STATISTIK

II

ÖKONOMEN
&
WIRTSCHAFTSINGENIEURE

WS 2014 / 2015

Aufgabe 1

A

Statistix ist begeisterter Fußball–Fan und kauft sich bisher für jede Spielsaison eine Dauerkarte für den Stadionbesuch. Allerdings ärgert er sich in letzter Zeit über die – seiner Meinung nach – zunehmend unfaire Spielweise vieler Spieler auf dem Fußballplatz, da die Schiedsrichter zu oft die gelbe Karte zeigen müssen. Daher möchte er seinem geliebten Fußballsport nur dann treu bleiben, wenn pro Spiel durchschnittlich höchstens 3 gelbe Karten vom Schiedsrichter gezeigt werden. Ansonsten will er sich in Zukunft vom Fußballsport abwenden und nur noch Dauerkarten für Rhönrad–Veranstaltungen kaufen. Diese schwierige Entscheidung will Statistix vom Ergebnis eines statistischen Tests ($\alpha = 0{,}05$) abhängig machen. Dabei will er das Risiko, seinem geliebten Fußballsport fälschlicherweise den Rücken zu kehren, so klein wie möglich halten.

Als Datenbasis für die Testdurchführung sollen ihm die am nächsten Samstag anstehenden 36 Profi–Spiele dienen (einfache Stichprobe).

Er geht davon aus, dass die Varianz der durchschnittlich pro Spiel gezeigten gelben Karten 0,25 beträgt.

1) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !
2) Wie lautet die Prüfgröße für diesen Test formal und verbal, und wie ist diese unter H_0 verteilt ?
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) Berechnen Sie die Wahrscheinlichkeit für eine Fehlentscheidung bei diesem Test, wenn in Wahrheit pro Spiel durchschnittlich
   a) 2,7
   b) 3,2
   c) 4,1

   gelbe Karten vom Schiedsrichter gezeigt werden !
6) Wodurch lässt sich (bei gleichbleibendem Signifikanzniveau) die Trennschärfe dieses Tests erhöhen ? Erläutern Sie kurz, was der Vorteil eines trennschärferen Tests ist !
7) Wie groß müsste der Stichprobenumfang mindestens gewählt werden, damit die Wahrscheinlichkeit für eine Fehlentscheidung um mindestens 69,435 Prozentpunkte geringer ausfallen würde als beim vorliegenden Testaufbau, wenn in Wahrheit pro Spiel durchschnittlich 3,2 gelbe Karten vom Schiedsrichter gezeigt werden ?
8) Handelt es sich beim vorliegenden Test um einen „ konservativen " Test ? Begründen Sie kurz Ihre Antwort !

B

Empirix ist ein Freund von Statistix und ebenfalls ein großer Fußball–Fan. Er stört sich weniger an den gezeigten gelben Karten als vielmehr an den wegen „ Meckerns " der Spieler vom Schiedsrichter gezeigten roten Karten. Seine Freundin Empirica will ihn schon lange vom Fußball weglotsen, damit er sie stattdessen bei ihren Töpfer–Kursen begleitet. Um diesbezüglich endlich eine Entscheidung herbeizuführen, soll Empirix einen statistischen Test ($\alpha = 3\ \%$) durchführen. Wenn sich mit Hilfe dieses Tests statistisch untermauern lässt, dass in Wahrheit bei mehr als 20 % der Spiele eine rote Karte wegen „ Meckerns " vom Schiedsrichter gezeigt werden muss, ist Empirix′ Schicksal besiegelt, d.h., er muss in Zukunft töpfern.

Als Datenbasis für die Testdurchführung sollen die am nächsten Sonntag anstehenden 30 Profi–Spiele dienen (einfache Stichprobe).

1) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !
2) Wie lautet die Testfunktion für diesen Test, und wie ist diese unter H_0 verteilt ?
3) Bestimmen Sie den Annahme– und Verwerfungsbereich dieses Tests !
4) Wie groß ist die Wahrscheinlichkeit, dass Empirix in Zukunft töpfern gehen muss, obwohl in Wahrheit bei nur 15 % der Spiele eine rote Karte wegen „ Meckerns " vom Schiedsrichter gezeigt wird ?

Am Sonntag wurde bei 12 Spielen vom Schiedsrichter die rote Karte wegen „ Meckerns " gezeigt.

5) Wie lautet die Testentscheidung ?
6) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt ! Wird Empirix in Zukunft töpfern gehen müssen ?
7) Wie hoch ist die Wahrscheinlichkeit, dass diese Testentscheidung falsch ist ?
8) Welcher Fehler ist bei der Testentscheidung unterlaufen ?

Aufgabe 2

Der Jung–Unternehmer Herr Maier betreibt einen Gartenzwerg–Großhandel. Das Sortiment besteht aus dem Basismodell „ parvi " und dem Premiummodell „ sumptuosus ". Dabei geht Herr Maier von folgenden betriebswirtschaftlichen Kenngrößen aus :

	Absatzmenge [Stück pro Jahr]	Nettoerlös [€ pro Stück]	variable Kosten [€ pro Stück]	Fixkosten [€ pro Jahr]	Gewinn [€ pro Jahr]
„ parvi "	A_P	**10.–**	**4,50**		
„ sumptuosus "	A_S	**20.–**	**9.–**		
gesamt	$A = A_P + A_S$			**11.000.–**	**G**

Die Absatzmengen A , A_P und A_S sowie der Gewinn G sind als Zufallsvariable aufzufassen.

1) Stellen Sie die Gleichung für den Gewinn G in Abhängigkeit der Absatzmengen A_P und A_S auf !

Bezüglich der Absatzmengen A , A_P und A_S gilt folgende Auffassung :

⇨ Die Absatzmengen A_P und A_S sind unabhängig voneinander, und es bestehen keine Kapazitätseinschränkungen.

⇨ Bei der Absatzmenge A_P sind 500 Stück fix plus durchschnittlich 3.922 Stück bei einer Standardabweichung von 803 Stück.

⇨ Bei der Absatzmenge A_S sind 250 Stück fix plus durchschnittlich 1.961 Stück bei einer Varianz von 42.436 Stück^2 .

2) Wie sind die Zufallsvariablen A und G verteilt ?

3) Berechnen Sie $E(A)$ und $Var(A)$!

4) Berechnen Sie $E(G)$ und $Var(G)$!

5) Wie groß ist die Wahrscheinlichkeit, dass die Absatzmenge A mindestens 3.317 Stück, aber höchstens 9.949 Stück beträgt ?

Herr Maier plant, den Großhandel 40 Jahre lang zu betreiben. Er geht davon aus, dass die Gewinne über diese Jahre hinweg unabhängig voneinander sein werden.
Sei

G_{ges} : „ Gesamtgewinn über die 40 Jahre "

6) Wie groß ist die Wahrscheinlichkeit, dass der Gesamtgewinn G_{ges} größer als 1.450.000 € sein wird ?

Im Rahmen einer Qualitätskontrolle möchte Herr Maier die <u>Anzahl</u> δ einwandfreier Gartenzwerge im Lagerbestand mit Hilfe einer einfachen Stichprobe vom Umfang $n = 200$ schätzen.

7) Geben Sie explizit das approximative Konfidenzintervall für den wahren <u>Anteil</u> π einwandfreier Gartenzwerge im Lagerbestand zum Konfidenzintervall $1 - \alpha$ an !

8) Geben Sie explizit das approximative Konfidenzintervall für die wahre <u>Anzahl</u> δ einwandfreier Gartenzwerge im Lagerbestand zum Konfidenzintervall $1 - \alpha$ an !

Angenommen, man würde feststellen, dass in der Stichprobe (aus einem Lagerbestand von 8.000 Stück) genau 32 Gartenzwerge mängelbehaftet sind.

9) Bestimmen Sie für die wahre <u>Anzahl</u> δ einwandfreier Gartenzwerge im Lagerbestand das approximative Schätzintervall zum Konfidenzniveau von 95 % !

<u>Hilfe</u> : $\sqrt{0,000672} \approx 0,026$

Aufgabe 3

A

In der Region Arithmetica ist ein Grippe–Virus ausgebrochen. Es gibt diese Krankheit betreffend vier verschiedene Möglichkeiten :

H : eine Person ist in den ersten 24 Stunden nach Infektion **hoch** ansteckend, zeigt aber noch keine Symptome.

M : bei einer Person, die Symptome zeigt, verringert sich die Gefahr, andere Personen anzustecken, auf ein **mittleres** Niveau.

G : hat eine Person die Symptome überstanden, ist die Gefahr, andere Personen anzustecken, nur noch **gering**.

I : eine Person ist **immun** gegen diese Krankheit, d.h. sie kann sich nicht anstecken.

Die Gesundheitsbehörde geht davon aus, dass die Wahrscheinlichkeiten für diese vier Möglichkeiten von einem Parameter π abhängen, der von Region zu Region variiert. Die Behörde gibt folgende Informationen bekannt :

⇨ die Wahrscheinlichkeit, **gering**fügig ansteckend zu sein, beträgt $2 \cdot \pi$.

⇨ die Wahrscheinlichkeit, gegen diese Krankheit **immun** zu sein, ist ein Fünftel so groß wie die Wahrscheinlichkeit, **hoch**gradig ansteckend oder **gering**fügig ansteckend zu sein.

⇨ die Wahrscheinlichkeit, **hoch**gradig ansteckend oder **immun** zu sein, ist doppelt so groß wie die Wahrscheinlichkeit, **gering**fügig ansteckend zu sein.

1) Bestimmen Sie (in Abhängigkeit von π) die Wahrscheinlichkeitsfunktion der Zufallsgröße
X : „ Ansteckungsgrad mit diesem Grippe–Virus "

x	h	m	g	i
P (X = x)				

Mit Hilfe einer einfachen Stichprobe vom Umfang n soll der Parameter π für die Region Arithmetica geschätzt werden.

2) Stellen Sie die Maximum–Likelihood–Funktion $L [\pi \mid (x_1 , . . . , x_n)]$ auf !

3) Stellen Sie die logarithmierte Maximum–Likelihood–Funktion $\ln L [\pi \mid (x_1 , . . . , x_n)]$ auf !

4) Bestimmen Sie den Maximum–Likelihood–Schätzwert $\hat{\pi}_{ML}$ für π !

B

Statistix bezweifelt die Angaben der Gesundheitsbehörde und möchte diese Angaben mit Hilfe eines statistischen Tests ($\alpha = 0{,}2$) auf Basis einer einfachen Stichprobe vom Umfang n = 50 überprüfen.

1) Wie heißt das hier zu verwendende Testverfahren ?

2) Formulieren Sie die Hypothesen für diesen Test ! Begründen Sie Ihre Wahl – falls nötig – mit Hilfe einer Risikoüberlegung !

3) Wie lautet die Testfunktion, und wie ist diese unter H_0 verteilt ?

Von den untersuchten 50 Personen waren 10 Personen **hoch**gradig ansteckend, 40 % der Personen zeigten Krankheitssymptome, während 5 Personen **immun** waren.

4) Bestimmen Sie den Verwerfungsbereich für diesen Test !

5) Wie lautet die Testentscheidung ?

6) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !
Sind die Zweifel von Statistix berechtigt ?

Aufgabe MC

Block A

1) Eine asymptotisch erwartungstreue Schätzfunktion ist auch konsistent.

2) Betrachtet sei der Fall einer (zweiseitig symmetrischen) Konfidenzschätzung für den Parameter μ einer normalverteilten Zufallsvariablen mit bekannter Varianz.
Das Konfidenzintervall wird umso breiter, je stärker die Werte in der Stichprobe streuen.

3) Eine erwartungstreue Stichprobenfunktion ist konsistent, falls ihre Varianz mit wachsendem Stichprobenumfang gegen Null konvergiert.

4) Die Stichprobenfunktion S^2 ist eine Schätzfunktion für die Varianz σ^2, die selbst auch eine Varianz besitzt.

Block B

1) Der Wert der Gütefunktion eines Tests an der Stelle $\vartheta \in H_1$ ist das Komplement der Wahrscheinlichkeit für den Fehler 2. Art an der Stelle ϑ.

2) Der Erwartungswert einer diskreten Zufallsvariablen X ist (summiert über alle Realisationsmöglichkeiten x) das Produkt der realisierten Ausprägungen mit den jeweiligen Wahrscheinlichkeiten, dass X den Wert x annimmt.

3) Sei X_1 verteilt nach $N(\mu_1; \sigma_1^2)$ und X_2 verteilt nach $N(\mu_2; \sigma_2^2)$. Dann ist $X_1 - X_2$ verteilt nach $N(\mu_1 - \mu_2; \sigma_1^2 + \sigma_2^2)$.

4) Sei X eine normalverteilte Zufallsvariable, d.h. $X \sim N(\mu; \sigma^2)$.
Dann ist der Maximalwert der Dichtefunktion gleich $\frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}}$.

Block C

1) Eine Zufallsauswahl, die uneingeschränkt ist, muss eine gleichgewichtete Zufallsauswahl sein.

2) Nach dem „ Mean Square Error – Konzept " soll der (quadrierte) theoretische Schätzfehler minimiert werden.

3) Sei X eine stetige Zufallsvariable mit der Verteilungsfunktion F (x), die für $x \leq 0$ gleich Null ist. Dann gilt :

$$F(x_{0,5}) - F(-x_{0,5}) = 1$$

4) Bei der Intervallschätzung bezieht sich die Wahrscheinlichkeitsaussage in Form des Konfidenzniveaus $1 - \alpha$ nicht auf die Situation nach der Beobachtung.

Block D

1) Auch bei den sog. „ verteilungsfreien“ Tests benötigt man eine Testfunktion und folglich auch eine Verteilung derselben, um den Test durchzuführen zu können.

2) Die Varianz der Differenzfunktion $X - Y$ bei verbundenen Stichproben kann größer sein als bei unverbundenen Stichproben.

3) Bei einem proportional geschichteten Auswahlverfahren hat jede Stichprobe $(X_1 , \dots , X_n)$ die gleiche Wahrscheinlichkeit realisiert zu werden.

4) Die Varianz einer t – verteilten Zufallsvariablen kann auch kleiner sein als die Varianz einer standardnormalverteilten Zufallsvariablen.

Block E

1) Sei X eine beliebige Zufallsvariable mit $E (X) = \mu$ und $Var (X) = 3$. Dann gilt :

$$P (| X - \mu | \geq 2) \leq \frac{3}{4}$$

2) Eine erwartungstreue Schätzfunktion $\hat{\Theta}$ kann einen Schätzwert $\hat{\vartheta}$ liefern, der gleich dem zu schätzenden Parameter ϑ ist.

3) Eine Zufallsvariable X , für die gilt : $E (X) = 0$ und $Var (X) = 1$, ist standardnormalverteilt.

4) Das „ Likelihood – Prinzip “ ist frequentistisch, d.h. die Bewertung der ϑ – Werte durch die Likelihoodfunktion richtet sich allein nach der einen beobachteten Stichprobe. Nach der Beobachtung wird nicht mehr berücksichtigt, was sonst noch alles hätte beobachtet werden können.