Binomialtests, Chiquadratverfahren

1. a. Bestimmen Sie den **kritischen Bereich für die Anzahl** der Dreier bei 12 Würfen für folgende Behauptungen über den Anteilswert (der Dreier):

H_O: $p \leq 1/6$ gegen die logische H_1

wenn die WK dafür, bei einem fairen Würfel in diesen kritischen Bereich zu kommen, höchstens 12% betragen darf.

b. Sie haben mit einem Würfel drei Mal gewürfelt und genau 4 (genau 7) Dreier erhalten. Was schließen Sie daraus?

Lösung:a. K = {5, 6, …, 12} b. „Würfel ist nicht (mit hoher Sicherheit) unfair“

2. Unter den 22 WELSCH-Läufern einer früheren Aufgabe (die Stichprobenrealisation von Bsp.2, Blatt 3) waren nur wenige aus dem Ausland.

a. ***Lässt sich*** damit die Behauptung (Also die Gegenhypothese $\mathbf{H_1}$):
"Der Ausländeranteil liegt unter 30 Prozent“ signifikant ***bestätigen***? (Niveau $\alpha = 0.05$).
Bestimmen Sie den Kritischen Bereich für die Testgröße. Wie ist auf Grund der vorliegenden Daten zu entscheiden?

b. Wie groß ist die WK für die ungerechtfertigte Akzeptanz der H_O, **(die WK für einen Fehler 2. Art)**, wenn der tatsächliche Ausländeranteil **20** Prozent beträgt, H_1 also wahr ist?

Lösung: a: Antwort: Nein; b: 84.5% (!)

3. Eine Serie von 72 Würfen mit einem (fairen?) Würfel ergab die Augenzahlen 1 bis 6 mit den angegebenen Häufigkeiten::

1	2	3	4	5	6
19	8	10	12	5	18

a. Formulieren sie in Form einer Verteilung die *Nullhypothese* für "Der Würfel ist fair"
b. Berechnen sie die *Testgröße* für den χ^2 -Test aus den beobachteten Häufigkeiten
c. Geben sie den *Kritischen Bereich* für die Testgröße zu den Testniveaus $\alpha = 0.1$ und $\alpha = 0.01$ an und treffen sie die *Entscheidung* über die „Fairness“ des Würfels!

Lösung: Testgröße = 12.83

Zusatzaufgabe: Würfeln sie selbst tatsächlich 72 mal und prüfen Sie Ihren Würfel analog auf Fairness!

4. Die folgenden 36 Zahlenwerte sollen mittels **Chiquadrat-Test** daraufhin untersucht werden, ob sie einer **normalverteilten Grundgesamtheit** entstammen.

1.2, 2.0, 1.5, 1.8, 1.6, 1.9, 2.1, 1.1, 1,3, 1.4, 1.6, 2.0, 2.1, 1.4, 1.6, 1.8, 1.9, 2.0
1.9, 1.8, 1.6, 1.9, 1.3, 0.9, 1.5, 1.6, 1.2, 1.7, 1.4, 0.8, 1.1, 1.6, 2.2, 1.8, 1.9, 1.9

a. Schätzen sie dazu vorerst Erwartungswert und Standardabweichung
b. Bilden sie nun **vier** Klassen mit den **Klassengrenzen: 1.422; 1.622; 1.822**
(Diese Grenzen liegen symmetrisch um den geschätzten Erwartungswert)
und berechnen sie die somit zu erwartenden Anzahlen an Beobachtungen je Klasse
c. Führen sie den *(modifizierten) Anpassungstest* durch; Testniveaus $\alpha = 0.01$ und 0.1
Berechnen Sie die Testgröße und beantworten Sie:
„Kann die Normalverteilung in der Grundgesamtheit akzeptiert werden?“

Lösung c: Testgröße =1.365

Nicht vergessen: rechtzeitig zur Prüfung anmelden! Nur über UGO! Sie erhalten die online-Anmeldebestätigung!
SIE ALLEINE SIND DAFÜR VERANTWORTLICH!

5. Unter 220 Studierenden der Soziologie waren Raucher/Nichtraucher wie angegeben verteilt:

	Raucher	Nichtraucher
Männer	46	41
Frauen	61	72

Lässt sich daraus auf unterschiedliches Rauchverhalten von Männern und Frauen schließen? Führen Sie dazu einen Vierfeldertest mit einem Testniveau Ihrer Wahl durch und beantworten Sie die gestellte Frage präzise!

Lösung: Testgröße = 1.034

6. Zum Zusammenhang zwischen Geschlecht und sportlicher Aktivität wurden Studierende befragt: „Wie oft treiben Sie Sport" und man erhielt Antworten gemäß folgende Tabelle:

	Nie	selten	ziemlich oft	täglich
Männer	34	79	234	43
Frauen	50	61	158	41

a. Bestimmen Sie eine Kennzahl, um die Stärke der Abhängigkeit der beiden Merkmale zu beschreiben!

b. Lässt sich zeigen, dass Geschlecht und sportliche Aktivität voneinander abhängig sind? ***Testen*** Sie zu den Niveaus $\alpha = 0.1$ und $\alpha = 0.005$ und ***interpretieren*** Sie genau!

Lösung: c = 0.117   Testgröße = 11.1

7. Um eine möglicherweise vorhandene Abhängigkeit zwischen höchstem Schulabschluss und regionaler Herkunft aufzudecken, wurden dreißigjährige Personen befragt und man erhielt folgende Ergebnisse in Form einer Tabelle: (Hinweis: vgl. Aufgabe 12, Blatt 1)

	Schulabschluss		
Herkunft:	Universität oder FH	Matura	Hauptschule plus Lehrabschluss
Wien, Linz, Graz	46	78	53
Andre Bezirkshauptstädte	40	30	79
Sonstige Gemeinden	24	72	68

a. Berechnen sie **die quadratische Kontingenz**

b. Sind generell Herkunft und Schulabschluss voneinander abhängig? Testen Sie zu $\alpha = 0.05$! Formulieren sie jeweils Null- und Gegenhypothese und interpretieren sie die Ergebnisse!

Lösung b: Testgröße = 25.22 Antwort: JA

8. Zu den Daten der WELSCH-Läufer (Blatt 3, Aufgabe 2):
Kann man die Behauptung, das Einkommen und die Herkunft seien voneinander unabhängig, beibehalten? Testen Sie zum Niveau 0.1. Warum dürfen Sie diesen Test mit den vorliegenden Daten eigentlich nicht durchführen?

Antwort: Der Test sagt: NEIN, ist aber nicht zulässig

Nicht vergessen: rechtzeitig zur Prüfung anmelden! Nur über UGO! Sie erhalten die online-Anmeldebestätigung!
SIE ALLEINE SIND DAFÜR VERANTWORTLICH!

Zweistichprobentests, parametrisch und nichtparametrisch

9. Wieder zu den Daten der WELSCH-Läufer (die Stichprobenrealisation von Bsp.2, Blatt 3).
Fassen Sie die erhobenen ***Gewichte der Steirer*** (erste Datenreihe) und der ***anderen Teilnehmer*** (zweite Datenreihe) als zwei (unabhängige) Stichproben auf und testen Sie, *unter Annahme normalverteilter Merkmale und Varianzhomogenität*, folgende Hypothesen (Testniveau jeweils $\alpha = 0.1$):

a. H_0: $\mu_1 = \mu_2$ (d. h.. H_1: die beiden Erwartungswerte sind verschieden)
b. H_1: $\mu_1 > \mu_2$ (d. h. also H_0: $\mu_1 - \mu_2 \geq 0$)

c. H_1: „Die Steirer sind um mindestens zehn (mindestens zwölf) kg schwerer"
Formulieren und **interpretieren Sie die Ergebnisse in Worten!**

Lösung: Testgröße zu a. und b.: 3.77 Testgröße zu c: 1.62 (1,19)

10. Nochmals die Daten der Läufer:
Nun wird das ***Alter*** der ***Steirischen Teilnehmer*** (X) mit jenem der ***Ausländer*** (Y) verglichen
Es soll getestet werden, ob/welcher Unterschied in den Erwartungswerten der zugrunde-liegenden Normalverteilungen vorliegt.

a. Ist die Voraussetzung für den t-Test gegeben? D. h.: Kann man „Gleichheit der Varianzen" akzeptieren?
b Lässt sich die Behauptung, die Steirer seien jünger ("Der Erwartungswert von X ist kleiner als jener von Y") signifikant nachweisen?
c. Macht es Sinn, die Behauptung:
"Die Steirer sind um mindestens drei Jahre jünger" nachweisen zu wollen?

Beachten Sie immer die Wahl von H_0 und H_1 und testen Sie unter Annahme normalverteilter Grundgesamtheiten zum Niveau $\alpha = 0.1$

Lösung a: Testgröße = 0.47, JA b. |Testgröße| = 0.98, c. NEIN

11. Nochmals Daten von Läufern. Jetzt die ***Zeiten*** der ***steirischen Läufer 2001 und 2002*** (In Minuten!)

a. Führen sie den *zweiseitigen t-Test für die Differenz der beiden Erwartungswerte* unter Annahme von Normalverteilungen zu $\alpha = 0.1$ durch.
b. Führen Sie einen *einseitigen* Unterschiedstest (Niveau 0.1) durch
c. Formulieren sie jeweils Null- und Gegenhypothese und interpretieren sie die Ergebnisse!

Lösung: t_0 = 1.1028, keine Unterschiede nachweisbar

12. An 10 Messstellen im Stadtbereich wurde die Staubbelastung gemessen, und zwar wurden sowohl die Montagsmittelwerte als auch die Sonntagsmittelwerte erhoben. Man erhielt folgende Daten (PM10 in µg/m³, zum einfacheren Rechnen ganzzahlig gerundet):

Messstelle Nr	1	2	3	4	5	6	7	8	9	10
Montagsmittel	12	44	47	9	15	51	22	12	38	30
Sonntagsmittel	13	25	26	12	12	41	14	15	20	22

a. Lässt sich aus diesen Daten mit einer Signifikanz von 90 Prozent, d. h. zum Testniveau $\alpha = 0.1$, zeigen, dass die Staubbelastung sonntags geringer ist?

Hat eine Veränderung des Testniveaus auf 0.01 einen Einfluss auf Ihre Entscheidung? Interpretieren Sie Ihre Ergebnisse präzise!

b. Macht es Sinn, zum Niveau $\alpha = 10$ Prozent zu prüfen, ob die Staubbelastung montags um mehr als fünf µg/m³ größer ist als an Sonntagen?
Lässt sich diese Vermutung nun signifikant bestätigen?
(Rechnen sie ***unter Annahme normalverteilter*** Belastung)

Lösung: a. Testgröße = 2.786 b. Sinnvoll Ja, Testgröße 0 1.045

13. In einem Betrieb wurden für 7 Lehrlinge jeweils die Durchschnittsnote des Abschlusszeugnisses aus der Schule und eine im Betrieb ermittelte Leistungskennzahl miteinander verglichen. (gleiche Daten wie Aufgabe 8, Blatt 1)
Testen Sie ***unter Normalverteilungsannahme*** („parametrisch“):

Lehrling	1	2	3	4	5	6	7
Notenschnitt	2.6	2.2	1.9	2.4	3.3	1.3	1.8
L-Kennzahl	61	79	82	72	66	89	71

Lässt sich mit diesen Daten folgende Behauptung „bestätigen“ (Testniveau $\alpha = 0.05$)
Gute Noten entsprechen auch guten Leistungen im Betrieb“

Antwort: Ja

****

14. Wieder zu den Daten der WELSCH-Läufer (die Stichprobenrealisation von Bsp.2, Blatt 3).
Fassen Sie die erhobenen ***Gewichte der Steirer*** (erste Datenreihe) und der ***anderen Teilnehmer*** (zweite Datenreihe) als zwei (unabhängige) Stichproben auf und testen Sie folgende Hypothesen (Testniveau jeweils $\alpha = 0.1$):

a. H_0: Die Gewichte sind gleich
b. H_1: Die Steirer sind schwerer (d. h. also H_0: $\mu_1 - \mu_2 \geq 0$)

Führen sie jetzt jeweils den ***Wilcoxon-Rangsummentest*** mit der geeigneten Hypothesenformulierung durch (Niveau 0.05)

15. In fünf bzw. sieben Gemeinden im Burgenland bzw. in der Steiermark wurden die Pro-Kopf-Ausgaben der Gemeinde für Straßenreinigung/Streudienst ermittelt.
Folgende Datenzeilen liegen vor:

Burgenland	182	240	291	280	220		
Steiermark	302	270	240	301	282	288	283

Es soll getestet werden, ob/welcher Unterschied in den Ausgaben zur Straßenreinigung vorliegt
Lässt sich die Behauptung "Die Ausgaben in der Steiermark sind höher " signifikant nachweisen?

Beachten Sie immer die Wahl von H_0 und H_1 und testen Sie zum Niveau $\alpha = 0.1$!

a. Unter ***Normalverteilungsannahme*** bei vorausgesetzter ***Varianzhomogenität***
b. ***Ohne*** Annahme normalverteilter Grundgesamtheiten(„nichtparametrisch“)

Lösung: a. |Testgröße|= 2.004 b. |Testgröße|= 1.624

16. Bei zehn Blutproben wurde je einmal durch den neuen „Schnell-Alkomat“ und einmal durch einen Labortest der Alkoholgehalt bestimmt. Man erhielt folgende Werte:

Schnell-Alkomat	.58	.73	.84	.89	1.10	.60	.69	.85	.79	1.03
Labor	.82	.83	.98	1.10	1.22	.38	.71	.77	.81	1.60

a. Testen Sie zum Niveau $\alpha = 0.05$, ***ohne*** ***Annahme von Normalverteilungen***, die Nullhypothese: „Die beiden Verfahren liefern die selben Ergebnisse“

b. Zeigt der Schnell-Alkomat signifikant zu niedrige Werte an?

Lösung: b.: Ja

17. In Graz wurden 23, in Wien 45 dreiköpfige Familien nach der Größe ihrer Wohnung befragt. Die errechnete mittlere Wohnungsgröße betrug in Graz 103.5 und in Wien nur 86.2 m². Das legt den Schluss nahe, dass die Wohnungen in Graz im Wesentlichen größer sind als in Wien. Ist dieser Schluss gerechtfertigt?
Da die erhobenen Zahlen einem Test auf Normalverteilung nicht standhielten, ist ein ***nichtparametrisches Verfahren*** zu verwenden.

a. Welcher Test ist also durchzuführen? Formulieren Sie die Hypothesen formal und auch in Worten.

b. Werden alle 68 Wohnungen der Größe nach geordnet und die Ränge der Wiener Wohnungen addiert, ergibt sich die Zahl **1612**
Lässt sich damit die Behauptung "Die Wohnungen in Graz sind größer" signifikant (Niveau 0.05) bestätigen?

Lösung: **(Obwohl die mittlere Größe der Grazer Wohnungen deutlich über jener in Wien liegt)** ***Nein!***

18. In einem Betrieb wurden für 7 Lehrlinge jeweils die Durchschnittsnote des Abschluss-zeugnisses aus der Schule und eine im Betrieb ermittelte Leistungskennzahl miteinander verglichen.

Lehrling	1	2	3	4	5	6	7
Notenschnitt	2.6	2.2	1.9	2.4	3.3	1.3	1.8
L-Kennzahl	61	79	82	72	66	89	71

Lässt sich mit diesen Daten folgende Behauptung „bestätigen“ (Testniveau $\alpha = 0.05$)
„Gute Noten entsprechen auch guten Leistungen im Betrieb“ ?
Testen Sie unter ***Verwendung der Rangkorrelation***!

Lösungshinweis: Testgröße = –1.837

Hinweis: *Benützen Sie für alle nichtparametrischen Aufgaben obwohl nicht korrekt, die Methode, die erst bei größeren Probenumfängen angewandt werden sollte (Benützung der Tabelle der Normalverteilung!)*

Nicht vergessen: rechtzeitig zur Prüfung anmelden! Nur über UGO! Sie erhalten die online-Anmeldebestätigung!
SIE ALLEINE SIND DAFÜR VERANTWORTLICH!