Kapitel 5

FRAGESTELLUNG 1

Öffne die Datei "alctobac.sav".

Zuerst werden wir ein Streudiagramm erstellen, um einen grafischen Überblick von diesem Datensatz zu erhalten.

Gehe dazu auf "Grafiken / Streudiagramm"



wähle "Einfach" aus.

en
en

Mit *"Definieren"* kommen wir ins nächste Menü, in dem wir als **Y-Variable** "alcohol" und als **X-Variable** "tobacco" auswählen. Bestätige mit *"OK".*



Als nächsten Schritt wollen wir den Korrelationskoeffizienten berechnen. Dazu müssen wir auf "Analysieren / Korrelation / Bivariat" gehen.

Analysieren	<u>G</u> rafiken	E <u>x</u> tras	E	enster	<u>H</u> ilfe
Beri <u>c</u> hte			۲		
Deskriptive Statistiken			۲		
<u>T</u> abellen			۲		
<u>M</u> ittelwerte vergleichen			۲	000	
Allgemeines lineares Modell			۲		
<u>K</u> orrelatio	n		Þ	<u>B</u> i	variat

Als Variablen sind "alcohol" und "tobacco" anzugeben. Weiters sollte "Pearson" und bei "Test auf Signifikanz" der Punkt "zweiseitig" ausgewählt sein. Bestätige mit "*OK*".

♦ alcohol > 4.02 (FILTE	<u>Variablen:</u>	ОК
	alcohol	Einfügen
	ล	Zurücksetzen
		Abbrechen
		Hilfe
Korrelationskoeffizienten ▼ Pearso <u>n</u>	<u> </u>	
• Zweiseitig C F	inseitig	

Wir erhalten nun den Korrelationskoeffizienten nach Pearson für alle Werte (also **mit** dem "outlier" NorthernIreland) $r_{xy} = 0.224$ und dem Signifikanzwert (zweiseitig) p = 0.509.

Korrelationen

		ALCOHOL	TOBACCO
ALCOHOL	Korrelation nach Pearson	1,000	,224
	Signifikanz (2-seitig)		,509
	N	11	11
TOBACCO	Korrelation nach Pearson	,224	1,000
	Signifikanz (2-seitig)	,509	
	N	11	11

Im nächsten Schritt berechnen wir den Korrelationskoeffizienten **ohne** NorthernIreland. Dazu müssen mir mittels des Menüpunkts "*Daten / Fälle auswählen"* den Ausreißer rausfiltern.

Daten	Transformieren Ana			
Da <u>t</u>	um definieren			
Vari	ia <u>b</u> le einfügen			
<u>F</u> all	einfügen			
<u>G</u> eł	ne zu Fall			
Fälle <u>s</u> ortieren				
Transponieren				
Dateien <u>z</u> usammenfügen				
Aggregieren				
Ort <u>h</u> ogonales Design				
Datej aufteilen				
Fäll	e aus <u>w</u> ählen			

Klicke danach rechts im Feld "Auswählen" auf "Falls Bedingung zutrifft".

🔒 Fälle auswählen		X
 	- Auswählen C Alle Fäll <u>e</u> Falls <u>B</u> edingung zutrifft	
rtobacco	Falls	

Jetzt müssen wir auf "Falls…" drücken und als Formel **region~=**"**NorthernIreland**" eingeben. Mit *"Weiter"* und *"OK"* bestätigen

Fälle auswählen: Falls		×
 ▲; region (#) alcohol (#) tobacco 	region ~= "NorthernIreland"	•
 	+ < > 7 8 9 Funktionen:	
	ABS(numausor) ANY(test,wert,wert,) ANY(test,wert,wert,) ARSIN(numausor) ARTAN(numausor) ARTAN(numausor)	1
	** ``()] Löschen CDF.BERNOULLI(q,p) Weiter Abbrechen Hilfe	_

"region" ist als Stringvariable definiert, daher müssen wir Northerlreland unter Hochkomma setzen (Hinweis: es ist auch auf Gross/Kleinschreibung zu achten). Zum Befehl Fälle auswählen siehe Kapitel 3 Fragestellung 3.

	region	alcohol	tobacco	ralcohol	rtobacco	filter_\$
1	North	6,47	4,03	11,000	9,000	Ausgewählt
2	Yorkshire	6,13	3,76	9,000	7,000	Ausgewählt
3	Northeast	6,19	3,77	10,000	8,000	Ausgewählt
4	EastMidlands	4,89	3,34	4,000	4,000	Ausgewählt
5	WestMidlands	5,63	3,47	6,000	5,000	Ausgewählt
6	EastAnglia	4,52	2,92	2,000	2,000	Ausgewählt
7	Southeast	5,89	3,20	7,000	3,000	Ausgewählt
8	Southwest	4,79	2,71	3,000	1,000	Ausgewählt
9	Wales	5,27	3,53	5,000	6,000	Ausgewählt
10	Scotland	6,08	4,51	8,000	10,000	Ausgewählt
11	NorthernIreland	4,02	4,56	1,000	11,000	Nicht ausgewählt

Wenn wir jetzt den Korrelationskoeffizienten berechnen (dabei müssen die gleichen Schritte machen wie eingangs erklärt gemacht werden), sehen wir, dass der Zusammenhang zwischen den zwei Variablen viel stärker ist. **Ohne** den "outlier" NorthernIreland beträgt $r_{xy} = 0.784$ und der Signifikanzwert (zweiseitig) ist p = 0.007.

		ALCOHOL	TOBACCO
ALCOHOL	Korrelation nach Pearson	1,000	,784**
	Signifikanz (2-seitig)	12	,007
	N	10	10
TOBACCO	Korrelation nach Pearson	,784**	1,000
	Signifikanz (2-seitig)	,007	1
	Ν	10	10

Korrelationen

**· Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

FRAGESTELLUNG 2

Öffne die Datei "2ndhandcar.sav".

Für den grafischen Überblick erstellen wir wie in Fragestellung 1 ein einfaches Streudiagramm.

Für die Y-Achse nehmen wir "preis", für die X-Achse "meilen". Das Streudiagramm sollte folgendermaßen aussehen:





Zum Berechnen des Regressionsmodells gehen wir auf "Analysieren / Regression / Linear".



Als "abhängige Variable" (Y-Variable) geben wir "preis" und als unabhängige Variable (X-Variable) "meilen" an.

Lineare Regressio	n	
🚸 meilen	Abhängige Variable	е:ОК
🔿 color	📄 👘 preis	Einfügen
ri1 A∵a		
#> 12 	Zurück Block 1 von 1	Weiter Zurücksetzen
#>µ	Unabhängige Varia	able(n): Abbrechen
service		Hilfe
	Methode: Einschluß	
	Aus <u>w</u> ahlvariable:	
		Bedingung
	Fall <u>b</u> eschriftungen:	:
WLS >>	Statistiken Diagramme S	Speichern

Um die vorhergesagten Werte in die Datenansicht zu bekommen, müssen wir auf *"Speichern"* drücken. Folgende Auswahlmöglichkeiten sind anzukreuzen:

Lineare Regression: Speichern		×
Vorhergesagte Werte	Residuen	Weiter
☑ Nicht standardisiert	🔽 Nicht standardisiert	Abbreshend
🗖 Standardisjert	☐ <u>S</u> tandardisiert	Abbrechen
☐ Korrigiert	☐ Studentisiert	Hilfe
☐ Standa <u>r</u> dfehler des Mittelwerts	🗖 Ausgeschlossen	
1	🗖 Studentisiert, ausgeschl.	

Sowohl bei "Vorhergesagte Werte", als auch bei den "Residuen" wählen wir "nicht standardisiert" aus. Mit "*Weiter"* und "*OK"* bestätigen.

Wir erhalten in der Datenansicht die vorhergesagten Werte (Variable "**pre_1**") und die Residuen (Variable "**res_1**"). Im Ausgabefenster werden wir mit viel Output konfrontiert, von dem für uns aber vorläufig nur die Tabelle **Koeffizienten** interessant ist.

		Nicht standardisierte Koeffizienten		Standardi sierte Koeffizien ten		
			Standardf	27 37 70 1 2 4	20	
Modell		B	ehler	Beta	Т	Signifikanz
1	(Konstante)	6533,383	84,512		77,307	,000
	MEILEN	-3,12E-02	,002	-,806	-13,495	,000

Koeffizienten^a

a. Abhängige Variable: PREIS

Die Konstante liegt bei **6533.383** und der "Anstieg" beträgt –**0.031**. Daraus ergibt sich die Regressionsgleichung **Y** = **6533.383** - **0.031** * **X**

Zur Überprüfung dieses Modells betrachten wir im Menü *"Ausgabe / Modellzusammenfassung"*. Hier wird der R^2 – Wert angegeben, mit dem ausgesagt wird, wie viel Prozent der Daten durch das Modell erklärt wird (in unserem Fall also **65%**).

			-	
				Standardf
			Korrigiertes	ehler des
Modell	R	R-Quadrat	R-Quadrat	Schätzers
1	,806ª	,650	,647	151,57

Modellzusammenfassung^b

a. Einflußvariablen : (Konstante), MEILEN

b. Abhängige Variable: PREIS

Grafisch eignet sich ein Residuen – Plot sehr gut, um das Modell zu überprüfen. Dazu gehen wir auf "*Grafiken / Streudiagramm"* und wählen *"einfach"* aus.

Danach geben wir als Y-Variable "res_1" und als X-Variable "pre_1" an. Diese Variablen wurden durch die obige Auswahl von **Residuen** und **Vorhergesagte Werte,** (engl. "predicted", daher "pre_") von SPSS erzeugt und in der Datenmatrix hinten angefügt.

Einfaches Streud	diagramm		×
🛞 preis		Y-Achse:	ОК
meilen		∉ res_1	Ein <u>f</u> ügen
		X-Achse:	Zurücksetzen
(#) i2 (#) □		(₩) pre_1	Abbrechen
* 0		Markierungen festlegen durch:	Hilfe
* service			
		Fallbeschriftung:	
1			

Mit "OK" bestätigen.



Unstandardized Predicted Value

Der Residuenplot zeigt kein deutliches Muster, daher dürften die Voraussetzungen zur Berechnung eines Regressionsmodells erfüllt sein.

FRAGESTELLUNG 3a

Öffne die Datei "bicycle.sav".

Als deskriptive Methode eignet sich ein Boxplot. Dazu gehen wir auf "Grafiken / Boxplot". Im anschließenden Menü "Einfach" und "Auswertung über verschiedene Variablen" auswählen und mit "Definieren" bestätigen.

3oxplots	
dial Finfach	Definieren
	Abbrechen
∯ # Gruppiert	Hilfe
Daten im Diagramm	egorien einer Variablen
 Auswertung über vers 	:chiedene <u>V</u> ariablen

Im folgenden Menü beide Variablen in die rechte Box schieben und mit "OK" bestätigen.

Einfachen Boxplot definieren	Auswertung über verschi	edene Variablen 📃 🕨
	Box entspricht:	OK Einfügen
(Sec. 1)	THE KINYOI	Zurücksetzen
		Abbrechen
		Hilfe

Wir erhalten den Boxplot:



Gehe nun auf "Analysieren / Mittelwerte vergleichen / T-Test bei gepaarten Stichproben" aus.



Gehe wie folgt vor:

<u>A · · · · · · · · · · · · · · · · · · ·</u>		Genaarte Variablen:	
#> kmnach		kroppole krouor	OK
₩ kmvor		KIINGCHINAKKIIVO	Einfügen
			Zurücksetzer
			Abbrechen
			Hilfe
Aktuelle Auswahl	1		
Variable 1:			
Variable 2:			Optionen

Zuerst "kmnach" und dann "kmvor" markieren und die Variablen hinüber schieben.

Mit "OK" bestätigen.

Test bei gepaarten Stichproben

	1	Gepaarte Differenzen						· · · · ·	
			Standardab	Standardfe bler des	95% Konfid der Dif	enzintervall ferenz	8		
		Mittelwert	weichung	Mittelwertes	Untere	Obere	T	ďf	Sig. (2-seitig)
Paaren 1	KMNACH - KMVOR	,7700	3,0650	,2167	,3426	1,1974	3,553	199	,000

Wir erhalten den **T-Wert 3.553** und ein Signifikanzwert, der mit 0,000 angegeben ist. Dieser Wert ist natürlich nicht wirklich 0 aber so klein, dass bei Rundung auf 3 Stellen hinter dem Komma noch immer 0,000 stehen bleibt.

FRAGESTELLUNG 3b

Öffne die Datei "alcattrneu.sav".

Gehe auf "Grafiken / Boxplot". Im anschließenden Menü wiederum "Einfach" und "Auswertung über verschiedene Variablen" auswählen und mit "Definieren" bestätigen.

Boxplots	
dBall Finfach	Definieren
	Abbrechen
∯ # Gruppiert	Hilfe
Daten im Diagramm C Auswertung über Kateg C Auswertung über verso	gorien einer Variablen shiedene <u>V</u> ariablen

Schieben Sie die zwei variablen wiederum nach rechts und drücken dann auf "OK".

	Box entspricht:	ОК
	Sperrstunde [nachher] A Stunden vorher [vorher]	Einfügen
(m)		Zurücksetzen
×.		Abbrechen
		Hilfe

Der Boxplot:



Gehe nun auf "Analysieren / Nichtparametrische Tests / Zwei verbundene Stichproben".

Analysieren	<u>G</u> rafiken	E <u>x</u> tras	Eenster	<u>H</u> ilfe	
Beri <u>c</u> hte					
<u>D</u> eskriptr <u>T</u> abellen	ve statistiki	en	; -		
<u>M</u> ittelwer	te vergleicł	nen	• ar	Var	Var
<u>A</u> ligemeir Korrelatio	nes lineares)n	Modell			
 <u>R</u> egressi	on		•		
L <u>og</u> linear Klassifizie	ren		1		
Dim <u>e</u> nsio	nsreduktion	n	•		
Skalierer Niebtear) motrische	Tooto		hi Quadrat	
Zejtreihe	n n	Tests	► <u>B</u> i	nomial	
Ü <u>b</u> erlebensanalyse			<u>S</u> equenzen		
Mehr <u>f</u> ach Analyse f	hantworten Te <u>h</u> lender W	/erte	► K 	-S bei <u>e</u> iner Stich wei unabhängige unabhängige Sti	probe Stichproben chproben
			Z	wei <u>v</u> erbundene :	Stichproben

Als nächstes markiere die Variable "vorher" und dann "nachher". Jetzt können die Variablen in das Feld "Ausgewählte Variablen" verschoben werden.

🐞 vorher		Ausgewählte Variablenpaare:	OK
nachher		vorher nachher	Ein <u>f</u> ügen
			Zurücksetzer
			Abbrechen
			Hilfe
Aktuelle Auswahl	1	, Welche Tests durchführen?	
Variable 1:		🛛 🔽 🖳 Wilcoxon 🗖 🛛 Vorzeichen 🗖 🕅	<u>[</u> cNemar
Variable 2:		🗖 Rand-Homogenität	

Drücke auf "OK".

Wir erhalten nun als Ausgabe den Wilcoxon – Signifikanzwert:

Ränge

		N	Mittlerer Rang	Rangsumme
Sperrstunde - 3	Negative Ränge	39ª	61,67	2405,00
Stunden vorher	Positive Ränge	138 ^b	96,72	13348,00
	Bindungen	3°		
	Gesamt	180		

a. Sperrstunde < 3 Stunden vorher

b. Sperrstunde > 3 Stunden vorher

C. 3 Stunden vorher = Sperrstunde

Statistik für Test^b

	Sperrstunde - 3 Stunden vorher
Z	-8,015ª
Asymptotische Signifikanz (2-seitig)	,000

a. Basiert auf negativen Rängen.

b. Wilcoxon-Test