

Statistik 3

Übung 3

Einleitung und Ziele der Übung

Herzlich Willkommen zur dritten Übung Statistik 3. Diese Übung führt Sie schrittweise zur multiplen linearen Regression (ohne Interaktionen mit intervallskallierten Variablen) und wiederholt einige Aspekte der deskriptiven Statistik.

Agenda

- Visualisierung und deskriptive Statistiken der Daten
- Korrelationsanalysen
- Einfache lineare Regression
- Multiple lineare Regression

1. Visualisierung und deskriptive Statistiken der Daten - Einführung

In der Stadt Luzern wurde eine Parkanlage neugestaltet. Um zu sehen, wie gut die Neugestaltung von der Bevölkerung angenommen wird, wurde eine Umfrage durchgeführt. Neben der Zufriedenheit (Skala von 1 bis 100) wurden weitere Variablen erfasst: Das Alter (in Jahren), die durchschnittliche Verweildauer pro Woche (in Minuten) und die Offenheit für neue Erfahrungen der Befragten. Sie können davon ausgehen, dass jede Variable mindestens intervallskaliert ist. Im Vorabbericht steht, dass die Befragten umso zufriedener mit der Neugestaltung waren, je jünger sie waren. Offenheit und Verweildauer konnten ebenfalls die Zufriedenheit vorhersagen. Sie wurden damit beauftragt diese Aussagen nochmals zu überprüfen, bevor der Bericht publiziert wird.

Zunächst wollen wir die Daten inspizieren. Laden Sie den Datensatz **Uebung_Statistik3_3A.csv** herunter und importieren Sie ihn mit R.

Überlegen Sie anschliessend wie die Befehle für Mittelwert, Standardabweichung, das Zeichnen von Histogrammen und das Zeichnen von Streudiagrammen lauteten.

1. Visualisierung und deskriptive Statistiken der Daten - Übung

Bestimmen Sie die die Werte für die angegebenen Statistiken. Visualisieren sie anschliessend die Verteilungen sowie die Zusammenhänge der vier Variablen.

Statistik		Wert
$\overline{x}_{Alter}$	=	
s_{Alter}	=	
$\overline{x}_{Verweildauer}$	=	
$s_{Verweildauer}$	=	
$\overline{x}_{Offenheit}$	=	
$s_{Offenheit}$	=	
$\overline{x}_{Zufriedenheit}$	=	
$s_{Zufriedenheit}$	=	

2. Korrelationsanalysen - Einführung

Jetzt wollen wir zunächst die bivariaten Zusammenhänge zwischen den Variablen analysieren. Die Voraussetzungen haben wir bereits visuell überprüft und auf den ersten Blick nichts problematisches entdeckt.

Überlegen Sie kurz, mit welchem Befehl Sie Pearson-Korrelationen in R bestimmen können. Dann sind Sie bereit für die nächste Übung.

2. Korrelationsanalysen - Übung

Vor statistischen Analysen (strenggenommen sogar schon vor den deskriptiven Statistiken), sollten zunächst die Hypothesen klar formuliert werden. Formulieren Sie für den Zusammenhang von Zufriedenheit mit den drei anderen Variablen jeweils die statistischen und inhaltlichen Null- und Alternativhypothesen. Nutzen Sie dafür die Informationen aus der Studienbeschreibung als Arbeitshypothesen.

Bestimmen Sie anschliessend die Pearson-Korrelationen, den zugehörigen Determinationskoeffizient und deren statistische Signifikanz ($\alpha = .05$) mittels R. Achten Sie darauf, die p -Werte korrekt für die zugehörigen Hypothesen zu bestimmen.

r_{ZA}	=	
R^2_{ZA}	=	
p_{ZA}	=	
r_{ZV}	=	
R^2_{ZV}	=	
p_{ZV}	<	
r_{ZO}	=	
R^2_{ZO}	=	
p_{ZO}	=	

Welche Hypothesen sollten verworfen/angenommen werden?

3. Einfache lineare Regression - Einführung

Jetzt soll das gleiche nochmals mit der einfachen linearen Regression durchgeführt werden. Mit welchen Befehlen führen Sie die einfache lineare Regression durch und mit welchem Befehl prüfen Sie die Voraussetzungen? Dann sind Sie bereit für die nächste Übung

3. Einfache lineare Regression - Übung 1

Wir wollen statt der drei Korrelationen drei einfache lineare Regressionen durchführen. Dabei soll die Zufriedenheit als Regressand dienen. Wie ändern sich die Hypothesen? Führen Sie anschliessend die linearen Regressionen durch und berichten Sie die angegebenen Werte.

Zufriedenheit und Alter

b_0	=		p_{b_0}	<	
b_1	=		p_{b_1}	=	
R^2	=				

Zufriedenheit und Verweildauer

b_0	=		p_{b_0}	<	
b_1	=		p_{b_1}	<	
R^2	=				

Zufriedenheit und Offenheit

b_0	=		p_{b_0}	<	
b_1	=		p_{b_1}	=	
R^2	=				

3. Einfache lineare Regression - Übung 2

Vergleichen Sie die Werte von Regression und Korrelation. Was fällt Ihnen auf?

Welche Werte aus der Regression brauchen Sie um dieselben Hypothesen wie die für die Korrelation zu prüfen?

Was verraten Ihnen die übrigen Werte?

Prüfen Sie noch schnell die Voraussetzungen in R. Gibt es Auffälligkeiten?

4. Multiple lineare Regression - Einführung

Jetzt wollen wir eine multiple lineare Regression (noch ohne Interaktionen) durchführen. Die in R benötigten Befehle ändern sich nicht. Sie können direkt in die Übung einsteigen.

4. Multiple lineare Regression - Übung 1

Stellen Sie zunächst die Regressionsgleichung für die multiple lineare Regression der Zufriedenheit auf Alter, Verweildauer und Offenheit auf. Hinweis (Bitte stellen Sie ihr Modell auch in der gleichen Reihenfolge auf.)

Die Hypothesen ändern sich jetzt. Stellen Sie die statistischen und inhaltlichen Null- und Alternativhypothesen auf. Vergessen Sie dabei nicht die Hypothesen für den Achsenabschnitt und für das Gesamtmodell.

Welche Voraussetzung ist neu bei der multiplen linearen Regression und muss zusätzlich geprüft werden?

4. Multiple lineare Regression - Übung 2

Bestimmen Sie die angegebenen Werte mittels R.

b_0	=		p_{b_0}	<	
b_1	=		p_{b_1}	=	
b_2	=		p_{b_2}	<	
b_3	=		p_{b_3}	=	
R^2	=		R^2_{adj}	=	
F	=		p	<	

Was sagt dieses Ergebnis inhaltlich aus, was bedeutet es für die Hypothesen? Prüfen Sie anschliessend noch die Voraussetzungen der multiplen linearen Regression.

Damit ist dieser Termin abgeschlossen