

Handreichung für Dozierende zum Themenpaket Statistik

OER.Stochastik.nrw

Materialien in diesem Paket

1 Video, 6 Interaktive Anwendungen, 13 Aufgaben

Fachbereiche

Mathematische Statistik und Wahrscheinlichkeitsrechnung, Mathematik, Physik, Informatik, Ingenieurwissenschaften, Wirtschaftswissenschaften, Pharmazie, Humanmedizin und Gesundheitswissenschaften

Kurzbeschreibung der Materialien

Video: Was ist ein Histogramm und was eine empirische Verteilungsfunktion?

Am Beispiel der Verteilung des Gewichts von Schokolinsen werden das Histogramm und die empirische Verteilungsfunktion eingeführt.

Interaktive Anwendung: Der Einfluss der Klassenbreite auf das Histogramm

In dieser interaktiven Anwendung können die Studierenden den Einfluss der Klassenbreite auf das Histogramm untersuchen. Dabei können ein voreingestellter Datensatz, eigene Datensätze oder zufällig erzeugte Datensätze verwendet werden.

Interaktive Anwendung: Zufallsexperiment „Würfeln“

In dieser interaktiven Anwendung würfeln die Studierenden digital mit einem sechsseitigen fairen Würfel. Der Fokus liegt auf dem Vergleich verschiedener Darstellungs- bzw. Visualisierungsformen des durch mehrfaches Würfeln entstehenden Datensatzes, der die Ergebnisse der Durchführungen enthält.

Interaktive Anwendung: Lineare Regression

In dieser interaktiven Anwendung wird die Residuenquadratsumme als Maß für die Güte eines Regressionsmodells thematisiert. Zudem erkunden die Studierenden, welche Auswirkungen die Veränderung von Daten auf ein angepasstes Modell haben kann.

Interaktive Anwendung: Verteilung von Schätzern

In dieser interaktiven Anwendung werden jeweils zwei verschiedene Schätzer für den Erwartungswert und die Varianz normalverteilter Daten diskutiert. Um die Verteilung dieser Schätzer zu veranschaulichen, können Daten mit verschiedenen Parametern simuliert werden.

Interaktive Anwendung: Vergleich von Schätzern

In dieser interaktiven Anwendung werden jeweils zwei verschiedene Schätzer für den Erwartungswert und die Varianz normalverteilter Daten diskutiert. Die Schätzer können dann in Hinblick auf die Kenngrößen MSE und Bias verglichen werden. Dazu kann der Prozess der Stichprobenziehung mit verschiedenen Parametern simuliert werden.

Interaktive Anwendung: Der Binomialtest

Diese Anwendung behandelt die Grundlagen der Testtheorie im Sachkontext der Qualitätskontrolle. Es werden die beiden Begriffe *Fehler 1. Art* und *Fehler 2. Art* anhand zweier unterschiedlicher Testverfahren thematisiert. Es wird zunächst ein naives Testverfahren vorgeschlagen und danach der Binomialtest untersucht.

Aufgabe: Boxplot

In dieser Aufgabe ist eine Grafik mit einem Boxplot gegeben. Die Studierenden sollen einen Datensatz angeben, der zum gegebenen Boxplot passt.

Aufgabe: Empirische Verteilungsfunktion zeichnen

In dieser Aufgabe sollen die Studierenden eine empirische Verteilungsfunktion aus einem gegebenen Datensatz zeichnen, indem sie in einer Grafik Punkte platzieren.

Aufgabe: Regressionsgerade (Schritt für Schritt)

In dieser Aufgabe berechnen die Studierenden aus gegebenen Datenpunkten Schritt für Schritt eine Regressionsgerade nach der Kleinste-Quadrate-Methode. Dazu füllen sie sukzessive eine Tabelle aus, berechnen die Spaltensummen und bestimmen schließlich den Korrelationskoeffizienten sowie die Parameter der Regressionsgeraden.

Aufgabe: Regressionsgerade (Grafik)

In der ersten Teilaufgabe ist der Scatterplot eines Datensatzes gegeben und es soll zu einem x -Wert, zu dem kein Datenpunkt existiert, ein passender y -Wert geschätzt werden, sodass sich der resultierende Punkt plausibel in die Punktwolke einfügt. Im zweiten Aufgabenteil soll aus bereits gegebenen Kenngrößen eine Regressionsgerade nach der Kleinste-Quadrate-Methode bestimmt werden. In der dritten Teilaufgabe soll der zuvor geschätzte y -Wert mithilfe der Geradengleichung rechnerisch bestimmt werden.

Aufgabe: Verkehrstote in Bochum (ML-Schätzer)

In dieser Aufgabe wird die jährliche Anzahl der Verkehrstoten in der Stadt Bochum als Realisierung einer Poisson-verteilten Zufallsvariablen modelliert. Aus Daten von 12 aufeinanderfolgenden Jahren sollen der Maximum-Likelihood Schätzer sowie ein Konfidenzintervall für den unbekannt Parameter der Poisson-Verteilung bestimmt werden. Die Studierenden werden schrittweise an die Berechnung des ML-Schätzers herangeführt.

Aufgabe: Das Gewicht einer Großpackung Schokolinsen

Mitarbeiter:innen eines Lehrstuhls haben ein Stichprobe von Schokolinsenpackungen gewogen. Wir modellieren das Gewicht durch eine normalverteilte Zufallsvariable mit unbekanntem Erwartungswert und unbekannter Varianz. In dieser Aufgabe wird nach einem Konfidenzintervall

für den Erwartungswert und für die Varianz gefragt. Außerdem wird eine Verständnisfrage zu Konfidenzintervallen im Allgemeinen gestellt.

Aufgabe: Qualitätskontrolle im Sauerland

In dieser Aufgabe wird eine typische Fragestellung aus der Qualitätskontrolle vorgestellt. Die Studierenden untersuchen in dieser Aufgabe das Verfahren mithilfe des Binomialtests. In den fünf Aufgabenteilen werden die Studierenden nach dem statistischen Modell, nach dem Fehler 1. Art, nach der Macht und dem p-Wert gefragt. Außerdem sollen sie eine geeignete Hypothese und Alternative für das vorgestellte Verfahren formulieren.

Aufgabe: IQ-Stichprobe

Anhand einer IQ-Stichprobe soll den Studierenden ein möglicher Anwendungsfall für den Gaußtest vorgestellt werden. In vier Aufgabenteilen leitet die Aufgabe durch die Vorgehensweise bei diesem Test.

Aufgabe: Die tägliche Radfahrt zur Universität (F-Test)

In dieser Aufgabe soll ein F-Test zum Vergleich von Varianzen von Fahrzeiten durchgeführt werden. Es wurden in zwei verschiedenen Monaten im Jahr Fahrzeiten mit dem Rad für eine Strecke von der Wohnung zur Universität ermittelt. Den Studierenden werden die Kennzahlen Mittelwert und Stichprobenvarianz für die zwei Stichproben präsentiert. Es soll die Hypothese, dass die beiden Varianzen gleich sind, gegen die Alternative, dass die Varianz in einem Monat größer war, getestet werden.

Aufgabe: Die tägliche Radfahrt zur Universität (t-Test)

In dieser Aufgabe soll ein t-Test zum Vergleich von Erwartungswerten von Fahrzeiten durchgeführt werden. Es wurden in zwei verschiedenen Monaten im Jahr Fahrzeiten mit dem Rad für eine Strecke von der Wohnung zur Universität ermittelt. Den Studierenden werden die Kennzahlen Mittelwert und Stichprobenvarianz für die zwei Stichproben präsentiert. Es soll die Hypothese, dass die beiden Erwartungswerte gleich sind, gegen die Alternative, dass der Erwartungswert in einem Monat größer war, getestet werden.

Aufgabe: Farbverteilung in einer Packung Schokolinsen

Im Zentrum dieser Aufgabe steht eine Behauptung über die Farbverteilung von Schokolinsen. Den Studierenden wird ein Datensatz von 180 Packungen präsentiert. Mithilfe des Chi-Quadrat-Tests sollen die Studierenden entscheiden, ob die Daten die behauptete Farbverteilung stützen. Die Studierenden werden schrittweise durch dieses Vorhaben geleitet.

Aufgabe: Test auf einen fairen Würfel

Den Studierenden wird ein Datensatz eines Würfelexperiments präsentiert. Mithilfe eines geeigneten statistischen Tests und dieses Datensatzes sollen die Studierenden entscheiden, ob der Würfel fair ist. Die Studierenden werden schrittweise durch dieses Vorhaben geleitet.

Aufgabe: Energieverbrauch Fernwärme

Professor D. hat den Energieverbrauch seines Hauses in Abhängigkeit der Außentemperatur gemessen. Den Studierenden werden in dieser Aufgabe die arithmetischen Mittel, die empirischen Standardabweichungen sowie der Korrelationskoeffizient für diesen Datensatz gegeben.

Die Daten sollen nun mithilfe der linearen Regression untersucht werden. In sieben Aufgabenteilen sollen die Kleinste-Quadrate Regressionsgerade bestimmt werden, eine Vorhersage für den Energieverbrauch gemacht werden, eine Schätzung für die Varianz des Zufallsterms im Modell gegeben werden, 95%-Konfidenzintervalle für die Steigung, den y -Achsenabschnitt und die Varianz des Zufallsterms angegeben werden und die Hypothese, dass die Außentemperatur keinen Einfluss auf den Energieverbrauch hat, getestet werden.

Schlagwörter

*Mit einem * markierte Begriffe werden als bekannt vorausgesetzt.*

Lageparameter, Mittelwert, Median, Quartil, Streuungsmaß, Stichprobenvarianz, Interquartilsabstand, *Boxplot, Histogramm, *empirische Verteilungsfunktion, *gemeinsame Wahrscheinlichkeitsfunktion, *gemeinsame Dichtefunktion, Schätzer, Schätzwert, Maximum-Likelihood-Schätzer, Konfidenzintervall, Hypothese und Alternative, Fehler 1. und 2. Art, Niveau und Macht eines statistischen Tests, Binomialtest, Tests bei normalverteilten Stichproben, t -Test, χ^2 -Test (bei diskreten Daten), F -Test

Einsatzszenarien

Das Themenpaket ist modular aufgebaut, alle Materialien können einzeln und unabhängig voneinander eingesetzt werden.

- Das Video bietet einen motivierenden Einstieg in das Thema Histogramm und empirische Verteilungsfunktion. Sie wird parallel zum entsprechenden Vorlesungsabschnitt gezeigt oder den Studierenden zur eigenständigen Vorbereitung im Flipped Classroom Szenario zur Verfügung gestellt.
- Mit den interaktiven Anwendungen erkunden die Studierenden eigenständig und mithilfe von Beobachtungsaufträgen verschiedene Themen der Statistik. Sie wiederholen und vertiefen auf diese Weise den Vorlesungsstoff. Die interaktiven Anwendungen können auch direkt in der Vorlesung gezeigt werden.
- Die Aufgaben ergänzen im wöchentlichen Übungsbetrieb die schriftlichen Hausaufgaben zur Statistik. Die Studierenden nutzen sie zur Wiederholung des Vorlesungsstoffs und zur Prüfungsvorbereitung. Weitere Informationen befinden sich in separaten Handreichungen zu den einzelnen Aufgaben.

Ausblick

Das Projekt OER.Stochastik.nrw stellt weiterführende Materialien zur Verfügung.

- Das Themenpaket **Diskrete Verteilungen** beinhaltet Materialien zu diskreten Zufallsvariablen, ihren Verteilungen und Kennzahlen wie Erwartungswert und Varianz sowie zu konkreten Anwendungen bei der Modellierung diskreter Daten.
- Das Themenpaket **Stetige Verteilungen** gibt eine Einführung in die Theorie dichteverteilter Zufallsvariablen und deren Anwendung bei der Modellierung stetiger Daten.

Literaturhinweise

- [1] H. Dehling, B. Haupt (2004) Einführung in die Wahrscheinlichkeitstheorie und Statistik. 2. Auflage. Springer-Verlag Berlin Heidelberg. Kapitel 11, 12
- [2] N. Henze (2019) Stochastik: Eine Einführung mit Grundzügen der Maßtheorie. 1. Auflage. Springer Spektrum Berlin, Heidelberg. Kapitel 7.1 – 7.4
- [3] N. Henze (2021) Stochastik für Einsteiger. 13. Auflage. Springer Spektrum Berlin Heidelberg. Kapitel 5, 29, 30, 34



„Handreichung für Dozierende zum Themenpaket Statistik“ wurde entwickelt von Riko Kelter, Jonas Lache, Daniel Meißner, Christian Müller (OER.Stochastik.nrw)

Dieses Werk ist lizenziert nach CC BY SA 4.0. Ausgenommen aus der Lizenz sind alle Logos. Der Lizenzvertrag ist hier abrufbar: <https://creativecommons.org/licenses/by-sa/4.0/>

Das Werk ist online verfügbar unter: <https://www.orca.nrw>