

Abschlussprojekt

von

Tat Dat Tran

Thema: Statistik mit R

Dozentin: Julianne Wawerda

Projektzeitraum: 15.12.2021-17.12.2021

Partner: Irene Poczka, Mohanad Abdaldayem

Inhaltsverzeichnis

[Aufgabe 1: Grundlagen 3](#_Toc41292296)

[Aufgabe 2: Multiple Choice 6](#_Toc41292297)

[Aufgabe 3: Zusammenhangshypothese 7](#_Toc41292298)

[Aufgabe 4: Unterschiedshypothese 8](#_Toc41292299)

[Aufgabe 5: Unterschiedshypothese 9](#_Toc41292300)

[Aufgabe 6: Unterschiedshypothese 10](#_Toc41292301)

# Aufgabe 1: Grundlagen

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SAP vorher | 2 | 5 | 2 | 7 | 5 | 6 | 1 | 3 | 7 | 3 |
| SAP nachher | 10 | 10 | 8 | 6 | 4 | 9 | 4 | 8 | 7 | 5 |

1. Berechne die Mittelwerte, Modus/Modi und die Mediane (SAPvorher und SAPnachher)

SAPvorher:

* Mittelwerte: (2+5+…+3) = 4.1
* Modi: 1:1, 2:2, 3:2, 5:2, 6:1, 7:2 🡪 {2,3,5,7}
* Mediane: 1 2 2 3 3 5 5 6 7 7 🡪 4

SAPnachher:

* Mittelwerte: (10+10+…+5) = 7.1
* Modi: 4:2, 5:1, 6:1, 7:1, 8:2, 9:1, 10:2 🡪 {4,8,10}
* Mediane: 4 4 5 6 7 8 8 9 10 10 🡪 7.5

1. Berechne die Varianzen und Standardabweichungen

(SAPvorher und SAPnachher)

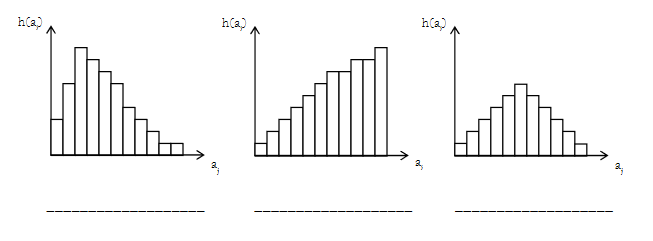
SAPvorher:

* Varianzen: ( + … ) = 4.29
* Standardabweichungen: = 2.07

SAPnachher:

* Varianzen: ( + … ) = 4.69
* Standardabweichungen: = 2.17

1. Ist der Graph recht-, linksschief und symmetrisch?



Rechtsschief Linksschief symmetrisch

1. Ordne der Daten das Skalenniveau zu: (Nominal, Ordinal, Intervall, Ratio, Absolut), welcher Operation ist erlaubt.

|  |  |  |
| --- | --- | --- |
| Art der Variable | Skalenniveau | Operation |
| Militärdienstgrad | Ordinal | =/ ≠  </> |
| Alter | Ratio | =/ ≠  </>  +/-  ×/÷ |
| Verkehrsdichte | Ordinal | =/ ≠  </> |
| Geschlecht | Nominal | =/ ≠ |
| Fahrpreise | Ratio | =/ ≠  </>  +/-  ×/÷ |
| Nationalität | Nominal | =/ ≠ |
| Schulbildung | Ordinal | =/ ≠  </> |
| Intelligenzquotient | Intervall | =/ ≠  </>  +/- |
| Studienfach | Nominal | =/ ≠ |
| Semesterzahl | Ordinal | =/ ≠  </> |
| Klausurpunkte | Ordinal | =/ ≠  </> |
| Tarifklassen bei der Kfz-Haftpflicht | Ordinal | =/ ≠  </> |

5) Ordne den Daten die folgenden Variablen das Variablenniveau zu (stetig vs. diskret).

|  |  |  |  |
| --- | --- | --- | --- |
| Nr. | Wert | Variable | |
| diskret | stetig |
| 1 | Steuerklasse |  |  |
| 2 | Geschlecht |  |  |
| 3 | soziale Schicht |  |  |
| 4 | Einkommenssteuer |  |  |
| 5 | Temperatur in Kelvin |  |  |
| 6 | Windstärke in Meter/Sekunde |  |  |
| 7 | Körpergewicht |  |  |
| 8 | Schulnote (1-6) |  |  |
| 9 | Klausurpunkte |  |  |
| 10 | Einwohnerzahl |  |  |
| 11 | Semesterzahl |  |  |
| 12 | Handelsklasse (Obst) |  |  |

1. Beschreibe in Sätzen, was der Unterschied und Gemeinsamkeiten zwischen Standardnormalverteilung und der Normalverteilung ist. Verwenden Sie die Formeln.

Die Normalverteilung und die Standardnormalverteilung haben folgende Eigenschaften:

- Symmetrisch

- Glockenförmig

- Mittelwert und Median sind gleich; beide befinden sich im Zentrum der Verteilung

- Der Mittelwert der Normalverteilung bestimmt ihre Lage und die Standardabweichung ihre Streuung.

Die Standardnormalverteilung aber ist eine besondere Form der Normalverteilung und liegt dann vor, wenn wir eine Normalverteilung mit einem Mittelwert von μ = 0 und einer Standardabweichung von σ = 1 haben.

**Formeln:**

Die Normalverteilung: z.B. Formel für die Verteilung einer Variablen:

f(x)=.

Formel für die Standardnormalverteilung:

f(x)=.

# Aufgabe 2: Multiple Choice

Der Sonderpunkt gilt nur innerhalb dieser Aufgabe. Maximal können Sie 10 Punkte erreichen.

|  |  |
| --- | --- |
| 1. Ein Bravais-Pearson-Korrelationskoeffizient von 0,85 deutet auf eine schwache lineare Korrelation hin. | |
| Richtig | Falsch (x) |
| 1. Der Interquartilsabstand (IQR) ist der doppelte Abstand zwischen Median und Modus. | |
| Richtig | Falsch (x) |
| 1. Die Modi lassen sich nur bestimmen, wenn eine unimodale Verteilung vorliegt. | |
| Richtig | Falsch (x) |
| 1. Nominalskalierte Daten können in eine natürliche Reihenfolge gebracht werden. | |
| Richtig | Falsch (x) |
| 1. Ausreißer wirken sich auf die Ergebnisse nicht robuster Analyseverfahren besonders stark aus. | |
| Richtig (x) | Falsch |
| 1. Die Standardabweichung berechnet sich nicht als positive Wurzel aus der Varianz. | |
| Richtig | Falsch (x) |
| 1. Die Kurtosis ist ein Maß für die Wölbung einer Verteilung. | |
| Richtig (x) | Falsch |
| 1. Die Berechnung der Varianz setzt mindestens metrisch skalierte Daten voraus. | |
| Richtig (x) | Falsch |
| 1. Die Spannweite ist der absolute Abstand zwischen dem kleinsten und dem größten Wert. | |
| Richtig (x) | Falsch |
| 1. Der Bravais-Pearson-Korrelationskoeffizient kann nur Werte zwischen 0 und 1 annehmen. | |
| Richtig | Falsch (x) |
| 1. Der statistische Ersatz fehlender Werte setzt mindestens metrisch skalierte Daten voraus. | |
| Richtig (x) | Falsch |

# Aufgabe 3: Zusammenhangshypothese

Datensatz: cleanhospital.xlsx

Var 1 = systolic\_blood\_pressure

Var 2 = weight

Aufgabenstellung

1. Hypothese
2. Voraussetzungen
3. Grundlegende Konzepte: Was ist Pearson?
4. Grafische Veranschaulichung des Zusammenhangs
5. Deskriptive Statistik
6. Ergebnisse der Korrelationsanalyse
7. Berechnung des Bestimmtheitsmasses
8. Berechnung der Effektstärke
9. Eine Aussage

# Aufgabe 4: Unterschiedshypothese

Datensatz: cleanhospital.xlsx

Var 1 = comorbidities (UV)

Var 2 = hemoglobin (AV)

Aufgabenstellung

1. Hypothese
2. Voraussetzungen des t-Tests für unabhängige Stichproben
3. Grundlegende Konzepte: Was ist t-Test für unabhängige Stichproben?
4. Deskriptive Statistiken
5. Test auf Varianzhomogenität (Levene-Test)
6. Ergebnisse des t-Tests für unabhängige Stichproben
7. Berechnung der Effektstärke
8. Eine Aussage

# Aufgabe 5: Unterschiedshypothese

Datensatz: foodDiet.csv

Var 1 = weight (AV)

Var 2 = time (pre,6weeks)

Aufgabenstellung

1. Hypothese
2. Voraussetzungen des t-Tests für abhängige Stichproben
3. Grundlegende Konzepte: Was ist t-Test für abhängige Stichproben?
4. Deskriptive Statistiken und Korrelation
5. Ergebnisse des t-Tests für abhängige Stichproben
6. Berechnung der Effektstärke
7. Eine Aussage

# Aufgabe 6: Unterschiedshypothese

Datensatz: cleanhospital.xlsx

Var 1 = fever (Faktor)

Var 2 = age (AV)

Aufgabenstellung

1. Hypothese
2. Voraussetzungen für die einfaktoriellen Varianzanalyse ohne Messwiederholung
3. Grundlegende Konzepte: Was ist die einfaktoriellen Varianzanalyse ohne Messwiederholung
4. Deskriptive Statistiken
5. Prüfung der Varianzhomogenität (Levene-Test)
6. Ergebnisse der einfaktoriellen Varianzanalyse ohne Messwiederholung
7. Post-hoc-Tests
8. Profildiagramm
9. Berechnung der Effektstärke
10. Eine Aussage