STATISTIK

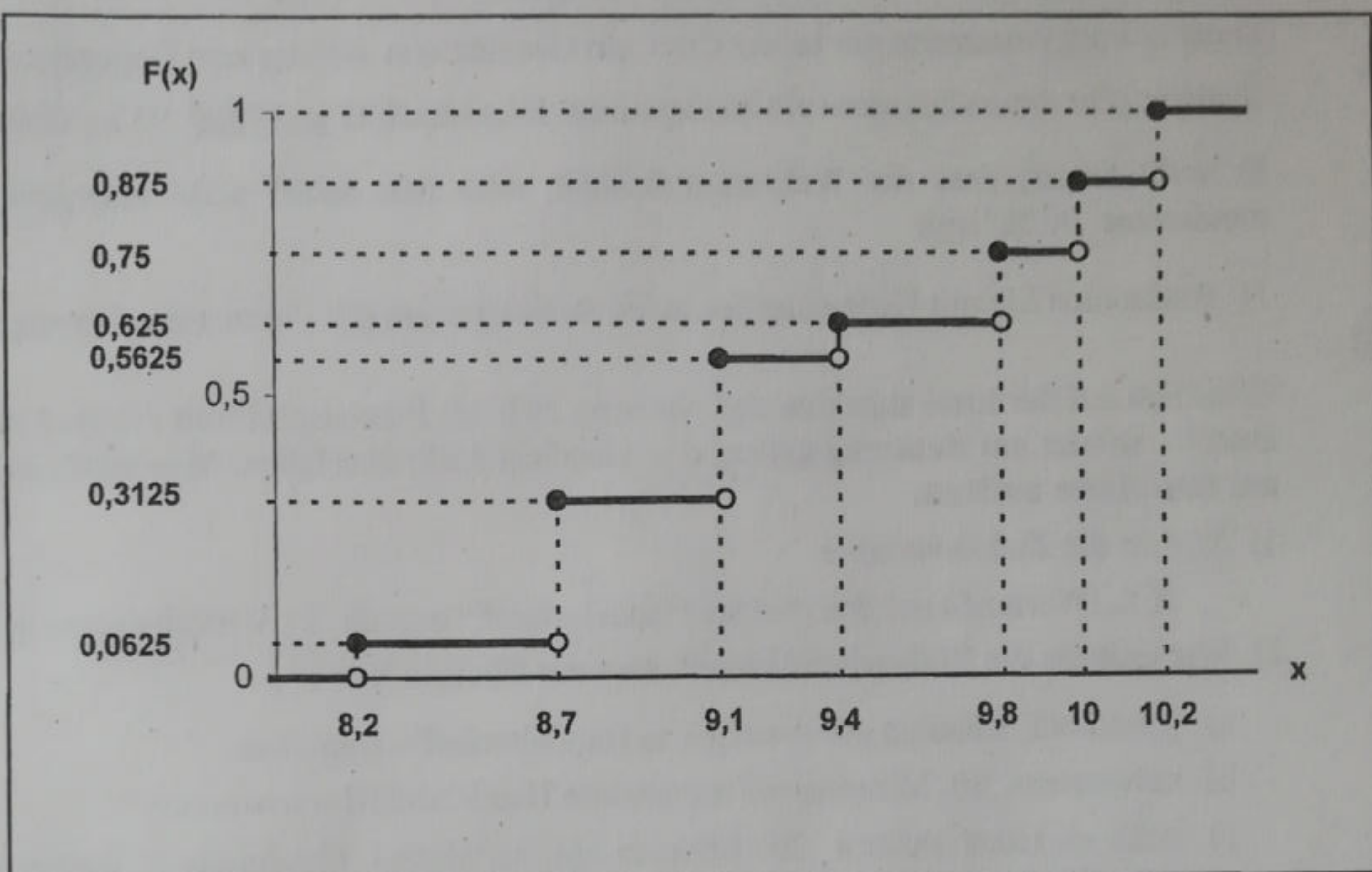
II

ÖKONOMEN
&
WIRTSCHAFTSINGENIEURE

WS 2015 / 2016

Aufgabe 1

Statistix absolviert ein Praktikum in einem Werk, das Automotoren herstellt. Alarmiert vom derzeitigen Abgasskandal will Statistix die hergestellten Motoren bezüglich ihrer Abgaswerte mit Hilfe eines statistischen Tests ($\alpha = 5\ \%$) überprüfen. Als Stichprobenbasis wählt er aus der Produktion zufällig 4 Paletten à 4 Motoren aus (einfache Stichprobe). Gemäß der gängigen Euro–6–Norm soll die CO_2 – Emission eines Motors den Richtwert von 9,5 $g/_{10\ km}$ nicht überschreiten. Sollte der Test ergeben, dass der wahre Median der CO_2 – Emission eines Motors höher als der Richtwert ist, hätte dies für das Werk weitreichende Konsequenzen zur Folge (Produktionsstopp, Rückrufaktionen, Imageverlust etc.). Daher will Statistix den Test so aufbauen, dass das Risiko für einen unberechtigten Produktionsstopp so klein wie möglich gehalten wird.
Er geht zwar nicht von einer bestimmten Verteilungsannahme für die Emissionswerte aus, nimmt aber an, dass diese symmetrisch verteilt sind.



Sei

X_i : „ CO_2 – Emission des i–ten Motors $[g/_{10\ km}]$ " $[\, i = 1, \ldots, n \,]$

Die Auswertung seiner Daten ergab die abgebildete (empirische) Verteilungsfunktion.

1) Bestimmen Sie tabellarisch die zugehörige absolute Häufigkeitsverteilung !
2) Wie lautet das (unter Ausschöpfung aller Informationen) hier durchzuführende Testverfahren ?
3) Wie lauten die Hypothesen für diesen Test ? Begründen Sie Ihre Wahl – <u>falls nötig</u> – mit Hilfe einer Risikoüberlegung !
4) Wie lautet die Testfunktion, und wie ist diese unter H_0 verteilt ?
5) Bestimmen Sie den Annahme– und Verwerfungsbereich dieses Tests !
6) Wie lautet die Testentscheidung ?
7) Interpretieren Sie das Testergebnis inhaltlich und statistisch exakt !

Angenommen, Statistix würde davon ausgehen, dass die Emissionswerte normalverteilt sind.

8) Wie lautet das nun zu verwendende Testverfahren ?
9) Wie lauten die Hypothesen für diesen Test ?
10) Wie lautet die Testfunktion, und wie ist diese unter H_0 verteilt ?
11) Bestimmen Sie den Annahme– und Verwerfungsbereich dieses Tests !

Angenommen, Statistix hätte den Testfunktionswert $v = -1{,}2734$ ermittelt.

12) Bestimmen Sie den Wert der (empirischen) Standardabweichung (Runden Sie auf 4 Nachkommastellen) !

Aufgabe 2

A

Statistix ist Kapitän der Piratencrew „ Statistik–Bande ". Er will mit seinem Schiff zur Insel „ Isla de Pólya " fahren, um dort Proviant und Rum kaufen. Dafür werden mindestens 850 kg Gold und Diamanten benötigt, welche durch das Kapern von 7 Handelsschiffen erbeutet werden sollen. Das Schiff der Piratencrew hat ein Eigengewicht von 1500 kg und würde bei einem Gesamtgewicht über 3000 kg untergehen.

Statistix trägt zusammen mit seiner Crew ein Gewicht von 600 kg zum Gesamtgewicht bei.

Statistix geht davon aus, dass ein zu kaperndes Handelsschiff im Mittel 95 kg Gold und 30 kg Diamanten mit sich führt.

Er weiß ferner, dass die Wahrscheinlichkeit, dass sein Schiff nicht untergeht, aber genug Beute tragen kann, bei mindestens 96 % liegt.

1) Bestimmen Sie mit Hilfe eines formalen Rechenganges die Varianz des Gesamtgewichts des Piratenschiffs !

B

Glücklich auf der Insel angekommen, versorgt sich das Piratenschiff mit Proviant und Rum, und will – versteckt in einer Bucht – wieder auf Beutezug gehen, d.h. Handelsschiffe überfallen. Man weiß, dass alle 2 Stunden ein Handelsschiff aus dem Hafen ausläuft.

1) Wie ist die Zufallsvariable

   X : „ Wartezeit auf das nächste Handelsschiff " verteilt ? [Verteilungstyp und –parameter !]

2) Wie groß ist die Wahrscheinlichkeit, dass das Piratenschiff

   a) genau 45 Minuten auf das nächste Handelsschiff warten muss ?
   b) mindestens 80 Minuten auf das nächste Handelsschiff warten muss ?
   c) noch maximal weitere 30 Minuten auf das nächste Handelsschiff warten muss, wenn man schon eine Stunde vergeblich wartet ?

Angenommen, bei einem drohenden Unwetter würden die Handelsschiffe nur unregelmäßig aus dem Hafen auslaufen. Pro Tag (≅ 24 Stunden) verlassen dann durchschnittlich 12 Schiffe den Hafen.

3) Wie ist die Zufallsvariable

   Y : „ Wartezeit auf das nächste Handelsschiff " verteilt ? [Verteilungstyp und –parameter !]

4) Wie groß ist die Wahrscheinlichkeit, dass das Piratenschiff

   a) länger als 3 Stunden und 25 Minuten auf das nächste Handelsschiff warten muss ?
   b) zwischen 2 und 6 Stunden auf das nächste Handelsschiff warten muss ?
   c) weniger als 45 Minuten auf das nächste Handelsschiff warten muss ?
   d) noch maximal weitere 2 Stunden auf das nächste Handelsschiff warten muss, wenn man schon eine halbe Stunde vergeblich wartet ?

C

Statistix hat die Idee, dass er vielleicht noch mehr verdienen könnte, wenn er für ein gekapertes Schiff – sowohl vom Kapitän als auch vom Besitzer des Schiffes – Lösegeld verlangen würde. Mit Hilfe eines statistischen Tests ($\alpha = 5\,\%$) will Statistix überprüfen, ob sich die Bereitschaft, Lösegeld zu bezahlen, zwischen Kapitänen und Besitzern von Schiffen unterscheidet. Als Datenbasis zur Durchführung des Tests sollen ihm die diesbezüglichen Erfahrungen der nächsten 50 gekaperten Schiffe dienen (einfache Stichprobe).

1) Wie lautet das zu verwendende (nichtparametrische) 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 für diesen Test !

Angenommen, es würde sich folgendes ergeben :

- bei 25 gekaperten Schiffen zahlten sowohl die Kapitäne als auch die Besitzer Lösegeld.
- bei nur 5 gekaperten Schiffen zahlten weder die Kapitäne noch die Besitzer Lösegeld.
- bei 12 gekaperten Schiffen zahlten die Kapitäne Lösegeld, die Besitzer aber nicht.

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

Aufgabe 3

Statistix ist Hobby–Zoologe und beschäftigt sich in seiner Freizeit mit Insektenzucht. Durch Genmanipulation hat er eine neue Insektenart gezüchtet, deren Fortpflanzungsverhalten er jedoch noch nicht kontrollieren kann. So bekommen die Weibchen unvorhersehbar in unregelmäßigen Abständen Nachwuchs. Statistix will nun den Sachverhalt statistisch untersuchen und betrachtet dazu das Fortpflanzungsverhalten in n voneinander getrennten Terrarien, die gleichartig besetzt sind.
Er nimmt zunächst an, dass die Zufallsvariablen

X_i : „ Anzahl der Nachwuchs bekommenden Weibchen im i–ten Terrarium “ $[\, i = 1, \ldots, n \,]$

beliebig verteilt sind mit $E(X_i) = \lambda$.

1) Ermitteln Sie mit Hilfe eines geeigneten Konstruktionsverfahrens einen Schätzwert für λ !
Wie lautet demnach die zugehörige Schätzfunktion $\hat{\Theta}_1$?

Nach reiflicher Überlegung geht Statistix nun davon aus, dass die betrachteten Zufallsvariablen X_i poissonverteilt sind und betrachtet die Zufallsvariable

T : „ Wartezeit auf die nächste Insektengeburt “

mit $f(t) = \lambda \cdot e^{-\lambda t}$; $E(T) = \frac{1}{\lambda}$; $Var(T) = \frac{1}{\lambda^2}$; $P(T > t) = e^{-\lambda t}$

2) Zeigen Sie mit Hilfe eines geeigneten Konstruktionsverfahrens, dass für den Schätzwert der Standardabweichung σ der Zufallsvariablen T gilt : $\hat{\sigma} = \overline{t}$, d.h. dass die zugehörige Schätzfunktion lautet : $\hat{\Theta}_2 = \overline{T}$!

3) Zeigen Sie mit Hilfe eines Rechenganges, dass es sich bei $\hat{\Theta}_2 = \overline{T}$ um eine erwartungstreue Schätzfunktion für σ handelt !

4) Zeigen Sie mit Hilfe eines Rechenganges, dass es sich bei $\hat{\Theta}_2 = \overline{T}$ um eine konsistente Schätzfunktion für σ handelt !

Von einem Kollegen erfährt Statistix, dass es beim Schätzen einer Wartezeit manchmal passiert, dass das Experiment beendet werden muss, bevor das erwartete Ereignis eingetreten ist. Dieser Fall tritt ein, wenn in einem Terrarium zur Abbruchzeit z noch kein Nachwuchs geboren ist. Dies will er nun berücksichtigen und eine Schätzung für die Standardabweichung σ durchführen.

Zum Abbruchzeitpunkt z hat sich in k Terrarien Nachwuchs eingestellt, die restlichen $n-k$ Terrarien blieben ohne Nachwuchs, d.h. die Wartezeit bis zur ersten Geburt ist in diesen $n-k$ Terrarien größer als z .
Statistix stellt folgende Maximum–Likelihood–Funktion auf :

$$L\left[\sigma \,\middle|\, (t_1, \ldots, t_n)\right] = \prod_{i=1}^{k} f(t_i \mid \sigma) \cdot \prod_{j=k+1}^{n} P(T_j > z \mid \sigma)$$

5) Zeigen Sie mit Hilfe eines Rechenganges, dass für den Maximum–Likelihood–Schätzwert für σ gilt :

$$\hat{\sigma} = \frac{1}{k} \cdot \sum_{i=1}^{k} t_i + \frac{(n-k)}{k} \cdot z \quad !$$

Nach drei Tagen und drei Nächten beendet Statistix das Experiment.
Der Anteil der Terrarien <u>ohne</u> Nachwuchs beträgt $\frac{50}{122}$, in den übrigen 36 Terrarien <u>mit</u> Nachwuchs ergab sich eine Gesamtwartezeit von 1800 Stunden.

6) Bestimmen Sie aus diesen Angaben den Maximum–Likelihood–Schätzwert $\hat{\sigma}$!

Aufgabe MC

Block A

1) Das „Mean Square Error – Konzept" betrachtet den mittleren tatsächlichen Schätzfehler einer Schätzfunktion.

2) Ein „ Niveau – α – Test " heißt konservativ, falls mit wachsendem Stichprobenumfang die Ablehnung einer unzutreffenden Nullhypothese immer sicherer wird.

3) Die Varianz einer t – verteilten Zufallsvariablen kann nicht kleiner sein als die Varianz einer standardnormalverteilten Zufallsvariablen.

4) Das Schätzintervall für den unbekannten Parameter μ einer normalverteilten Grundgesamtheit wird bei bekanntem σ^2 (unter sonst gleichen Bedingungen) mit zunehmendem Stichprobenumfang stets kleiner.

Block B

1) Bei den sog. „ verteilungsfreien" Tests benötigt man keine Testfunktion und folglich auch keine Verteilung derselben, um den Test durchzuführen.

2) Die Wahrscheinlichkeit für das Ereignis „ $\bar{x} = \mu$ " ist gleich Null, falls die Zufallsvariable $\overline{X}$ stetig ist.

3) Eine uneingeschränkte Zufallsauswahl ist immer auch eine gleichgewichtete Zufallsauswahl.

4) Um mit Hilfe des Vorzeichentests den unbekannten Median einer Verteilung zu testen, setzt man nur voraus, dass das untersuchte Merkmal stetig verteilt ist.

Block C

1) Erhöht man beim Einstichproben–Gaußtest (unter sonst gleichen Bedingungen) den Stichprobenumfang n auf n^* , so gilt für die entsprechenden Verwerfungsbereiche B bzw. B^* : $B \subseteq B^*$.

2) Die Summe von mehr als 30 Zufallsvariablen ist approximativ normalverteilt.

3) Entscheidet man sich bei einem statistischen Test zum Signifikanzniveau $\alpha = 5$ % für die Gegenhypothese H_1 , so ist diese Entscheidung mit einer Wahrscheinlichkeit von $P(\text{„}H_1\text{"} / H_0) \cong \alpha = 5$ % falsch.

4) Ein „ Niveau – α – Test " heißt **gleichmäßig bester** ($\cong$ trennscharfer) Test, falls kein anderer „ Niveau – α – Test " auf dem gesamten Hypothesenbereich H_1 eine steilere Gütefunktion besitzt.

Block D

1) Die Tschebyscheffsche Ungleichung lässt sich auch dann anwenden, wenn bekannt ist, welchem stochastischen Verteilungstyp die betrachtete Zufallsvariable folgt.

2) Ein Schätzintervall überdeckt mit $(1 - \alpha)$ %–iger Wahrscheinlichkeit den zu schätzenden Parameter.

3) Bei der Auswahl einer einfachen Stichprobe hat jedes Element der Grundgesamtheit die gleiche Wahrscheinlichkeit, in die Stichprobe zu gelangen.

4) Ein Test heißt „ Niveau – α – Test “ , falls gilt : $g(\vartheta) \leq \alpha$ für alle $\vartheta \in H_1$.

Block E

1) Sei $(X_1, \ldots, X_n)$ eine einfache Stichprobe mit $E(X_i) = \mu$ und $Var(X_i) = \sigma^2$.
Nach dem (schwachen) Gesetz der großen Zahlen gilt :
„ Für großes n nimmt $\overline{X}_n$ mit hoher Wahrscheinlichkeit Werte nahe bei μ an. “

2) Der Erwartungswert einer Zufallsvariablen ist nicht das arithmetische Mittel aus allen realisierten Ausprägungen.

3) Das Klumpenverfahren ist eine uneingeschränkte Zufallsauswahl.

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