

Explorative Datenanalyse

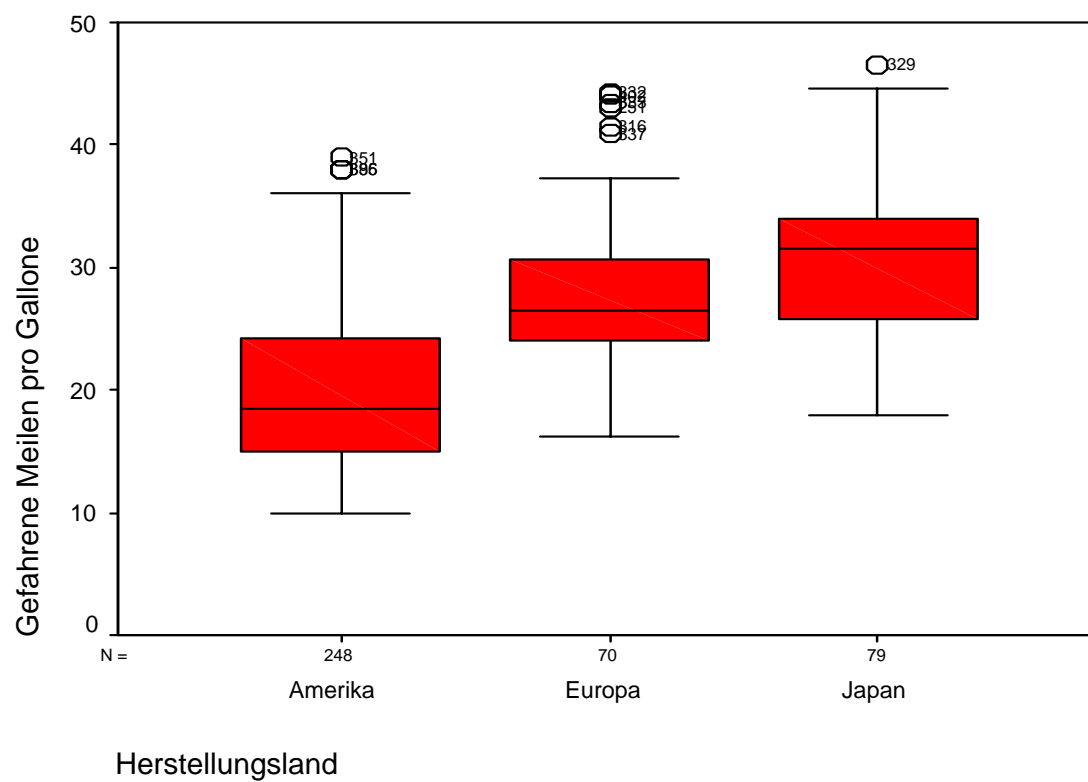
Um Euch noch einen schönen SPSS Output an die Hand zu geben betrachten wir noch mal das Beispiel mit den Gefahrenen Meilen und den Herkunftsländern. Zum generieren von Arbeitshypothesen schaut man erstmal auf einige Grafiken.

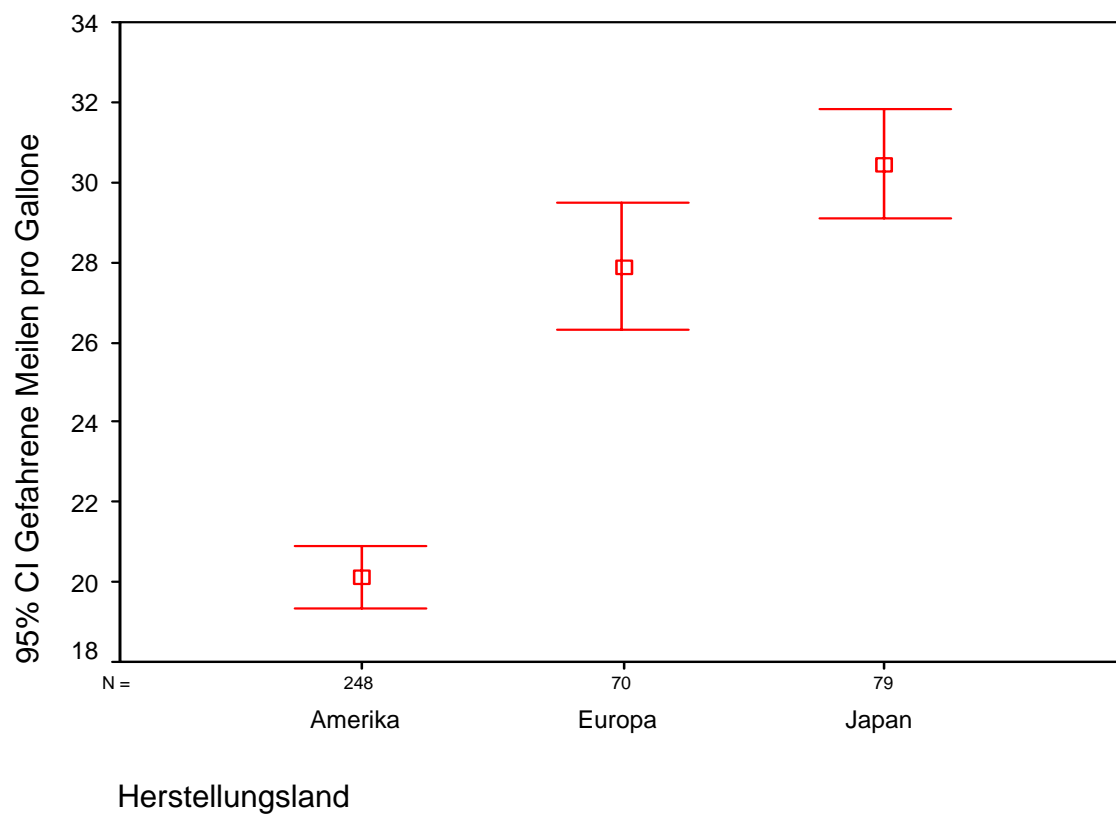
Herstellungsland

Verarbeitete Fälle

Herstellungsland		Fälle					
		Gültig		Fehlend		Gesamt	
		N	Prozent	N	Prozent	N	Prozent
Gefahrene Meilen pro Gallone	Amerika	248	98.0%	5	2.0%	253	100.0%
	Europa	70	95.9%	3	4.1%	73	100.0%
	Japan	79	100.0%	0	.0%	79	100.0%

Gefahrene Meilen pro Gallone





Die obere Grafik zeigt die Boxplots und die untere stellt die 95% Konfidenzintervalle dar. Man sieht, dass sich Autos aus Amerika deutlich von Autos aus Europa und Japan bezüglich der gefahrenen Meilen unterscheiden. Also kommt man zu dem Testproblem unterscheiden sich die Gefahrenen Meilen nach Herkunftsland? Die Stichproben sind hier unabhängig. Es werden also folgende Hypothesen getestet:
 $H_0: \mu_{Landi} = \mu_{Landj}$ für $i, j = \text{Japan, Amerika, Europa}$ und $i \neq j$
 $H_1: \text{nicht } H_0$
 Beginnen wir mit dem Vergleich von Japan und Amerika, wobei wir Normalverteilung unterstellen: Anmerkung zur Normalverteilung, die Boxplots sehen alle recht symmetrisch aus aber weisen auch alle Ausreisser auf. Dies im Hinterkopf müssen wir die Normalverteilungsannahme etwas skeptisch sehen. In der Praxis würde man auch Nichtparametrische Tests durchführen. Dies ist hier aber nicht unser Problem wir betrachten die Normalverteilung als gesetzt.

T-Test

Gruppenstatistiken

	Herstellungsland	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Gefahrenere Meilen pro Gallone	Amerika	248	20.13	6.377	.405
	Japan	79	30.45	6.090	.685

Hier sehen wir die Mittelwerte und Standardabweichungen beider Länder, sowie wie viel Autos der Stichprobe aus den jeweiligen Ländern kommen(N).

Die Frage ist nun ob sich die Mittelwerte statistisch unterscheiden. Was wir zeigen wollen wird in der Gegenhypothese formuliert (erstes Prinzip zur Formulierung von Hypothesen): Das sich die Leistungen unterscheiden steht also in H_0 .

Da man bei einem Zweistichprobenproblem zwei verschiedene T-Tests machen kann, wobei einer Gleichheit der Varianzen unterstellt während der andere dies nicht tut schaltet SPSS hier einen Test auf Gleichheit der Varianzen vor.

Gleichheit steht immer in der Nullhypothese (zweites Prinzip zur Formulierung von Hypothesen).

Somit lauten die Hypothesen hier:

H_0 : $\sigma_{Landi} = \sigma_{Landj}$ für $i, j = \text{Japan, Amerika, Europa}$ und i ungleich j

H_1 : nicht H_0

Dieser Test gibt uns einen Hinweis, in welcher Zeile sich das korrekte Ergebnis für unseren T-Test befindet.

Test bei unabhängigen Stichproben

	Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
	F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
								Untere	Obere
Gefahrenere Meilen pro Gallone	.198	.656	-12.664	325	.000	-10.32	.815	-11.926	-8.719
			-12.970	136.731	.000	-10.32	.796	-11.896	-8.749

Betrachten wir erst die ersten beiden Spalten. Sie zeigen das Ergebnis eines Testes auf Gleichheit der Varianzen. F ist dabei der Wert der Teststatistik. Die Signifikanz ist im Normalfall die Überschreitungswahrscheinlichkeit (p-value). Sprich es ist die Wahrscheinlichkeit, dass man, unter H_0 , den Wert der Teststatistik und noch extremere Werte beobachtet. Wir lehnen H_0 ab wenn die Signifikanz kleiner ist als $\alpha/2$ (zweiseitiger Test, siehe Hypothesen), siehe dazu Buch Seite 157 ff. Das bedeutet unter H_0 ist die Wahrscheinlichkeit der Teststatistik so klein, dass wir den Glauben an H_0 verlieren. Man kann α setzen wie man will, im Prinzip sagt es aus ab welcher Wahrscheinlichkeit man nicht mehr an H_0 glaubt. Im Normalfall setzen wir α immer auf 0.05. Wenn wir also Werte der

Teststatistik beobachten deren Wahrscheinlichkeit kleiner als 0.05 (einseitig) bzw. 0.025 (zweiseitig) ist glauben wir nicht mehr an H_0 und lehnen H_0 ab. SPSS ist ein äusserst nettes Programm, um Problemen vorzubeugen, die immer mal so entstehen, z.B. die Frage muss ich den p-Wert jetzt mit alpha oder alpha/2 vergleichen, nimmt uns SPSS diese Wahl ab. Es berechnet die Wahrscheinlichkeiten immer so, dass wir sie einfach nur mit alpha vergleichen müssen, im zweiseitigen Fall nimmt SPSS also die doppelte Überschreitungswahrscheinlichkeit als Signifikanz. Bei diesem Test ist 0.198 so gross (viel grösser als alpha), dass wir H_0 nicht ablehnen können. Also können wir von Gleichheit der Varianzen der gefahrenen Meilen beider Länder ausgehen. Also wenden wir den T-Test an der Gleichheit der Varianzen unterstellt. Sprich wir betrachten die erste Zeile der Tabelle und ignorieren die zweite Zeile. T ist der Wert der Teststatistik des T-Test, den können sie per Hand nachrechnen, df sind die Freiheitsgrade, auch diese sollten sie mit ihrem Wissen erklären können. Der interessante Teil ist die Signifikanz, SPSS sagt uns sogar, dass zweiseitig getestet wurde. Und wieder vergleichen wir die Signifikanz mit 0.05 und stellen fest, dass wir H_0 ablehnen. Die Wahrscheinlichkeit für die Nullhypothese ist so klein, dass wir nicht mehr daran glauben können. Sprich die gefahrenen Meilen der Autos die aus Japan kommen unterscheiden sich von denen die aus Amerika kommen. Als nächstes wollen wir Europa und Japan vergleichen. Überlegen Sie sich was sie aufgrund unserer Grafiken für ein Ergebnis erwarten würden.

T-Test

Gruppenstatistiken

	Herstellungsland	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Gefahrenen Meilen pro Gallone	Europa	70	27.89	6.724	.804
	Japan	79	30.45	6.090	.685

Test bei unabhängigen Stichproben

	Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
	F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
								Untere	Obere
Gefahrenen Meilen pro Gallone	.140	.709	-2.438	147	.016	-2.56	1.050	-4.634	-.485
			-2.423	140.226	.017	-2.56	1.056	-4.647	-.471

Der Test auf Gleichheit der Varianzen kann wiederum nicht abgelehnt werden, also erste Zeile betrachten. Analog zu den Ergebnissen oben unterscheiden sich die gefahrenen Meilen von Japan und Europa signifikant. Würden wir aber alpha auf 0.01 setzten, so würden wir H_0 nicht mehr ablehnen können.

Vergleichen wir noch schnell Amerika und Europa. Wenn Amerika ungleich Japan ist und Japan ungleich Europa, so sollte auch Europa ungleich Amerika sein. Schauen wir auf das Ergebnis:

T-Test

Gruppenstatistiken

	Herstellungsland	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Gefahrenere Meilen pro Gallone	Amerika	248	20.13	6.377	.405
	Europa	70	27.89	6.724	.804

Test bei unabhängigen Stichproben

	Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
	F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
								Untere	Obere
Gefahrenere Meilen pro Gallone	.002	.961	-8.887	316	.000	-7.76	.874	-9.482	-6.045
			-8.627	106.562	.000	-7.76	.900	-9.547	-5.979

Auch hier kann von Gleichheit der Varianzen ausgegangen werden. Und auch hier kann man davon ausgehen, dass sich die gefahrenen Meilen der beiden Länder unterscheiden.

Anmerkung: Wenn man mit Hilfe von SPSS einseitig Testen will muss man nur die Signifikanz halbieren und mit alpha vergleichen.