# Aufgabe 1

**A**

Statistix plant mit seiner Freundin Statistika im nächsten Sommer eine 32–tägige Reise nach China zu machen mit der Option, die Reise eventuell um eine Woche zu verlängern. Beide wissen noch nicht, ob sie sich die Verlängerungswoche finanziell überhaupt leisten können. Die Reise soll nur dann verlängert werden, wenn die Reisekosten pro Tag durchschnittlich weniger als 30 € betragen. Sie wollen auf jeden Fall vermeiden, durch eine Verlängerung der Reise in finanzielle Schwierigkeiten zu kommen. Statistix möchte die Entscheidung, ob die Reise verlängert werden kann, vom Ergebnis eines statistischen Tests ( $\alpha = 2{,}5$ % ) abhängig machen.
Als Datenbasis zur Testdurchführung sollen die täglich anfallenden Kosten der Reise dienen ( einfache Stichprobe ).

Weiterhin geht Statistix davon aus, dass die Varianz der Reisekosten pro Tag 2 $€^2$ beträgt.

1) Wie lautet das hier zu verwendende ( parametrische ) Testverfahren ?
2) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !
3) Wie lautet formal und verbal die Prüfgröße für diesen Test, und wie ist diese unter $H_0$ verteilt ?
4) Wie lautet die Testfunktion für diesen Test, und wie ist diese unter $H_0$ verteilt ?
5) Bestimmen Sie den Annahme– und Verwerfungsbereich dieses Tests !
6) Berechnen Sie die Wahrscheinlichkeit einer Fehlentscheidung bei diesem Test, wenn in Wahrheit die durchschnittlichen Reisekosten pro Tag
   a) 30 Euro und 50 Cent betragen !
   b) 29 Euro und 30 Cent betragen !
7) Bei welchem Eurobetrag der wahren durchschnittlichen Reisekosten pro Tag beträgt die Wahrscheinlichkeit einer berechtigten Ablehnung der Nullhypothese 87,9 % ?

Angenommen, am Ende der 32–tägigen Reise würden die beiden feststellen, dass sie insgesamt 944 Euro ausgegeben hätten.

8) Wie lautet die Testentscheidung ?
9) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt ! Werden die beiden also die Reise um eine Woche verlängern ?

**B**

Während ihrer Reise verbringen Statistix und seine Freundin 8 Tage in einem Reservoir für Pandabären und helfen dort bei der dreimal täglich stattfindenden Fütterung der Pandas. Kung Fu ist ein männlicher Panda und bekannt für seine spontanen „ Beiß–Attacken “ während mancher Fütterungen. Die verantwortlichen Tierpfleger überlegen, ob sie diesen Panda in ein anderes Reservoir umsiedeln, falls sich das Verhalten von Kung Fu nicht verbessert.
Statistix will den Tierpflegern bei ihrer Entscheidung mit Hilfe eines statistischen Tests ( $\alpha = 5$ % ) helfen. Wenn sich statistisch absichern lässt, dass Kung Fu in Wahrheit bei weniger als 40 % der Fütterungen beißt, darf er in seiner gewohnten Umgebung bleiben.
Als Datenbasis zur Testdurchführung sollen alle Fütterungen während des Aufenthalts von Statistix und Statistika in diesem Reservoir dienen ( einfache Stichprobe ).

1) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !
2) Wie lautet formal und verbal die für diesen Test geeignete Stichprobenfunktion, und wie ist diese unter $H_0$ verteilt ?
3) Bestimmen Sie den Annahme– und Verwerfungsbereich dieses Tests !

Angenommen, Kung Fu würde bei 5 dieser Fütterungen um sich beißen.

4) Wie lautet die Testentscheidung ? Darf Kung Fu demnach in seiner gewohnten Umgebung bleiben ?
5) Welcher Fehler ist bei der Testentscheidung unterlaufen ?

Angenommen, Statistix und seine Freundin würden ihren Aufenthalt im Reservoir um 7 Tage verlängern und die Fütterungen ihrer gesamten Aufenthaltsdauer als Datenbasis zur Testdurchführung nutzen.

6) Wie ist jetzt die für diesen Test geeignete Stichprobenfunktion unter $H_0$
   a) exakt verteilt ? [ Verteilungstyp und –parameter ! ]
   b) approximativ verteilt ? [ Verteilungstyp und –parameter ! ]

# Aufgabe 2

Jeder Schüler eines Internats wird nach einem Eingewöhnungszeitraum vom Internatsleiter Statistix gemäß bestimmter Eigenschaften eingeschätzt und einer der vorhandenen vier Unterkünfte ( **Löwen**–Haus, **Adler**–Haus, **Dachs**–Haus und **Schlangen**–Haus ) zugeteilt. Dort verbleibt dann der Schüler bis zum Ende seiner Schulzeit. Im Rahmen des Unterrichts und auch durch außerschulische Aktivitäten können die Schüler durch gute Leistungen und vorbildhaftes Verhalten Pluspunkte für das ihnen zugewiesene Haus sammeln.
Nach einem bestimmten Zeitraum will sich Statistix die erreichten Punkte von einigen Schülern in jeder der vier Haus–Kategorien anschauen. Dabei geht er davon aus, dass die so erhaltene Datenbasis aus vier untereinander unabhängigen und jeweils einfachen Stichroben ( mit den Umfängen : $n_L = 16$ ; $n_A = 18$ ; $n_D = 18$ ; $n_S = 17$ ) besteht.
Mit Hilfe eines statistischen Test ( $\alpha = 5\ \%$ ) möchte Statistix untersuchen, ob sich die erreichten Punktzahlen der Haus–Kategorien signifikant voneinander unterscheiden.
Statistix erhält folgendes Datenmaterial :

| Schüler Nr. pro Haus | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Löwe** | 5 | 5 | 5 | 6 | 6 | 7 | 7 | 7 | 7 | 8 | 8 | 10 | 10 | 13 | 14 | 14 | | |
| **Adler** | 1 | 1 | 3 | 3 | 4 | 4 | 5 | 5 | 5 | 6 | 7 | 10 | 10 | 11 | 11 | 12 | 12 | 14 |
| **Dachs** | 0 | 1 | 1 | 1 | 2 | 3 | 4 | 4 | 5 | 5 | 6 | 7 | 10 | 10 | 10 | 11 | 11 | 12 |
| **Schlange** | 2 | 2 | 3 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 6 | 6 | 8 | 14 | 14 | 14 | |

**A**

Für die folgende Testdurchführung werden die erreichten Punktzahlen in jeder Haus–Kategorie in drei Gruppen zusammengefasst : [ 0 – 5 ) : [ 5 – 10 ) ; [ 10 – 15 )

1) Wie lautet das hier zu verwendende ( nicht–parametrische ) Testverfahren ?
2) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !
3) Wie lautet die Testfunktion für diesen Test, und wie ist diese unter $H_0$ verteilt ?
4) Bestimmen Sie den Annahme– und Verwerfungsbereich dieses Tests !
5) Wie lautet die Testentscheidung ?
6) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

**B**

Auf Basis des erhaltenen Datenmaterials der Urliste möchte Statistix weiterhin mit Hilfe eines statistischen Tests ( $\alpha = 5\ \%$ ) versuchen, statistisch nachzuweisen, dass die durchschnittlich erreichte Punktzahl eines Schülers des **Löwen**–Hauses signifikant größer als die durchschnittlich erreichte Punktzahl eines Schülers des **Schlangen**–Hauses ist.
Dabei geht er davon aus, dass die erreichte Punktzahl eines Schülers normalverteilt und die Varianz der erreichten Punktzahl eines Schülers für jede Haus–Kategorie gleichgroß ist.

1) Wie lautet das hier zu verwendende ( parametrische ) Testverfahren ?
2) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !
3) Wie lautet die Testfunktion für diesen Test, und wie ist diese unter $H_0$ verteilt ?
4) Bestimmen Sie den Annahme– und Verwerfungsbereich dieses Tests !

**Rechenhilfe :** Die Stichprobenvarianz der erreichten Punktzahlen für das **Löwen**–Haus beträgt : 9,533

Die Stichprobenvarianz der erreichten Punktzahlen für das **Schlangen**–Haus beträgt : 16,029

5) Wie lautet die Testentscheidung ?
6) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

# Aufgabe 3

**A**

Statistix befindet sich derzeit in der Tschebyscheffschen Republik und hat dort ein interessantes Glücksspiel kennengelernt. Bei diesem Spiel kann man pro Durchführung entweder 10 € verlieren oder 5 € verlieren oder 5 € gewinnen oder aber 10 € gewinnen. Nach Beobachtung einiger Spiele hat er folgende Einschätzung :

- die Ausgänge der einzelnen Spiele sind voneinander unabhängig.
- die Wahrscheinlichkeiten für die jeweiligen Ereignisse bleiben von Spiel zu Spiel gleich.
- die Wahrscheinlichkeit für das Gewinnen von 10 € ist gleich $\pi$.
- die Wahrscheinlichkeit, 5 € zu verlieren <u>oder</u> zu gewinnen, ist viermal so hoch, wie die Wahrscheinlichkeit, 10 € zu gewinnen.
- die Wahrscheinlichkeit, 5 € zu verlieren, ist genauso hoch, wie die Wahrscheinlichkeit, 5 € zu gewinnen.

Sei

$X_i$ : „ Gewinn beim i–ten Spiel “ $( i = 1 , \dots , n )$

1) Stellen Sie die Wahrscheinlichkeitsfunktion der Zufallsvariablen $X_i$ tabellarisch dar !
2) Für welchen Wert von $\pi$ wäre das Spiel „ fair “ ?

Statistix möchte den unbekannten Parameter $\pi$ anhand einer einfachen Stichprobe schätzen und definiert folgende Zufallsvariablen :

A : „ Anzahl der Spiele, bei denen man 10 € verliert “

B : „ Anzahl der Spiele, bei denen man 5 € verliert “

C : „ Anzahl der Spiele, bei denen man 5 € gewinnt “

D : „ Anzahl der Spiele, bei denen man 10 € gewinnt “

3) Stellen Sie die Maximum–Likelihood–Funktion für den unbekannten Parameter $\pi$ ( in Abhängigkeit von a, b, c und d ) auf !
4) Ermitteln Sie mit Hilfe eines formalen Rechenganges den Maximum–Likelihood–Schätzwert $\hat{\vartheta}_{ML}$ für $\pi$ !
5) Zeigen Sie mit Hilfe eines formalen Rechenganges, dass $\hat{\Theta}_{ML}$ eine erwartungstreue Schätzfunktion für $\pi$ ist !

**B**

Da Anton und seine Freundin Berta die nächsten 7 Tage verreisen werden, müssen sie die Versorgung ihrer sechs Hamster während dieser Woche organisieren. Die zwei männlichen Hamster sind eher friedliche Gesellen, während die vier weiblichen Hamster leider zickig und aggressiv sind, so dass jede Hamsterdame in einem Einzelkäfig untergebracht werden muss. Antons Mutter wird die Versorgung der fünf Käfige übernehmen. Unabhängig vom Geschlecht benötigt jeder Hamster im Mittel 7 Gramm Futter pro Tag bei einer Standardabweichung von 1,4 Gramm. Seiner Mutter stellt Anton für die Versorgung eine bestimmte Futtermenge bereit.

Aus Erfahrung weiß Anton, dass die gefressenen Futtermengen sowohl zwischen den einzelnen Käfigen als auch von Tag zu Tag unabhängig voneinander, jedoch zwischen mehreren Hamstern innerhalb eines Käfigs <u>nicht</u> unabhängig voneinander sind.

Der Erwartungswert des Produkts der täglich von zwei Hamstern gefressenen Futtermengen beträgt 48,12 [ Gramm$^2$ ].

Sei

X : „ Gefressene Futtermenge pro Woche der beiden männlichen Hamster “

Y : „ Gefressene Futtermenge pro Woche der vier weiblichen Hamster “

X + Y = G : „ Gefressene Gesamtfuttermenge aller Hamster pro Woche “

1) Wie ist die Zufallsvariable G verteilt ? [ Verteilungstyp und –parameter ! ]
2) Wie viel Gramm Futter werden mit einer Wahrscheinlichkeit von 90,32 % pro Woche mindestens gefressen ?

Nach dem Urlaub muss sich Anton seiner Semesterarbeit im Rahmen des Studiums der Veterinärmedizin widmen. Hierbei möchte er die Anzahl $\delta$ der übergewichtigen Hamster unter den sehr vielen im Universitätslabor lebenden Hamster mit Hilfe einer einfachen Stichprobe vom Umfang $n = 50$ schätzen.

3) Geben Sie explizit das approximative [ zweiseitig symmetrische ] Konfidenzintervall für die wahre Anzahl $\delta$ zum Konfidenzniveau $1 - \alpha$ an !

Angenommen, Anton würde erfahren, dass im Universitätslabor 1000 Hamster zu Forschungszwecken gehalten werden, und dass in seiner Stichprobe genau 28 Hamster Übergewicht hätten.

4) Bestimmen Sie ( approximativ ) das Schätzintervall für $\delta$ zum Konfidenzniveau 99 % !

## Aufgabe MC

### Block A

1) Das Maximum der Gütefunktion eines zweiseitigen Tests auf $\mu$ bei bekannter Varianz $\sigma^2$ zum Signifikanzniveau $\alpha$ ist gleich $1-\alpha$.

2) Die Stichprobenfunktion $\hat{\Theta}_n = \frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$ ist eine erwartungstreue Schätzfunktion für $\mu^2$.

3) Die Länge eines Konfidenzintervalls für $\mu$ wird bei bekannter Varianz $\sigma^2$ – unter sonst gleichen Bedingungen – mit wachsendem Konfidenzniveau größer.

4) Das ( schwache ) Gesetz der großen Zahlen lässt sich mit Hilfe der Tschebyscheffschen Ungleichung nachweisen.

### Block B

1) Von den Schätzfunktion $\hat{\Theta}_1$ und $\hat{\Theta}_2$ für $\vartheta$ erfülle $\hat{\Theta}_1$ mehr Gütekriterien als $\hat{\Theta}_2$. Dann liegt ein Schätzwert $\hat{\vartheta}_1$ näher am Parameter $\vartheta$ als ein Schätzwert $\hat{\vartheta}_2$.

2) Beim Schichtungsverfahren liegt Homogenität innerhalb der Schichten und Heterogenität zwischen den Schichten vor.

3) Die Tschebyscheffsche Ungleichung gilt auch für eine normalverteilte Zufallsvariable.

4) Um die Hypothesen bei einem einseitigen Parametertest festzulegen, ist eine Risikoüberlegung erforderlich.

### Block C

1) Bei einem Test auf $\pi$ ist es möglich, dass das Signifikanzniveau nicht vollständig ausgeschöpft wird.

2) Der einseitige Test auf $\mu$ ist – unter sonst gleichen Bedingungen – trennschärfer als der zweiseitige Test auf $\mu$.

3) Unter der Voraussetzung, dass der Verteilungstyp der Prüfgröße unter $H_0$ und auch unter $H_1$ bekannt ist, kann man den Fehler 2.Art beschränken.

4) Der Mean–Square–Error einer Schätzfunktion setzt sich additiv zusammen aus der Varianz der Schätzfunktion und der Differenz zwischen dem Erwartungswert der Schätzfunktion und dem unbekannten Parameter.



### Block D

1) Bei einem beliebig festgesetzten Konfidenzniveau $1-\alpha$ ist es nicht möglich, ein Konfidenzintervall für $\mu$ ( bei bekannter Varianz $\sigma^2$ ) mit beliebig kleiner Länge zu erzeugen.

2) Sei $\hat{\Theta}_{KQ}$ eine Schätzfunktion zum Schätzen des Parameters $\vartheta$, die mit Hilfe der Methode der kleinsten Quadrate konstruiert wurde. Dann ist es nicht immer möglich, auch mit Hilfe der Maximum–Likelihood–Methode eine Schätzfunktion $\hat{\Theta}_{ML}$ für den Parameter $\vartheta$ zu konstruieren.

3) Der Hypothesenbereich $H_1$ und der Ablehnbereich eines Parametertests auf $\mu$ sind disjunkte Mengen.

4) Die Maximum–Likelihhod–Schätzfunktion zum Schätzen von $\sigma^2$ einer normalverteilten Zufallsvariablen ( bei unbekanntem $\mu$ ) lautet : $\frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$ .

### Block E

1) Bei einer verbundenen Stichprobe ist Var ( X + Y ) immer dann kleiner als bei zwei unabhängigen Stichproben, wenn die Korrelation zwischen X und Y negativ ist.

2) Bei einem Parametertest auf $\mu$ ist der Annahmebereich eine Teilmenge des Hypothesenbereichs $H_0$.

3) Der Vorzeichentest lässt sich nicht nur für Hypothesen bezüglich des Medians, sondern auch für Hypothesen bezüglich anderer p–Quantile anwenden.

4) Die Fraktile der t – Verteilung werden mit wachsenden Freiheitsgraden immer größer.