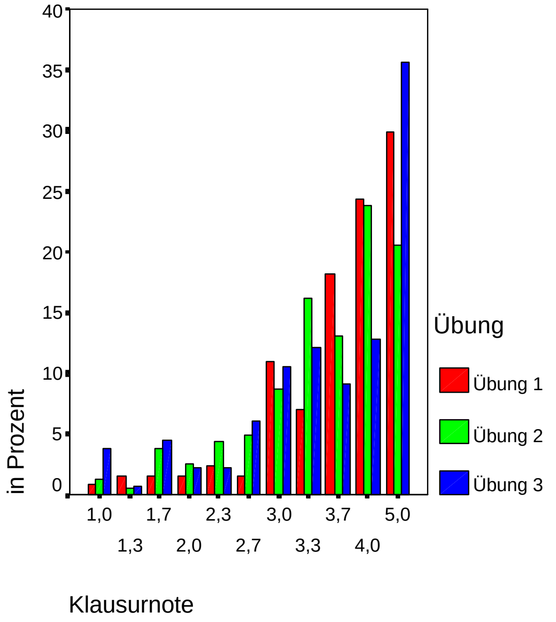


Abbildung 4-9 Balkendiagramm der Klausurnoten nach Übung

Note * Übung Crosstabulation

			Übung			Total
			Übung 1	Übung 2	Übung 3	
Note	1,0	% within Übung	,8%	1,3%	3,8%	1,9%
	1,3	% within Übung	1,6%	,6%	,8%	1,0%
	1,7	% within Übung	1,6%	3,8%	4,5%	3,3%
	2,0	% within Übung	1,6%	2,5%	2,3%	2,1%
	2,3	% within Übung	2,4%	4,4%	2,3%	3,1%
	2,7	% within Übung	1,6%	5,0%	6,1%	4,3%
	3,0	% within Übung	11,0%	8,8%	10,6%	10,0%
	3,3	% within Übung	7,1%	16,3%	12,1%	12,2%
	3,7	% within Übung	18,1%	13,1%	9,1%	13,4%
	4,0	% within Übung	24,4%	23,8%	12,9%	20,5%
	5,0	% within Übung	29,9%	20,6%	35,6%	28,2%
Total		% within Übung	100,0%	100,0%	100,0%	100,0%
		% of Total	30,3%	38,2%	31,5%	100,0%

Abbildung 4-10 Übersicht der Klausurnoten nach Übung

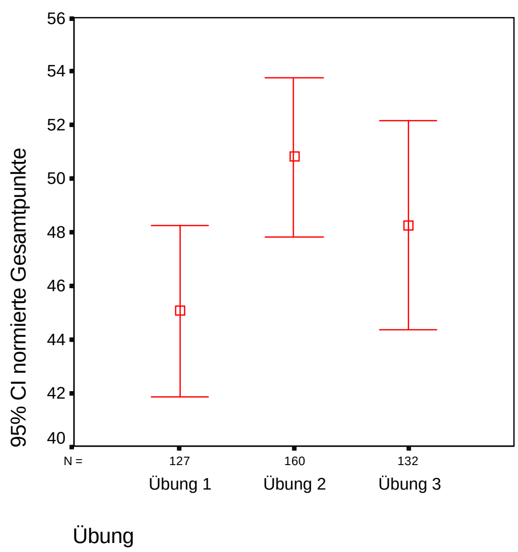


Abbildung 4-11 Fehlerbalkendiagramm der normierten Gesamtpunkte nach Übung

Die Durchführung eines Kruskal-Wallis-Tests (Tabelle A-17) nach Ablehnung der Normalverteilung (Tabelle A-16) zeigt, dass die Unterschiede der beobachteten Mittelwerte der Übungen signifikant sind. Damit ist klar, dass der Besuch der Übung 2 sich positiv auf das Klausurergebnis ausgewirkt hat.

Wie kann man dieses Ergebnis erklären? Aus dem Kaptitel 4.1. weiß man, dass die Männer im Mittel besser die Klausur bestehen. Kann es nun sein, dass aus dieser Übung so gute Klausurnoten hervorgegangen sind, weil sie von mehr Männern besucht wurde?
Die Abbildung 4-12 bestätigt den Verdacht, da nur an der Übung 2 mehr Männer teilgenommen haben. Jedoch impliziert diese Aussage, dass Übungen mit einem hohen Frauenanteil sehr schlechte Klausurergebnisse haben müssen. Nach dieser These hätte die Übung 3 sehr schlecht ausfallen müssen, doch Abbildung 4-11 widerlegt dies jedoch. Daher kann der Anteil der Frauen bzw. Männer an einer Übung nicht ausschlaggebend für das Klausurergebnis sein.

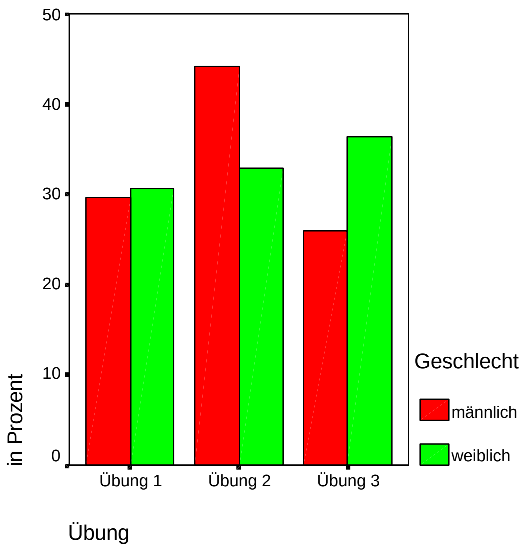


Abbildung 4-12 Anzahl der Studenten nach Geschlecht in den jeweiligen Übungen

Eine wichtige Rolle spielt dagegen die Erfahrung des Übungsleiters. Ein Übungsleiter, der schon viele Übungen gehalten hat, weiß auf welche Themen er verstärkt eingehen muss und bei welchen Aufgabenstellungen die Studenten Probleme haben. Nach Auskunft von Dr. Klinke hat der Übungsleiter der Übung 2 schon mehrere Semester die Übung gehalten. Die wenigsten Erfahrungen dagegen hatte der Übungsleiter der Übung 1. Bei nochmaliger Betrachtung der Abbildung 4-11 wird diese These bestätigt.

4.1.3 Einfluss des Merkmals „Fachrichtung“

Im Nachfolgenden wird der Einfluss der gewählten Fachrichtung eines Studenten auf sein Klausurergebnis untersucht. Dabei werden BWL-Studenten, VWL-Studenten und sonstige Studenten mit Zweitfach unterschieden.
Dazu wurden in der Statistik I - Klausur (Sommersemester 2002, 30. Juli und 09. Oktober) die Studenten nach ihrer Studienrichtung befragt. Es ergaben sich folgende Ergebnisse :

1. 196 Studenten studieren BWL.

2. 162 Studenten gehören der Fachrichtung VWL an.
3. 103 Studenten gehören einer sonstigen Studienrichtung an und studieren BWL bzw. VWL im Zweitfach.

Die BWL-Studenten sind also am stärksten vertreten. Weitere Analysen ergeben :

4. 20.9 % der BWL-Studenten sind durchgefallen (Abbildung 4-13).
5. Bei den VWL-Studenten sind es 31.5 %.
6. Bei den Studenten mit sonstiger Fachrichtung sind es sogar 34 %.

Die Mittelwerte der normierten Gesamtpunkte lauten wie folgt :

7. Die BWL-Studenten erzielten im Mittel ein Ergebnis von 50.26 Punkten.
8. 46.33 Punkte wurden von den VWL-Studenten im Mittel erreicht.
9. Die sonstigen Studenten erzielten einen Mittelwert von 47.75.

Das Fehlerbalkendiagramm (Abbildung 4-14) der normierten Gesamtpunkte nach Studienrichtung zeigt jedoch keine signifikanten Unterschiede.

Zusammenfassend ist also die These aufzustellen, dass die gewählte Fachrichtung keine Rolle für das Klausurergebnis spielt.

Da drei Stichproben vorliegen, bietet es sich an einen ANOVA-Test durchzuführen. Dabei wird zuerst getestet, ob die Voraussetzungen für die Durchführung gegeben sind. Falls dies verneint wird, wird der Kruskal-Wallis-Test herangezogen.

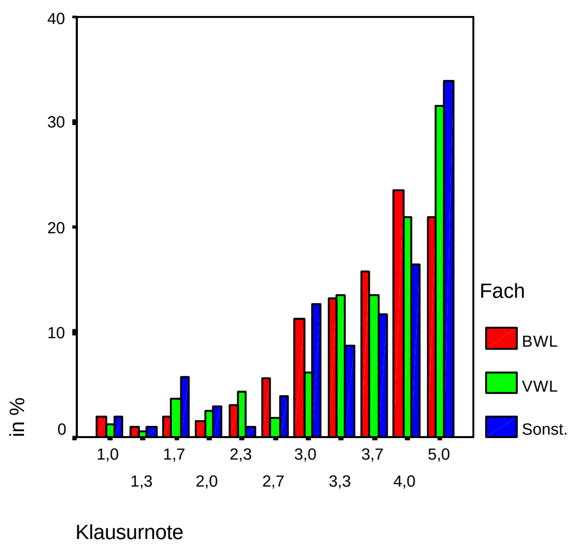


Abbildung 4-13 Balkendiagramm der Klausurnoten nach Studienrichtung

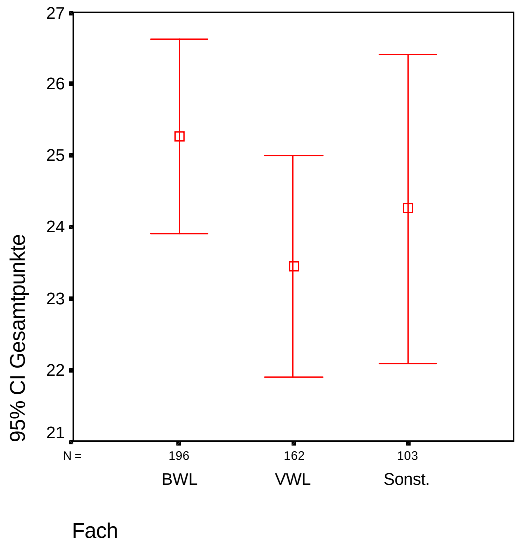


Abbildung 4-14 Fehlerbalkendiagramm der normierten Gesamtpunkte nach Studienrichtung

Die Zufallsvariablen müssen in den Grundgesamtheiten normalverteilt sein, dazu wird der Kolmogorov-Smirnov-Test angewendet. Die Auswertung (Tabelle A-19) zeigt, dass keine Normalverteilung vorliegt.

Da das empirische Signifikanzniveau (Tabelle A-20) größer ist als $\alpha = 0{,}05$ wird die Homogenitätshypothese angenommen. Also lässt sich für den Studenten die These aufstellen, dass es bezüglich der Klausurnote nicht von Bedeutung ist, aus welcher Studienrichtung er kommt.

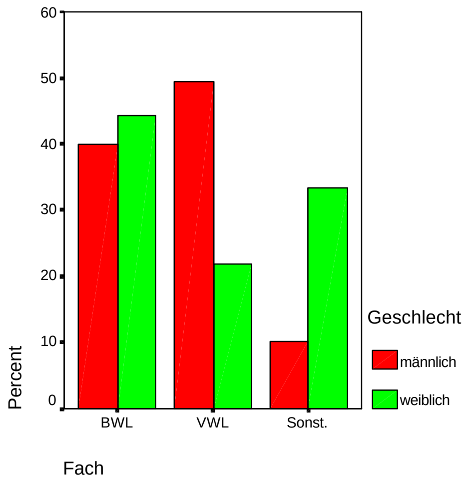


Abbildung 4-15 Balkendiagramm der Studienrichtung nach Geschlecht

Dies kann daran liegen, dass für sich für alle Studenten die Lehrveranstaltung im Grundstudium nicht unterscheiden. Somit haben alle Studenten die gleichen Voraussetzungen.
Im Kapitel 4.1.1 wurde gezeigt, dass die Frauen im Mittel die Klausuren schlechter bestehen als die Männer. Setzt sich diese Tendenz in den Fachrichtungen weiter? Kann behauptet werden, dass eine VWL-Studentin im Mittel schlechter ist als ein VWL-Student?
Abbildung 4-15 zeigt die Geschlechterzusammensetzung bezüglich der Studienrichtung. Anschließend soll der Einfluss des Geschlechts auf die Klausurnote getrennt nach Studienrichtung überprüft werden.

4.1.3.1 Die BWL-Studenten

Von den insgesamt 196 BWL-Studenten sind 87 männlich und 106 weiblich, 3 Studenten haben keine Angaben gemacht. In der Grafik (Abbildung 4-16) ist folgende Tendenz zu erkennen :

1. Von 39 durchgefallenen Studenten sind 27 weiblich.
2. Auch unter den schlechteren Noten (4.0 und 3.7) sind mehr Frauen vertreten.
3. Die männlichen Studenten sind bei den besseren Klausurnoten (1.3, 1.7, 2.0 und 2.7) stärker vertreten.

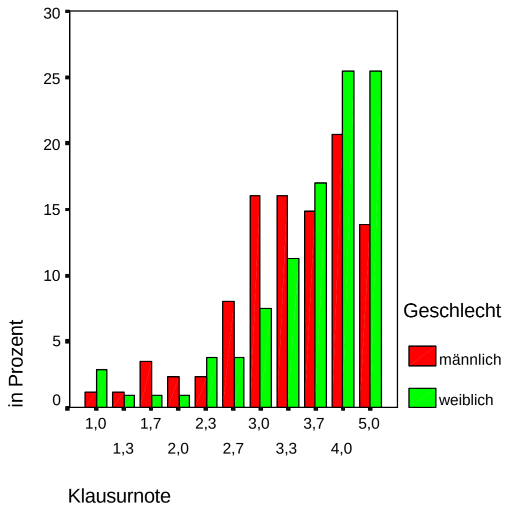


Abbildung 4-16 Balkendiagramm der Klausurnoten der BWL-Studenten nach Geschlecht

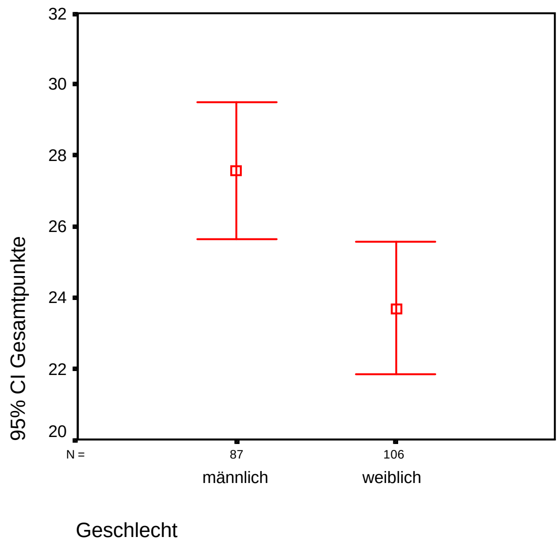


Abbildung 4-17 Fehlerbalkendiagramm der normierten Gesamtpunkte der BWL-Studenten nach Geschlecht

Zusätzlich lässt das Fehlerbalkendiagramm (Abbildung 4-17) signifikante Unterschiede im Mittelwert vermuten, da sich die Konfidenzintervalle nicht überlappen.

Wie in Kapitel 4.1. gezeigt, erhärtet sich auch hier die These, dass weibliche BWL-Studenten im Mittel schlechtere Klausurergebnisse erzielen.

Der Kolmogorov-Smirnov-Test (Tabelle A-22) zeigt, dass die Grundgesamtheiten nicht normalverteilt sind. Also wird der Mann-Whitney-Test angewendet. Dieser wird (Tabelle A-23) abgelehnt.

Die oben gemachte These wird damit statistisch bewiesen.

4.1.3.2 Die VWL-Studenten

Nun soll die analoge Untersuchung für die VWL-Studenten durchgeführt werden. Von den 160 VWL-Studenten sind 108 männlich und 52 weiblich. Damit sind die Männer in dieser Studienrichtung am stärksten vertreten. Aus dem Diagramm (Abbildung 4-18) sind folgende Angaben zu entnehmen :

1. Von den Frauen sind 36.5 % durchgefallen, von den Männer 28.7 %.
2. 25 % der Frauen und 19.4 % der Männer haben mit 4.0 bestanden.
3. Ansonsten sind die Männer prozentual bei den besseren Klausurnoten (1.0, 2.3 bis 3.7) stärker vertreten.

Das Fehlerbalkendiagramm (Abbildung 4-19) zeigt keine signifikanten Unterschiede in den mittleren normierten Gesamtpunkten.

Damit kann die These aufgestellt werden, dass das Geschlecht der VWL-Studenten keinen Einfluss auf das Klausurergebnis hat.

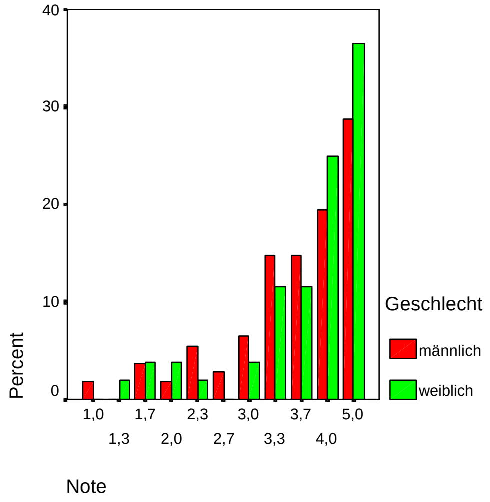


Abbildung 4-18 Balkendiagramm der Klausurnoten der VWL-Studenten nach Geschlecht

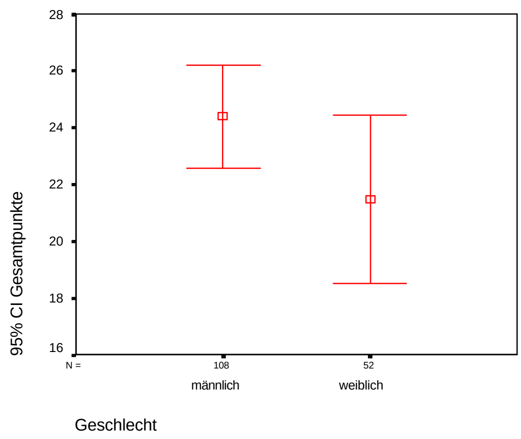


Abbildung 4-19 Fehlerbalkendiagramm der normierten Gesamtpunkte der VWL-Studenten nach Geschlecht

Da die Nullhypothese auf Normalverteilung (Tabelle A-24) und die Varianzhomogenität (Tabelle A-25) ebenfalls angenommen wird, kann der t-Test durchgeführt werden. Dieser zeigt (Tabelle A-25), dass die Nullhypothese nicht zutrifft. Damit wird die oben gestellte These statistisch widerlegt.

4.1.3.3 Sonstige

Die letzten 102 Studenten, 22 Männer und 80 Frauen, werden unter der Rubrik „sonstige Fachrichtung“ zusammengefasst. Es kann nicht genau festgestellt werden, aus welcher Studienrichtung diese Studenten kommen. Es ergaben sich nach Abbildung 4-20 folgende Ergebnisse :

1. Von den Frauen sind 36.3 % und von den Männern 27.3 % durchgefallen.
2. Die Klausurnoten 1.3, 1.7 und 2.0 wurden nur von weiblichen Studenten geschrieben.
3. In den Klausurnoten 1.0, 2.7 bis 3.7 und 4.0 sind die Männer stärker vertreten.

Das Fehlerbalkendiagramm (Abbildung 4-21) zeigt keine signifikanten Unterschiede im Mittelwert der normierten Gesamtpunkte.

Somit lässt sich vermuten, dass das Geschlecht der Studenten der sonstigen Fachrichtung keinen Einfluss auf die Klausurnote hat. Laut Kolmogorov-Smirnov-Test (Tabelle A-26) liegen normalverteilte Grundgesamtheiten vor. Die Varianzhomogenität wird nach dem Levené-Test (Tabelle A-27) angenommen. Die Durchführung des t-Tests (Tabelle A-27) zeigt, dass die

Homogenitätshypothese nicht verworfen werden kann. Damit muss die oben gestellte These abgelehnt werden.

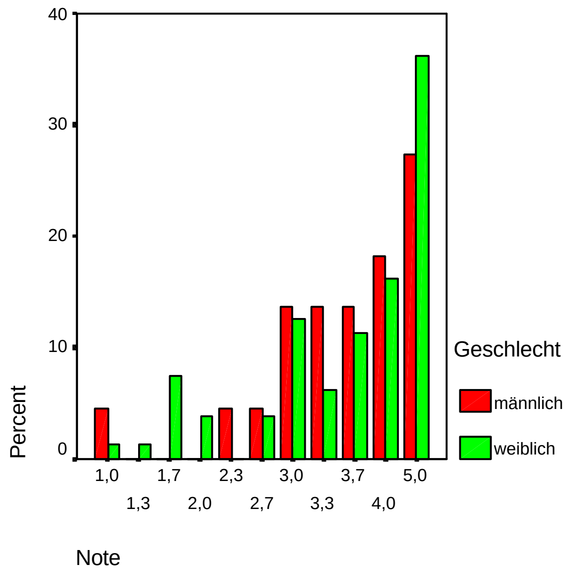


Abbildung 4-20 Balkendiagramm der Klausurnoten des sonst. Studienfachs nach Geschlecht

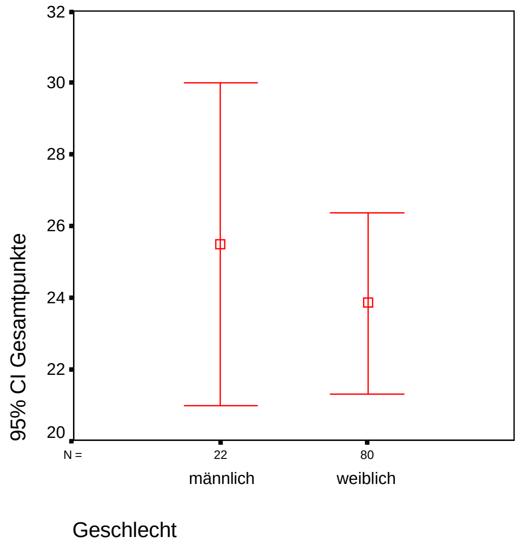


Abbildung 4-21 Fehlerbalkendiagramm der normierten Gesamtpunkte des sonst. Studienfachs nach Geschlecht

Wie oben in

Abbildung 4-15 gezeigt, dominieren in dieser Studienrichtung die Frauen sehr stark. Jedoch erreichen sie im Mittel bessere Klausurnoten als die Studentinnen der anderen Fachrichtungen (Abbildung 4-20). Es kann also sein, dass die Studentinnen aus den sonstigen Fachrichtungen bessere Vorkenntnisse hatten, dies kann jedoch aufgrund der fehlenden Daten in dieser Arbeit nicht weiter vertieft werden.

4.1.4 Einfluss des Merkmals MM-Stat

In den zwei Klausuren des Wintersemesters 2002/2003 war es möglich, den Studenten die Frage zu stellen, ob sie MM-Stat als Vorbereitung für die Klausur bzw. zum Lernen und Vertiefen des Stoffes benutzt haben. Die interessante Frage hierbei ist, ob es einen erkennbaren Einfluss auf die Klausur hat, falls sich der Student mit MM-Stat vorbereitet hat oder nicht.

MM-Stat ist ein interaktives Multimedia-Tool, das eine Einführung in die Statistik geben soll. Dabei behandelt es Themengebiete der deskriptiven und induktiven Statistik. Ein Kapitel setzt sich aus mehreren Lektionen zusammen, die das Grundkonzept, die Definitionen, die Formeln und die grafischen Darstellungen des jeweiligen Themas näher erläutern. Durch viele Beispiele und die Möglichkeit des interaktiven Wechsels zu einzelnen Themengebiete kann jeder Nutzer des MM-Stat seinen Lernprozess individuell gestalten. (www.md-stat.com)

Das Balkendiagramm (Abbildung 4-22) und Fehlerbalkendiagramm (Abbildung 4-23) ermöglichen nachstehende Angaben :

1. Von insgesamt 368 Studenten benutzten 207 MM-Stat, 31 Studenten machten keine Angaben.
2. Es ist den Abbildungen nicht zu entnehmen, ob die Gruppe von Studenten, die MM-Stat benutzt hat, bessere Klausurnoten erzielt.
3. Die Überlappung der Konfidenzintervalle deutet auf keinen signifikanten Unterschied hin.

Daher wird die These aufgestellt, dass die Benutzung von MM-Stat keinen Einfluss auf das Klausurergebnis hat.

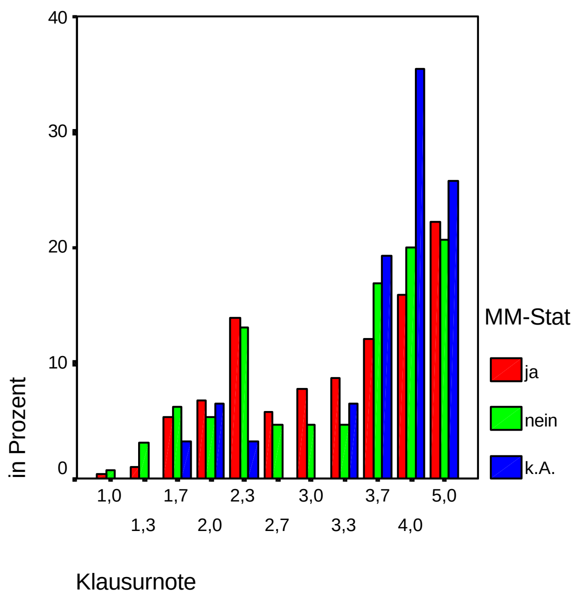


Abbildung 4-22 Balkendiagramm der Klausurnoten nach Benutzung von MM-Stat

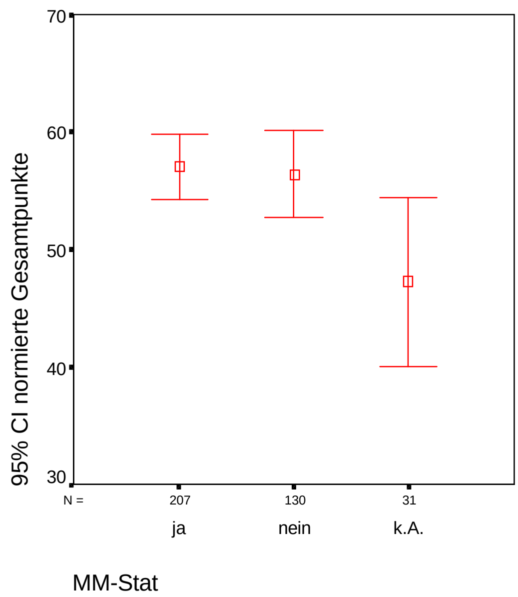


Abbildung 4-23Fehlerbalkendiagramm der normierten Gesamtpunkte nach Benutzung von MM-Stat

Da der Kolmogorov-Smirnov-Test (Tabelle A-28) die Normalverteilung der Grundgesamtheiten ablehnt, wird der in Kapitel 3 beschriebene Kruskal-Wallis-Test angewendet. Dieser verwirft die Homogenitätshypothese (Tabelle A-29) nicht. Also unterscheiden sich die drei Mittelwerte nicht signifikant voneinander. Die oben gemachte These ist damit statistisch nicht widerlegt werden.

Wie kann man sich dieses Ergebnis nun erklären. Ist es etwa unsinnig mit MM-Stat zu arbeiten, da der Student keine Vorteile daraus ziehen kann oder gibt es eine andere Erklärung? Es lässt sich vermuten, dass die Studenten noch viele andere Lehrmaterialien benutzen. Es könnte daher sein, dass diese vergleichbar mit MM-Stat sind, so dass daraus keine Unterschiede ableitbar sind. Jedoch fehlen für weitere Untersuchungen die Daten, welche anderen Lehrmaterialien von den Studenten noch genutzt wurden, um weitere Konsequenzen daraus ableiten zu können.

Da es jedoch zwischen den Geschlechtern deutliche Unterschiede, wie in Kapitel 4.1.1 gezeigt, gibt, existieren eventuell nach dessen Unterscheidung wesentliche

MM-Stat * Geschlecht Crosstabulation

% of Total

		Geschlecht		
		männlich	weiblich	Total
MM-Stat	ja	30,1%	26,7%	56,8%
	nein	17,3%	17,5%	34,8%
	k.A.	3,9%	4,5%	8,4%
Total		51,3%	48,7%	100,0%

Tabelle 4-1 Übersicht Benutzung von MM-Stat nach Geschlecht

Unterschiede in der Benutzung von MM-Stat. Könnte es daher sein, dass eine Frau, die MM-Stat benutzt hat besser in der Klausurnote abgeschnitten hat, als eine Studentin, die MM-Stat nicht benutzt hat?

4.1.4.1 Die Frauen

Zuerst werden die Voraussetzungen für die Durchführung einer ANOVA wie oben geprüft. Der Kolgomorov-Smirnov-Test auf Normalverteilung (Tabelle A-30) wird angenommen und der Levené-Test (Tabelle A-31) auf Varianzhomogenität ebenfalls.

Note * MM-Stat Crosstabulation[a]

			MM-Stat			
			ja	nein	k.A.	Total
Note	1,3	% within MM-Stat		3,2%		1,1%
	1,7	% within MM-Stat	4,2%	3,2%	6,3%	4,0%
	2,0	% within MM-Stat	4,2%	1,6%	6,3%	3,4%
	2,3	% within MM-Stat	14,6%	14,3%	6,3%	13,7%
	2,7	% within MM-Stat	3,1%	4,8%		3,4%
	3,0	% within MM-Stat	7,3%	4,8%		5,7%
	3,3	% within MM-Stat	8,3%	3,2%	12,5%	6,9%
	3,7	% within MM-Stat	16,7%	25,4%	18,8%	20,0%
	4,0	% within MM-Stat	17,7%	22,2%	18,8%	19,4%
	5,0	% within MM-Stat	24,0%	17,5%	31,3%	22,3%
Total		% within MM-Stat	100,0%	100,0%	100,0%	100,0%
		% of Total	54,9%	36,0%	9,1%	100,0%

a. Geschlecht = weiblich

Tabelle 4-2 Übersicht Klausurnote nach Benutzung von MM-Stat für die Frauen

Die Nullhypothese der ANOVA wird (Tabelle A-32) angenommen. Dadurch wird gezeigt, dass es keine Mittelwertunterschiede in den normierten Gesamtpunkten in Bezug auf die Nutzung von MM-Stat gibt.
Das oben gezeigte Ergebnis wird damit also für die Frauen statistisch bestätigt.

4.1.4.2 Die Männer

Analog zu den Frauen wird die These an den Männern getestet. Da der Kolmogorov-Smirnov-Test (Tabelle A-33) die Hypothese auf Normalverteilung ablehnt, wird der Kruskal-Wallis-Test verwendet. Dieser zeigt (Tabelle A-34), dass es keinen Grund gibt, die Hypothese auf gleiche Mittelwerte abzulehnen.

Note * MM-Stat Crosstabulation[a]

			MM-Stat			
			ja	nein	k.A.	Total
Note	1,0	% within MM-Stat	,9%	1,6%		1,1%
	1,3	% within MM-Stat	1,9%	3,2%		2,2%
	1,7	% within MM-Stat	6,5%	9,7%		7,1%
	2,0	% within MM-Stat	9,3%	8,1%	7,1%	8,7%
	2,3	% within MM-Stat	13,9%	12,9%		12,5%
	2,7	% within MM-Stat	8,3%	4,8%		6,5%
	3,0	% within MM-Stat	8,3%	4,8%		6,5%
	3,3	% within MM-Stat	9,3%	4,8%		7,1%
	3,7	% within MM-Stat	8,3%	9,7%	21,4%	9,8%
	4,0	% within MM-Stat	13,0%	16,1%	57,1%	17,4%
	5,0	% within MM-Stat	20,4%	24,2%	14,3%	21,2%
Total		% within MM-Stat	100,0%	100,0%	100,0%	100,0%
		% of Total	58,7%	33,7%	7,6%	100,0%

a. Geschlecht = männlich

Tabelle 4-3 Übersicht Klausurnote nach Benutzung von MM-Stat für Männer

Damit hat sich auch bei den männlichen Studenten statistisch bestätigt, dass die Nutzung von MM-Stat keinen signifikanten Einfluss auf die Klausurnote hat.

4.2 Weitere Ergebnisse

In diesem Kapitel sollen weitere Ergebnisse untersucht werden, die sich anhand der erhobenen Daten ableiten lassen.

4.2.1 Die Durchfallquote

Die Studenten haben nach der alten Prüfungsordnung drei Versuche, um die Klausur zu bestehen. Sollte ein Student im dritten Versuch nicht bestanden haben, so wird er exmatrikuliert. Das Problem in der Untersuchung der Durchfallquote liegt darin, dass nur ein kleiner Zeitabschnitt vom Wintersemester 2000/2001 bis zum Wintersemester 2002/2003 betrachtet wurde. Damit ist nicht klar, ob Studenten die im Wintersemester 2000/2001 geschrieben haben, schon vorher durchgefallen sind und nun ein zweites oder drittes Mal schreiben und wie die Studenten, die im Wintersemester 2002/2003 durchgefallen sind in den nächsten Semestern erneut die Prüfung bestehen oder ob sie überhaupt nicht mehr antreten werden. Somit kann nur annähernd eine Durchfallquote ermittelt werden.

4.2.1.1 Die Klausur Statistik I

In diesem Kapitel soll die Durchfallquote der Statistik I untersucht werden. Die Tabelle A-36 gibt an, dass von 853 Studenten

- im ersten Versuch 23,9 %,
- im zweiten Versuch 2,8% und
- im dritten Versuch 0,5 % durchgefallen sind.

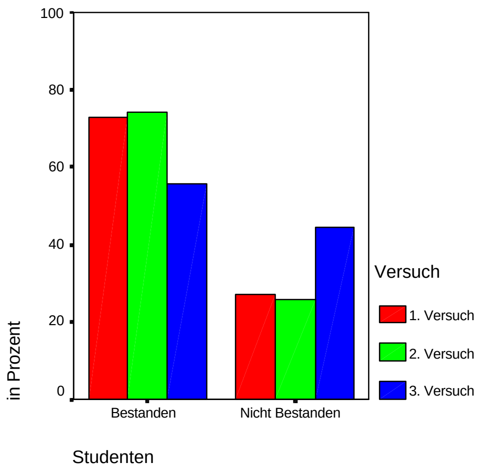


Abbildung 4-24 Balkendiagramm der Durchfallquote nach Anzahl der Versuche in Statistik I

Nun stellt sich die Frage, wie sich ein Student, der im ersten Versuch durchgefallen ist weiter verhalten soll. Dabei wird als erstes untersucht, wie sich die Mittelwerte tendenziell verhalten. Dazu wird die Behauptung aufgestellt, dass der Mittelwert des 2. Versuches höher ist als der Mittelwert des 1. Versuches (bei den Studenten, die im ersten Versuch durchgefallen sind). Tabelle A-37 bestätigt dies.

Da für die Anwendung eines parametrischen Tests keine Unabhängigkeit der Stichproben vorliegt, wird an dieser Stelle der Wilcoxon-Test durchgeführt. Dieser lehnt die Nullhypothese ab (Tabelle A-38). Damit ist die Behauptung statistisch bewiesen. Somit hat ein Student tendenziell die Chance sich im nächsten Versuch zu verbessern.

Auf den dritten Versuch soll hier und im Folgenden nicht weiter eingegangen werden, da dieser nur von neun Studenten durchgeführt wurde.

Zwar sollte immer ein zweiter Versuch unternommen werden, jedoch könnte der Zeitabstand zwischen den Versuchen eine große Rolle spielen. So stellt sich die Frage, ob es für den Studenten besser ist den zweiten Versuch gleich oder im nächsten Semester zu unternehmen.

Die Abbildung 4-25 und die Tabelle A-39 zeigen dabei nachstehende Ergebnisse :

1. 65 % der Studenten schreiben die 2. Klausur zum nächst möglichen Termin.
2. Je später der zweite Versuch unternommen wird, desto schlechter wird die Klausurnote. (Zwar ist der Mittelwert wenn die Klausur zwei Klausuren später geschrieben wird höher als wenn die Klausur zum nächsten Termin geschrieben wird, jedoch konzentrieren sich die Klausurergebnisse bei der späteren Wiederholung stärker um die 3.7 und ist im Bereich der Klausurnoten eins bis zwei fast gar nicht vertreten.)

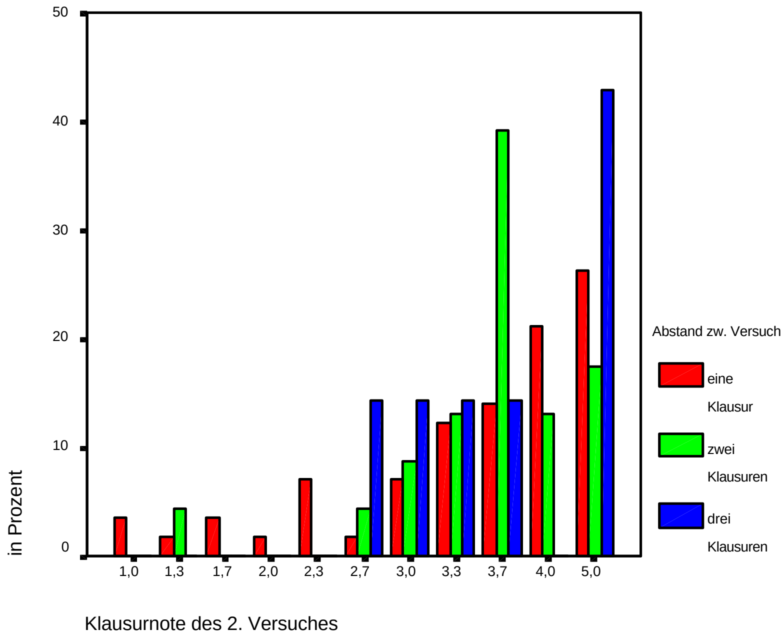


Abbildung 4-25 Balkendiagramm der Klausurnoten nach Anzahl der Versuche in Statistik I

Die These unter Punkt 2 soll statistisch untersucht werden. Da Normalverteilung (Tabelle A-40) und Varianzhomogenität (Tabelle A-41) angenommen werden, zeigt die Durchführung einer ANOVA (Tabelle A-42), dass diese Annahme nicht bestätigt werden kann. Zwar tendiert der einzelne Student dazu die Klausur schnellst möglichst zu wiederholen, es spielt jedoch keine Rolle wann die Wiederholung der Klausur stattfindet.

4.2.1.2 Die Klausur Statistik II

Die analogen Untersuchungen zur Durchfallquote werden für die Statistik II – Klausur durchgeführt. Im Ganzen haben 718 Studenten die Klausur geschrieben. Dabei sind 19.4 % Studenten im ersten Versuch und 1.8 % Studenten im zweiten Versuch durchgefallen. Im dritten Versuch haben alle Studenten die Klausur bestanden. Weitere Angaben sind der Tabelle A-43 zu entnehmen. Diese Durchfallquoten sind geringer als die Durchfallquoten der Statistik I. Im Kapitel 2.4 wurde bereits bewiesen, dass die Statistik II – Klausur im Mittel bessere Ergebnisse erzielt als die Statistik I. Dieses Ergebnis ist also plausibel.

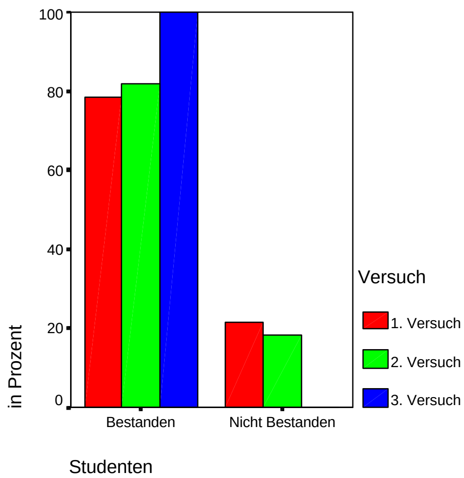


Abbildung 4-26 Balkendiagramm der Durchfallquote nach Anzahl der Versuche in Statistik II

Es wird wiederum die These aufgestellt, dass der Mittelwert der normierten Gesamtpunkte im zweiten Versuch höher ist als der im ersten Versuch. Die Tabelle A-44 bestärkt diese These. Dies wird mit dem Wilcoxon-Test (Tabelle A-45) bestätigt.

Wiederum kann es von Interesse sein, wann ein Student die Klausur ein zweites Mal schreibt. Wird auch hier die Annahme bestätigt, dass es für ihn nicht von Bedeutung ist und gibt es eine Tendenz, darin, dass der Student die Klausur zu nächstmöglichen Termin wiederholt?

In Tabelle A-46 und Abbildung 4-27 erkennt man :

1. 72.86 % der Studenten schreiben den zweiten Versuch zum nächstmöglichen Termin.
2. Es entsteht wieder der Eindruck, dass es besser ist die Klausur so schnell wie möglich zu wiederholen.

Übrigens ein Student hat vier Klausurtermine verstreichen lassen und ist bei seinem Versuch dann durchgefallen.

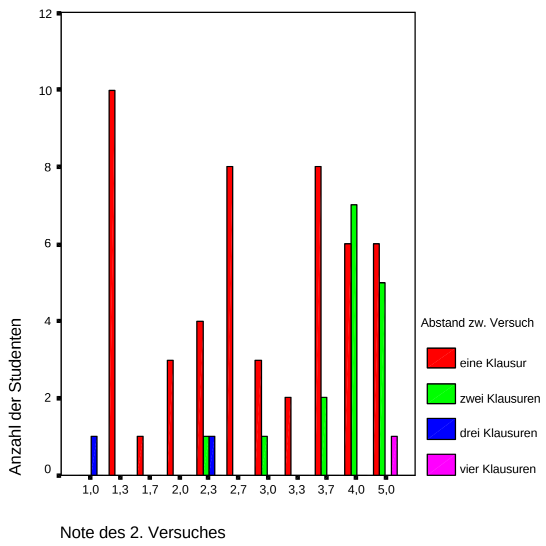


Abbildung 4-27 Balkendiagramm der Klausurnoten nach Anzahl der Versuche in Statistik I

Nach Ablehnung der Normalverteilung (Tabelle A-47), zeigt der Kruskal-Wallis-Test (Tabelle A-48), dass die Nullhypothese nicht aufrechterhalten werden kann. Die Mittelwertunterschiede sind signifikant verschieden.

Anders als in Statistik I ist es in Statistik II also von Bedeutung, wann die Klausur wiederholt wird. Für den Studenten ist es am besten, die Klausur zum nächst möglichen Termin zu wiederholen. Dies wird von den meisten Studenten auch intuitiv gemacht.

4.2.2 Zusammenhang zwischen Statistik I und Statistik II

Kann behauptet werden, dass ein Student der Statistik I sehr gut bestanden hat, diesen Trend in der Statistik II – Klausur fortsetzen kann oder besteht diesbezüglich kein Zusammenhang? Es soll also untersucht werden, ob eine positive Korrelation zwischen den Gesamtpunkten von Statistik I und Statistik II besteht. Dazu wird eine Stichprobe herangezogen, in der alle Studenten erfasst sind, die an beiden Klausuren teilgenommen haben. Nach Kapitel 3.2.2 eignet sich der bivariate Korrelationskoeffizient Kendall's tau_b am besten dafür. Die Korrelation von +0,174 (entsprechend der (Tabelle A-49) ist jedoch nicht signifikant, so dass daraus geschlossen werden kann, dass es keinen Zusammenhang zwischen den Gesamtpunkten der Statistik I und Statistik II gibt. Es kann also sein, dass ein Student, der Statistik I mit sehr gut bestanden hat, Statistik II sehr schlecht besteht und umgekehrt.

Wie ist dies nun zu erklären, dass ein Student in Statistik I mit gut bestehen kann, aber in Statistik II total versagt? Die Statistik II – Klausur wird bei den Studenten zu einem Zeitpunkt geschrieben an dem auch viele andere Klausuren aus anderen Fachrichtungen (wie Ökonometrie) eine große Rolle spielen. Daher könnte es sein, dass die Vorbereitungszeit auf diese Klausuren das Ergebnis der Statistik – Klausur negativ beeinflussen. Weiterhin existiert noch eine Vielzahl von Einflüssen, die eine Klausur beeinflussen kann. Jedoch werden diese Einflüsse im Rahmen dieser Arbeit nicht weiterbehandelt, da sie meistens nicht konkret erfassbar sind.

4.2.3 Alte gegen neue Prüfungsordnung

Nun ist es sicherlich interessant die Durchfallquoten bezüglich alter und neuer Prüfungsordnung mit einander zu vergleichen und eventuelle Rückschlüsse zu ziehen. Als erstes sollte untersucht werden, ob es überhaupt einen Unterschied zwischen den beiden Prüfungsordnungen gibt. Daraus kann dann geschlossen werden, ob es für den Studenten besser ist das Fach Statistik in einer (alte PO) oder in zwei (neue PO) Klausuren abzuschließen.

Die Grafik (Abbildung 4-28) zeigt, dass nach der alten Prüfungsordnung weniger Studenten durchgefallen sind. Auch der Boxplot (Abbildung 4-29) zeigt, dass die Studenten nach alter Prüfungsordnung im Mittel besser gewesen sind (Tabelle A-50).

.

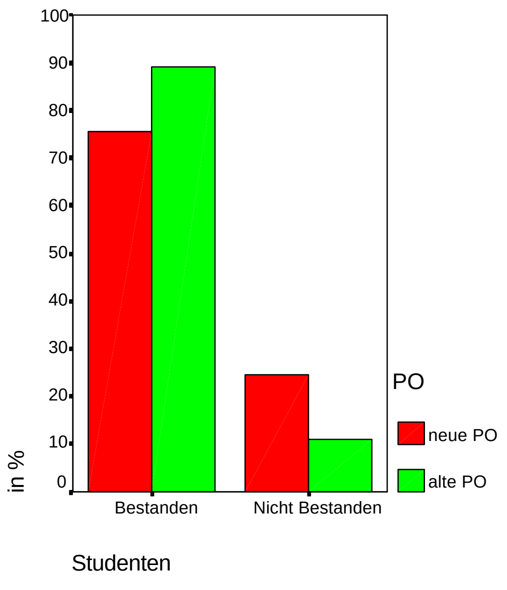


Abbildung 4-28 Balkendiagramm der Durchfallquote nach alter und neuer PO

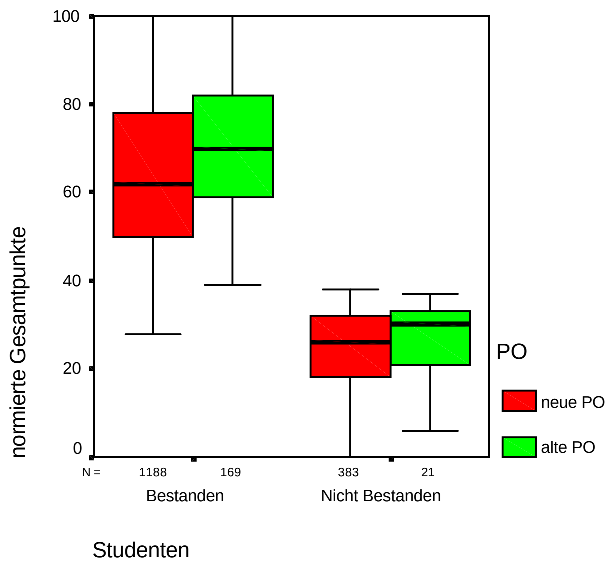


Abbildung 4-29 Boxplot der normierten Gesamtpunkte nach alter und neuer PO

Da die beiden Stichproben unabhängig sind, kann geprüft werden, ob ein parametrischer Test über die Mittelwertunterschiede durchgeführt werden kann. Der Kolmogorov-Smirnov-Test auf Normalverteilung wird abgelehnt (Tabelle A-51*)*. Daher wird auf den nichtparametrischen Mann-Whitney-Test zurückgegriffen. Die Anwendung dieses Tests (Tabelle A-52) zeigt, dass die Nullhypothese verworfen werden muss. Die beobachteten Mittelwertunterschiede sind also nicht rein zufällig. Daher kann behauptet werden, dass Studenten, die nach alter Prüfungsordnung geschrieben haben, im Mittel eine höhere normierte Gesamtpunktzahl erreicht haben und damit auch eine bessere Klausurnote.

Wie kann nun dieses Ergebnis erklärt werden? Nach der alten PO wurde der Bereich Statistik I und Statistik II zusammen geschrieben, es ist bekannt (wie im Kapitel 2.4 gezeigt), dass Statistik II - Klausur von den Studenten besser bestanden wird. So können nach alter PO Schwächen in Statistik I durch die Stärken in Statistik II ausgebessert werden. Nach neuer PO ist dies jedoch nicht mehr möglich.

4.2.4 Punktevergabe bei den Aufgaben

Der Student erhält für eine Klausuraufgabe entweder die volle Punktezahl (bestanden) oder gar keine Punkte (nicht bestanden). Dabei sollte die Punkteverteilung so erfolgen, dass schwere Aufgaben mit höheren Punkten bedacht werden als leichtere. Daraus resultiert, dass Aufgaben mit hohen Punkten von den Studenten schlechter gelöst werden. Dies deutet auf eine negative Korrelation zwischen der Sollpunktezahl einer Aufgabe und dem prozentualen Bestehen einer Aufgabe hin. Ab dem Sommersemester 2002 erfolgte eine neue Systematik in der Punktevergabe. Es wurden Klausuraufgaben gestellt, die mehrere Unteraufgaben enthalten, so dass ein Student die Möglichkeit hat, für eine Klausuraufgabe doch noch Punkte zu erhalten, auch wenn er eine Teilaufgabe nicht richtig beantwortet hat. Diese Aufgaben werden aus der Analyse herausgenommen, da sie das Ergebnis verfälschen könnten. Der Grund ist, dass sich diese Unteraufgaben (Sollpunktezahl 1 bis 5) in ihrer Aufgabenstellung von den bisher gestellten Aufgaben unterscheiden. Sie enthalten kurze Verständnis- bzw. Wissensfragen und bereiteten den Studenten daher Probleme. Der daraus resultierende Kendall’s tau_b Koeffizient von -0,073 zeigt eine geringe negative Korrelation. Es kann ein starker Zusammenhang daher nicht bejaht werden (Tabelle A-53). Bei Berücksichtigung der oben genannten Unteraufgaben würde sich ein r_K von 0,091 (Tabelle A-54) ergeben.

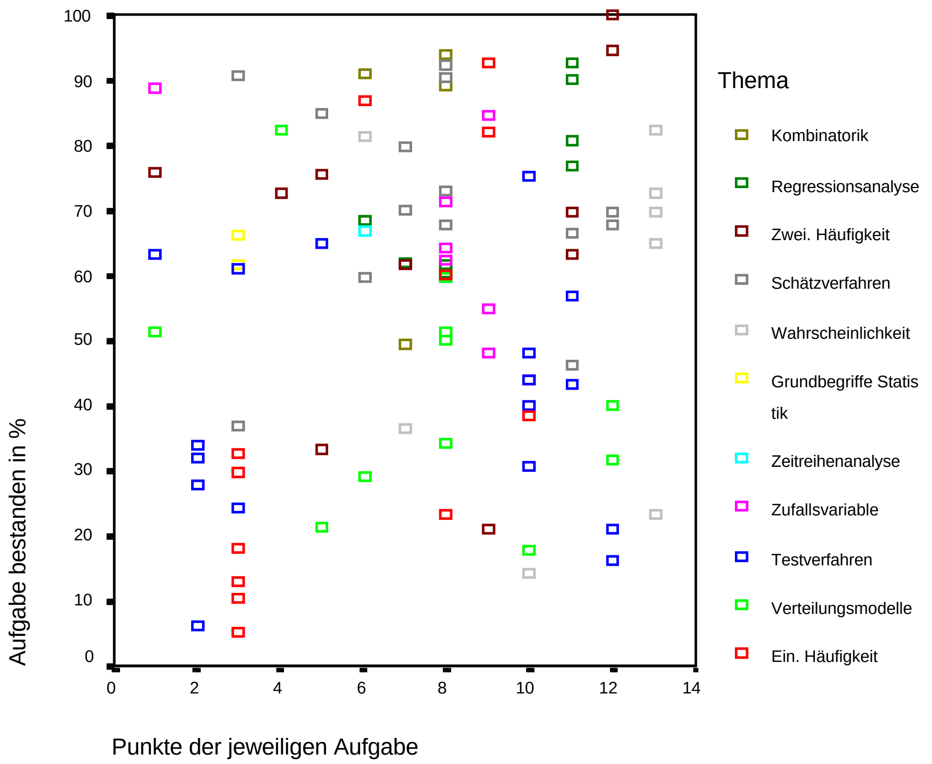


Abbildung 4-30 Scatterplot für bestandene Aufgabe und Sollpunkte nach Thema

Also kann der Annahme, dass Aufgaben mit hohen Punkten bei den Studenten Probleme bereiten, nicht zugestimmt werden. Es ist jedoch so, dass der Aufgabentyp mit hohen Punkten oft großen rechnerischen Aufwand erfordert, der nach einem Schema abgearbeitet werden kann. Dieses Schema muss dann von den Studenten in der Vorbereitungszeit auf die Klausur auswendig gelernt und dann angewendet werden. Klausuraufgaben mit geringeren Punkten verlangen dagegen nicht diesen Aufwand, sondern oft statistisches Verständnis. Daher bereiten diese Aufgabenstellungen den Studenten Probleme.

Außerdem soll eine möglichst homogene Sollpunkteverteilung in den Klausuren den Studenten die Chance geben, die Klausur auch mit „einfachen“ Aufgaben zu bestehen. Denn würden „schwierige“ Aufgaben mit hohen Sollpunkten versehen, könnte ein Student der diese „schwierige“ Aufgabe nicht gelöst hat (zum Beispiel aus Zeitgründen), sehr wohl aber alle „einfachen“ Aufgaben, in der Klausur durchfallen. Dies wäre dann nicht gerecht gegenüber dem Studenten.

4.2.5 Das Thema

In jeder Klausur wurden die Aufgabenstellungen erfasst und einem Thema zugeordnet. Es soll untersucht werden, wie die Punkteverteilung von den Themengebieten abhängt, insbesondere wo die Studenten Probleme haben. Um jedoch dies auswerten zu können, musste zuerst eine Einordnung der Aufgaben in eine Schwierigkeitsstufe (1 – leicht, 2 – mittel, 3 – schwer) erfolgen. Anschließend sollen nur Aufgaben berücksichtigt werden, die einen mittleren Schwierigkeitsgrad haben. Für die Einteilung in die Schwierigkeitsstufen wurden Herr Prof. Dr. Rönz und Herr Dr. Klinke, die an der Aufstellung der Klausuren maßgeblich beteiligt sind und die Vorlesungen Statistik I und Statistik II halten, gebeten, sich die Klausuraufgaben anzusehen und eine Einstufung von leicht bis schwierig zu geben.

Schwierigkeit nach Klinke * Schwierigkeit nach Rönz Crosstabulation

			Schwierigkeit nach Rönz			
			leicht	mittel	schwer	Total
Schwierigkeit nach Klinke	leicht	Count	15	16		31
		% of Total	19,5%	20,8%		40,3%
	mittel	Count	10	18	4	32
		% of Total	13,0%	23,4%	5,2%	41,6%
	schwer	Count	1	5	8	14
		% of Total	1,3%	6,5%	10,4%	18,2%
Total		Count	26	39	12	77
		% of Total	33,8%	50,6%	15,6%	100,0%

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Measure of Agreement	Kappa	,252	,095	3,013	,003
N of Valid Cases		77			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Tabelle 4-4 Kontingenztabelle für die Bewertung der Klausuraufgaben durch Prof. Rönz und Dr. Klinke und Kappa-Koeffizient

In der Tabelle 4-4 ist zu erkennen, dass in 53.3 % der Klausuraufgaben Prof. Rönz und Dr. Klinke übereinstimmen. Daraus ergibt sich dann ein Kappa-Koeffizient von $\kappa = 0.252$. (Die entsprechenden Formeln für Herleitung des Ergebnisses sind dem Kapitel 3.2.3 zu entnehmen.) Somit besteht eine geringfügige Übereinstimmung in der Beurteilung der Klausuraufgaben.

Für die Untersuchung werden nur die Klausuraufgaben herangezogen, die einen mittleren Schwierigkeitsgrad, d.h. zwischen 1.5 und 2.5, besitzen. Bei unterschiedlicher Einteilung der

Klausuraufgaben von Prof. Rönz und Dr. Klinke wurde nämlich das arithmetische Mittel gebildet.

Anhand des Boxplotes (Abbildung 4-31) ist zu erkennen in welchem Themengebiet die Studenten die größten Schwierigkeiten haben. Die Themen „Eindimensionale Häufigkeitsverteilung“ und „Wichtige Verteilungsmodelle“ aus der Statistik I sind offenbar am problematischsten (unter der Beachtung, dass die leichtesten und schwersten Klausuraufgaben herausgenommen worden sind). Dagegen fallen den Studenten die Themen „Regressionsanalyse“ und „Kombinatorik“ am leichtesten.

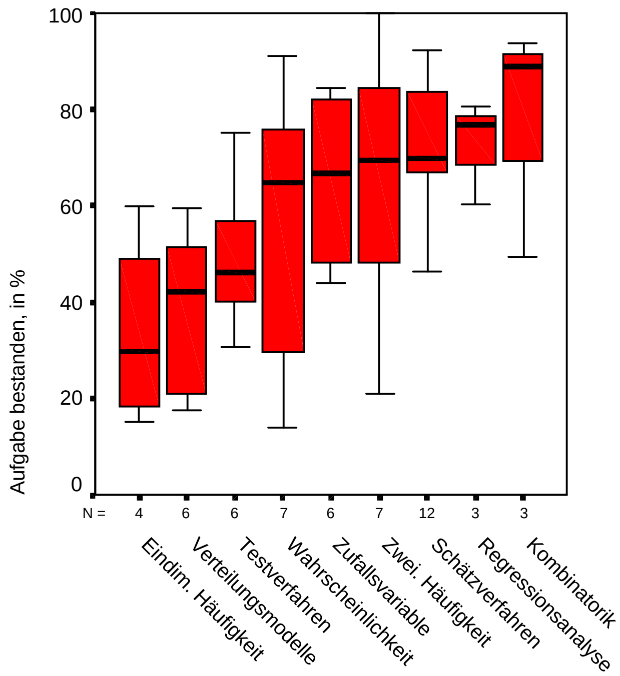


Abbildung 4-31 Boxplot für bestandene Aufgaben nach Thema

Dieses Ergebnis ist nicht weiter verwunderlich, da diese Themen für die Lösung der Aufgaben rein rechnerisches Verständnis verlangt, dass in der Vorbereitungszeit für die Klausur leicht erlernt werden kann. Dagegen erfordern die Themen „Eindimensionale Häufigkeitsverteilung“ und „Wichtige Verteilungsmodelle“ oft ein weitergehendes statistisches Verständnis, welches nicht auswendig gelernt werden kann. Zusätzlich stammen diese Themen aus der Statistik I- Vorlesung, wo ein Student oft das erste Mal auf die Statistik trifft und noch Probleme hat.

4.2.6 Die Clusteranalyse

Im Folgenden soll eine Clusteranalyse getrennt nach Statistik I und Statistik II durchgeführt werden. Ziel ist es die Studenten in mögliche Gruppen einzuteilen und daraus Rückschlüsse zu ziehen. Der Vergleich der Studenten erfolgt dabei über das Bestehen der Aufgaben, nach Themengebieten gegliedert. Dabei werden ähnliche Studenten in einer Gruppe zusammengefasst, d.h. Studenten die die Klausur ähnlich gelöst haben (mit den gleichen gelösten und nicht gelösten Aufgaben). Für die Clusteranalyse wird prozentual erfasst, wie ein Student die Aufgabe des jeweiligen Themengebietes gelöst hat.

4.2.6.1 Die Klausur Statistik I

Für die Untersuchung wurden die Klausuren im Sommersemester 2002 herangezogen, denn in diesen Klausuren wurden von den Studenten zusätzlich das Merkmal „Geschlecht“, „Anzahl der Versuche“, „Studienfach“ und „teilgenommene Übung“ erfragt.
465 Studenten haben die zwei Klausuren (30. Juli und 09. Oktober 2002) mit den folgenden Themengebieten geschrieben :

- „Statistische Grundlagen“,
- „Eindimensionale Häufigkeitsverteilung“,
- „Zweidimensionale Häufigkeitsverteilung“,
- „Wahrscheinlichkeitsrechung“,
- „Zufallsvariable“,
- „Wichtige Verteilungsmodelle“.

Es soll untersucht werden, ob die in Gruppen zusammengefassten Studenten bestimmte Eigenschaften besitzen. Die Frage ist, ob z.B. eine Gruppe gebildet werden kann, in der die Studenten der Studienrichtung „VWL“ dominieren.
Um die Clusterbildung durchführen zu können, werden für alle 465 Studenten die Distanzwerte für alle Paare berechnet, die sich aus den 465 Studenten bilden lassen (wie in Kapitel 3.7 beschrieben). Da die Beobachtungswerte der Themengebiete die gleiche Dimension haben ist eine Transformation nicht notwendig.
Abbildung A-1 zeigt einen Auszug des Dendogramms für die Clusterbildung. Bei der Bildung von zwei bis 25 Clustern bildet ein bestimmter Student immer eine eigene Gruppe. Er wird nie mit anderen Studenten zusammengefasst. Schaut man sich seine Klausur genauer an, so fällt auf, dass dieser eine Student die „Statistischen Grundlagen“, der „Eindimensionale Häufigkeitsverteilung“ und der „Zweidimensionale Häufigkeitsverteilung“ hundertprozentig gelöst hat. Die anderen Themengebiete hat er nicht bestanden. Teilt man nun die restlichen St u-

denten in drei Gruppen ein, so erhält man eine gute, eine mittlere und eine schlechtere Gruppe. Dies ist besonders gut in der Abbildung 4-32 zu sehen.

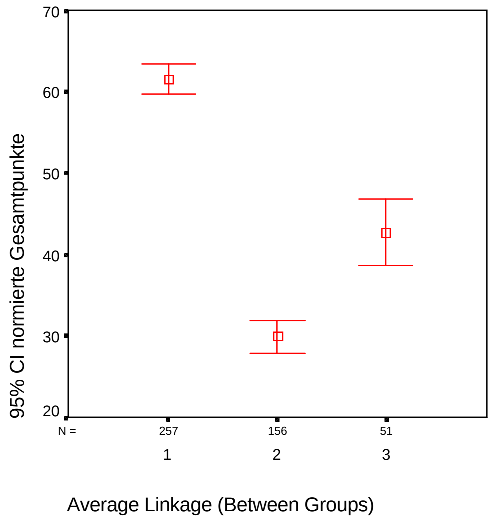


Abbildung 4-32 Fehlerbalkendiagramm der normierten Gesamtpunkten nach den Gruppen 1 bis 3

Die Gruppe 1 hat Probleme in den Gebieten der „Wahrscheinlichkeitsrechnung“ (wenn auch nicht so stark wie die Gruppe 3) und der „Ein- und Zweidimensionalen Häufigkeitsverteilung“ (Abbildung 4-33). Ansonsten wurden von dieser Gruppe alle anderen Themengebiete gut gelöst. Das Themengebiet der „Wichtigen Verteilungsmodelle“ wurde sogar von allen Studenten der Gruppe 1 gelöst.

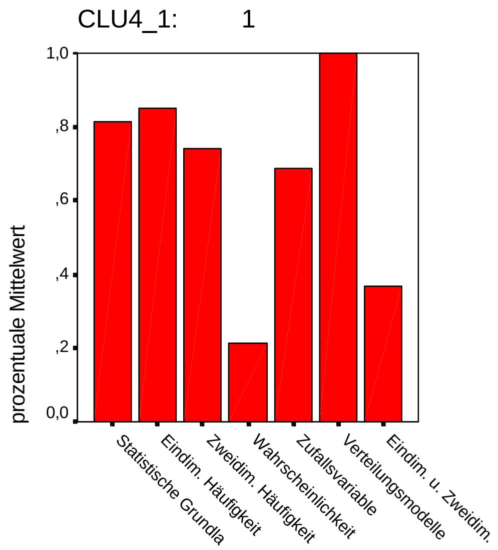


Abbildung 4-33 Balkendiagramm der Gruppe 1 nach Thema

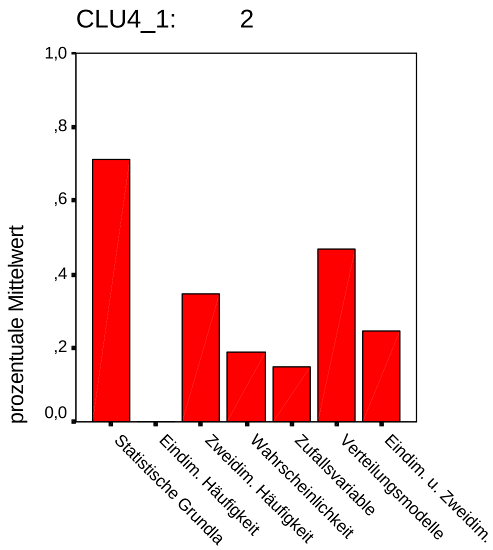


Abbildung 4-34 Balkendiagramm der Gruppe 2 nach Thema

Die Gruppe 2 hat im Gebiet der „Eindimensionalen Häufigkeitsverteilung“ die größten Probleme, denn kein Student dieser Gruppe konnte die Aufgabe lösen. Ansonsten sind dieser

Gruppe noch die „Wahrscheinlichkeitsrechnung“ und die „Zufallvariable“ sehr schwer gefallen. Die restlichen Themengebiete wurden im Mittel jedoch gut gelöst (Abbildung 4-34).
Die Gruppe 3 ist sehr gut in dem Gebiet der „Eindimensionalen Häufigkeitsverteilung“. Starke Probleme bereiten ihnen dagegen die „Wichtigen Verteilungsmodelle“ (im Gegensatz zu Gruppe 1). Weiterhin problematisch ist die „Wahrscheinlichkeitsrechnung“ (Abbildung 4-35).

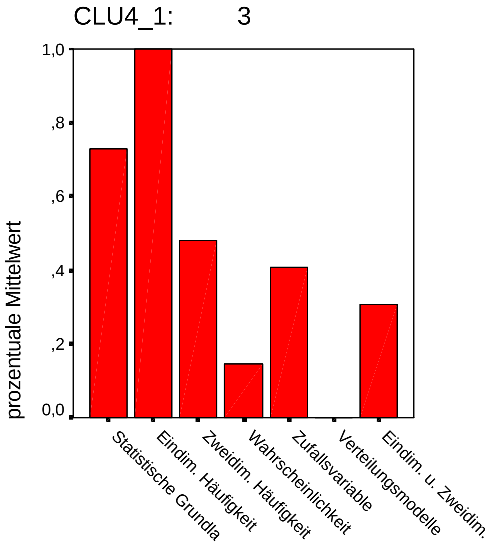


Abbildung 4-35 Balkendiagramm der Gruppe 3 nach Thema

Das Thema „Wahrscheinlichkeitsrechnung“ ist allen drei Gruppen schwer gefallen.
Keine der drei Gruppen wird von einem der oben genannten Merkmale (Geschlecht, Fachrichtung, usw.) dominiert.

4.2.6.2 Die Klausur Statistik II

Diese Untersuchung wird analog zur Statistik I durchgeführt. Die Daten werden aus dem Wintersemester 2002/2003 genommen. An den Klausuren vom 18. Februar und 11. April haben 367 Studenten teilgenommen. Zusätzlich konnte das Geschlecht der Studenten, die Anzahl der Versuche und die Benutzung von MM-Stat erfragt werden.
Die Studenten hatten die Klausuren mit den folgenden Themengebieten zu lösen :

- „Wichtige Verteilungsmodelle“,
- „Statistischen Schätzverfahren“,
- „Statistische Testverfahren“,
- „Regressionsanalyse“,
- „Zeitreihenanalyse“.

Abbildung A-2 zeigt einen Auszug des Dendogramms für die Clusterbildung der Statistik II. Eine Einteilung der Studenten in vier Gruppen hat ergeben, dass die Gruppe 1 (Abbildung

4-36) mit 174 Studenten die Themengebiete „Regressionsanalyse“ und „Zeitreihenanalyse“ am besten beherrscht (alle Studenten dieser Gruppe haben die Aufgaben gelöst). Die Gruppe 2 (Abbildung 4-37) beherrscht dagegen nur die „Zeitreihenanalyse“ sehr gut, hat aber das Thema „Regressionsanalyse“ nicht gelöst. In der Gruppe 3 (Abbildung 4-38) werden diese zwei Themengebiete überhaupt nicht gut gelöst. Die Gruppe 4 (Abbildung 4-39) dagegen hat Probleme in der „Zeitreihenanalyse“ und löst die „Regressionsanalyse“ zu 100 %. Die „Wichtigen Verteilungsmodelle“, „Statistische Schätz- und Testverfahren“ werden von allen vier Gruppen in etwa gleich gut gelöst.

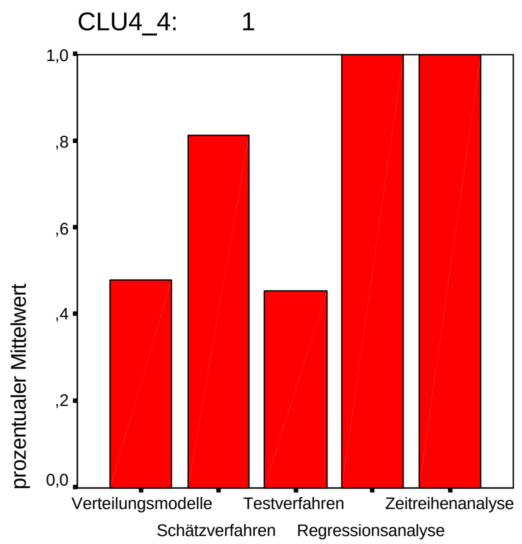


Abbildung 4-36 Balkendiagramm der Gruppe 1 nach Thema

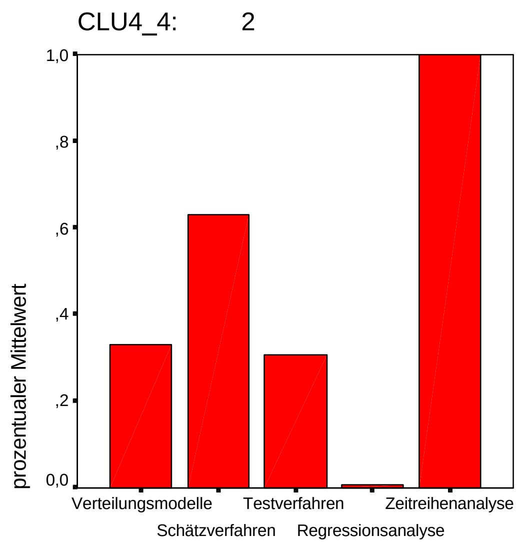


Abbildung 4-37 Balkendiagramm der Gruppe 2 nach Thema

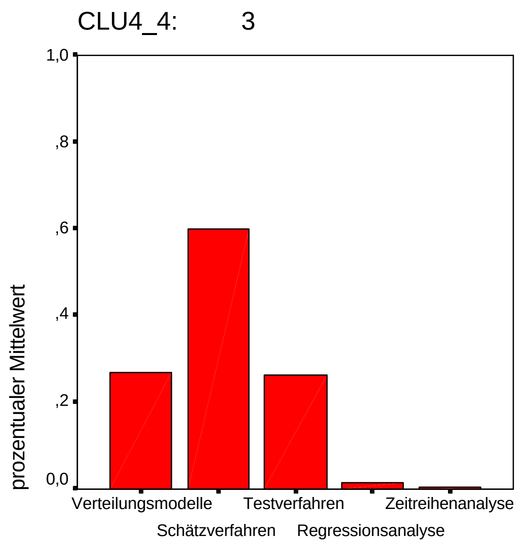


Abbildung 4-38 Balkendiagramm der Gruppe 3 nach Thema

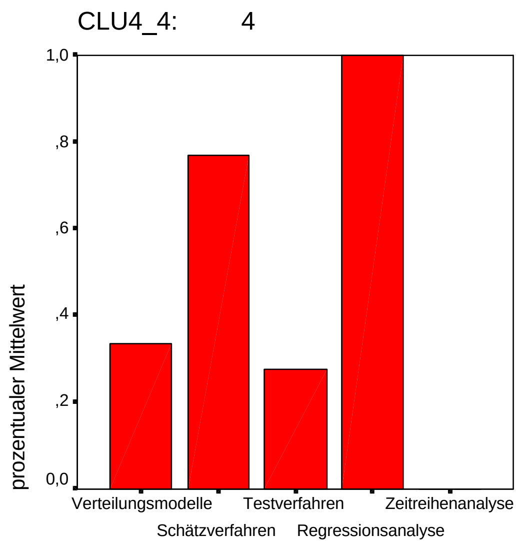


Abbildung 4-39 Balkendiagramm der Gruppe 4 nach Thema

Eine Unterscheidung der Gruppen nach Geschlecht, Anzahl der Versuche bzw. ob MM-Stat benutzt wurde, ist nicht erkennbar.

4.3 Zusammenfassung

Rückblickend stellen sich folgende Ergebnisse dar:

1. Die Statistik II – Klausur wird von den Studenten im Mittel besser geschrieben als die Statistik I – Klausur.
2. Das Geschlecht hat einen Einfluss auf das Klausurergebnis der Statistik I und der Statistik II. Beide Klausuren werden im Mittel von Männern besser geschrieben.
3. Der Besuch der Übung 2 hat sich positiv auf das Klausurergebnis ausgewirkt.
4. Für das Bestehen der Klausur spielt die Studienrichtung des Studenten keine Rolle.
5. Die unter Punkt 1 gestellte Aussage kann speziell für die Fachrichtung VWL und BWL aufrechterhalten werden, jedoch nicht für die Studenten die aus sonstigen Fachrichtungen stammen.
6. Die Benutzung von MM-Stat hat weder bei den Frauen noch bei den Männern einen Einfluss auf das Klausurergebnis.
7. Im ersten Versuch fallen in Statistik I 23.9 % der Studenten durch, in Statistik II nur noch 19.4 % der teilnehmenden Studenten.
8. Die sofortige Wiederholung der Klausur zum nächstmöglichen Termin hat bei der Statistik I – Klausur keinen Einfluss auf das Klausurergebnis. Jedoch bei der Statistik II – Klausur die Folge, dass zum nächstmöglichen Termin im Mittel eine bessere Klausurnote geschrieben wird.
9. Es besteht kein linearer Zusammenhang zwischen dem Klausurergebnis der Statistik I – Klausur und dem der Statistik II – Klausur.
10. Die Studenten nach alter Prüfungsordnung haben im Mittel ein besseres Klausurergebnis erzielt.
11. Der Zusammenhang zwischen Punktevergabe und Schwierigkeitsgrad der Aufgaben ist sehr gering.
12. Die Themen „Eindimensionale Häufigkeitsverteilung“, „Wichtige Verteilungsmodelle“ und „Statistische Testverfahren“ bereiten den Studenten die größten Probleme. Ohne große Schwierigkeiten werden dagegen die „Regressionsanalyse“ und die „Kombinatorik“ gelöst.
13. Die Clusteranalyse teilt die Studenten in Gruppen, die in Statistik I in den Themen „Eindimensionaler Häufigkeitsverteilung“ und in „Wichtige Verteilungsmodelle“

Probleme haben. In der Statistik II werden den Studenten in Gruppen geteilt, die die „Regressionsanalyse“ und „Zeitreihenanalyse“ zu 100 % beherrschen oder nur eine der beiden Themen oder beide Themen nicht.

5 Schlussfolgerungen

Welche Konsequenzen können nun aus den Ergebnissen des zweiten und dritten Kapitels für zukünftige Statistikveranstaltungen gezogen werden?

1. Frauen sollten verstärkt gefördert werden. Vielleicht sollte eine Übung angeboten werden, die nur für Frauen vorgesehen ist.
2. Es sollte darauf geachtet werden, dass die Übungsleiter didaktische Erfahrungen haben.
3. Die Möglichkeit zur Wiederholung der Klausur im zweiten Prüfungsabschnitt sollte unbedingt beibehalten werden (gerade für die Statistik II – Klausur).
4. Die Vergabe der Sollpunkte in den Klausuren sollte überdacht werden, insbesondere ob nicht eine Gliederung in mehrere Unterpunkte angebracht ist, die gesondert bewertet werden. Stattdessen ist auf die Stellung von Einzelaufgaben mit hohen Punktzahlen zu verzichten.
5. Die Themen „Eindimensionale Häufigkeitsverteilung“, „Wichtige Verteilungsmodelle“ und „Statistische Testverfahren“ sollten in der Vorlesung und in den Übungen besondere Beachtung finden und eventuell noch vertieft werden.

Diese Anregungen können durchaus noch vertieft werden, dies sollte jedoch Thema einer weiteren Diplomarbeit sein.

Literaturverzeichnis

Bohley, P. (1996). *Statistik – Einführendes Lehrbuch für Wirtschafts- und Sozialwissenschaftler*, R. Oldenbourg Verlag, München

Bortz, J. (1999). *Statistik für Sozialwissenschaftler*, Springer Verlag

Brosius, F. (1998). *SPSS 8*, International Thomson Publishing

Brunner, E. und Munzel, U. (2002). *Nichtparametrische Datenanalyse – Unverbundene Stichproben*, Springer Verlag

Büning, H. und Trenkler, G. (1994). *Nichtparametrische statistische Methoden*, Walter de Gruyter

Dür, M. (2002). *Vorlesung und Proseminar aus Mathematik / Statistik 2*, Vorlesungsskript, Abteilung für mathematische Methoden der Statistik, Institut für Statistik, Wirtschaftsuniversität Wien

Eckstein, H. (1997). *Angewandte Statistik mit SPSS – Praktische Einführung für Wirtschaftswissenschaftler*, Gabler Wiesbaden

Hartung, J. (1999). *Statistik – Lehr- und Handbuch der angewandten Statistik*, R. Oldenbourg Verlag, München

Rönz, B. (2001). *Skript - Computergestützte Statistik I*, Vorlesungsskript, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin

Rönz, B. (2000). *Skript - Computergestützte Statistik II*, Vorlesungsskript, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin

Sachs, L. (2003). *Angewandte Statistik – Anwendung statistischer Methoden*, Springer Verlag

Zech, F. (1996). *Grundkurs Mathematikdidaktik – Theoretische und praktische Anleitung für das Lehren und Lernen von Mathematik*, Beltz Verlag, Weinheim und Basel

verwendete Statistiksoftwaresysteme :

SPSS 11.0.1 for Windows (SPSS Inc.)

Internetadressen :

www.md-stat.com

www.wiwi.uni-hannover.de/statistik

A Anhang

Tabelle A-1Zusammenhang zwischen der Klausurnote und den normierten Gesamtpunkten

Klausurnote	normierte Gesamtpunkte
1,0	98 - 100
1,3	90 - 97
1,7	86 - 89
2,0	80 - 85
2,3	74 - 79
2,7	70 - 73
3,0	66 - 69
3,3	60 - 65
3,7	52 - 59
4,0	40 - 51
5,0	0 - 39

Tabelle A-2 Übersicht der Statistikklausuren

Report

normierte Gesamtpunkte auf 100

Art der Klausur	Klausur	N	Median	Variance	Std. Deviation
Statistik I	22.02.2001	11	96,0000	416,873	20,41746
	09.04.2001	56	80,0000	379,781	19,48796
	26.07.2001	187	52,0000	535,758	23,14645
	28.09.2001	134	44,0000	448,939	21,18818
	30.07.2002	273	56,0000	303,445	17,41966
	09.10.2002	192	38,0000	429,265	20,71872
	Total	853	50,0000	505,233	22,47739
Stastistik II	22.02.2001	9	56,0000	422,444	20,55345
	09.04.2001	38	58,0000	634,492	25,18913
	21.02.2002	141	58,0000	819,548	28,62775
	08.04.2002	164	74,0000	577,295	24,02697
	18.02.2003	222	58,0000	407,581	20,18863
	11.04.2003	144	54,0000	446,263	21,12493
	Total	718	60,0000	575,874	23,99737
Statistik I +II	22.02.2001	54	73,0000	335,195	18,30833
	09.04.2001	109	66,0000	415,697	20,38864
	08.04.2002	13	55,0000	246,077	15,68684
	Total	176	68,0000	409,558	20,23755

Tabelle A-3 Kolmogorov-Smirnov-Test nach Klausurtyp

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Art der Klausur	Statistic	df	Sig.	Statistic	df	Sig.
normierte Gesamtpunkte auf 100	Statistik I	,041	853	,002	,985	853	,000
	Stastistik II	,080	718	,000	,975	718	,000
	Statistik I +II	,075	176	,018	,972	176	,001

a. Lilliefors Significance Correction

Tabelle A-4 Mann-Whitney-Test nach Klausur typ

Ranks

	Art der Klausur	N	Mean Rank	Sum of Ranks
normierte Gesamtpunkte auf 100	Statistik I	853	713,84	608908,98
	Stastistik II	718	871,72	625897,00
	Total	1571		

Test Statistics[a]

	normierte Gesamtpunkte auf 100
Mann-Whitney U	244678,000
Wilcoxon W	608909,000
Z	-6,873
Asymp. Sig. (2-tailed)	,000

a. Grouping Variable: Art der Klausur

Tabelle A-5 Kruskal-Wallis-Test für Statistik I

Ranks

	normierte Klausur	N	Mean Rank
normierte Gesamtpunkte auf 100	Statistik I 22.2.01	11	736,41
	Statistik I 9.4.01	56	699,50
	Statistik I 26.7.01	187	428,06
	Statistik I 28.9.01	134	389,29
	Statistik I 30.7.02	273	463,31
	Statistik I 9.10.02	192	303,46
	Total	853	

Test Statistics [a,b]

	normierte Gesamtpunkte auf 100
Chi-Square	143,247
df	5
Asymp. Sig.	,000

a. Kruskal Wallis Test

b. Grouping Variable: normierte Klausur

Tabelle A-6 Kruskal-Wallis-Test für Statistik II

Ranks

	normierte Klausur	N	Mean Rank
normierte Gesamtpunkte auf 100	Statistik II 22.2.01	9	375,17
	Statistik II 9.4.01	38	320,49
	Statistik II 21.2.02	141	355,13
	Statistik II 8.4.02	164	446,32
	Statistik II 18.2.03	222	342,86
	Statistik II 11.4.03	144	299,87
	Total	718	

Test Statistics [a,b]

	normierte Gesamtpunkte auf 100
Chi-Square	43,564
df	5
Asymp. Sig.	,000

a. Kruskal Wallis Test

b. Grouping Variable: normierte Klausur

Tabelle A-7 Kruskal-Wallis-Test für Statistik nach alter PO

Ranks

	normierte Klausur	N	Mean Rank
normierte Gesamtpunkte auf 100	Statistik I + II 22.2.01	55	105,35
	Statistik I + II 9.4.01	108	84,70
	Statistik I + II 8.4.02	13	48,81
	Total	176	

Test Statistics[a,b]

	normierte Gesamtpunkte auf 100
Chi-Square	14,512
df	2
Asymp. Sig.	,001

a. Kruskal Wallis Test

b. Grouping Variable: normierte Klausur

Tabelle A-8 Kolmogorov-Smirnov-Test nach Geschlecht für Statistik I

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Geschlecht	Statistic	df	Sig.	Statistic	df	Sig.
Gesamtpunkte	männlich	,066	218	,023	,989	218	,098
	weiblich	,031	239	,200*	,986	239	,018

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Tabelle A-9 Mann-Whitney-Test nach Geschlecht für Statistik I

Ranks

	Geschlecht	N	Mean Rank	Sum of Ranks
normierte Gesamtpunkte	männlich	218	246,53	53744,50
	weiblich	239	213,01	50908,50
	Total	457		

Test Statistics[a]

	normierte Gesamtpunkte
Mann-Whitney U	22228,500
Wilcoxon W	50908,500
Z	-2,711
Asymp. Sig. (2-tailed)	,007

a. Grouping Variable: Geschlecht

Tabelle A-10 Übersicht normierte Gesamtpunkte nach Geschlecht für Statistik I

normierte Gesamtpunkte

Geschlecht	Mean	N	Std. Deviation	Minimum	Maximum	% of Total N
männlich	51,04	218	18,827	9	100	47,7%
weiblich	46,08	239	20,693	6	98	52,3%

Tabelle A-11 Übersicht normierte Gesamtpunkte nach Geschlecht für Statistik II

Report

normierte Gesamtpunkte

Geschlecht	Mean	N	Std. Deviation	Minimum	Maximum	% of Total N
männlich	58,27	184	21,703	0	98	51,3%
weiblich	54,17	175	19,241	0	92	48,7%

Tabelle A-12 Kolmogorov-Smirnov-Test nach Geschlecht für Statistik II

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Geschlecht	Statistic	df	Sig.	Statistic	df	Sig.
Gesamtpunkte	männlich	,086	184	,002	,974	184	,001
	weiblich	,071	175	,029	,986	175	,090

a. Lilliefors Significance Correction

Tabelle A-13 Mann-Whitney-Test nach Geschlecht für Statistik II

Ranks

	Geschlecht	N	Mean Rank	Sum of Ranks
normierte Gesamtpunkte	männlich	184	190,68	35086,00
	weiblich	175	168,77	29534,00
	Total	359		

Test Statistics[a]

	normierte Gesamtpunkte
Mann-Whitney U	14134,000
Wilcoxon W	29534,000
Z	-2,001
Asymp. Sig. (2-tailed)	,045

a. Grouping Variable: Geschlecht

Tabelle A-14 Kolmogorov-Smirnov-Test bei Übungsteilnahme

Tests of Normality

	Vergleich Teilnahme an Übung und keine	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
normierte Gesamtpunkte	Teilnahme an einer Übung	,047	419	,027	,990	419	,004
	k.A.	,061	48	,200*	,983	48	,728

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Tabelle A-15 Mann-Whitney Test auf Übungsteilnahme

Ranks

	Vergleich Teilnahme	N	Mean Rank	Sum of Ranks
normierte Gesamtpunkte	Teilnahme an einer Übung	419	234,42	98221,00
	k.A.	48	230,35	11057,00
	Total	467		

Test Statistics[a]

	normierte Gesamtpunkte
Mann-Whitney U	9881,000
Wilcoxon W	11057,000
Z	-,198
Asymp. Sig. (2-tailed)	,843

a. Grouping Variable: Vergleich Teilnahme an Übung und keine Teilnahme

Tabelle A-16 Kolmogorov-Smirnov-Test nach Übung

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Übung	Statistic	df	Sig.	Statistic	df	Sig.
normierte Gesamtpunkte	Übung 1	,059	127	,200*	,989	127	,435
	Übung 2	,060	160	,200*	,985	160	,081
	Übung 3	,091	132	,010	,974	132	,013

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Tabelle A-17 Kruskal-Wallis-Test nach Übung

Ranks

	Übung	N	Mean Rank
normierte Gesamtpunkte	Übung 1	127	189,09
	Übung 2	160	227,20
	Übung 3	132	209,26
	Total	419	

Test Statistics[a,b]

	normierte Gesamtpunkte
Chi-Square	7,021
df	2
Asymp. Sig.	,030

a. Kruskal Wallis Test

b. Grouping Variable: Übung

Tabelle A-18 Übersicht der Mittelwerte nach Geschlecht und Übung

Report

normierte Gesamtpunkte

Geschlecht	Übung	Mean	N	Std. Deviation
männlich	Übung 1	46,02	57	17,303
	Übung 2	56,07	85	18,041
	Übung 3	49,50	50	21,150
	Total	51,38	192	19,105
weiblich	Übung 1	44,56	68	19,377
	Übung 2	44,71	73	18,611
	Übung 3	47,35	81	23,764
	Total	45,63	222	20,804
Total	Übung 1	45,22	125	18,402
	Übung 2	50,82	158	19,112
	Übung 3	48,17	131	22,743
	Total	48,29	414	20,215

Tabelle A-19 Kolmogorov-Smirnov-Test nach Studienrichtung

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Fach	Statistic	df	Sig.	Statistic	df	Sig.
normierte Gesamtpunkte	BWL	,070	196	,022	,986	196	,052
	VWL	,064	162	,200*	,984	162	,066
	Sonst.	,083	103	,075	,978	103	,091

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Tabelle A-20 Kruskal-Wallis-Test nach Studienrichtung

Ranks

	Fach	N	Mean Rank
normierte Gesamtpunkte	BWL	196	244,65
	VWL	162	216,02
	Sonst.	103	228,59
	Total	461	

Test Statistics[a,b]

	normierte Gesamtpunkte
Chi-Square	4,142
df	2
Asymp. Sig.	,126

a. Kruskal Wallis Test

b. Grouping Variable: Fach

Tabelle A-21 Mittelwerte normierte Gesamtpunke nach Studienrichtung und Geschlecht

Report

normierte Gesamtpunkte

Fach	Geschlecht	Mean	N
BWL	männlich	54,39	87
	weiblich	46,88	106
	Total	50,26	193
VWL	männlich	48,20	108
	weiblich	42,42	52
	Total	46,33	160
Sonst.	männlich	50,27	22
	weiblich	47,05	80
	Total	47,75	102
Total	männlich	50,89	217
	weiblich	45,96	238
	Total	48,31	455

Tabelle A-22 Kolmogorov-Smirnov-Test bei BWL-Studenten nach Geschlecht

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Geschlecht	Statistic	df	Sig.	Statistic	df	Sig.
normierte Gesamtpunkte	männlich	,107	87	,016	,981	87	,237
	weiblich	,080	106	,088	,982	106	,161

a. Lilliefors Significance Correction

Tabelle A-23 Mann-Whitney-Test bei BWL-Studenten nach Geschlecht

Ranks

	Geschlecht	N	Mean Rank	Sum of Ranks
normierte Gesamtpunkte	männlich	87	110,08	9577,00
	weiblich	106	86,26	9144,00
	Total	193		

Test Statistics[a]

	normierte Gesamtpunkte
Mann-Whitney U	3473,000
Wilcoxon W	9144,000
Z	-2,948
Asymp. Sig. (2-tailed)	,003

a. Grouping Variable: Geschlecht

Tabelle A-24 Kolmogorov-Smirnov-Test bei VWL-Studenten nach Geschlecht

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Geschlecht	Statistic	df	Sig.	Statistic	df	Sig.
normierte Gesamtpunkte	männlich	,081	108	,079	,978	108	,068
	weiblich	,085	52	,200*	,973	52	,291

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Tabelle A-25 Levené-Test und t-Test bei VWL-Studenten nach Geschlecht

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
normierte Gesamtpunkt	Equal variances assumed	,126	,723	1,750	158	,082	5,78	3,303	-,743	12,304
	Equal variances not assumed			1,696	93,088	,093	5,78	3,409	-,989	12,551

Tabelle A-26 Kolmogorov-Smirnov-Test bei sonst. Studenten nach Geschlecht

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Geschlecht	Statistic	df	Sig.	Statistic	df	Sig.
normierte Gesamtpunkte	männlich	,099	22	,200*	,967	22	,643
	weiblich	,090	80	,168	,974	80	,100

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Tabelle A-27 Levené-Test und t-Test bei sonst. Studenten nach Geschlecht

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
normierte Gesamtpunkte	Equal variances assumed	,599	,441	,610	100	,543	3,22	5,279	-7,251	13,696
	Equal variances not assumed			,647	36,428	,522	3,22	4,981	-6,876	13,322

Tabelle A-28 Kolmogorov-Smirnov-Test bei MM-Stat nach Benutzung

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	MM-Stat	Statistic	df	Sig.	Statistic	df	Sig.
normierte Gesamtpunkte	ja	,082	207	,002	,977	207	,002
	nein	,080	130	,039	,983	130	,109
	k.A.	,105	31	,200*	,968	31	,464

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Tabelle A-29 Kruskal-Wallis-Test bei MM-Stat nach Benutzung

Ranks

	MM-Stat	N	Mean Rank
normierte Gesamtpunkte	ja	207	190,01
	nein	130	186,25
	k.A.	31	140,40
	Total	368	

Test Statistics[a,b]

	normierte Gesamtpunkte
Chi-Square	5,922
df	2
Asymp. Sig.	,052

a. Kruskal Wallis Test

b. Grouping Variable: MM-Stat

Tabelle A-30 Kolmogorov-Smirnov-Test bei MM-Stat für Frauen

Tests of Normality

	MM-Stat	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
normierte Gesamtpunkte	ja	,085	96	,087	,971	96	,031
	nein	,084	63	,200*	,982	63	,474
	k.A.	,120	16	,200*	,967	16	,793

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Tabelle A-31 Levené-Test bei MM-Stat für Frauen

Test of Homogeneity of Variances

normierte Gesamtpunkte

Levene Statistic	df1	df2	Sig.
,945	2	172	,391

Tabelle A-32 ANOVA bei MM-Stat für Frauen

ANOVA

normierte Gesamtpunkte

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	517,857	2	258,929	,697	,499
Within Groups	63897,000	172	371,494		
Total	64414,857	174			

Tabelle A-33 Kolmogorov-Smirnov-Test bei MM-Stat für Männer

Tests of Normality

	MM-Stat	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
normierte Gesamtpunkte	ja	,108	108	,003	,962	108	,004
	nein	,115	62	,039	,959	62	,036
	k.A.	,223	14	,057	,896	14	,099

a. Lilliefors Significance Correction

Tabelle A-34 Kruskal-Wallis-Test bei MM-Stat für Männer

Ranks

	MM-Stat	N	Mean Rank
normierte Gesamtpunkte	ja	108	95,88
	nein	62	93,15
	k.A.	14	63,50
	Total	184	

Test Statistics[a,b]

	normierte Gesamtpunkte
Chi-Square	4,601
df	2
Asymp. Sig.	,100

a. Kruskal Wallis Test

b. Grouping Variable: MM-Stat

Tabelle A-35 Zusammenhang zwischen Statistik I - und Statistik II - Klausurnoten

Note aus Statistik I * Note aus Statistik II Crosstabulation

Count

		Note aus Statistik II												
		1,0	1,3	1,7	2,0	2,3	2,4	2,7	3,0	3,3	3,7	4,0	5,0	Total
Note aus Statistik I	1,0	4	2	1	1	1	1	4	1	1	2	1		19
	1,3	1	7	1	2	6		2	2		4		3	28
	1,7		1	2		4		2			1		1	11
	2,0	2	2	2	1	3		3			3	1		17
	2,3	2	3	2	1	3		3	4	2	5	4		29
	2,7		3	1	3	4		5	2	1	5	5	2	31
	3,0		4	3	3	6		2	4	3	4	3	3	35
	3,3		3	1	8	6		7	3	5	2	11	2	48
	3,7	1	8	4	5	4		7	3	3	13	9	6	63
	4,0	4	7	4	4	14		6	7	2	12	17	16	93
	5,0	4	12	2	1	10		6	2	5	10	19	19	90
Total		18	52	23	29	61	1	47	28	22	61	70	52	464

Tabelle A-36 Übersicht Klausurnote und Anzahl der Versuche in Statistik I

Klausurnote * versuch Crosstabulation

			versuch			Total
			1. Versuch	2. Versuch	3. Versuch	
Klausurnote	1,0	% within versuch	3,5%	2,2%		3,3%
	1,3	% within versuch	4,5%	2,2%		4,2%
	1,7	% within versuch	2,7%	2,2%		2,6%
	2,0	% within versuch	3,2%	1,1%		2,9%
	2,3	% within versuch	5,2%	4,3%		5,0%
	2,7	% within versuch	4,9%	4,3%	11,1%	4,9%
	3,0	% within versuch	6,3%	8,6%		6,4%
	3,3	% within versuch	9,6%	11,8%		9,7%
	3,7	% within versuch	12,1%	20,4%	22,2%	13,1%
	4,0	% within versuch	20,9%	17,2%	22,2%	20,5%
	5,0	% within versuch	27,2%	25,8%	44,4%	27,2%
Total		% within versuch	100,0%	100,0%	100,0%	100,0%
		% of Total	88,0%	10,9%	1,1%	100,0%

Tabelle A-37 Mittelwert der normierten Gesamtpunkte nach Versuchfür Statistik I

Descriptive Statistics

	N	Mean	Std. Deviation
Gesamtpunkte des 1. Versuches	87	26,21	8,220
Gesamtpunkte des 2. Versuches	87	50,66	18,031
Valid N (listwise)	87		

Tabelle A-38 Wilcoxon Signed Ranks Test nach Versuche für Statistik I

Ranks

		N	Mean Rank	Sum of Ranks
Gesamtpunkte des 2. Versuches - Gesamtpunkte des 1. Versuches	Negative Ranks	3 [a]	5,50	16,50
	Positive Ranks	84 [b]	45,38	3811,50
	Ties	0 [c]		
	Total	87		

a. Gesamtpunkte des 2. Versuches < Gesamtpunkte des 1. Versuches

b. Gesamtpunkte des 2. Versuches > Gesamtpunkte des 1. Versuches

c. Gesamtpunkte des 1. Versuches = Gesamtpunkte des 2. Versuches

Test Statistics[b]

	Gesamtpunkte des 2. Versuches - Gesamtpunkte des 1. Versuches
Z	-8,032[a]
Asymp. Sig. (2-tailed)	,000

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Tabelle A-39 Mittelwerte des 2. Versuches nach Abstand zwischen den Versuchen in Statistik I

Report

Gesamtpunkte des 2. Versuches

Abstand zwischen	Mean	N	Std. Deviation
eine Klausur	51,16	57	19,055
zwei Klausuren	52,09	23	14,441
drei Klausuren	41,86	7	20,145
Total	50,66	87	18,031

Tabelle A-40 Kolmogorov-Smirnov-Test bei den Gesamtpunkten des 2. Versuches nach den Zeitabständen zwischen den Klausuren in Statistik I

Tests of Normality

	Abstand zwischen Versuchen	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Gesamtpunkte des 2. Versuches	eine Klausur	,088	57	,200*	,964	57	,084
	zwei Klausuren	,123	23	,200*	,960	23	,461
	drei Klausuren	,228	7	,200*	,915	7	,429

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Tabelle A-41 Test auf Varianzhomogenität bei Gesamtpunkte des 2. Versuches nach Abstand zwischen den Versuchen in Statistik I

Test of Homogeneity of Variances

Gesamtpunkte des 2. Versuches

Levene Statistic	df1	df2	Sig.
1,645	2	84	,199

Tabelle A-42 ANOVA auf Gesamtpunkte des 2. Versuches nach Abstand zwischen den Versuchen in Statistik I

ANOVA

Gesamtpunkte des 2. Versuches

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	603,393	2	301,696	,926	,400
Within Groups	27356,262	84	325,670		
Total	27959,655	86			

Tabelle A-43 Übersicht der Klausurnote und Anzahl der Versuch in Statistik II

Klausurnote * versuch Crosstabulation

			versuch			
			1. Versuch	2. Versuch	3. Versuch	Total
Klausurnote	1,0	% within versuch	3,3%	1,4%		3,1%
	1,3	% within versuch	8,6%	15,3%		9,2%
	1,7	% within versuch	4,7%	2,8%		4,5%
	2,0	% within versuch	5,6%	4,2%		5,4%
	2,3	% within versuch	11,2%	6,9%	25,0%	10,9%
	2,7	% within versuch	7,6%	11,1%		7,9%
	3,0	% within versuch	4,8%	5,6%		4,9%
	3,3	% within versuch	5,5%	2,8%		5,2%
	3,7	% within versuch	12,5%	13,9%		12,5%
	4,0	% within versuch	11,1%	18,1%	75,0%	15,3%
	5,0	% within versuch	21,7%	18,1%		21,2%
Total		% within versuch	100,0%	100,0%	100,0%	100,0%
		% of Total	89,4%	10,0%	,6%	100,0%

Tabelle A-44 Mittelwerte Gesamtpunkte der 70 Studenten im 1. und 2. Versuch in Statistik II

Descriptive Statistics

	N	Mean	Std. Deviation
Gesamtpunkte des 1. Versuches	70	23,49	11,725
Gesamtpunkte des 2. Versuches	70	61,69	20,948
Valid N (listwise)	70		

Tabelle A-45 Wilcoxon Signed Ranks Test nach Versuche für Statistik II

Ranks

		N	Mean Rank	Sum of Ranks
Gesamtpunkte des 2. Versuches - Gesamtpunkte des 1. Versuches	Negative Ranks	0[a]	,00	,00
	Positive Ranks	70[b]	35,50	2485,00
	Ties	0[c]		
	Total	70		

a. Gesamtpunkte des 2. Versuches < Gesamtpunkte des 1. Versuches

b. Gesamtpunkte des 2. Versuches > Gesamtpunkte des 1. Versuches

c. Gesamtpunkte des 1. Versuches = Gesamtpunkte des 2. Versuches

Test Statistics[b]

	Gesamtpunkte des 2. Versuches - Gesamtpunkte des 1. Versuches
Z	-7,273[a]
Asymp. Sig. (2-tailed)	,000

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Tabelle A-46 Mittelwerte der Gesamtpunkte des 2. Versuches nach Abstand zwischen den Versuchen in Statistik II

Report

Gesamtpunkte des 2. Versuches

Abstand zw. Versuch	Mean	N	Std. Deviation
eine Klausur	66,24	51	20,070
zwei Klausuren	45,75	16	13,244
drei Klausuren	87,00	2	18,385
vier Klausuren	34,00	1	,
Total	61,69	70	20,948

Tabelle A-47 Kolmogorov-Smirnov-Test der Gesamtpunkte des 2. Versuchen nach Abstand zwischen den Versuchen in Statistik II

Tests of Normality[b]

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Abstand zw. Versuch	Statistic	df	Sig.	Statistic	df	Sig.
Gesamtpunkte des 2. Versuches	eine Klausur	,104	51	,200*	,957	51	,059
	zwei Klausuren	,131	16	,200*	,939	16	,342
	drei Klausuren	,260	2	,			

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

b. Gesamtpunkte des 2. Versuches is constant when Abstand zw. Versuch = vier Klausuren. It has been omitted.

Tabelle A-48 Kruskal-Wallis-Test der Gesamtpunkte des 2. Versuches nach Abstand zwischen den Versuchen in Statistik II

Ranks

	Abstand zw. Versuch	N	Mean Rank
Gesamtpunkte des 2. Versuches	eine Klausur	51	40,13
	zwei Klausuren	16	19,44
	drei Klausuren	2	60,00
	vier Klausuren	1	7,50
	Total	70	

Test Statistics[a,b]

	Gesamtpunkte des 2. Versuches
Chi-Square	17,437
df	3
Asymp. Sig.	,001

a. Kruskal Wallis Test

b. Grouping Variable: Abstand zw. Versuch

Tabelle A-49 Zusammenhang nach Kendall's tau_b für normierte Gesamtpunkte der Statistik I und der Statistik II

Correlations

			Gesamtpunkte Statstik I	Gesamtpunkte Statistik II
Kendall's tau_b	Gesamtpuntke Statstik I	Correlation Coefficient	1,000	,174**
		Sig. (2-tailed)	,	,000
		N	464	464
	Gesamtpunkte Statistik II	Correlation Coefficient	,174**	1,000
		Sig. (2-tailed)	,000	,
		N	464	464

**. Correlation is significant at the .01 level (2-tailed).

Tabelle A-50 Mittelwert der normierten Gesamtpunkte nach neuer und alter Prüfungs-ordnung

Report

normierte Gesamtpunkte

alte oder neue PO	Mean	N	Std. Deviation
neue PO	55,133	1571	23,5027
alte PO	65,047	190	19,7758
Total	56,203	1761	23,3282

Tabelle A-51 Kolmogorov-Smirnov-Test nach neuer und alter Prüfungsordnung

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	alte oder neue PO	Statistic	df	Sig.	Statistic	df	Sig.
normierte Gesamtpunkte	neue PO	,042	1571	,000	,984	1571	,000
	alte PO	,060	190	,088	,977	190	,003

a. Lilliefors Significance Correction

Tabelle A-52 Mann-Whitney-Test nach neuer und alter Prüfungsordnung

Ranks

	alte oder neue PO	N	Mean Rank	Sum of Ranks
normierte Gesamtpunkte	neue PO	1571	856,67	1345824,04
	alte PO	190	1082,19	205616,99
	Total	1761		

Test Statistics[a]

	normierte Gesamtpunkte
Mann-Whitney U	111018,0
Wilcoxon W	1345824
Z	-5,775
Asymp. Sig. (2-tailed)	,000

a. Grouping Variable: alte oder neue PO

Tabelle A-53 Kendall's tau_b für Punkte und bestandene Aufgabe mit Spezifikation

Correlations

			PUNKTE	JA
Kendall's tau_b	PUNKTE	Correlation Coefficient	1,000	-,073
		Sig. (2-tailed)	,	,402
		N	70	70
	JA	Correlation Coefficient	-,073	1,000
		Sig. (2-tailed)	,402	,
		N	70	70

Tabelle A-54 Kendall's tau_b für Punkte und bestandene Aufgabe ohne Spezifikation

Correlations

			JA	PUNKTE
Kendall's tau_b	JA	Correlation Coefficient	1,000	,091
		Sig. (2-tailed)	,	,206
		N	97	97
	PUNKTE	Correlation Coefficient	,091	1,000
		Sig. (2-tailed)	,206	,
		N	97	97

Abbildung A-1Auszug des Dendogramms für die Clusterbildung in Statistik I

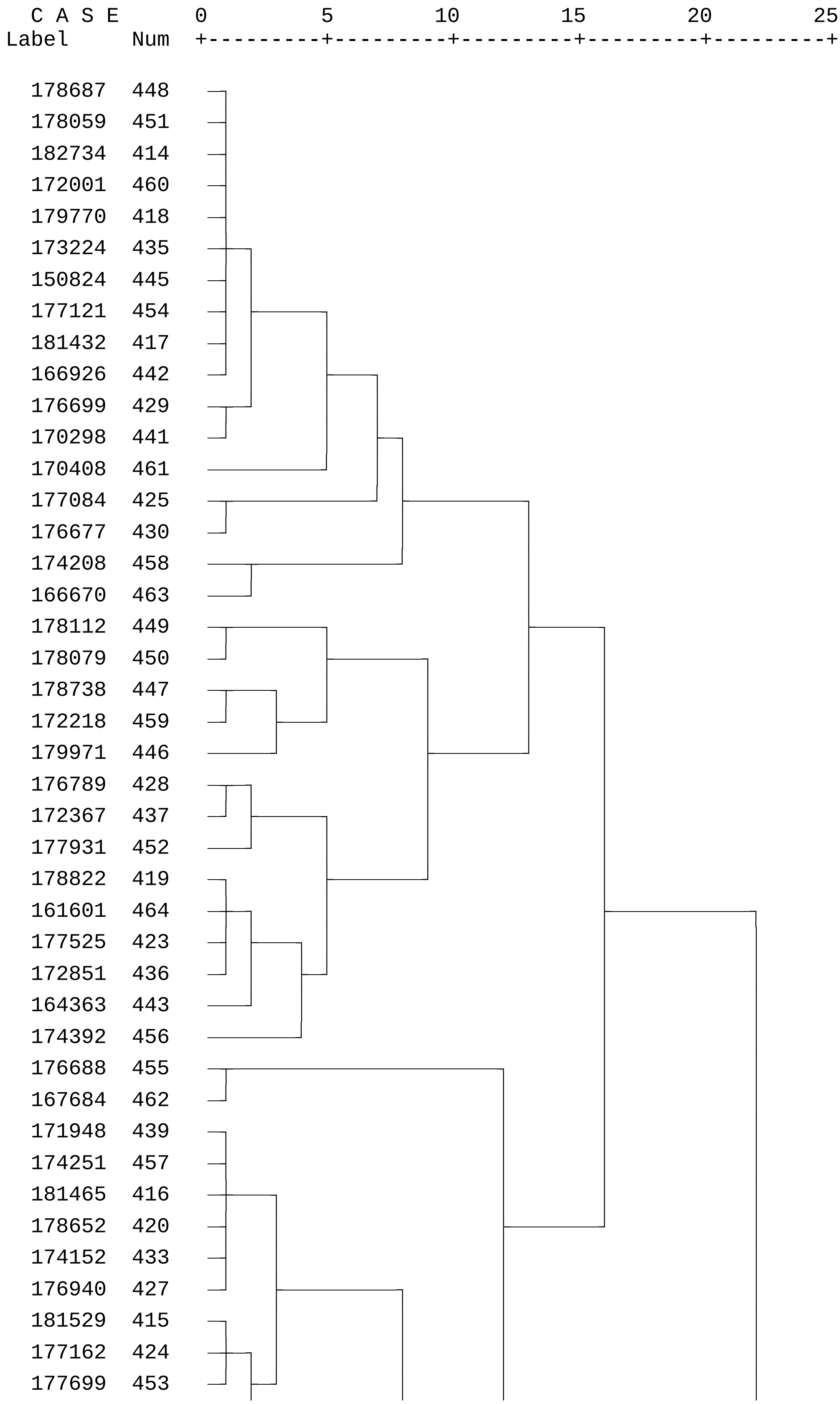
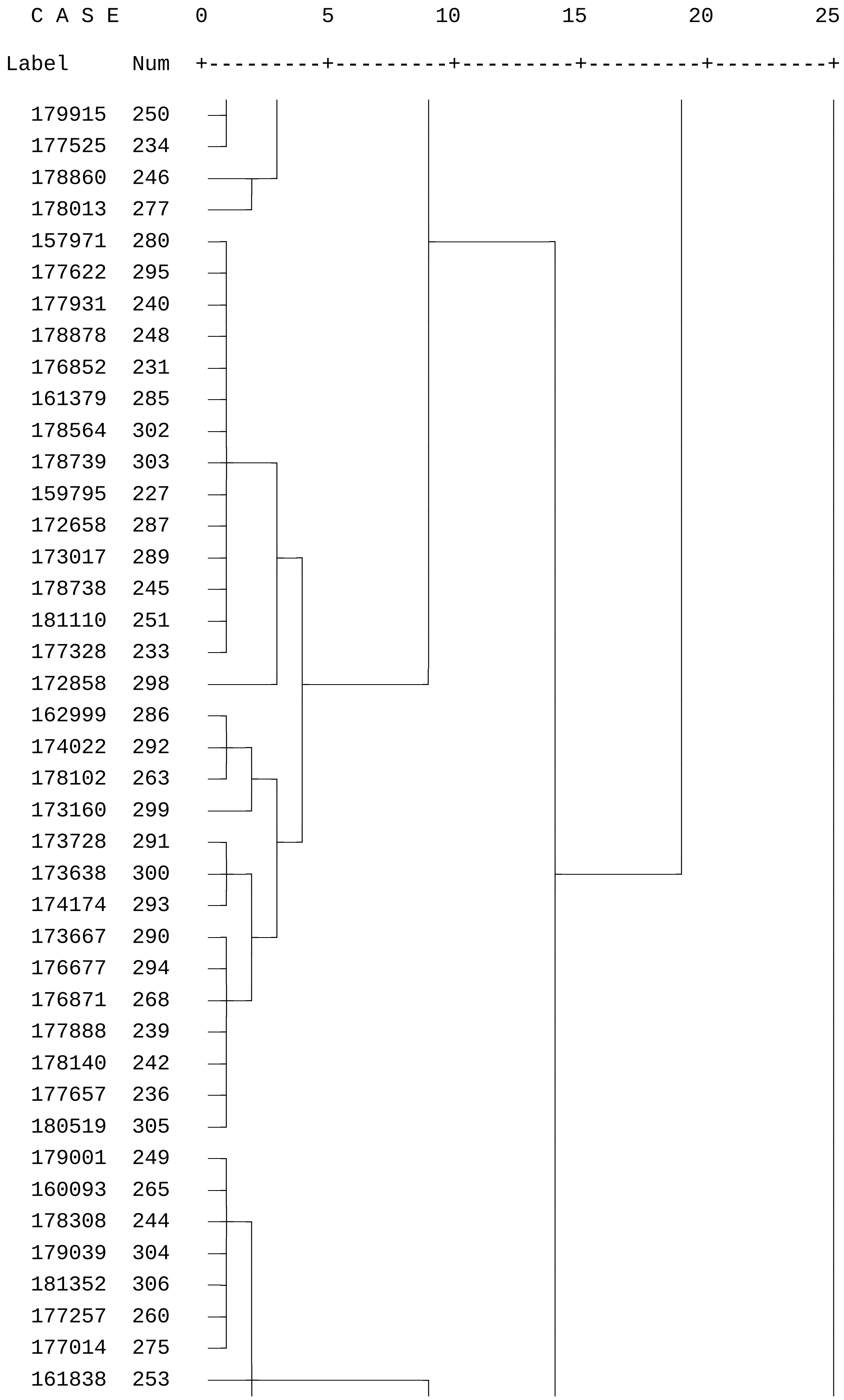


Abbildung A-2 Auszug des Dendogramms für die Clusterbildung in Statistik II



Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit zum Thema „Statistische Bewertung und Analyse der Klausurergebnisse Statistik (Grundstudium)“ selbständig verfasst und keine anderen als die angegebenen Quellen verwendet habe. Die Arbeit hat keiner anderen Prüfungsbehörde vorgelegen.

Ulrike Brandes
Berlin, 09. Januar 2004