# Statistik Vorlesung

**04. Juli 2025**
Arbeitszeit: 100 Minuten

| | | | |
|---|---|---|---|
| VORNAME: | | MATR.NR.: | |
| NACHNAME: | | | |

ERLAUBT: **Skriptum des Instituts, Taschenrechner gemäß Liste**

VERBOTEN: **alle sonstigen Unterlagen, Handys**

Aufgabe 1 und 2: Für jede vollständig richtig gelöste Frage gibt es 2 Punkte. Es gibt keine Teilpunkte und keine Minuspunkte.

Lösungswege müssen nachvollziehbar aufgeschrieben sein.

---

| Aufgabe | max. Punkte | erreichte Punkte | Teilsumme |
|---|---|---|---|
| 1 | 10 | | |
| 2 | 10 | | |
| 9 | 12 | | |
| 3 | 09 | | |
| 4 | 12 | | |
| 5 | 10 | | |
| 6 | 13 | | |
| 7 | 11 | | |
| 8 | 13 | | |
| Summe | 100 | | |
| Note | | | |

1. (10 Punkte)

a) Für welche der folgenden Merkmale ist der Median ein geeignetes Lagemaß? Kreuzen Sie die beiden Aussagen an, die auf jeden Fall zutreffen.

- [ ] Nationalität
- [ ] Geschlecht
- [X] Schulnote
- [X] Einkommen
- [ ] Familienstand

b) Ein Korrelationskoeffizient beträgt $\rho = 0,3$. Was folgt daraus für die Regressionsgerade? Kreuzen Sie die beiden Aussagen an, die auf jeden Fall zutreffen.

- [X] Die Steigung der Regressionsgeraden ist positiv
- [ ] Die Steigung der Regressionsgeraden ist negativ
- [ ] Die Steigung der Regressionsgeraden ist $0,3$
- [ ] Der $y$-Achsenabschnitt beträgt $0,3$
- [X] Die Punkte des Streudiagramms streuen um eine steigende Gerade

c) Ergänzen Sie die Textlücke im nachstehenden Satz so, dass jedenfalls eine wahre Aussage entsteht.

Beim Durchführen eines statistischen Tests wird die Nullhypothese

[verworfen], wenn der p-Wert kleiner als das Signifikanzniveau $\alpha$ ist.

d) Ergänzen Sie die Textlücken im nachstehenden Satz durch Ankreuzen des jeweils zutreffenden Ausdrucks aus den nachstehenden Tabellen so, dass jedenfalls eine wahre Aussage entsteht.

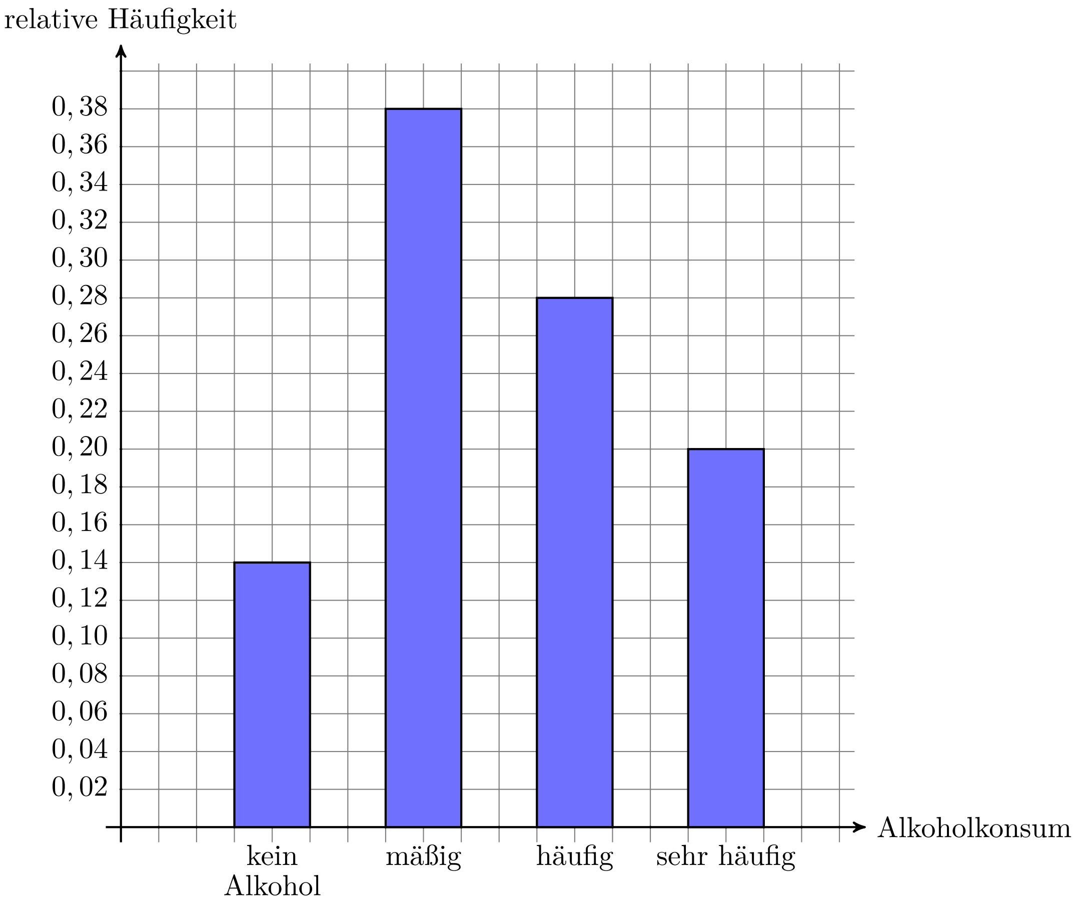
Als ____(1)____ einer **stetigen** Zufallsvariable $X$ bezeichnet man die Funktion, die die Wahrscheinlichkeit dafür angibt, dass die Zufallsvariable $X$ ____(2)____ annimmt.

| (1) | |
|---|---|
| Verteilungsfunktion $F(x)$ | ☒ |
| Dichtefunktion $f(x)$ | ☐ |
| Schätzfunktion $L(x)$ | ☐ |

| (2) | |
|---|---|
| genau den Wert $x$ | ☐ |
| mindestens den Wert $x$ | ☐ |
| höchstens den Wert $x$ | ☒ |

e) Eine Standardnormalverteilung hat immer einen Erwartungswert von [0] und eine Standardabweichung von [1].

2. (10 Punkte) In einer Umfrage wurde die Häufigkeit des Alkoholkonsum von 200 Männern erhoben. Es ergab sich folgende Verteilung:




a) Welches Merkmal wurde hier erhoben und welches Skalenniveau hat es in dieser Umfrage?

Häufigkeit des Alkoholkonsums; ordinal

b) Welche möglichen Ausprägungen besitzt es in dieser Umfrage?

kein Alkohol, mäßig, häufig, sehr häufig

c) Bestimmen Sie die Anzahl der Befragten, die angegeben haben keinen Alkohol zu trinken:

28

d) Kreuzen Sie die beiden Aussagen an, die auf jeden Fall zutreffen.

- [X] Der Median der Verteilung ist „mäßig“.
- [X] 48 % der Befragten gaben häufigen oder sehr häufigen Alkoholkonsum an.
- [ ] Drei Viertel der Befragten gaben an Alkohol zu konsumieren.
- [ ] Die Spannweite der Verteilung beträgt 0,24.

e) Geben Sie den Modalwert der Verteilung an: mäßig

3. Ein Unternehmen möchte wissen, ob es einen Zusammenhang zwischen der Kundenzufriedenheit (niedrig oder hoch) und dem Kundentyp (Neu- oder Stammkunde) gibt. Zu diesem Zweck wurden 140 Kunden befragt.

   25 der befragten Neukunden bewerteten ihre Zufriedenheit als hoch. Von den 80 Stammkunden bewerteten nur 15 ihre Zufriedenheit als niedrig.

   a) (3 Punkte) Tragen Sie die entsprechenden absoluten Häufigkeiten in die nachstehende Tabelle ein und bilden Sie alle Zeilen- und Spaltensummen.

   b) (3 Punkte) Geben Sie die Tabelle mit den unter Unabhängigkeit erwarteten Häufigkeiten an. Geben Sie dabei exemplarisch für ein Feld den vollständigen Rechenweg an.

   c) (3 Punkte) Angenommen der $\chi^2$-Wert für diese Tabelle beträgt $23,398$. Bestimmen Sie dann den korrigierten Kontingenzkoeffizienten und interpretieren Sie diesen mit Bezug zur Aufgabenstellung.

Lösung:

a)

| $H_{ij}$ | Zufriedenheit niedrig | Zufriedenheit hoch | $\Sigma$ |
|---|---|---|---|
| Stammkunde | 15 | 65 | 80 |
| Neukunde | 35 | 25 | 60 |
| $\Sigma$ | 50 | 90 | 140 |

b)

| | Zufriedenheit niedrig | Zufriedenheit hoch | $\Sigma$ |
|---|---|---|---|
| Stammkunde | 28,57 | 51,43 | 80 |
| Neukunde | 21,43 | 38,57 | 60 |
| $\Sigma$ | 50 | 90 | 140 |

$$\frac{80 \cdot 50}{140} = 28,57$$

c)

$C_{korr} = 0,535$

Ein korrigierter Kontingenzkoeffizient von 0,535 zeigt einen mittleren Zusammenhang zwischen Kundenzufriedenheit und Kundentyp. Es gibt also Unterschiede in der Zufriedenheit zwischen Neu- und Stammkunden – die beiden Merkmale sind nicht unabhängig voneinander.

4. In einer Stadt gibt es zehn Bäckereien, die sich bezüglich ihres Umsatzes in drei Gruppen einteilen lassen: kleine, mittlere und große Bäckereien (wobei innerhalb einer Gruppe der Umsatz jeweils gleich ist). Im Jahr 2023 erzielten alle Bäckereien zusammen einen Gesamtumsatz von 3 Millionen Euro. 40 Prozent davon entfielen auf die einzige große Bäckerei, während die fünf kleinen Bäckereien einen Umsatz von insgesamt 600.000 Euro erzielten.

   a) (8 Punkte) Bestimmen Sie die Punkte der Lorenzkurve und zeichnen Sie diese anschließend in nachstehendes Koordinatensystem.

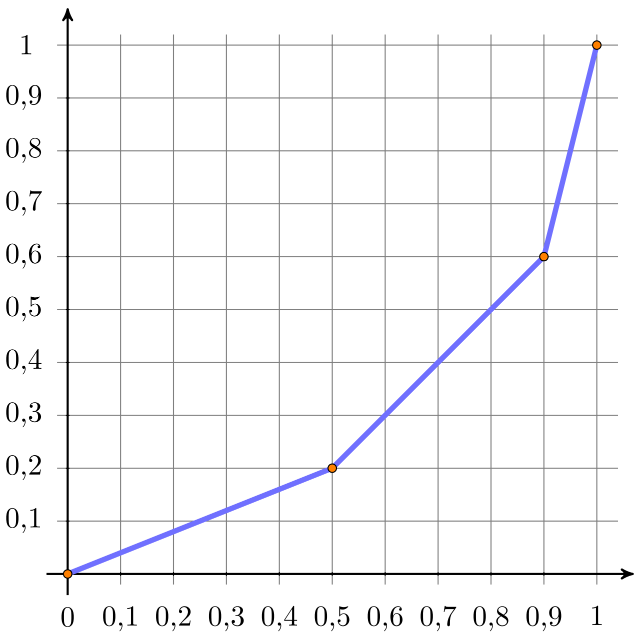
   b) (2 Punkte) Angenommen die Fläche unter der Lorenzkurve aus Aufgabe a) beträgt $0,29$. Berechnen Sie den Ginikoeffizienten.

   c) (2 Punkte) Die größte Bäckerei konnte im folgenden Jahr 2024 ihren Umsatz nochmals um 50 Prozent steigern, während der Umsatz der übrigen Bäckereien gleich bliebe. Erklären Sie, ohne Neuberechnung, wie sich die Lorenzkurve ändert.

Lösung:

a)

| | klein | mittel | groß |
|---|---|---|---|
| Anteil MT | 0,5 | 0,4 | 0,1 |
| $u_i$ | 0,5 | 0,9 | 1 |
| Anteil MS | 0,2 | 0,4 | 0,4 |
| $v_i$ | 0,2 | 0,6 | 1 |




b) 0,42

c) Wenn die größte Bäckerei ihren Umsatz um 50 % steigert und die Umsätze der anderen Bäckereien gleich bleiben, erhöht sich der Anteil der größten Bäckerei am Gesamtumsatz weiter. Die Lorenzkurve wird dadurch noch weiter von der Diagonalen (Gleichverteilung) entfernt liegen, insbesondere im letzten Abschnitt wird der Anstieg steiler, was eine noch ungleichere Umsatzverteilung widerspiegelt.

5. In einer gemeinnützigen Organisation haben sich 10 Freiwillige für die Mitarbeit an einem sozialen Projekt beworben, von denen 6 bereits Erfahrungen in ähnlichen Projekten gesammelt haben. Per Losentscheid werden 4 Freiwillige aus den Bewerbern für das Projekt ausgewählt.

   a) (4 Punkte) Durch welches Modell (d.h., durch welche Verteilung mit welchen Parametern) lässt sich beschreiben, wie viele erfahrene Freiwillige unter den per Los ausgewählten 4 Personen sind und warum? Erläutern Sie die Bedeutung der einzelnen Parameter.

   b) (2 Punkte) Wie groß ist die Wahrscheinlichkeit, dass bei der Auswahl von 4 Freiwilligen kein erfahrener Freiwilliger dabei ist?

   c) (2 Punkte) Wie groß ist die Wahrscheinlichkeit, dass mindestens ein erfahrener Freiwilliger in der Auswahl ist?

   d) (2 Punkte) Angenommen, das Auswahlverfahren würde sehr oft wiederholt. Wie viele der 4 ausgewählten Freiwilligen hätten im Durchschnitt bereits Erfahrung in ähnlichen Projekten?

   Lösung:

   a) Hypergeometrische Verteilung, Ziehen ohne Zurücklegen.

      $N = 10$: Gesamtzahl der Bewerber

      $M = 6$: Anzahl der erfahrenen Freiwilligen

      $n = 4$: Anzahl der gezogenen Freiwilligen

   b) 0,005

   c) 0,995

   d) $E(X) = 2,4$

6. Auf einer Maschine werden Ketchup-Flaschen abgefüllt. Ihr Inhalt $X$ ist normalverteilt mit dem Erwartungswert $E(X) = 400$ g und einer Standardabweichung von $\sigma = 2$ g.

   a) (4 Punkte) Skizzieren Sie den Graphen der Dichtefunktion dieser Normalverteilung in der nachstehenden Abbildung. Berücksichtigen Sie dabei den Erwartungswert und die Standardabweichung. Beschreiben Sie, wie sich eine kleinere Standardabweichung auf den Graphen der Dichtefunktion auswirken würde.

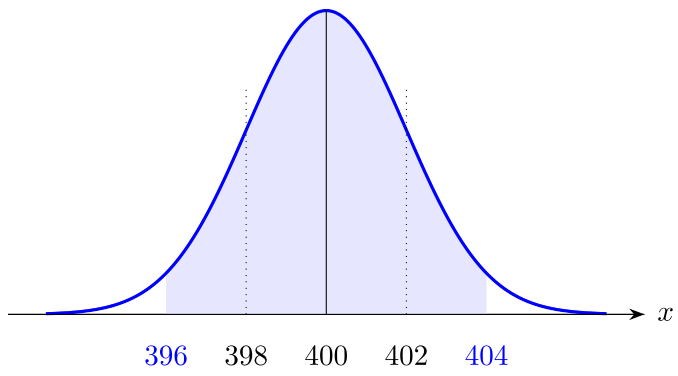
   b) (4 Punkte) Wie groß ist die Wahrscheinlichkeit, dass die Füllmenge vom Erwartungswert höchstens um 4 g abweicht? Veranschaulichen Sie Ihr Ergebnis in Ihrer Skizze.

   c) (3 Punkte) Welche Füllmenge wird nur von 5 % aller Flaschen unterschritten?

   d) (2 Punkte) Die Flaschen werden in Kartons zu je 12 Stück verpackt. Wie ist die Gesamtfüllmenge eines solchen Kartons verteilt? Geben Sie die Parameter dieser Verteilung an.

Lösung:

a)




Eine kleinere Standardabweichung bewirkt, dass die Glockenkurve schmaler und höher wird. Die Werte liegen dann näher am Mittelwert, die Kurve wird steiler und konzentriert sich stärker um den Erwartungswert.

b) $P(396 \leq 404) = 0,954$

c) 396,71

d) Die Summe von 12 unabhängigen normalverteilten Zufallsvariablen ist wieder normalverteilt. $S \sim N(4800; 6,928))$

7. Baustellen sorgen in Innsbruck für zahlreiche Verkehrsbehinderungen und Verspätungen im Berufsverkehr. Um die Auswirkungen auf die Pendelzeiten zu untersuchen, wurden dieselben Personen jeweils vor und nach Beginn der Bauarbeiten gefragt, wie lange (in Minuten) sie für ihren Weg zur Arbeit benötigen.

   Es soll zu einem Testniveau von $\alpha = 5$ % ohne Annahme von Normalverteilung nachgewiesen werden, dass sich durch die Baustellen signifikant längere Wegzeiten ergeben.

   a) (4 Punkte) Welcher Test ist anzuwenden? Begründen Sie Ihre Entscheidung.

   b) (2 Punkte) Formulieren Sie die Hypothesen dieses Tests.

   c) (3 Punkte) Der Testwert beträgt $t_0 = 2,52$. Bestimmen Sie den p-Wert und entscheiden Sie auf Basis des berechneten p-Wertes.

   d) (2 Punkte) Erläutern Sie, was Ihre Testentscheidung im Zusammenhang mit den Baustellen und den Pendelzeiten bedeutet.

   Lösung:

   a) Test auf Lage; Zwei Stichproben; Allgemeine Verteilung; verbunden $\Rightarrow$ Wilcoxon-Vorzeichen-Rang-Test

   b) $H_0 : d_{0,5} \leq 0$ gegen $H_1$: $d_{0,5} > 0$ mit $d = y - x$ (nach – vor)

   c) $p = 0,006 < \alpha = 0,05$ daher $H_0$ verwerfen.

   d) Es konnte auf dem 5 %-Niveau gezeigt werden, dass sich die Pendelzeiten durch die Baustellen signifikant verlängert haben.

8. Die zahlreichen Baustellen in Graz führen nicht nur zu längeren Wegzeiten, sondern zwingen viele Menschen, ihren Tagesablauf und ihr Mobilitätsverhalten anzupassen. Vor Beginn der Bauarbeiten nutzten die Bewohner des Zentrums von Graz die Verkehrsmittel Auto, Bus und Fahrrad wie folgt: 35 % Auto, 30 % Bus, 35 % Fahrrad. Nach Beginn der Bauarbeiten wurden 200 Personen befragt, welches Verkehrsmittel sie aktuell hauptsächlich nutzen. Die beobachteten Häufigkeiten dabei waren: 50 Personen nutzen das Auto, 70 den Bus und 80 das Fahrrad.

   Es soll auf einem Signifikanzniveau von 5 % untersucht werden, ob sich die Nutzung der Verkehrsmittel durch die Baustellen signifikant von den ursprünglich bekannten Anteilen unterscheidet.

   a) (2 Punkte) Welcher Test ist anzuwenden?

   b) (11 Punkte) Formulieren Sie die Hypothesen kontextbezogen. Geben Sie den Wert der Testgröße und die Testentscheidung an und interpretieren Ihr Ergebnis sowohl statistisch als auch inhaltlich.

   Lösung:

   a) Anpassungstest; Allgemeine Verteilung; Verteilung konkret gegeben $\Rightarrow$ Chi-Quadrat-Anpassungstest

   b) $H_0$: Die Verteilung der Verkehrsmittelnutzung (Auto, Bus, Fahrrad) hat sich durch die Baustellen nicht verändert, d.h. die aktuellen Anteile entsprechen den ursprünglichen Anteilen:

      $X$ unterliegt der Verteilung $V_0$:

      | Auto | Bus | Fahrrad |
      |---|---|---|
      | 35% | 30% | 35% |

      $H_1$: Die Verteilung der Verkehrsmittelnutzung hat sich durch die Baustellen verändert, d.h. die aktuellen Anteile unterscheiden sich von den ursprünglichen Anteilen:

      $X$ unterliegt nicht der Verteilung $V_0$

      $t_0 = 8,81 \in K$, $H_0$ verwerfen, es konnte also auf dem 5 %_Niveau gezeigt werden, dass es einen signifikanten Unterschied zwischen der aktuellen und der ursprünglichen Verteilung der Verkehrsmittelnutzung gibt.

      Die Baustellen haben das Mobilitätsverhalten der Bewohner im Zentrum von Graz verändert. Die aktuelle Nutzung der Verkehrsmittel unterscheidet sich signifikant von der ursprünglichen Verteilung.

9. Ein Schulpsychologe hat an 60 Vorschulkindern die folgenden Kennwerte eines Schuleignungstests ermittelt: Mittelwert $\bar{x} = 40$, Standardabweichung $s_x = 5$. Nach dem ersten Schuljahr wurden die tatsächlichen schulischen Leistungen derselben Kinder erfasst. Diese weisen folgende Kennwerte auf: Mittelwert $\bar{y} = 30$, Standardabweichung $s_y = 4$. Die Kovarianz zwischen den Ergebnissen des Eignungstest und der schulischen Leistung beträgt $Cov(x, Y) = 10$.

   a) (4 Punkte) Stellen Sie die Regressionsgleichung zur Vorhersage der schulischen Leistungen ($Y$) auf Basis des Ergebnisses im Schuleignungstest ($X$) auf und zeichnen Sie die Regressionsgerade in das unten stehende Koordinatensystem ein.

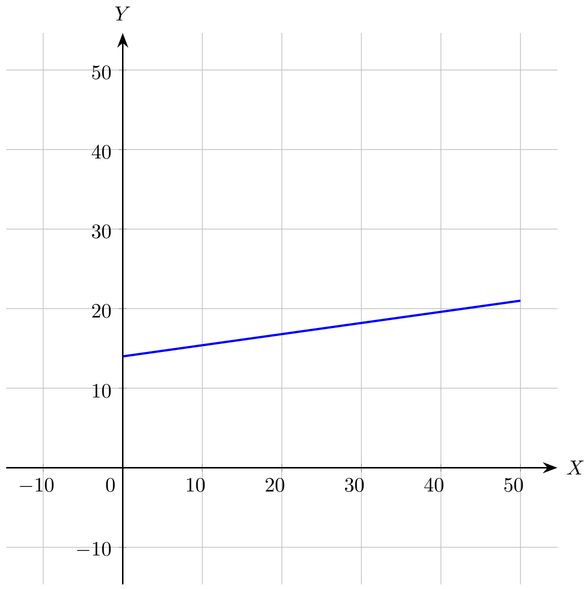
   b) (2 Punkte) Erläutern Sie, wie der Regressionskoeffizient der erklärenden Variable zu interpretieren ist.

   c) (2 Punkte) Mit welchem Wert für die schulische Leistung ist bei einer Schülerin zu rechnen, die im Schuleignungstest einen Wert von $x = 53$ erzielt hat?

   d) (4 Punkte) Ermitteln Sie das Bestimmtheitsmaß zur Beurteilung der Güte des Regressionsmodells und interpretieren Sie dessen Aussage.

Lösung:

a) $\hat{Y} = 14 + 0,4x$




b) Der Regressionskoeffizient $\hat{\beta}_1$ gibt an, um wie viele Einheiten sich die vorhergesagte abhängige Variable ($Y$) im Durchschnitt verändert, wenn die unabhängige Variable ($X$) um eine Einheit steigt. Im konkreten Fall bedeutet der Regressionskoeffizient von 0,4, dass die schulischen Leistungen der Kinder im Durchschnitt um 0,4 Punkte steigen, wenn ihr Ergebnis im Schuleignungstest um eine Einheit steigt.

c) $35,2$

d) $R^2 = 0,25$. Das Bestimmtheitsmaß von 0,25 bedeutet, dass 25 % der Varianz der schulischen Leistungen durch die Ergebnisse im Schuleignungstest erklärt werden können.