Statistik 3

Übung 8

Einleitung und Ziele der Übung

Herzlich Willkommen zur siebten Übung Statistik 3. In dieser Übung beginnen wir mit der Mehrebenenanalyse.

Agenda

- Ebenen und Variablen identifizieren
- Intercept-Only-Modell und Intraklassenkorrelation

Ebenen und Variablen identifizieren - Einführung

Zunächst wollen wir uns daran üben, Ebenen richtig zu bestimmen und ihnen die zugehörigen Variablen zuzuordnen. Bestimmen Sie für folgende Beispiele jeweils die zwei Ebenen und geben Sie für jede Ebene die zugehörigen Variablen an! Die Lösungen finden Sie im Anschluss an diese Übung.

Ebenen und Variablen identifizieren - Übung 1

Stellen Sie sich vor, Sie arbeiten im Bereich der Entwicklungs- und Schulpsychologie und interessieren sich (u.a.) für den Zusammenhang zwischen Motivation und mathematischen Fähigkeiten.

Sie verfügen über einen Datensatz, in dem für 210 Schulen und dort für je 100 Schüler*innen die Motivation und die Leistung in einem mathem.-nat. Test erhoben wurde. Damit Sie den Einfluss des Schwerpunkts der Schule berücksichtigen können, weisen jeweils 70 der Schulen einen gemeinsamen Schwerpunkt auf (mathemat-naturwissenschaftlich, sprachlich, künstlerisch). Für jede*n Schüler*in wurde zusätzlich eine (subjektive) Leistungsbeurteilung durch eine*n Lehrer*in vorgenommen.

Ebenen und Variablen identifizieren - Übung 2

Es sind harte Zeiten für den Weihnachtsmann, die Weltbevölkerung wächst und so auch die Anzahl der Kinder. Es stellt sich daher die Frage, ob die Spielzeugproduktion am Nordpol stärker mit Maschinen durchgeführt werden sollte.

Um die Kinderzufriedenheit sicherzustellen, fand im Jahr 2020 ein grossangelegter Test statt. Es wurden 600 zufällige Elfenwerkstätten ausgewählt, wobei 300 mit Maschinen ausgestattet wurden, während die 300 anderen bei Handarbeit blieben. In jeweils einem Drittel der Werkstätten wurden ungelernte Elfen eingesetzt, einem weiteren Drittel neu ausgebildete Elfen und im letzten Drittel Profielfen. Aus jeder Werkstatt wurden zufällig 200 Geschenke ausgewählt und jeweils der Aufwand in der Herstellung und die Kinderzufriedenheit je Geschenk erfasst.

Ebenen und Variablen identifizieren - Übung 3

Es gibt die Hypothese, dass Hirsche über mehrere Generationen kleinere Geweihe ausbilden, wenn Sie bejagt werden. Denn meist werden die Tiere geschossen, welche das schönste Geweih haben.

Eine Biologin möchte diese Theorie prüfen. Dazu bestimmt sie für 40 zufällig ausgewählte Wälder in der Schweiz, ob diese ein Naturschutzgebiet sind, saisonal bejagt werden dürfen, oder ganzjährig bejagt werden dürfen. Per Fotofalle macht sie in jedem Wald anschliessend Aufnahmen von mindestens 25 verschiedenen Hirschen um deren Geweihgrösse (in cm) zu ermitteln. Da grosse Geweihe bei Revierkämpfen von Vorteil sind, erfasst Sie ausserdem die Anzahl an Hirschen pro qm^2 für jeden Wald, sowie das geschätzte Alter jedes fotografierten Tieres in Jahren.

Ebenen und Variablen identifizieren - Übung 4

Ein befreundeter Bauer bittet Sie um Rat. Er möchte wissen, wie sich eine künstliche Bewässerung auf den Ertrag seiner Felder auswirkt.

Er besitzt insgesamt 50 Felder. Auf jedem Feld hat er je 5 bis 10 Bereiche als Testflächen abgesteckt. Jede Fläche wird entweder künstlich bewässert oder nicht. Da der Ertrag (in t) sicherlich nicht nur von der Bewässerung abhängig ist, sondern auch von anderen Faktoren, erfassen sie zusätzlich folgende Faktoren: Die Bodenqualität des Feldes (auf einer metrischen Skala), die angebaute Pflanze auf jeder Testfläche (Mais, Kartoffeln, Weizen oder Gerste) und die Grösse jeder Testfläche (in m^2).

Ebenen und Variablen identifizieren - Übung 5

Lösungen

- Übung 1
  - Ebene 1 = Schüler*innen: mathem. Fähigkeiten (AV), Motivation (UV), Beurteilung Lehrer*in (UV)
  - Ebene 2 = Schulen: Schwerpunkt (UV)
- Übung 2
  - Ebene 1 = Geschenk: Zufriedenheit (AV), Aufwand (UV)
  - Ebene 2 = Werkstatt: Fertigung (UV), Ausbildung (UV)
- Übung 3
  - Ebene 1 = Hirsch: Geweihgrösse (AV), Alter (UV)
  - Ebene 2 = Wald: Bejagung (UV), Anzahl (UV)
- Übung 4
  - Ebene 1 = Testfläche: Ertrag (AV), Bewässerung (UV), Pflanze (UV), Grösse (UV)
  - Ebene 2 = Feld: Bodenqualität (UV)

2. Intercept-Only-Modell und Intraklassenkorrelation - Einführung

Auf der nächsten Seite folgt zunächst eine Übungsaufgabe zum Intercept-Only-Modell anschliessend eine Aufgabe, die sie nochmal von der Studienbeschreibung bis zur Intraklassenkorrelation führen wird.

2. Intercept-Only-Modell und Intraklassenkorrelation - Übung 1

Stellen Sie für jedes Beispiel aus der Übung zuvor das zugehörige Intercept-Only-Modell auf.

Wofür wird dieses Modell hauptsächlich genutzt und was muss dazu berechnet werden?

2. Intercept-Only-Modell und Intraklassenkorrelation - Übung 2

Sie wollen die Zufriedenheit der Patient*innen in Spitälern der Schweiz evaluieren. Dazu wählen Sie zufällig 10 Spitäler der Schweiz und überprüfen die Zufriedenheit der Patient*innen vor Ort. Als Kontrollvariablen erfassen Sie die Schwere der Erkrankung der Patient*innen (zentriert) sowie ihr Alter (zentriert) sowie die allgemeine Belastung des Personals (zentriert) in jedem Spital.

Welche Ebenen liegen vor und welche Variablen befinden sich auf welcher Ebene?

2. Intercept-Only-Modell und Intraklassenkorrelation - Übung 3

Laden Sie den Datensatz Uebung_Statistik3_8A.csv und nutzen Sie den folgenden R-Code, um sich für jede der drei Kontrollvariablen den Zusammenhang mit der Zufriedenheit insgesamt und je Spital anzeigen zu lassen. Was stellen Sie fest?

```r
library(ggplot2) #Vorab installieren, falls notwendig
#Insgesamt
ggplot(Uebung_Statistik3_8A, aes(x = Belastung, y = Zufriedenheit)) +
geom_point() + geom_smooth(method = "lm", se = FALSE, color = "red")
ggplot(Uebung_Statistik3_8A, aes(x = Schwere, y = Zufriedenheit)) +
geom_point() + geom_smooth(method = "lm", se = FALSE, color = "red")
ggplot(Uebung_Statistik3_8A, aes(x = Alter, y = Zufriedenheit)) +
geom_point() + geom_smooth(method = "lm", se = FALSE, color = "red")
```

```r
#Je Spital
ggplot(Uebung_Statistik3_8A, aes(x = Belastung, y = Zufriedenheit)) +
geom_point() + geom_smooth(method = "lm", color = "red") +
facet_wrap(~ Spital)
ggplot(Uebung_Statistik3_8A, aes(x = Schwere, y = Zufriedenheit)) +
geom_point() + geom_smooth(method = "lm", color = "red") +
facet_wrap(~ Spital)
ggplot(Uebung_Statistik3_8A, aes(x = Alter, y = Zufriedenheit)) +
geom_point() + geom_smooth(method ="lm", color = "red") +
facet_wrap(~ Spital)
```

2. Intercept-Only-Modell und Intraklassenkorrelation - Übung 4

Betrachtet man die Daten insgesamt, sieht man für Belastung und Schwere einen negativen Zusammenhang, für Alter einen positiven. Je Spital sehen wir für Belastung keinen Trend, da je Spital nur ein Wert für Belastung erhoben wurde (alle Personen haben in einem Spital den gleichen Wert). Für Alter sehen wir in jedem Spital klar einen positiven Trend. Für die schwere der Erkrankung, sehen wir sehr unterschiedliche Trends. Für die meisten Spitäler eher negativ, aber auch einige relative neutrale und sogar ein leicht positiver Trend in Krankenhaus 3.

Die Regressionsgeraden dieser Diagramme stellen übrigens die Simple Slopes für die jeweilige Kontrollvariable in Abhängigkeit des Spitals dar. Die zugehörige Regressionsgleichung hätte als Regressand Zufriedenheit und als Regressoren eine Kontrollvariable, die Spitäler (z.B. dummykodiert) und die Interaktionen der Kontrollvariablen mit den kodierten Spitälern. Mit anderen Worten, je unterschiedlicher die beobachten Anstiege, desto grösser die Interaktion zwischen der Kontrollvariable und den Spitälern.

2. Intercept-Only-Modell und Intraklassenkorrelation - Übung 5

Stellen Sie Gleichung(en) für das Intercept-Only-Modell auf!

Schätzen Sie mit folgendem Code das Intercept-Only_Modell in R. Installieren Sie dafür zunächst die Packages lme4 und lmerTest.

```
#Je Spital
library(lmerTest) #lme4 wird von lmerTest automatisch geladen
ergebnis <- lmer(Zufriedenheit ~ 1 + (1|Spital), Uebung_Statistik3_8A)
summary(ergebnis)
```

2. Intercept-Only-Modell und Intraklassenkorrelation - Übung 6

Wir können uns die Voraussetzungen/Annahmen wieder mittels check_model() aus dem performance-Package anschauen. Allerdings sollten wir die Normalverteilung der Residuen auch immer je Element der 2. Ebene (also hier je Spital) anschauen. Nutzen Sie den nachfolgenden Code um sich einen Überblick über die Voraussetzungen/Annahmen unseres Modells zu verschaffen:

```
#Gesamtmodel
library(performance)
check_model(ergebnis)
shapiro.test(residuals(ergebnis))

#Normalverteilung je Element der 2. Ebene:
#Histogramme und Dichte:
ggplot(Uebung_Statistik3_8A, aes(x=residuals(ergebnis))) +
geom_histogram(aes(y=after_stat(density)), colour="black", fill="white")+
geom_density(alpha=.2, fill="#FF6666") +
facet_wrap(~ Spital)
```

```
#QQPlots
ggplot(Uebung_Statistik3_8A, aes(sample = residuals(ergebnis))) +
stat_qq() +
stat_qq_line() +
facet_wrap(~ Spital)

#Shapiro-Wilk-Test (p-Werte sind in letzter Spalte)
aggregate(residuals(ergebnis) ~ Spital, Uebung_Statistik3_8A,
function(x) shapiro.test(x)$p.value)
```

2. Intercept-Only-Modell und Intraklassenkorrelation - Übung 7

Auf den ersten Blick sehen die Annahmen recht gut aus.

Es fallen aber im Diagramm 5 (Normality of Residuals) die Abweichungen der zwei sehr kleinen Werte auf. Zusammen mit der grossen Stichprobe, wird dann auch der der Shapiro-Wilk-Test insgesamt signifikant.

Schauen wir uns die Normalverteilung je Spital an, sehen wir, dass unser Modell vor allem in Spital 4 und 9 Probleme hat. Dort wurden sehr niedrige (die einzigen negativen) Zufriedenheiten berichtet. Eventuell waren dies Fehleingaben (falls Zufriedenheit eigentlich keine negativen Werte enthalten sollte).

Streng genommen müssten wir uns jetzt überlegen, wie wir mit der Verletzung der Annahme der Normalverteilung umgehen wollen. Da dieses Modell aber für uns nur als Basis für die Intraklassenkorrelation interessant ist, ignorieren wir diese leichten Abweichungen hier an dieser Stelle. Alternativ könnte man z.B. überlegen, die Personen auszuschliessen (wenn ihre Werte unplausibel sind) oder die Daten zu transformieren. Es könnte auch sein, dass unser Modell besser passt, sobald wir die Kontrollvariablen hinzufügen.

2. Intercept-Only-Modell und Intraklassenkorrelation - Übung 8

Bestimmen Sie die Intraklassenkorrelation basierend auf dem Intercept-Only-Modell.

Welcher Anteil der Varianz der Zufriedenheit geht allein darauf zurück, dass die Patient*innen in verschiedenen Spitälern waren ohne eine Berücksichtigung der Kontrollvariablen?

$r_{ICC} =$