Vergleich der zentralen Tendenz in zwei unabhängigen Stichprobenerhebungen

Inhaltsverzeichnis

Vergleich der zentralen Tendenz in zwei unabhängigen Stichprobenerhebungen

Lernhinweise

Dieser Lernschritt gilt dem U-Test nach Mann-Whitney, mit dem die in zwei unabhängigen Stichproben erhobenen Daten bezüglich der zentralen Tendenz verglichen werden können. Im Rahmen eines Fallbeispiels werden dabei die theoretischen Grundlagen erarbeitet und anschaulich dargestellt.

Benötigte Vorkenntnisse

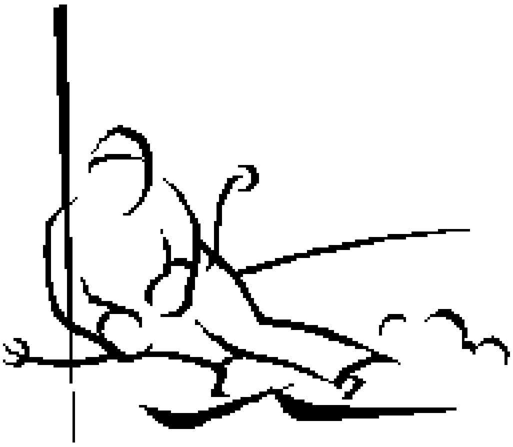
- Eigenschaften der Skalenniveaus
- In welchen Situationen werden nichtparametrische Verfahren eingesetzt?
- Grundprinzip entscheidungsstatistischer Verfahren
- Parametrische Prüfverfahren

Geschätzte Bearbeitungszeit
Sie können diesen Lernschritt in ca. 90 Minuten bearbeiten, wenn Sie die in ihrem Curriculum vorgesehene Vorbereitungslektüre konsultiert haben.

Theorie & Fallbeispiel (1-11)

Inhaltsübersicht

1. Ein Firmenskirennen wirft Fragen auf



Thomas und Helen, zwei befreundete Psychologie-Studierende, besuchen ein Firmenskirennen an dem Helens Vater zusammen mit allen Mitarbeitenden der Firma teilnimmt. Nach einem sonnigen, fröhlichen Tag im Schnee nimmt Thomas auf der Heimfahrt die Rangliste hervor.

Rangliste

Nachdem er sie studiert hat, vertritt er die Meinung, dass die Männer im 1. Lauf besser gefahren seien. Helen ist gar nicht dieser Ansicht und wehrt sich vehement: "Nur weil innerhalb der ersten 3 Ränge zwei Männer plaziert sind, heisst das noch lange nicht, dass die Männer die besseren Skifahrer sind!" Thomas lässt dieses Argument zwar nicht gelten, ist aber damit einverstanden, sich der Sache anzunehmen, sobald man wieder zuhause ist. Wenig später, als Helen die Rangliste studiert, behauptet sie: "Die Frauen sind im 2. Lauf auf den besseren Rängen plaziert als im 1. Lauf." Man beschliesst, sich auch dieser Frage zu Hause zu widmen.
Thomas und Helen wollen also zwei konkrete Fragestellungen überprüfen:

1. Sind die Männer im 1. Lauf des Skirennens gesamthaft besser gefahren als die Frauen?
2. Sind die Fahrerinnen im 2. Lauf auf den besseren Rängen plaziert als im 1. Lauf?

2. Vergleich der Rennresultate der Männer und Frauen im 1. Lauf

Zuhause angekommen, überlegen sich Thomas und Helen, anhand welcher Informationen aus der Rangliste die erste Frage beantwortet werden könnte. Nach einigen Überlegungen schlägt Thomas vor, dass man die Rennzeiten der Frauen mit denjenigen der Männer vergleichen könnte. Helen stimmt zu und resumiert: "Wir bilden also je eine Stichprobe der Männer und der Frauen und vergleichen diese Stichproben bezüglich der mittleren Rennzeiten."

Welches Verfahren fällt Ihnen zuerst ein, wenn Sie die mittleren Rennzeiten der Männer mit denjenigen der Frauen vergleichen möchten?

❍ 1. Korrelation

❍ 2. t-Test für abhängige Stichproben

❍ 3. Regression

❍ 4. t-Test für unabhängige Stichproben

Auch Thomas und Helen fällt als erstes ein t-Test für unabhängige Stichproben ein. Doch stellt sich sofort die Frage nach den Voraussetzungen, die die Daten erfüllen müssen, damit ein t-Test für unabhängige Stichproben zulässig ist. Welche Voraussetzungen müssen für einen t-Test für unabhängige Stichproben erfüllt sein?

❏ Normalverteilung der Daten in den Stichproben

❏ Normalverteilung der Daten in den Populationen

❏ Merkmal muss mindestens ordinal skaliert sein

❏ Merkmal muss mindestens intervall-skaliert sein

❏ Die Varianzen des Merkmals müssen in den Stichproben homogen sein

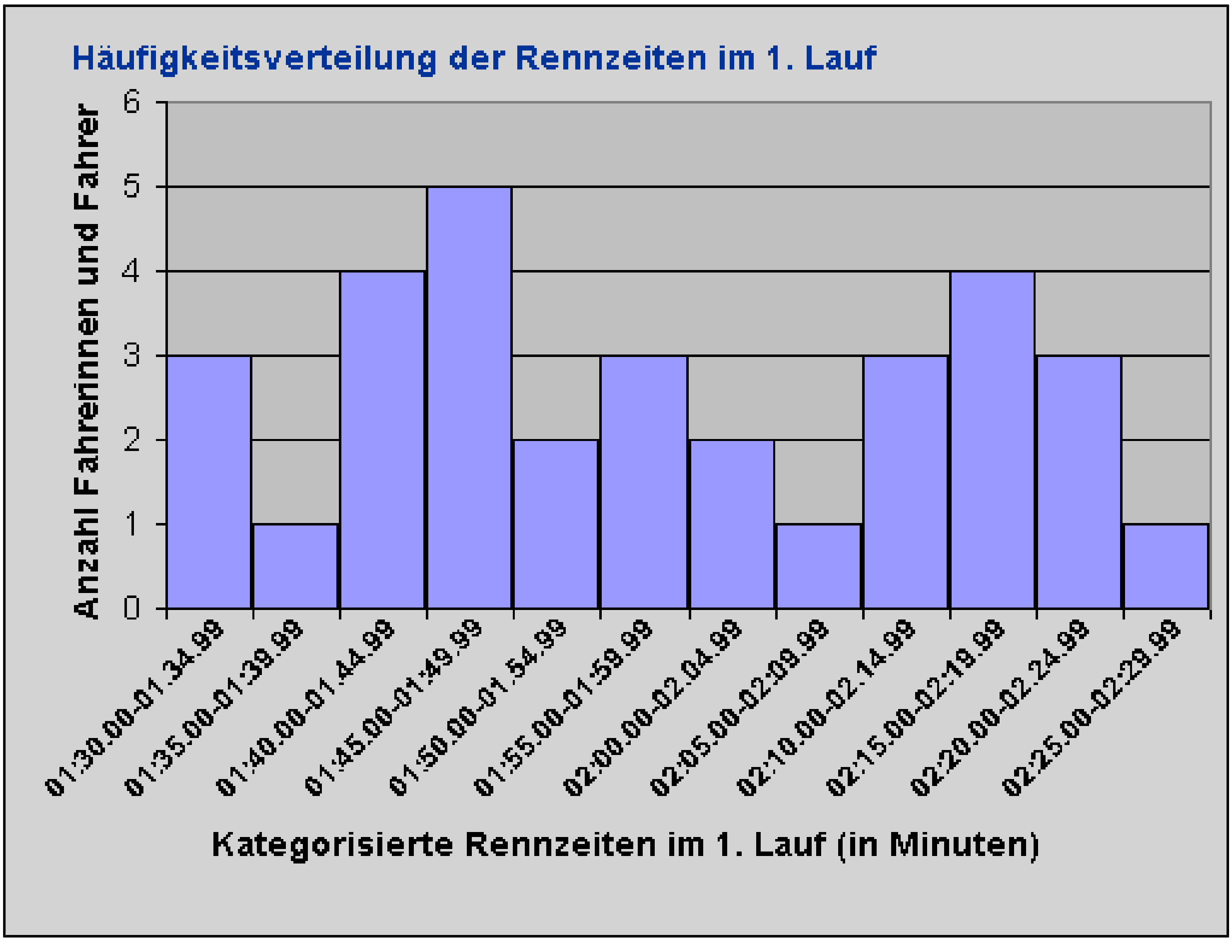
Thomas konsultiert sein Statistik-Skript und liest vor, dass der t-Test für unabhängige Stichproben nur angewendet werden darf, wenn:

1. das Merkmal intervall- oder proportional skaliert ist und
2. die Daten in den Populationen normalverteilt sind.

Helen und Thomas sind sich einig, dass die 1. Voraussetzung erfüllt ist. Die Rennzeiten im Firmenskirennen sind proportional skaliert. Hinsichtlich der 2. Voraussetzung aber, gerät Helen ins Stocken. Sie fragt Thomas: **"Was meinst du, sind die Rennzeiten normalverteilt?"** Thomas glaubt nicht so recht daran. Er argumentiert, dass die Rennzeiten wohl nicht normalverteilt sind, weil es in der Population der Arbeitnehmerinnen und

Arbeitnehmer Personen gibt, die seit ihrer Kindheit nie mehr auf den Brettern standen, während andere den Skisport noch heute aktiv ausüben. Thomas erwartet auch für das Firmenskirennen eine bimodale Verteilung, weil sich alle Mitarbeitenden der Firma an diesem "Freundschaftsrennen" beteiligen. Die beiden Freunde zeichnen die Verteilung der kategorisierten Rennzeiten auf, und es wird ersichtlich, dass eine bimodale Verteilung der Rennzeiten vorliegt.

Häufigkeitsverteilung der Rennzeiten



So muss davon ausgegangen werden, dass die 2. Voraussetzung nicht erfüllt ist. Ein Vergleich der Stichprobenmittelwerte mit einem t-Test für unabhängige Stichproben ist also nicht möglich. Glücklicherweise gibt es aber eine Gruppe von Verfahren, welche in solchen Fällen angewendet werden kann. Es sind dies die nichtparametrischen oder verteilungsfreien Verfahren.

FAQ

3. Gesucht ist ein verteilungsfreies nichtparametrisches Verfahren

Thomas und Helen suchen also ein Verfahren mit dem die Resultate der Frauen und Männer im 1. Lauf des Skirennens verglichen werden können, ohne dass bezüglich der Verteilung der Daten in den Populationen eine Voraussetzung gemacht werden muss.
Die Statistiker können weiterhelfen: "Wenn ihr die Resultate auf Ordinalskalenniveau auswertet, entfällt die Voraussetzung der Normalverteilung in den Populationen."
Nach kurzem Nachdenken wissen Helen und Thomas, was dies bedeutet:
Sie können ihre Frage beantworten, wenn nicht die mittleren Abfahrtszeiten, sondern die zentralen Tendenzen in den Rangverteilungen der beiden Stichproben verglichen werden. Als erstes erstellen sie für den 1. Lauf des Skirennens eine Rangliste, dann machen sie sich auf die Suche nach einem passenden Verfahren und finden den U-Test nach Mann-Whitney.

Rangliste

Wir gönnen Helen und Thomas eine verdiente Pause und wollen uns diese Prüfverfahren etwas genauer ansehen.
Vorerst rufen wir uns aber in Erinnerung, was ganz generell zu jedem entscheidungsstatistischen Prüfverfahren gehört.

FAQ

Welche Stichworte gehören aus formaler Sicht, d.h. ohne Berücksichtigung der Interpretation des Resultates, im einfachsten Fall zu jedem entscheidungsstatistischen Verfahren, mit dem ein Unterschied in der zentralen Tendenz der Daten aus zwei Stichproben nachgewiesen werden soll?

❍ Prüfgrösse
❍ Arbeitshypothese und Alternativhypothese
❍ Kenntnis, wie die Prüfgrösse bei Gültigkeit der Arbeitshypothese verteilt ist
❍ Kenntnis, wie die Prüfgrösse verteilt ist, wenn die Arbeitshypothese falsch ist
❍ a-Fehler-Risiko
❍ Kenntnis was die Überschreitung des a-Fehler-Risikos inhaltlich bedeutet

4. Arbeitshypothesen und Alternativhypothesen

Nachdem wir uns die Grundlagen der entscheidungsstatistischen Verfahren in Erinnerung gerufen haben, wollen wir uns dem U-Test nach Mann-Whitney zuwenden.
Was prüft der U-Test? Der U-Test prüft, ob bei der Erklärung des Unterschiedes in der zentralen Tendenz der rangierten, aus zwei unabhängigen Stichproben stammenden Daten, der Zufall ausgeschlossen werden kann.

Etwas anders formuliert: Es wird geprüft, ob angenommen werden darf, dass die beiden Stichproben aus Populationen stammen, in denen das interessierende Merkmal eine identische zentrale Tendenz aufweist.
Das Postulat dieser Identität der zentralen Tendenz dient uns als Arbeitshypothese H0. Sie setzt also voraus, dass die Stichproben tatsächlich aus Populationen stammen, deren zentrale Tendenzen im interessierenden Merkmal identisch sind. Da diese Arbeitshypothese nicht aufgrund von Verteilungsparametern formuliert werden kann, ist sie im Gegensatz zu den parametrischen Verfahren nur verbal beschreibbar. Sie lautet:
Arbeitshypothese H_0: Die Stichproben stammen aus Populationen mit identischer zentraler Tendenz in den rangierten Daten.
Dieser Arbeitshypothese stellen wir die Alternativhypothese H_1 gegenüber. Diese kann, je nach inhaltlicher Fragestellung, gerichtet oder ungerichtet und spezifisch oder unspezifisch formuliert sein.
Wir formulieren H_0 und H_1 für unser konkretes Beispiel.

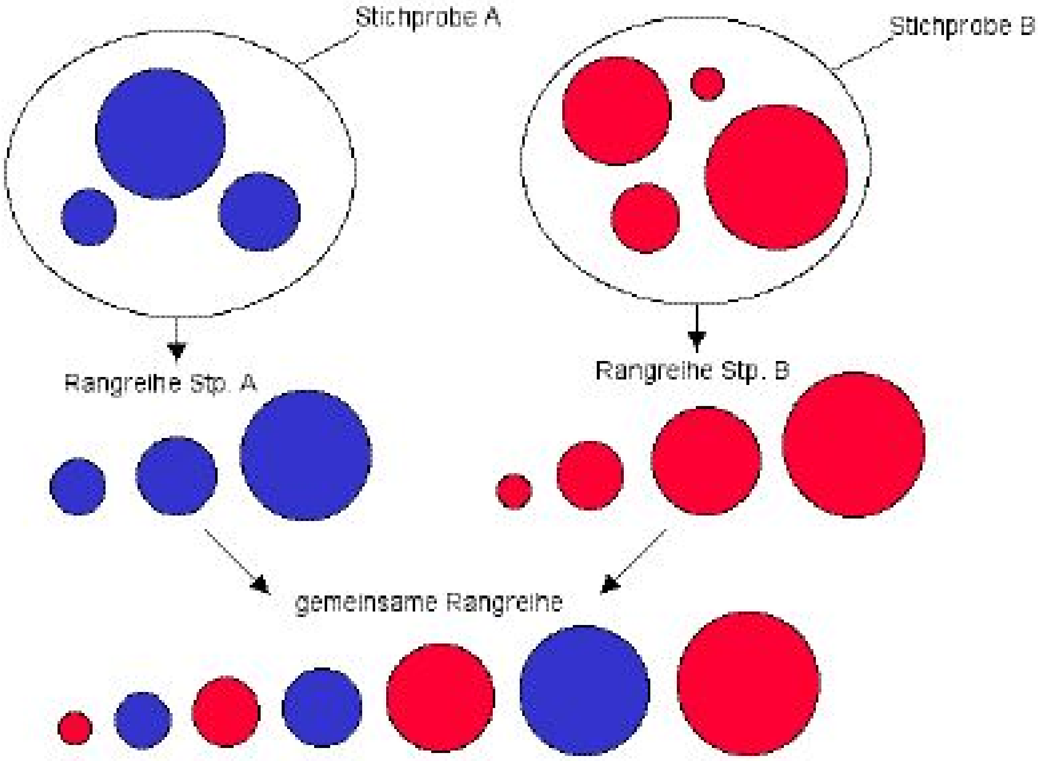
Arbeitshypothese H_0 : Die Stichproben von Damen und Herren stammen hinsichtlich des Merkmals "Rangplatz im 1. Lauf des Skirennens" aus Populationen mit identischer zentraler Tendenz.

Alternativhypothese H_1 : Die Stichprobe der Herren stammt bezüglich des Merkmals "Rangplatz im 1. Lauf des Skirennens" aus einer Population mit geringerer zentraler Tendenz als die Stichprobe der Frauen. Entsprechend der inhaltlichen Fragestellung ist diese Alternativhypothese gerichtet aber unspezifisch. Unspezifisch deshalb, weil nichts über die Grösse des Unterschieds in der zentralen Tendenz postuliert wird.
Wie bei jedem entscheidungsstatistischen Verfahren werden wir im Folgenden prüfen, inwieweit die Arbeitshypothese H_0 zu Gunsten der Alternativhypothese H_1 verworfen werden kann. Hierfür brauchen wir eine **Prüfgrösse** und die zugehörige **Prüfverteilung**.

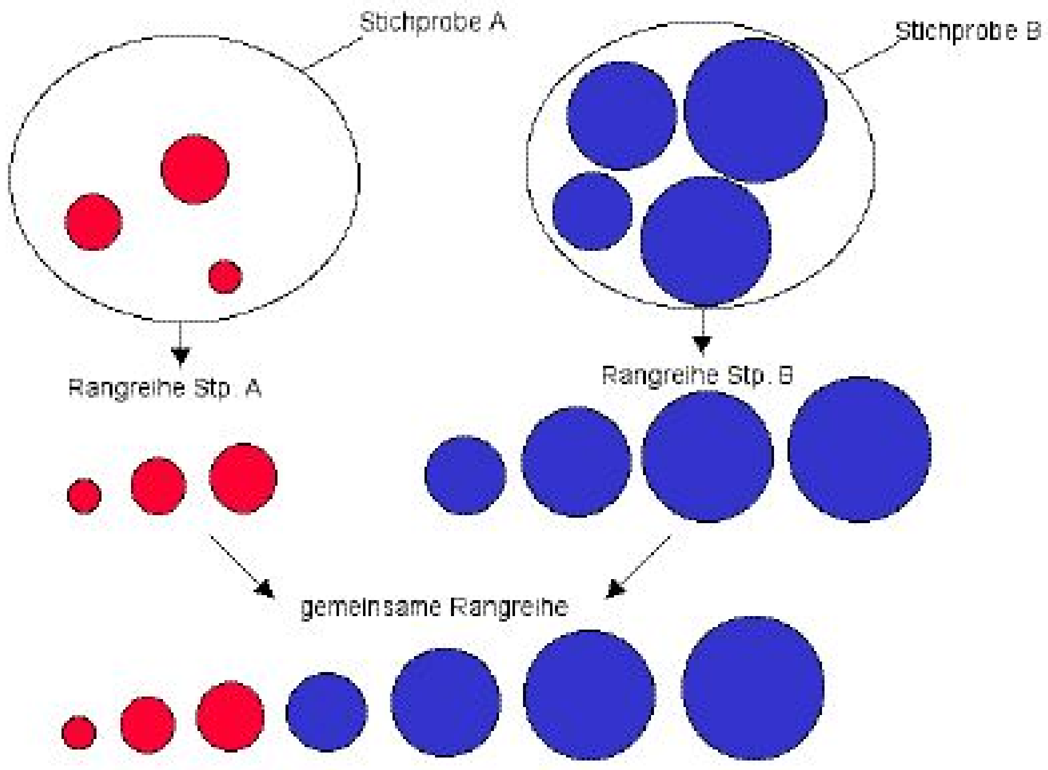
FAQ

5. Prüfgrösse U resp. U'

Als erstes widmen wir uns der **Herleitung der Prüfgrösse**. Wir suchen eine Prüfgrösse, deren Ausprägungsgrad die Unterschiedlichkeit der beiden zu vergleichenden Stichproben bezüglich der zentralen Tendenz der rangierten Daten beschreibt. Was für eine Prüfgrösse ist aber für ordinal skalierte Daten denkbar? Glücklicherweise hatten die Statistiker Mann und Whitney eine gute **Idee**, die wir im Folgenden nachvollziehen wollen.
Ausgangspunkt ist die Überlegung, dass sich die Daten aus zwei Rangreihen, die sich in ihrer zentralen Tendenz nicht unterscheiden, in einer gemeinsamen Rangreihe gleichmässig verteilen. Wir können dies graphisch veranschaulichen. Dabei symbolisiert die Grösse der Kreise die Rangausprägungen in den beiden Stichproben.



In der gemeinsamen Rangreihe der Daten sind die Stichprobendaten absolut gleichmässig verteilt. Intuitiv können wir nachvollziehen, dass in diesem Fall ausgeschlossen werden kann, dass sich die beiden Stichproben bezüglich der zentralen Tendenz der Daten signifikant unterscheiden.
Noch markanter ist wohl der zweite Extremfall. In diesem Fall unterscheiden sich die beiden Stichproben bezüglich der Ränge vollständig. In der gemeinsamen Rangreihe liegen dann die Daten der einen Stichprobe geschlossen am Anfang und diejenigen der anderen Stichprobe geschlossen am Ende. Wir wollen auch diesen Extremfall veranschaulichen.



Nun wissen wir, wie sich die beiden Extremfälle in der gemeinsamen Rangreihe der Daten zeigen:

Extrem grosser Unterschied in den Stichprobendaten:	Keine Vermischung der Daten in der gemeinsamen Rangreihe
Extrem geringer Unterschied in den Stichprobendaten:	Praktisch gleichmässige Verteilung der Daten in der gemeinsamen Rangreihe

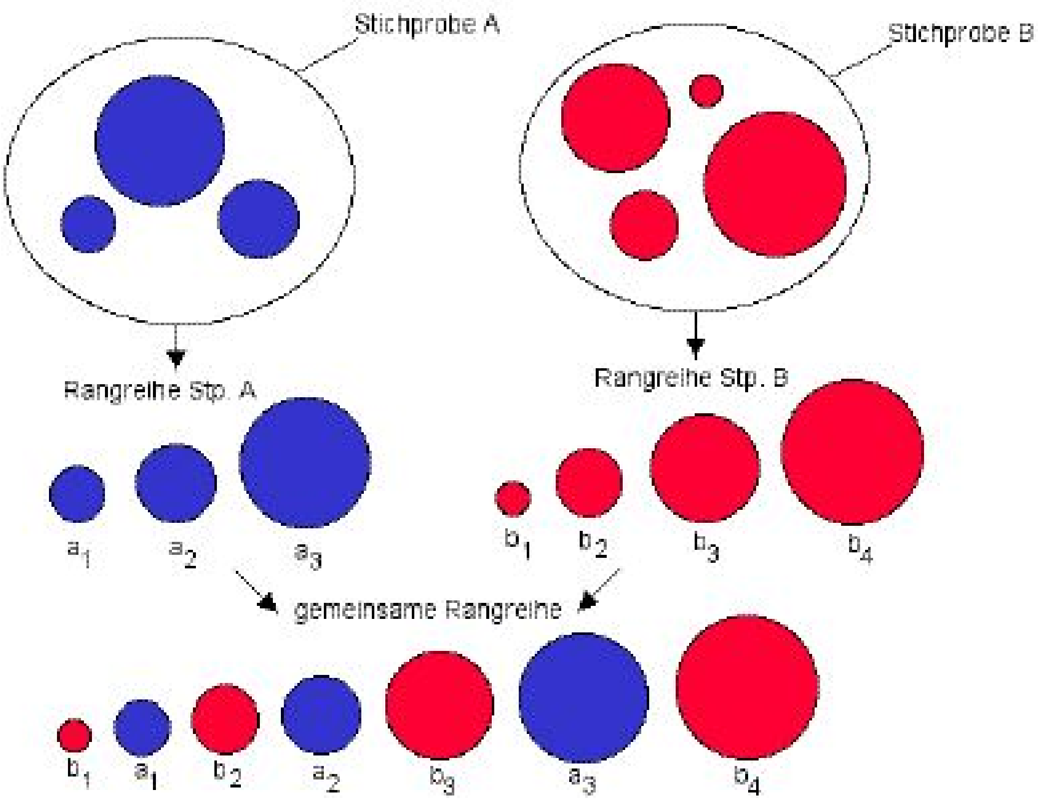
Dieser Sachverhalt bringt uns auf die Idee, die Gleichmässigkeit der Verteilung der Daten in der gemeinsamen Rangreihe für die Beschreibung der Unterschiedlichkeit der Daten in den beiden Stichproben zu benutzen. In der Forschungspraxis kommen die beiden Extremfälle nun aber praktisch nie vor. In der Regel sind die Daten in der gemeinsamen Rangreihe nur mehr oder weniger gleichmässig verteilt. **Gesucht** ist also ein **Mass für die Gleichmässigkeit der Verteilung der Daten in der gemeinsamen Rangreihe**. Mann und Whitney hatten die Idee, die "Anzahl Rangplatzüberschreitungen" und die "Anzahl Rangplatzunterschreitungen" als Mass heranzuziehen, und definierten so die Prüfgrösse U resp. U'.

Zur "Anzahl Rangplatzüberschreitungen" gelangt man, indem man in der gemeinsamen Rangreihe für jede Person der Stichprobe A auszählt, wieviele Personen der Stichprobe B vor ihr liegen. Summiert man diese Rangplatzüberschreitungen über alle Personen der Stichprobe A, so erhält man die gesuchte Prüfgrösse, die U genannt wird.

Zur Bestimmung der **"Anzahl Rangplatzunterschreitungen"** verfährt man in analoger Weise. Für jede Person aus der Stichprobe A wird ausgezählt, wieviele Personen der Stichprobe B hinter ihr liegen. Zählt man diese Rangplatzunterschreitungen aller Personen der Stichprobe A zusammen, so erhält man die "Anzahl Rangplatzunterschreitungen". Diese Kenngrösse bezeichnet man mit U'.

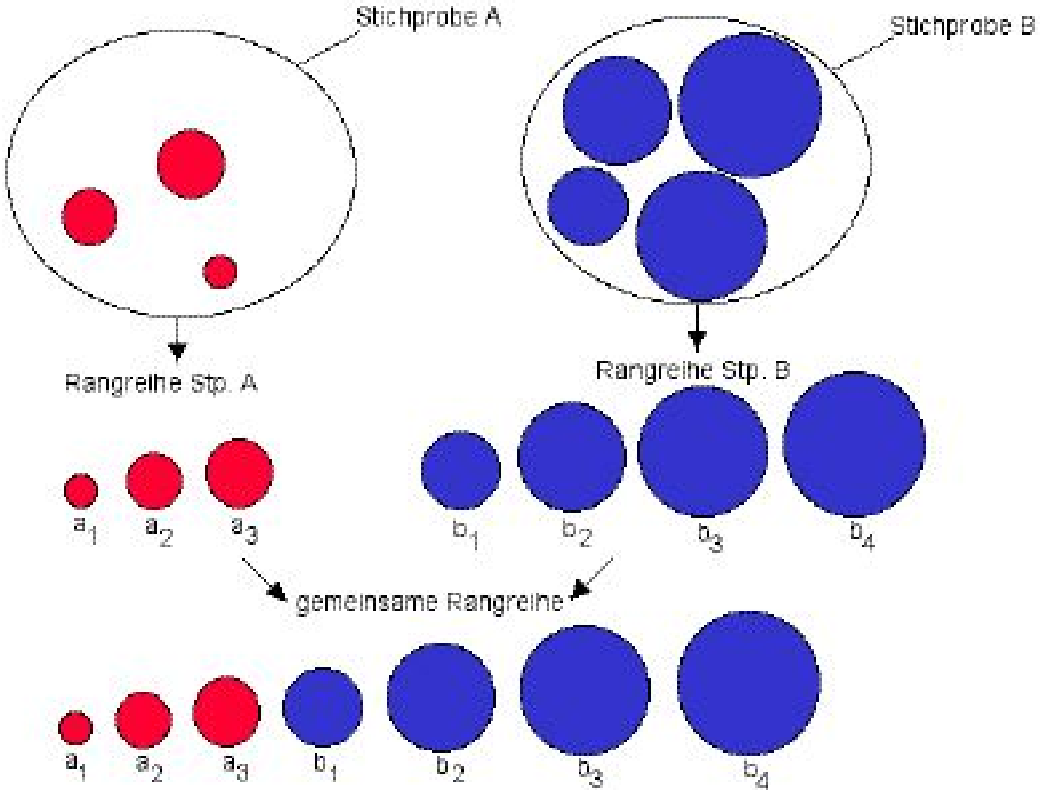
Ohne auf die Beweisführung einzugehen, glauben wir den Statistikern, dass U' auch aus U berechnet werden kann:

$U' = n_A * n_B - U$, wobei n_A die Grösse der Stichprobe A und n_B diejenige der Stichprobe B bezeichnet.
Als erstes wollen wir U und U' für die beiden besprochenen Extremfälle bestimmen:
Minimaler Unterschied in den Stichproben:



a_1 wird von einem Rang aus B übertroffen	1 Überschreitung
a_2 wird von zwei Rängen aus B übertroffen	2 Überschreitungen
a_3 wird von drei Rängen aus B übertroffen	3 Überschreitungen
Total "Anzahl Rangplatzüberschreitungen"	U = 6 Überschreitungen
Total "Anzahl Rangplatzunterschreitungen"	U' = 3 * 4 - 6 = 6

Maximaler Unterschied in den Stichproben:



a_1 wird von keinem Rang aus B übertroffen	0 Überschreitungen
a_2 wird von keinem Rang aus B übertroffen	0 Überschreitungen
a_3 wird von keinem Rang aus B übertroffen	0 Überschreitungen
Total "Anzahl Rangplatzüberschreitungen"	U = 0 Überschreitungen
Total "Anzahl Rangplatzunterschreitungen"	U' = 3 * 4 - 0 = 12

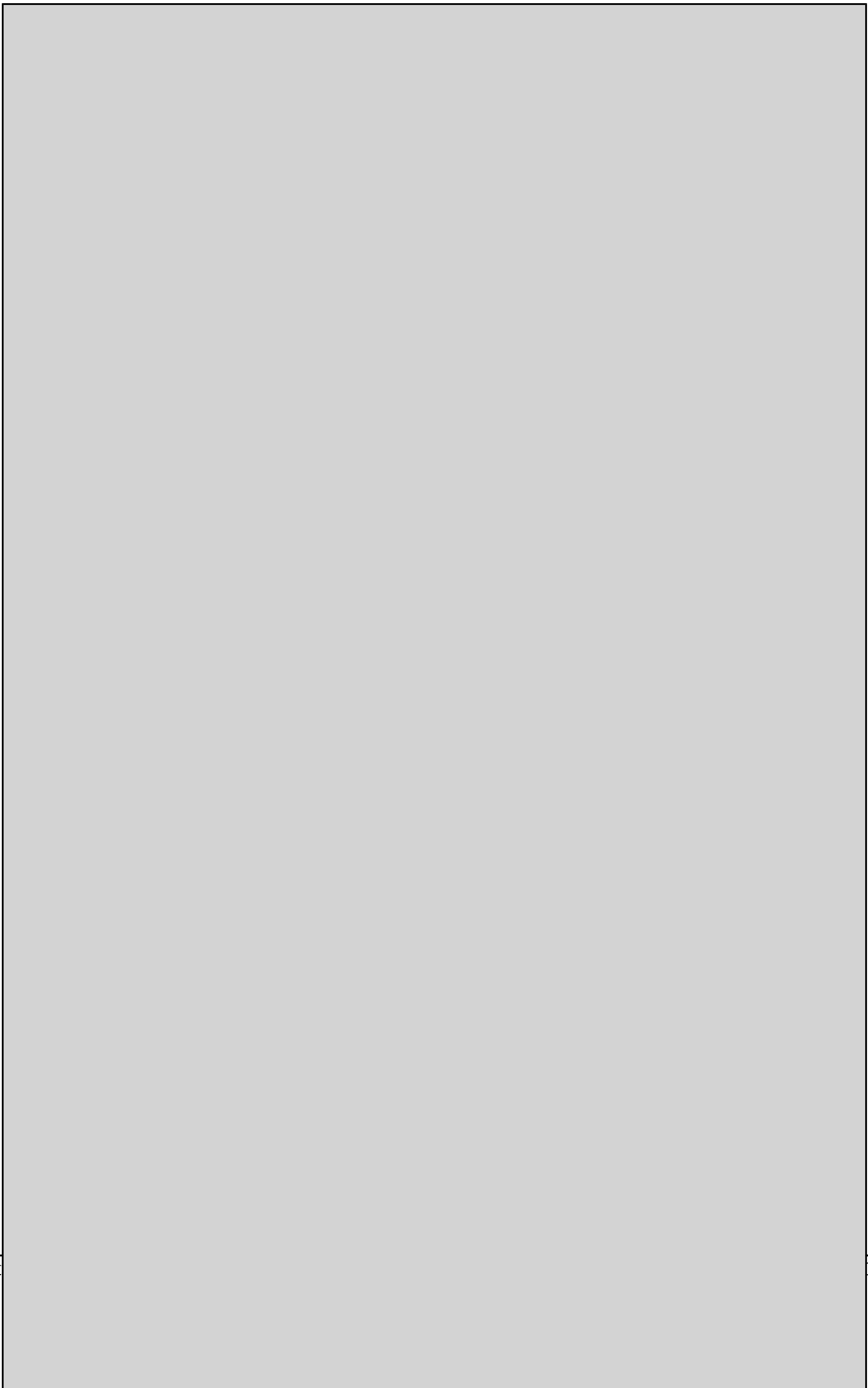
Die Unterschiedlichkeit von zwei Rangverteilungen kann also mit zwei verschiedenen Kennzahlen, mit U und U' beschrieben werden. Zur Vereinheitlichung und - so werden wir später sehen - zur Vereinfachung der vorhandenen Tabellen haben sich die Statistiker darauf geeinigt, dass U als Prüfgrösse dient, falls U kleiner als U' ist. Ansonsten ist U' die Prüfgrösse.
Als nächstes wollen wir U und U' für einen in der Realität üblichen Fall bestimmen. Hierfür können uns die Daten unseres Skirennens dienen.

FAQ

6. Bestimmung der Prüfgrösse für die Resultate des Skirennens

Als erstes erstellen wir die gemeinsame Rangreihe, aus welcher die Rangplatzüberschreitungen ermittelt werden kann.

Gemeinsame Rangreihe



Wir haben also eine Rangreihe mit Skifahrern in der Anordnung:
Mann, Frau, Mann, Mann, Mann, Mann, Frau, Frau, Frau, Mann, Frau…………..
Wir zählen nun für jede Frau aus, von wievielen Männern sie übertroffen wurde. Die erste Frau wird von einem Mann übertroffen, die zweite schon von 5 Männern, die dritte ebenfalls von 5 Männern usw. Schliesslich werden diese Rangplatzüberschreitungen zusammengezählt und man gelangt zur Prüfgrösse U = 172. U' bestimmt sich dann wie folgt:
U' = 18 * 14 - 172 = 80.

Berechnung von U

Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]

Weil U grösser ist als U', dient uns U' als Prüfgrösse.
Die Bestimmung von U und U' durch Auszählen der Rangplatzüberschreitungen ist wohl sehr leicht verständlich, im Rahmen realer Datenauswertungen aber etwas mühsam und unpraktisch. So soll noch kurz und ohne Beweisführung auf eine etwas einfachere Berechnungsmethode für U und U' eingegangen werden.
Die Prüfgrösse U resp. U' kann über die folgenden zwei Schritte aus den sogenannten Rangsummen der beiden Stichproben ermittelt werden:
1. Erstellen der Ranglisten für die beiden Stichproben und Bestimmung der Rangsummen T_A und T_B.
2. Bestimmung von U und U' aus den Rangsummen nach den Formeln:

$$U = n_A \cdot n_B + \frac{n_B \cdot (n_B + 1)}{2} - T_B; \qquad U' = n_A \cdot n_B - U$$

Wir wollen U und U' nach dieser Methode für die Daten unseres Skirennens bestimmen.
Als erstes erstellen wir für die Stichproben der Frauen und der Männer je eine Rangliste und bestimmen die zugehörige Rangsumme.

Rangliste der Frauen	2	7	8	9	11	18	23	24	25	28	29	30	31	32					Rangsumme T_A = 277 n_A = 14
Rangliste der Männer	1	3	4	5	6	10	12	13	14	15	16	17	19	20	21	22	26	27	Rangsumme T_B = 251 n_B = 18

Die berechneten Rangsummen lassen sich über folgende, hier nicht hergeleitete, Formel kontrollieren:

$$T_A + T_B = \frac{n \cdot (n+1)}{2};$$

; wobei

$$n = (n_A + n_B),$$

d.h. n = Summe der beiden Stichprobengrössen.
In unserem Fall:

$$277 + 251 = \frac{32 \cdot (32+1)}{2} = 528; \quad \text{OK!}$$

Aus den Rangsummen bestimmen wir U und U' mit den obenstehenden Formeln:

$$U = n_A \cdot n_B + \frac{n_B \cdot (n_B + 1)}{2} - T_B = 14 \cdot 18 + \frac{18 \cdot (18+1)}{2} - 251 = 172$$

$$U' = n_A \cdot n_B - U = 80$$

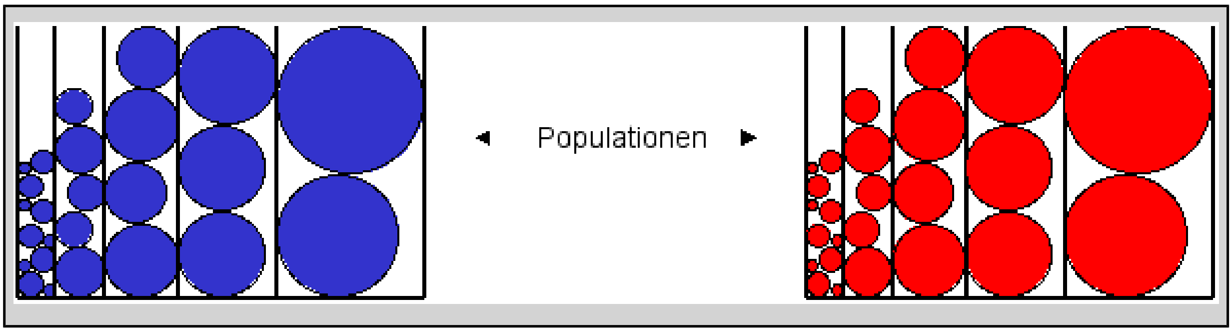
Die so über die Rangsummen bestimmten Ausprägungen von U und U' sind mit den durch die Auszählung der Rangplatzüberschreitungen ermittelten Werte identisch.
Da U grösser ist als U', dient uns U' als Prüfgrösse.
Nun kennen wir den Ausprägungsgrad der Prüfgrösse. Als nächstes müssen wir entscheiden, mit welcher Wahrscheinlichkeit dieser Ausprägungsgrad erreicht wird, wenn die beiden Stichproben aus Populationen mit identischer zentraler Tendenz stammen, d.h. wenn unsere Arbeitshypothese H_0 tatsächlich gültig ist. Dazu brauchen wir eine Prüfverteilung.

7. Bestimmung der Prüfverteilung

Die Prüfverteilung zeigt, wie die Prüfgrösse unter der Annahme der Gültigkeit der Arbeitshypothese H0, d.h. bei einer identischen zentralen Tendenz in den Populationen verteilt ist.
Die grundlegende Idee zur Bestimmung der Prüfverteilung lässt sich an einem einfachen Gedankenexperiment über 7 Schritte verdeutlichen.
Alternativ zum Text können Sie die Punkte 1-5 auch als Animation ansehen und bei Punkt 6 weiterfahren.

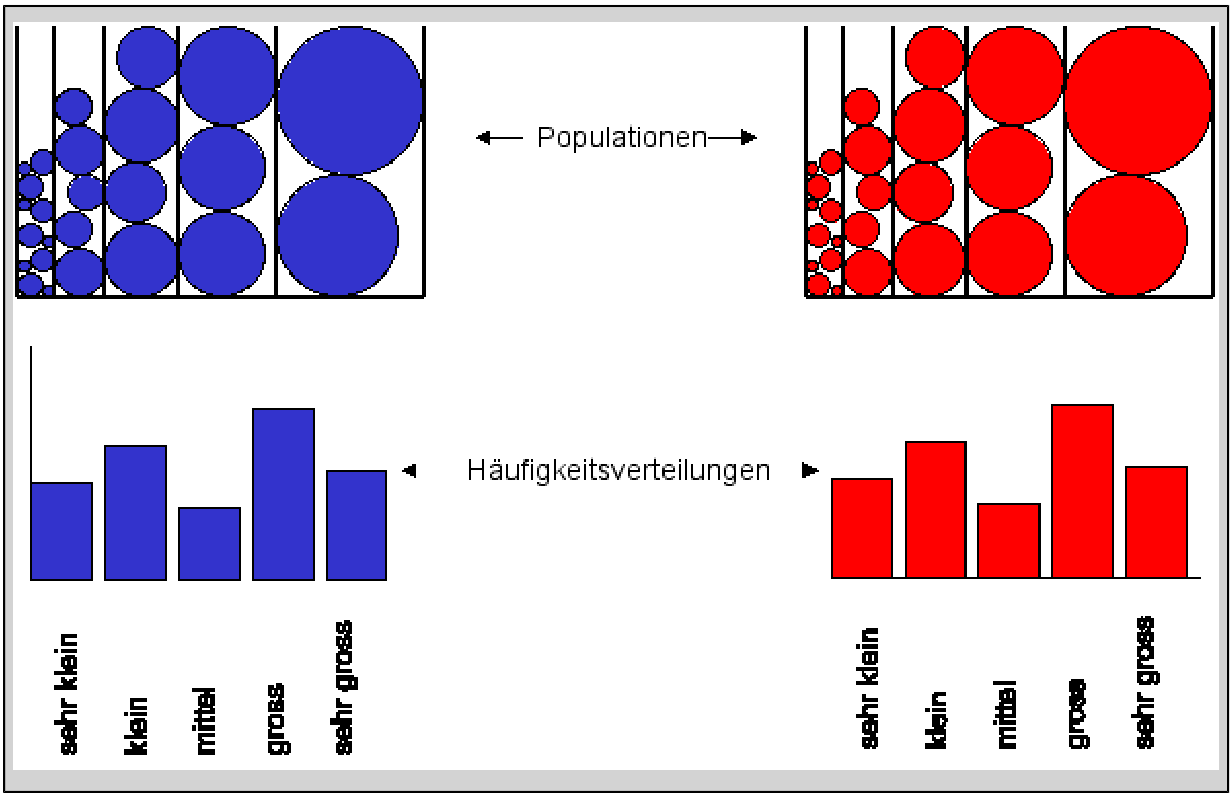
1. Ausgangspunkt sind 2 Populationen, in denen ein ordinal skaliertes Merkmal in identischer Weise verteilt ist. Das Merkmal hat damit in beiden Populationen sicher eine identische zentrale Tendenz. Die Daten werden als Kugeln dargestellt, wobei Daten mit gleichem Ausprägungsgrad in gleiche Gefässe eingeordnet werden.

   Daten der beiden Populationen



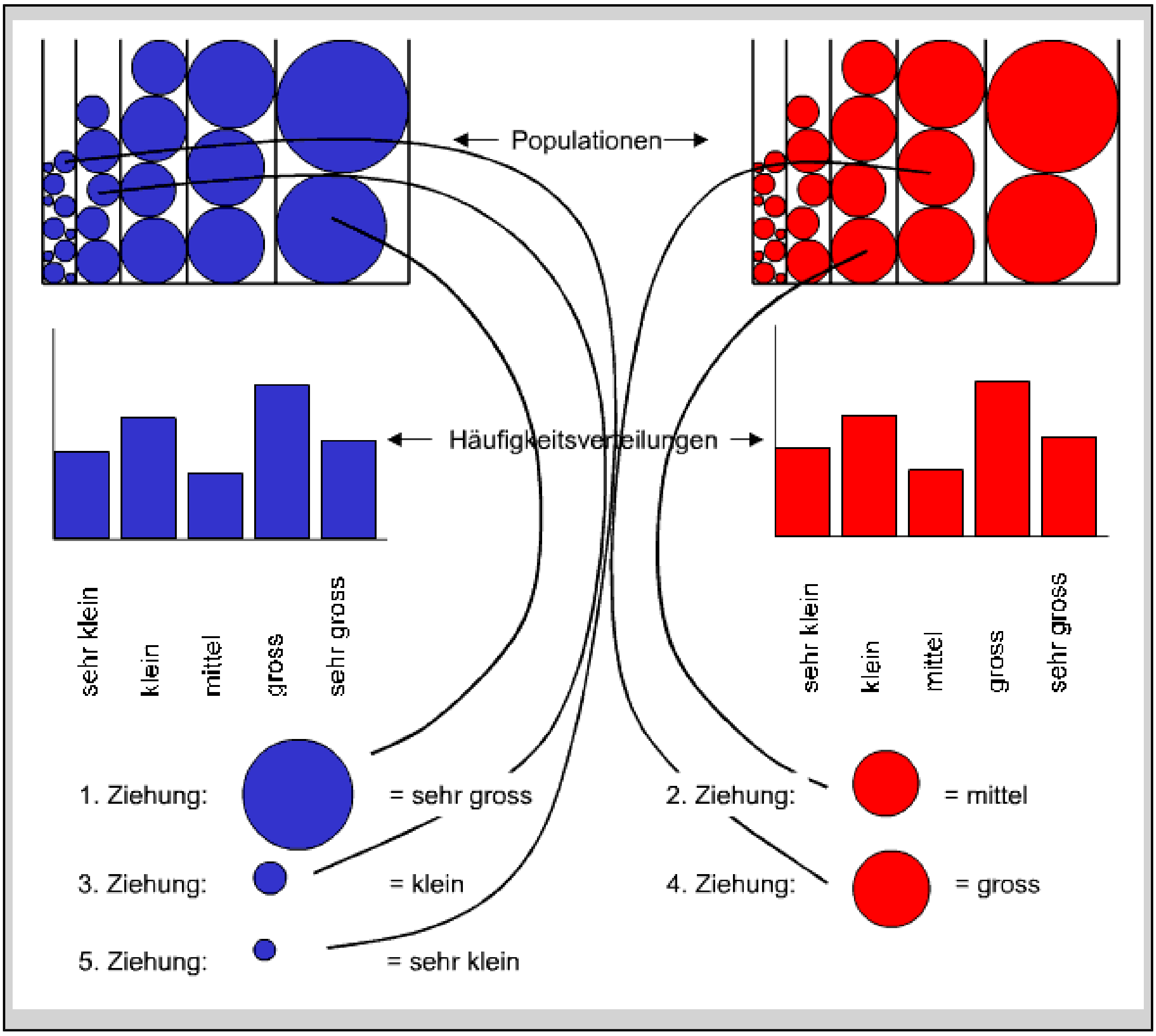
2. Werden die Kugeln pro Ausprägungskategorie ausgezählt, so erhalten wir die Häufigkeitsverteilungen.

Häufigkeitsverteilungen



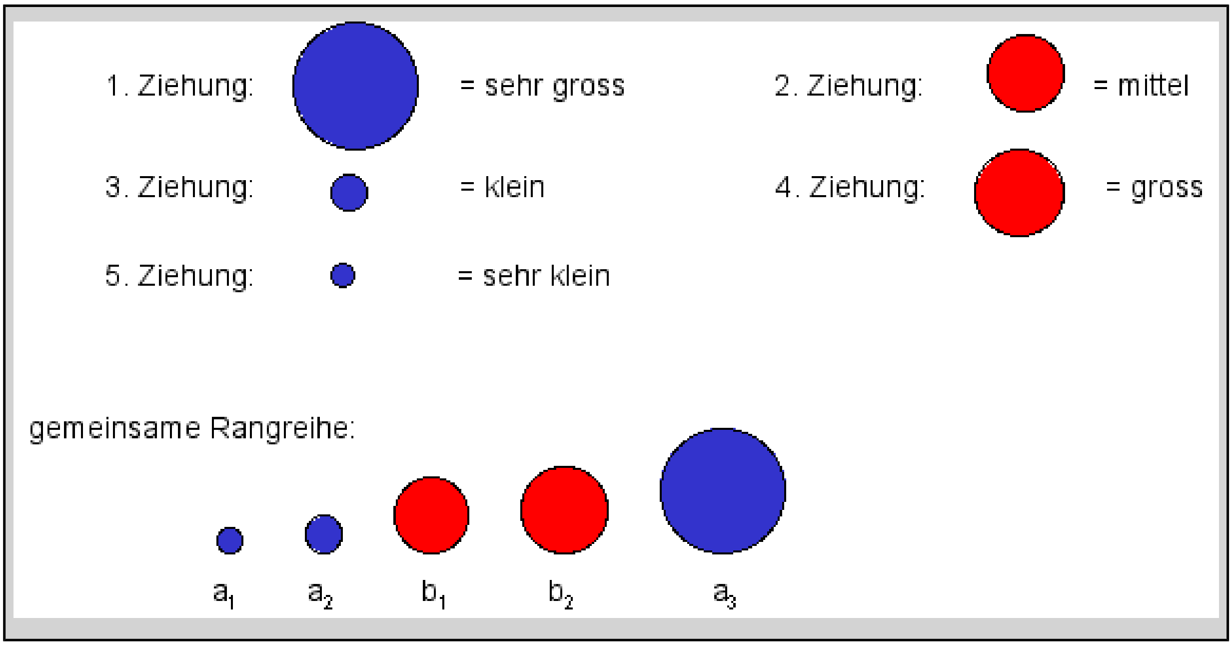
3. Aus den beiden Populationen wird nun je eine Zufallsstichprobe gezogen, aus der einen Population eine Stichprobe von 3 Elementen, aus der anderen eine Stichprobe von 2 Elementen. Die Ausprägungsgrade der zufällig gezogenen Elemente werden in einer Tabelle zusammengetragen.

Daten der beiden Zufallsstichproben



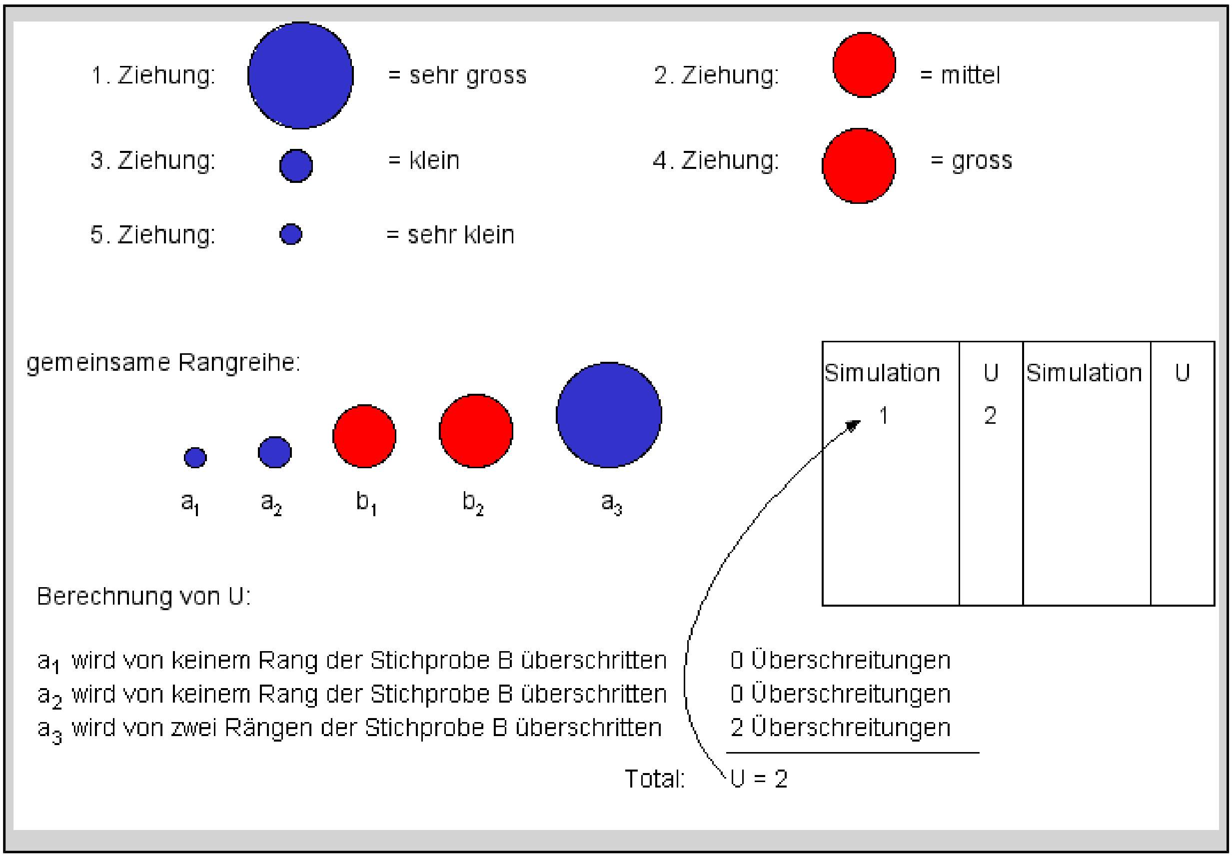
Nun bringen wir die Kugeln gemäss ihren Ausprägungsgraden in eine gemeinsame Rangreihe.

Gemeinsame Rangreihe der Stichprobendaten



Anhand dieser gemeinsamen Rangreihe bestimmen wir die Rangplatzüberschreitungen U und notieren den Ausprägungsgrad von U in einer Tabelle.

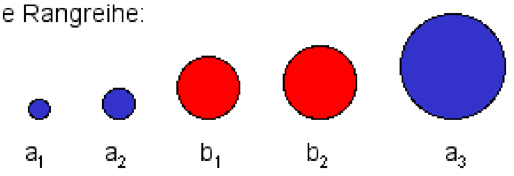
Berechnung von U



4. Wir legen alle Kugeln in die Gefässe zurück aus denen sie entnommen wurden, und wiederholen Schritt 3, d.h. wir ziehen wiederum zwei Zufallsstichproben.
5. Spielt man die Simulation mehrmals (theoretisch unendlich oft) durch und berechnet jedesmal U, so füllt sich unsere Tabelle mit den Ausprägungsgraden von U.

Tabelle der ermittelten Ausprägungen von U nach 12 Simulationen

gemeinsame Rangreihe:



Simulation	U	Simulation	U
1	2	7	5
2	3	8	4
3	6	9	3
4	2	10	6
5	3	11	0
6	4	12	1

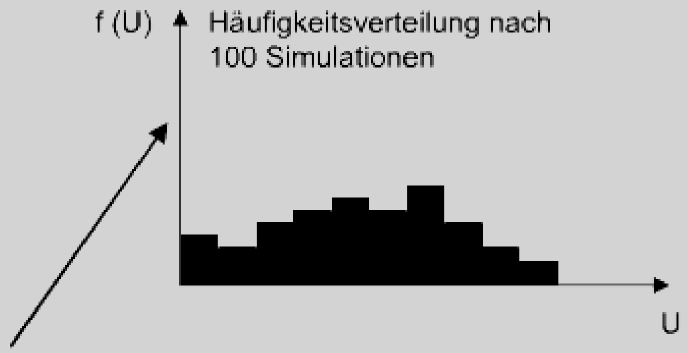
Berechnung von U:

a_1 wird von keinem Rang der Stichprobe B überschritten 0 Überschreitung
a_2 wird von keinem Rang der Stichprobe B überschritten 0 Überschreitung
a_3 wird von zwei Rängen der Stichprobe B überschritten 2 Überschreitungen
----
Total: U = 2

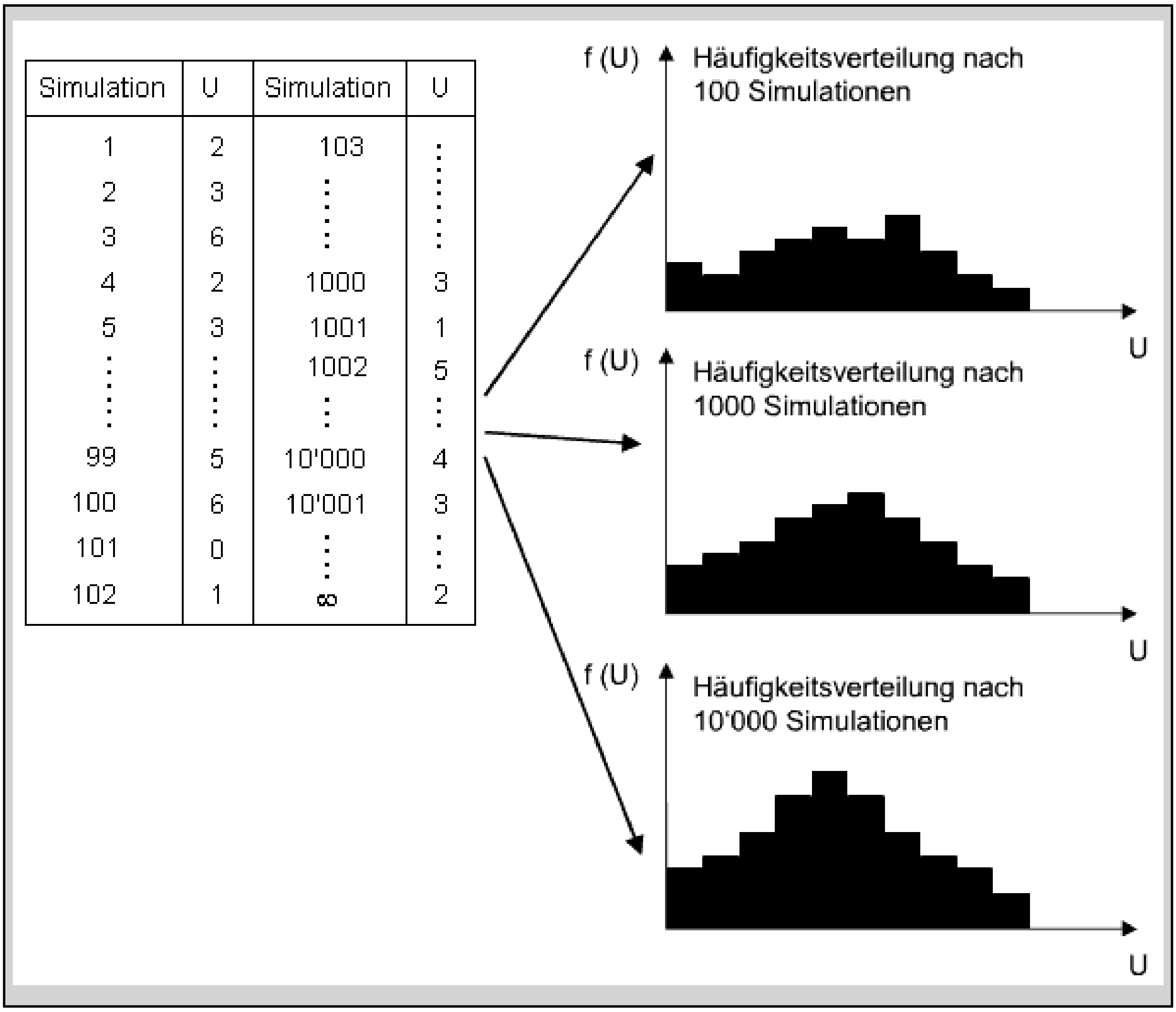
6. Die Daten unserer Tabelle können als Häufigkeitsverteilung dargestellt werden, wobei sich diese Häufigkeitsverteilung mit der Zahl der Simulationsdurchgänge verändert.

Häufigkeitsverteilung von U nach 100 Simulationen

Simulation	U	Simulation	U
1	2	103	⋮
2	3	⋮	⋮
3	6	⋮	⋮
4	2	1000	3
5	3	1001	1
⋮	⋮	1002	5
⋮	⋮	⋮	⋮
99	5	10'000	4
100	6	10'001	3
101	0	⋮	⋮
102	1	∞	2



Häufigkeitsverteilung von U nach 1000 resp. 10000 Simulationen



7. Dem aufmerksamen Beobachter wird kaum entgangen sein, dass bei einer sehr grossen Zahl von Simulationen die Ausprägungsgrade von U eingipflig und symmetrisch verteilt sind.
Die Statistiker nennen diese Verteilung eine U-Verteilung. Freundlicherweise haben sie für uns auch die Verteilungsparameter

μ_U

und

σ_U

bestimmt:

$$\mu_U = \frac{n_A \cdot n_B}{2}$$

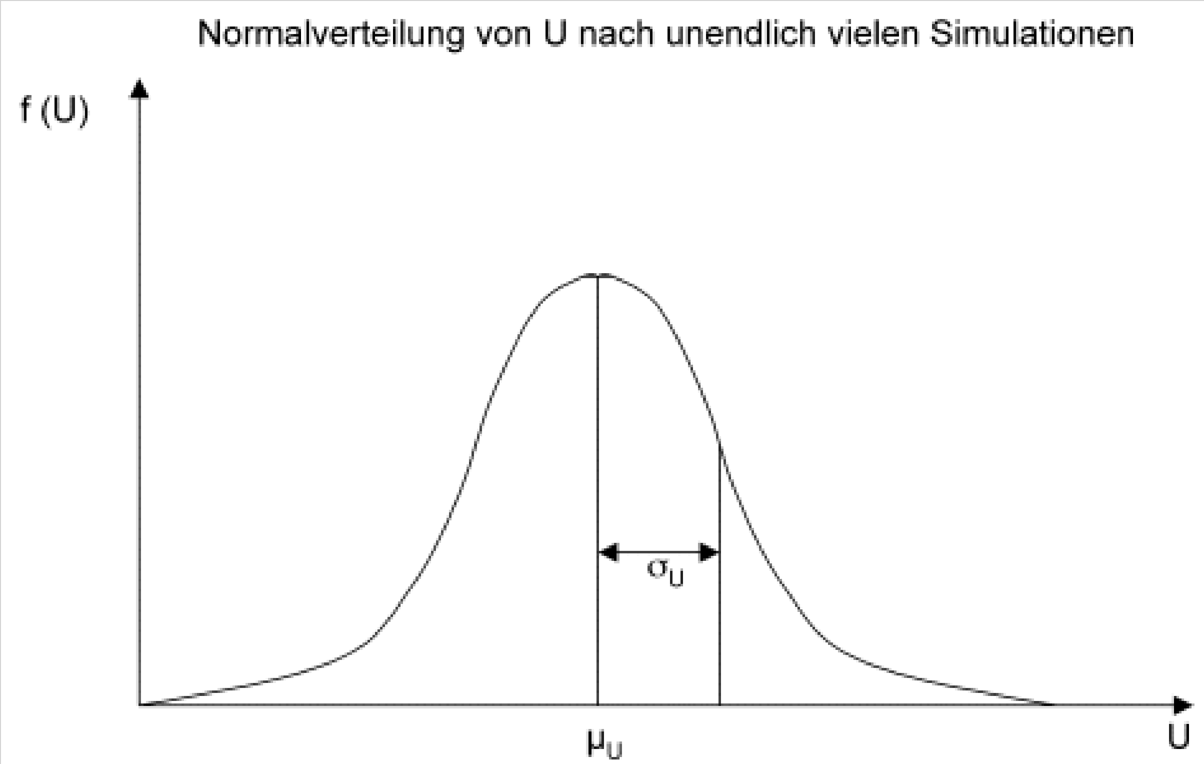
$$\sigma_U = \sqrt{\frac{n_A \cdot n_B \cdot (n_A + n_B + 1)}{12}}$$

Bei grösseren Stichproben, d.h. wenn mindestens eine der zu vergleichenden Stichproben mehr als 10 Elemente umfasst, geht die U-Verteilung über in eine Normalverteilung N

$$(\mu_{U,} \sigma_U)$$

Die Prüfverteilung für grössere Stichproben

Die Prüfverteilung für grössere Stichproben n_A # 10 oder n_B # 10



Damit haben wir unser Ziel erreicht! Wir wissen, wie die Prüfgrösse U resp. U' verteilt ist, wenn die beiden Stichproben aus Populationen stammen, in denen das Merkmal dieselbe zentrale Tendenz aufweist.
Liegt nun von einem Stichprobenvergleich ein konkreter Ausprägungsgrad von U resp. U' vor, so können wir anhand der Prüfverteilung entscheiden, mit welcher Wahrscheinlichkeit dieser Ausprägungsgrad rein zufällig auftreten könnte, wenn die beiden Stichproben aus Populationen mit identischer zentraler Tendenz stammen würden.
Diese Beurteilung eines konkreten Ausprägungsgrades der Prüfgrösse anhand der Prüfverteilung wollen wir im Rahmen unseres Beispiels kurz besprechen.

FAQ

8. Prüfverteilung und Unterschreitungswahrscheinlichkeit

Nachdem wir im letzten Abschnitt die Herleitung der Prüfverteilung in einem Gedankenexperiment nachvollzogen haben, möchten wir nun die für unsere Frage zum Skirennen relevante Prüfverteilung ermitteln. Wie wir gesehen haben, ist die Prüfgrösse U bzw. U' mit den Parametern

μ_U

und

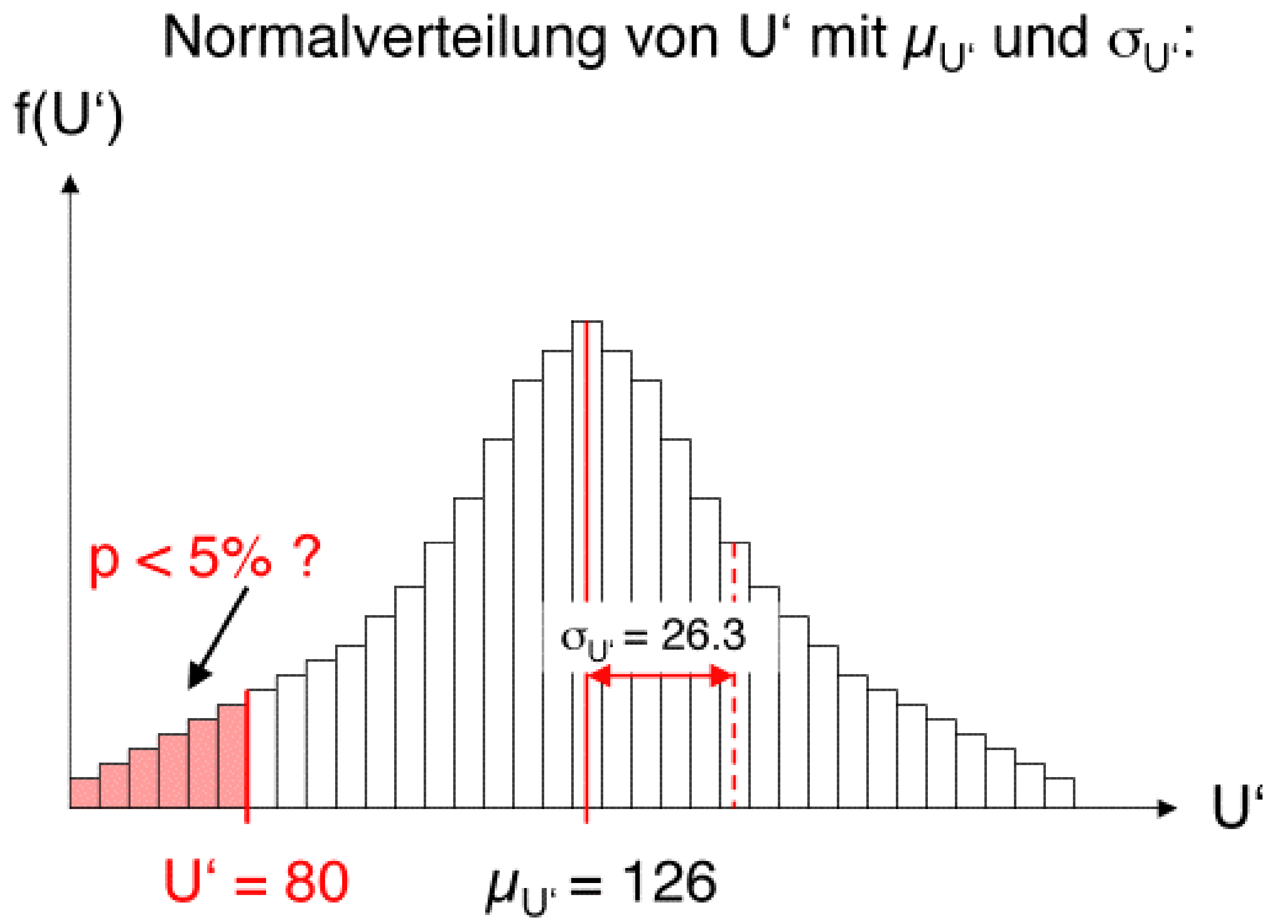
σ_U

U-verteilt. Falls mindestens eine der beiden Stichproben mehr als 10 Elemente umfasst, ist U resp U' mit denselben Parametern normalverteilt. In unserem Beispiel ist dies der Fall, beteiligten sich doch 14 Frauen und 18 Männer am 1. Lauf des Skirennens.
Wir bestimmen die Parameter **der Prüfverteilung**:

$$\mu_U = \frac{n_A \cdot n_B}{2} = \frac{14 \cdot 18}{2} = \frac{252}{2} = 126$$

$$\sigma_U = \sqrt{\frac{n_A \cdot n_B \cdot (n_A + n_B + 1)}{12}} = \sqrt{\frac{252 \cdot 33}{12}} = 26.3$$

Mit diesen Parametern ergibt sich die folgende Prüfverteilung:



Tragen wir den ermittelten Ausprägungsgrad unserer Prüfgrösse U´ in diese Verteilung ein (U´ = 80), so kann die gesuchte Unterschreitungswahrscheinlichkeit visualisiert werden.
Doch wie gross ist diese Unterschreitungswahrscheinlichkeit?
Da die Prüfgrösse bei unseren Stichprobengrössen mit den bekannten Parametern

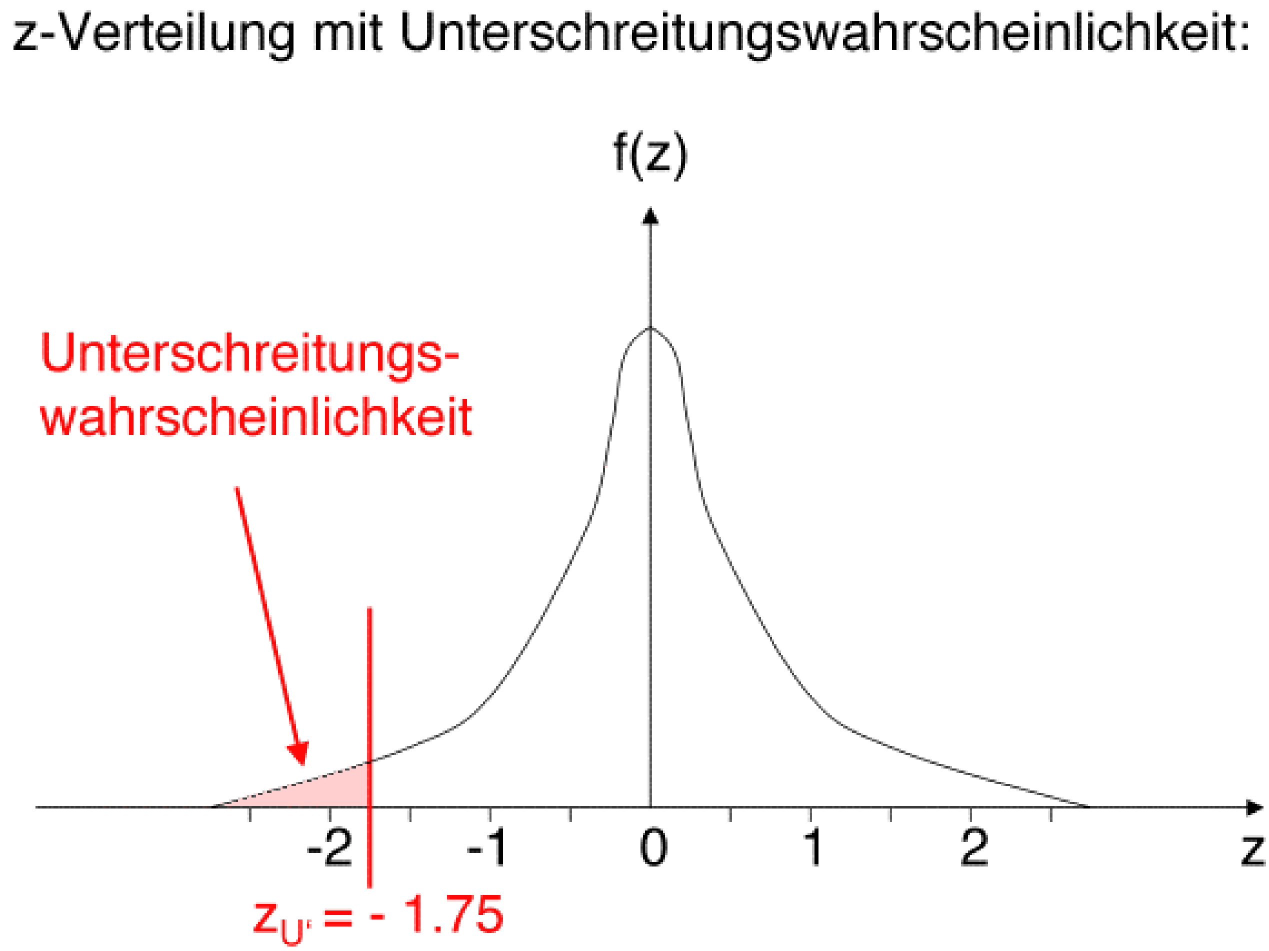
$\mu_{U'}$

und

σ_U

normalverteilt ist, klären wir diese Frage anhand der z-Verteilung, zu der Tabellen mit den Über- und Unterschreitungswahrscheinlichkeiten existieren.
Wir transformieren also U´ in die z-Verteilung.

$$z_U = \frac{U' - \mu_U}{\sigma_U} = \frac{80 - 126}{26.3} = -1.75$$



Unsere Prüfgrösse hat den Ausprägungsgrad U' = 80, der in die z-Verteilung transformierte Ausprägungsgrad beträgt

z_U

= - 1,75. Bestimmen Sie anhand der z-Verteilung die einseitige Unterschreitungswahrscheinlichkeit dieses Ausprägungsgrades.

❍ $p > 5$ %
❍ $p < 5$ %
❍ $p < 1$ %
❍ $p < 0.1$ %

9. Interpretation

Die Unterschreitungswahrscheinlichkeit unserer z-transformierten Prüfgrösse (

$z_{U'}$

) ist kleiner als 4,5 %. Dies bedeutet, dass unsere Prüfgrösse U' unter Annahme der Gültigkeit der Arbeitshypothese H_0 nur noch mit einer Wahrscheinlichkeit $p < 5\%$ erreicht oder unterschritten wird.

Diese Wahrscheinlichkeit ist so gering, dass wir die Arbeitshypothese H_0 auf dem 5 % - Signifikanzniveau zu Gunsten der Alternativhypothese H_1 ablehnen.

Wir wollen dieses Resultat stichwortartig interpretieren:

Hinsichtlich des Merkmals “Rangplatz im 1. Lauf” stammt die Stichprobe der Männer aus einer Population, die eine geringere zentrale Tendenz aufweist als die Population aus der die Stichprobe der Frauen stammt. Die Irrtumswahrscheinlichkeit p dieser Aussage ist kleiner als 5%.

Die Antwort auf die ursprüngliche Frage von Helen und Thomas lautet:

Mit einer Irrtumswahrscheinlichkeit kleiner 5% darf angenommen werden, dass die Stichprobe der Männer im 1. Lauf des Skirennens gesamthaft bessere Ränge erzielt hat als die Stichprobe der Frauen.

10. Auswertung der Daten mit SPSS

Zum Abschluss wollen wir unsere Daten auch mit Hilfe von SPSS auswerten. Zu diesem Zweck erstellen wir ein Datenfile mit dem Namen skirennen.sav, wobei wir die folgenden Variablen-Namen und Variablen-Labels wählen:

Variablen-Name	Variablen-Label	Code
geschl	Geschlecht	1=weiblich 2=männlich
name	Name	
ztlauf1	Rennzeit im 1. Lauf	
ztlauf2	Rennzeit im 2. Lauf	
ralauf1	Rangplatz im 1. Lauf	
ralauf2	Rangplatz im 2. Lauf	

Für die Auswertung geben wir die folgenden Befehle ein:

1. GET FILE 'X:\\SPSS_DAT**skirennen.sav**'.
   LIST **geschl, name, ztlauf1, ztlauf2, ralauf1, ralauf2**.
   Wir lesen unser Datenfile mit dem Namen **skirennen.sav** ein und lassen uns eine Liste der Rohdaten ausgeben. (**geschl, name, ztlauf1, ztlauf2, ralauf1, ralauf2**).

SPSS output 1

2. NPAR TESTS
   /M-W= **ralauf1** BY **geschl(1 2)**
   /MISSING ANALYSIS.

Wir verlangen einen nichtparametrischen Test (NPAR), konkret den U-Test nach Mann-Whitney (M-W). Das interessierende Merkmal ist der "Rangplatz im 1. Lauf" (**ralauf1**) und das die Stichproben definierende Merkmal ist das "Geschlecht" (**geschl**) mit den beiden möglichen Ausprägungen 'weiblich' und 'männlich' (1 und 2).

SPSS output 2

Der SPSS-Ausgabe entnehmen wir die für uns relevanten Kenngrössen: Es sind dies die Grössen der beiden Stichproben, die mittleren Ränge in den beiden Stichproben, der Ausprägungsgrad von U' und die zweiseitige Unter- / Überschreitungswahrscheinlichkeit des ermittelten Ausprägungsgrades von U' im Rahmen der Prüfverteilung.

Die mittleren Ränge der Stichproben geben einen Hinweis, welche Stichprobe tendenziell die besseren Ränge erzielt hat. Der mittlere Rang der Frauen beträgt 19.79, während die Männer mit 13.94 im Mittel besser plaziert sind.

Die Prüfgrösse hat denselben Ausprägungsgrad wie in unserer manuellen Berechnung, die zweiseitige Unter- / Überschreitungswahrscheinlichkeit beträgt 0.081.

Da wir aber eine einseitige Fragestellung bearbeiten (ist die zentrale Tendenz der Ränge in der Stichprobe der Männer geringer als in der Stichprobe der Frauen?), interessiert uns die einseitige Unterschreitungswahrscheinlichkeit. Für eine symmetrische Prüfverteilung - und die Normalverteilung ist symmetrisch - entspricht die einseitige Unterschreitungswahrscheinlichkeit der Hälfte der zweiseitigen Unter- / Überschreitungswahrscheinlichkeit. Sie beträgt somit 0.04, ist also geringer als 5%.

11. Zusammenfassung zum Lernschritt

Lassen Sie uns nun kurz überdenken, was wir in diesem Lernschritt bearbeitet haben:

1. Ausgangspunkt war eine konkrete Fragestellung: Sind die Männer im 1. Lauf eines Firmenskirennens besser gefahren als die Frauen?
2. Es stellte sich heraus, dass diese Frage nicht über einen Vergleich der mittleren Rennzeiten in den Stichproben von Männern und Frauen, d.h. anhand eines t-Tests, beantwortet werden kann, da die Voraussetzung des t-Tests "Normalverteilung der Daten in den Populationen" nicht erfüllt ist. Gesucht ist damit ein verteilungsfreies / nichtparametrisches Verfahren.
3. Nachdem wir uns in Erinnerung gerufen haben, was ganz grundsätzlich zu jedem entscheidungsstatistischen Verfahren gehört, folgten wir Mann und Whitney bei der Erfindung der Prüfgrösse U resp. U' und der Ermittlung der zugehörigen Prüfverteilung. Eingeprägt haben wir uns dabei die Tatsache, dass der U-Test nur die Ranginformationen der Daten auswertet, d.h. dass er für ordinal, proportional und intervall-skalierte Daten eingesetzt werden kann und bezüglich der Verteilung der Daten in den Populationen nichts voraussetzt.
4. Mit Hilfe des U-Tests, den wir zum Schluss auch noch mit Hilfe von SPSS durchführten, konnten wir die eingangs gestellte Frage beantworten

Wir schliessen diesen Abschnitt zum U-Test mit einem kleinen Quiz.

1. Der U-Test kann nur für intervall- oder proportional skalierte Daten eingesetzt werden.

❍ Stimmt

❍ Stimmt nicht

2. Der U-Test kann für alle Datenniveaus eingesetzt werden

❍ Stimmt

❍ Stimmt nicht

3. Der U-Test setzt bezüglich der Verteilung der Daten in den Populationen nichts voraus.

❍ Stimmt

❍ Stimmt nicht

Der U-Test 'verwertet' nur die Ranginformation der Daten. Dies auch dann, wenn das interessierende Merkmal intervall- oder proportional skaliert ist.

❍ Stimmt

❍ Stimmt nicht

5. Die Parameter μ_U und $\#_U$ beschreiben die Verteilung des interessierenden Merkmals in der Population.

❍ Stimmt

❍ Stimmt nicht

$N(\mu_U, \#_U)$ ist die Prüfverteilung für grössere Stichproben (d.h. wenn mindestens eine der Stichproben mehr als 10 Elemente umfasst).

❍ Stimmt

❍ Stimmt nicht