60-minütige Klausur zur Vorlesung „Statistik II für Studierende der Wirtschaftswissenschaften“ aus dem Sommersemester 2017

Prof. Dr. Helmut Küchenhoff

Ludwig-Maximilians-Universität München, Institut für Statistik

08.08.2017, 8:30 – 9:30 Uhr

- **Überprüfen Sie bitte, ob Ihre Angabe vollständig ist.** Diese Angabe sollte (inklusive dieser Seite) 14 Seiten mit insgesamt 4 Aufgaben umfassen. Alle 4 Aufgaben sind zu bearbeiten.
- Füllen Sie bitte das untenstehende Formular umgehend aus. Schreiben Sie bitte leserlich.
- Halten Sie für die Ausweiskontrolle bitte einen Lichtbildausweis (Personalausweis, Reisepass, Führerschein) und Ihren Studentenausweis bereit.
- Die Bearbeitungszeit beträgt **60 Minuten**. Es ist keine vorzeitige Abgabe möglich. Es können maximal 60 Punkte erreicht werden.
- Verwenden Sie für Ihre Notizen und Lösungen ausschließlich die Ihnen zur Verfügung gestellten Papierbögen und die Angabe. Kennzeichnen Sie jeden zur Abgabe vorgesehenen Bogen mit Namen und Matrikelnummer. Geben Sie außerdem die jeweilige Aufgabennummer (auch die der Teilaufgaben) an.
- Alle Lösungen müssen nachvollziehbar sein. Dazu gehört auch lesbare Schrift. Die Ergebnisse müssen klar erkennbar und zuzuordnen sein. Nur die in den Kästen eingetragenen Ergebnisse werden bepunktet; falls der Platz in einem der Kästen nicht ausreicht, muss im betreffenden Kasten ein Verweis zur Lösung zu finden sein. Runden Sie Zwischenergebnisse, Maßzahlen und Anteile auf die vierte Nachkommastelle, absolute Häufigkeiten auf ganze Zahlen. Wenn Formeln gefordert sind, werden diese auch bewertet.
- Es darf nicht mit Bleistift oder ähnlichen Stiften geschrieben werden.
- Zugelassene Hilfsmittel: Taschenrechner (nicht programmierbar, ohne Graphik-Funktion) und alle Unterlagen (open–book) in ausgedruckter, nicht-elektronischer Form.
- Bei Unterschleif erfolgt eine Meldung an das Prüfungsamt. Sie sind verpflichtet, durch Ihr Verhalten jegliche Missverständnisse diesbezüglich auszuschließen. Sorgen Sie insbesondere dafür, dass sich keinerlei Mobiltelefone und Uhren an Ihrem Arbeitsplatz befinden.
- Ein Toilettenbesuch ist nicht vorgesehen. In sehr dringenden Ausnahmefällen wenden Sie sich bitte an die Aufsicht.
- Verlassen Sie den Prüfungsraum erst, nachdem Sie der Aufsicht die Klausur persönlich übergeben haben. Für den Eingang der kompletten Klausur (Notizen, Angabe mit Ihren Lösungen, dieses Deckblatt) bei der Aufsicht sind Sie selbst verantwortlich.
- Bitte verlassen Sie nach der Klausur den Hörsaaltrakt zügig und leise, damit Sie die Teilnehmer anderer Klausuren nicht stören.

Ich bestätige, dass ich obige Hinweise zur Kenntnis genommen habe und sie befolgen werde.

Name: ________________ Vorname: ________________

Matrikelnummer: ________________ Unterschrift: ________________

Punkteverteilung (bitte für Korrektoren frei lassen!)

1	2	3	4	$\sum$

Aufgabe 1 12 Punkte

Ein Computerhersteller bestellt Festplatten von vier verschiedenen Firmen: F(utzi), I(KM), S(eetor), W(estig). Aus Erfahrung ist bekannt, dass die Lieferungen der Firmen F, I und S jeweils in 98% der Fälle keinen Grund zur Beanstandung geben. Bei Firma W ist dies nur in 93% der Fälle der Fall. Aus diesem Grund werden 90% der Bestellungen zu jeweils gleichen Teilen an die Firmen F, I und S gegeben und nur 10% an die Firma W. P(G) bezeichne die Wahrscheinlichkeit dafür, dass es keinen Grund zur Beanstandung gibt.

a) Berechnen Sie den zu erwartenden Anteil (in %) der nicht zu beanstandenden Lieferungen unter Angabe der Formel.

Formel zur Berechnung:

Ergebnis:

b) Der Computerhersteller erhält eine ordnungsgemäße Lieferung. Wie groß ist die Wahrscheinlichkeit, dass die Lieferung von der Firma W kommt? Geben Sie auch Formel zur Berechnung an.

Formel zur Berechnung:

Ergebnis:

c) Angenommen sie betrachten eine Zufallsstichprobe von Lieferungen der oben genannten Firmen. Sie wollen nun überprüfen, ob die Beanstandung eines Produktes unabhängig von der liefernden Firma ist. Begründen Sie, welchen Test Sie hierzu verwenden würden und geben Sie die zugehörigen Hypothesen an.

Name des zu verwendenden Tests:

Begründung für den Test:

Hypothesen:

d) Nun sei Ihnen zusätzlich bekannt, dass diese Stichprobe den Umfang $n = 200$ umfasst und je 50 Lieferungen pro Firma enthält. Erläutern Sie ein potenzielles Problem, dass bei dem Test in diesem Fall auftreten könnte.

Erläuterung des potenziellen Problems:

Aufgabe 2	14 Punkte

Sie möchten die Ausfälle einer größeren Heizungsanlage untersuchen. Sei X die Anzahl der Ausfälle der Heizung zwischen einschließlich April bis einschließlich September (Sommer), welche poissonverteilt mit Erwartungswert von 5 ist.

a) Argumentieren Sie, warum die Annahme der Poissonverteilung in diesem Fall aus formaler Sicht gerechtfertigt werden kann. Geben Sie dabei zwei Gründe an.

Begründungen:

b) Bestimmen Sie die Wahrscheinlichkeit, dass im Sommer mehr als zwei Ausfälle auftreten (mit Berechnung):

Berechnung der Wahrscheinlichkeit:

c) Sei Y die Anzahl der Ausfälle im restlichen Jahr, also zwischen Oktober bis März (Winter), welche weiterhin poissonverteilt ist. Allerdings ergibt sich witterungsbedingt ein höherer Erwartungswert von 8. Sie dürfen annehmen, dass die Ausfälle im Winter unabhängig vom Sommer sind! Geben Sie die Verteilung von $Z = X + Y$ an und begründen Sie ihr Vorgehen!
(Sie müssen dabei noch keine Parameterwerte angeben!)

Verteilung von Z:

Begründung:

d) Bestimmen Sie den Verteilungsparameter der Verteilung von Z, sowie den Erwartungswert von Z.

Verteilungsparameter der Verteilung von Z:

Erwartungswert von Z:

e) Bestimmen Sie nun für Z die Wahrscheinlichkeit, dass insgesamt 7 Ausfälle im Jahr auftreten.
Hinweis: Falls Sie unter Teilaufgabe d) zu keinem Ergebnis kommen, können Sie für den Verteilungsparameter einen Wert von 25 *annehmen.*

Wahrscheinlichkeit:

f) Sei W die Wartezeit (in Monaten) zwischen zwei Ausfällen der Heizungsanlage im Winter. Wie ist W verteilt? Geben Sie die Verteilung von W inklusive der Parameter an! Bestimmen Sie außerdem den Erwartungswert von W!

Verteilung von W:

Erwartungswert von W:

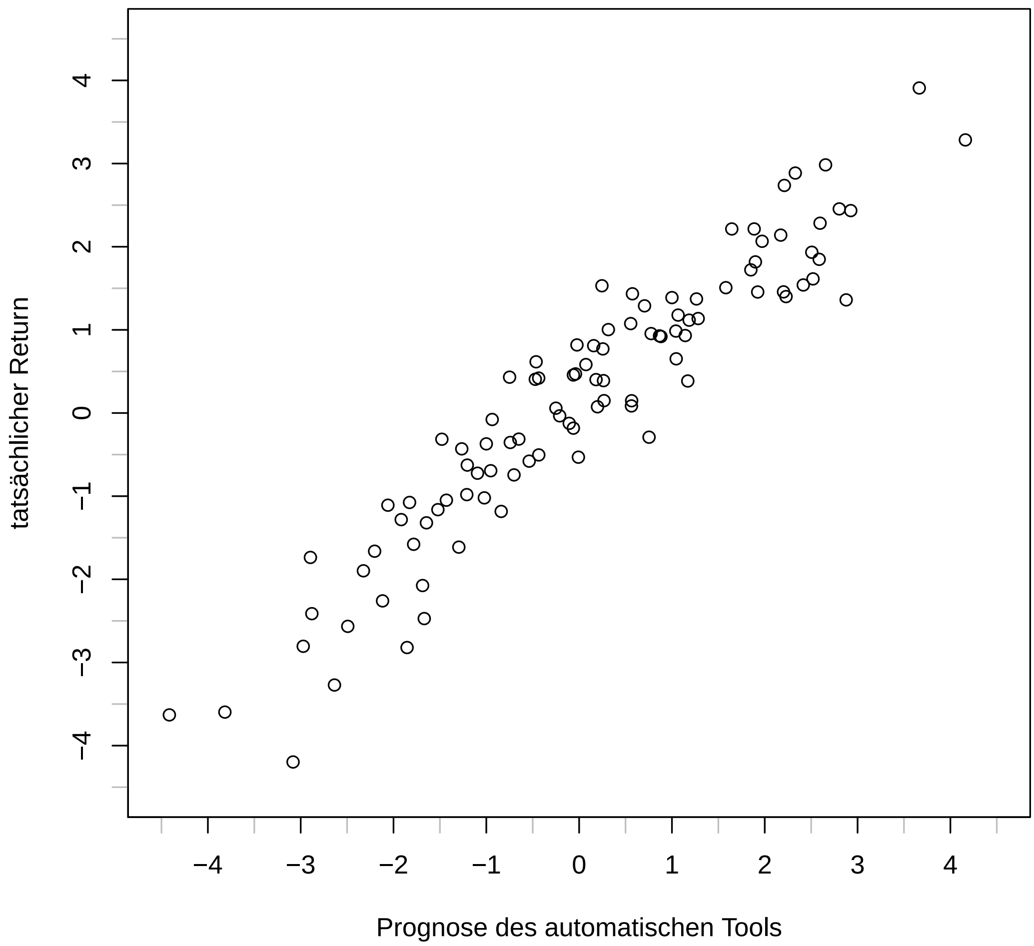
g) Bestimmen Sie die Wahrscheinlichkeit für W, dass die Wartezeit genau den Wert 4 Monate annimmt.

Wahrscheinlichkeit:

Aufgabe 3	14 Punkte

Sie beginnen als Analyst in einer Anlagefirma zu arbeiten. Bisher beruhen die Anlageentscheidungen ihrer Abteilung vor allem auf der Prognose eines automatischen Tools (AT), das Informationen aus Sozialen Medien nutzt, um den erwarteten Return einer Anlage (RE) vorherzusagen.
Nun soll die Prognose des automatischen Tools mit dem tatsächlichen Return einer Anlage verglichen werden, um die Prognosegüte zu untersuchen. Dazu werden die letzten hundert Analyseentscheidungen des Tools und die jeweiligen tatsächlichen Returns herangezogen.

a) Beschreiben Sie kurz den, in der Grafik dargestellten, Zusammenhang.



Beschreibung des Zusammenhangs:

b) Zum Bewertung der Prognosegüte des automatischen Tools wurde eine lineare Regression durchgeführt. Sie erhalten den folgenden Output:

```
Call: lm(formula = RE ~ AT)
 Residuals:
 Min         1Q         Median    3Q        Max
 -1.60315    -0.31640   0.03382   0.40298   1.20460

Coefficients:
              Estimate   Std. Error   t value   Pr(>|t|)
 (Intercept)  0.11061    0.05446      2.031     0.045 *
 EG           0.87736    0.02995      29.290    <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5435 on 98 degrees of freedom
Multiple R-squared: 0.8975, Adjusted R-squared: 0.8964
F-statistic: 857.9 on 1 and 98 DF, p-value: < 2.2e-16
```

(i) Geben Sie den, im Modell angenommenen, Zusammenhang zwischen der Prognose des Tools und dem tatsächlichen Return für eine Analyseentscheidung als Gleichung an.

Angenommener Zusammenhang:

(ii) Geben Sie die Korrelation zwischen der Prognose des Tools und dem tatsächlichen Return an. Begründen Sie, ob die Korrelation zum 5%-Niveau signifikant ist oder nicht?

Formel zur Berechnung der Korrelation:

Ergebnis für die Korrelation:

Argumentation bezüglich der Signifikanz:

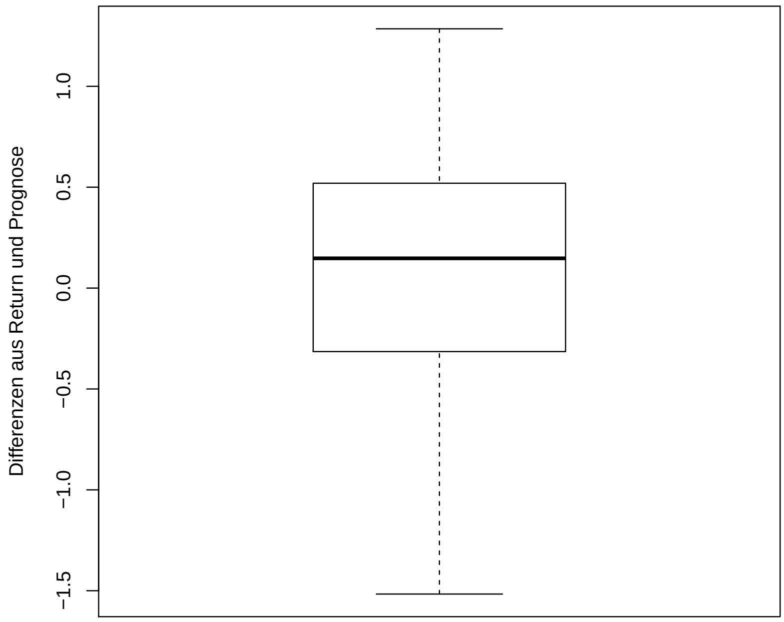
(iii) Bewerten Sie den Zusammenhang zwischen der Prognose des Tools und dem tatsächlichen Return.

Bewertung des Zusammenhangs:

c) Welche beiden Annahmen sind beim Regressionsmodell nötig, damit die p-Werte aus dem Modell interpretiert werden können?

Annahmen:

d) Nun werden die Differenzen der tatsächlichen Returns und der Prognosen des Tools betrachtet. Sie erhalten für die Differenzen folgenden Boxplot:



Können Sie anhand dieser Grafik nachweisen, ob im Allgemeinen ein systematischer Unterschied zwischen den tatsächlichen Returns und den Prognosen des Tools vorliegt? Geben Sie Ihre Entscheidung an und begründen Sie diese kurz.

Antwort:

e) Sie möchten nun mittels eines Tests entscheiden ob sich die tatsächlichen Returns signifikant von den Prognosen des Tools unterscheiden. Geben Sie an, welchen Test sie dafür verwenden würden und begründen jeweils Sie Ihre Entscheidung.

Antwort und Begründung:

Aufgabe 4	20 Punkte

Zum Lösen dieser Aufgabe stehen Ihnen folgende Quantile zur Verfügung:

$\mathbf{z_{0.9999}} = 3.72$ $\mathbf{z_{0.9995}} = 3.29$ $\mathbf{z_{0.999}} = 3.09$ $\mathbf{z_{0.995}} = 2.58$

$\mathbf{z_{0.99}} = 2.33$ $\mathbf{z_{0.975}} = 1.96$ $\mathbf{t^{999}_{0.9999}} = 3.73$; $\mathbf{t^{999}_{0.9995}} = 3.30$

$\mathbf{t^{999}_{0.999}} = 3.10$ $\mathbf{t^{999}_{0.995}} = 2.58$ $\mathbf{t^{999}_{0.99}} = 2.33$ $\mathbf{t^{999}_{0.975}} = 1.962$

Eine Studie zum Antrieb von Personenkraftwagen möchte den Anteil an Dieselmotoren in München mit Hilfe einer Stichprobe schätzen. Zum Jahresende 2015 waren 683433 PKW in München zugelassen. In einer unabhängigen Stichprobe werden nun 1000 PKW-Besitzer in München gefragt, ob sie ein Dieselfahrzeug besitzen. Von den befragten Personen haben dies 45.9% bestätigt.
(*Hinweis*: Eine Berücksichtigung des endlichen N ist bei der Berechnung nicht notwendig)

a) Berechnen Sie ein geeignetes Konfidenzintervall ($\gamma = 0,95$) für den Anteil an Dieselfahrzeugen in München.
(i) Definieren sie zunächst die Zufallsvariable und geben deren Verteilung (mit Parameterwerten) an.

Definition der Zufallsvariable und Angabe der Verteilung:

(ii) Bestimmen Sie das geeignete Quantil zur Berechnung des Konfidenzintervalls und begründen Sie Ihre Wahl kurz.

Gewähltes Quantil mit Begründung:

(iii) Berechnen Sie das Konfidenzintervall!

Konfidenzintervall für Anteil der Dieselfahrzeuge in München:

b) Ist es möglich, dass der wahre Anteil an Dieselfahrzeugen in München *nicht* im Konfidenzintervall liegt? Begründen Sie Ihre Antwort kurz.

Antwort mit Begründung:

Im Moment wird in vielen Städten darüber diskutiert, Dieselmotoren zu verbannen. Für Politiker ist eine solche Regelung einfacher durchzusetzen, wenn nur *wenige* PKW betroffen sind. Sie möchten jedoch sicher gehen, dass eine Verbannung nicht zu viele PKW betrifft und sie so bei vielen Wählern in Missgunst fallen.
Deshalb soll ein Hypothesentest durchgeführt werden, um zu überprüfen, ob der Anteil der PKW mit Dieselmotoren in München unter 50% liegt. Nur für den Fall, dass der Anteil laut Test unter 50% liegt, soll die Verbannung in Kraft treten.

c) Stellen Sie die Nullhypothese und die Alternativhypothese für das Testproblem auf. Erläutern Sie außerdem, was der Fehler erster Art inhaltlich bedeutet und was seine Folgen wären.

Hypothesen:

Erläuterung der Fehlers 1. Art und seiner Folgen:

d) Sie vermuten, dass ein Befürworter des Dieselverbots an der Durchführung des Tests beteiligt ist und dass dieser ein, für seine Interessen strategisch günstiges α wählt.
Zur Debatte stehen folgende Werte: $\alpha = 0.1$, $\alpha = 0.05$ und $\alpha = 0.01$. Erläutern Sie kurz, welche Wahl für den Befürworter strategisch am günstigsten ist.

Wahl mit Begründung:

e) Ein einseitiger (`alternative = ('less')`) approximativer Test wird nun mit R für $\alpha = 0.01$ durchgeführt. Sie erhalten den folgenden Output:

```
> t.test(x, mu = 0.5, alternative = c('less'), conf.level = 0.99)


One Sample t-test


data: x
t = -2.6005, df = 999, p-value = ???
alternative hypothesis: true mean is less than 0.5
99 percent confidence interval:
-Inf 0.4957362
sample estimates:
mean of x
0.459
```

(i) Wie lautet die Testentscheidung (mit Begründung)? Interpretieren Sie diese bezüglich des Dieselverbots.

Testentscheidung und Interpretation:

(ii) Gehen Sie kurz darauf ein, ob die Approximation, welche für den Test benötigt wurde, hier gerechtfertigt ist und begründen Sie Ihre Entscheidung.

Entscheidung mit Begründung:

f) Bei der obigen Durchführung des approximativen Binomialtests erhalten Sie einen Wert für die Teststatistik. Geben Sie ein möglichst kleines Intervall an, in dem der p-Wert des Tests liegt und interpretieren Sie diesen kurz.

Angabe des Intervalls:

Interpretation des Intervalls für den p-Wert:

g) In der Aufgabestellung wurde der geschätzte Anteil angegeben, der auch dem Maximum-Likelihood-Schätzer für π entspricht. Der Anteilsschätzer aus dieser Aufgabe ist erwartungstreu. Gilt dies auch im Allgemeinen für den Maximum-Likelihood-Schätzer?
(Hinweis: Ohne Begründung werden keine Punkte vergeben.)

Lösung mit Begründung:

Viel Erfolg!