

Erhebung und Analyse von Immobilienmarktinformationen per Bürgerbefragung

Bachelorübung, FK Raumplanung, Modul 11

Donnerstags 09:30 bis 14:00, GB 1 335 PC-Pool

Sommersemester 2024

M.Sc. Alexander Schuhmacher

TU Dortmund, Fakultät Raumplanung

Lehrstuhl Immobilienentwicklung



Agenda

- **Ziele der Präsentation**
- **Einführung in die Analysearten**
- **Mittelwertanalysen**
- **Varianzanalysen**

Ziele

- **Wissensaufbau in Mittwertanalysen**
- **Wissensaufbau in Varianzanalysen**
- **Auswertung und Interpretation des Datensatzes**

3. Mittelwert- und Varianzanalysen



Ausgangspunkt:

- Analyse und Vergleich verschiedener Stichproben z.B.
Verteilungsannahmen oder spezifischer Merkmale

Bewertungskriterium:

- Arithmetisches Mittel und/oder Median

Abhängig von Anzahl der Stichproben:

1 Stichprobe	2 Stichproben	n-Stichproben
• Kolmogorov-Smirnov-Test [Chi-Quadrat-Anpassungstest] • Einfacher t-Test	• Mann-Whitney-Test/Wilcoxon-Test • Doppelter t-Test (ggfs. für verbundene Stichpr.)	• Kruskal-Wallis-Test • Multiple Mittelwerttests • Ein-/Mehrfaktorielle Varianzanalyse

Parametrische Testverfahren (t-Tests) vs. nicht parametrische Testverfahren (MW-Test; Wilcoxon-Test)

Unterscheidung:

- **Annahmen** der Verteilung oder Gleichheit der Varianz innerhalb der Gruppen vs. keine/wenige Annahmen
- **Datentypen:** metrische/kontinuierliche vs. nominal/ordinal

Ziel der induktiven Statistik: „Schluss vom Teil (Stichprobe) auf die (unbekannte) Grundgesamtheit“ unter der Verwendung von Hypothesen

Grundidee des statistischen Tests:

- statistische Vermutungen, Annahmen oder Behauptungen mittels Parameter und/oder Verteilungen zu überprüfen.

Notwendige Bedingungen:

1. Eine (oder mehrere) Grundgesamtheit(en)
2. Eine (oder mehrere) Zufallsstichprobe(n)
3. Formulierung von (zu überprüfenden) Hypothesen
4. Vorab vereinbartes Signifikanzniveau

Signifikanzniveau (α):

- Schwellenwert zur Bestimmung, ob ein statistisches Ergebnis als signifikant zu bezeichnen ist. Er beschreibt die Wahrscheinlichkeit, ob ein beobachteter Effekt zufällig auftritt.
- Sofern p-Wert **< 0,01/0,05** (klassische Signifikanzniveaus):

Ablehnung der Nullhypothese (H_0) -> Effekt ist nicht signifikant

Annahme von Alternativhypothese (H_1) -> Effekt ist signifikant

Ziel:

- Ablehnung von H_0 und Annahme von H_1

Varianz & Standardabweichung:

- Maße, welche angeben, wie weit einzelne Datenpunkte im Datensatz vom Mittelwert entfernt sind

Varianz (σ^2):

- Durchschnittliche quadratische Abweichung einzelner Datenpunkte vom Mittelwert

Standardabweichung (σ):

- Quadratwurzel der Varianz, mit identischer Einheit der ursprünglichen Daten
- Hohe Varianz/Standardabweichung = Datenpunkte sind weit (vom Mittelwert) verstreut
- Niedrige Varianz/Standardabweichung = Datenpunkte liegen (vom Mittelwert) nah beieinander



Arbeitsschritte:

1. Aufstellung Nullhypothese H_0 & Alternativhypothese H_1 . Festlegung von Signifikanzniveau (α)
2. Bedingt durch Signifikanztest eine oder mehrere Zufallsstichproben erheben und statistische Auswertung dieser
3. Statistischer Vergleich des festgelegten Signifikanzniveaus (α) mit dem empirisch ermittelten Signifikanzniveaus (α^*) und sachlogische Interpretation

Beispielhypothesen:

H_0 = Es gibt **keinen** signifikanten Unterschied in den Immobilienpreisen zwischen den unterschiedlichen Stadtteilen Berlins.

H_1 = Es gibt **einen** signifikanten Unterschied in den Immobilienpreisen zwischen den unterschiedlichen Stadtteilen Berlins.

Alternativformulierung:

H_0 = Es besteht **kein** Zusammenhang. Es ist **kein** Effekt zu beobachten.

H_1 = Es besteht **ein** Zusammenhang. Es ist **ein** Effekt zu beobachten.

H_0 = Beschreibt den Zustand, wovon zunächst ausgegangen wird.

- Kann erst nach der Durchführung des statistischen Tests bestätigt oder widerlegt werden

Wiederholung:

- einfacher t-Test bei **einer Stichprobe**, doppelter t-Test bei **zwei Stichproben** (unabhängig oder verbunden)

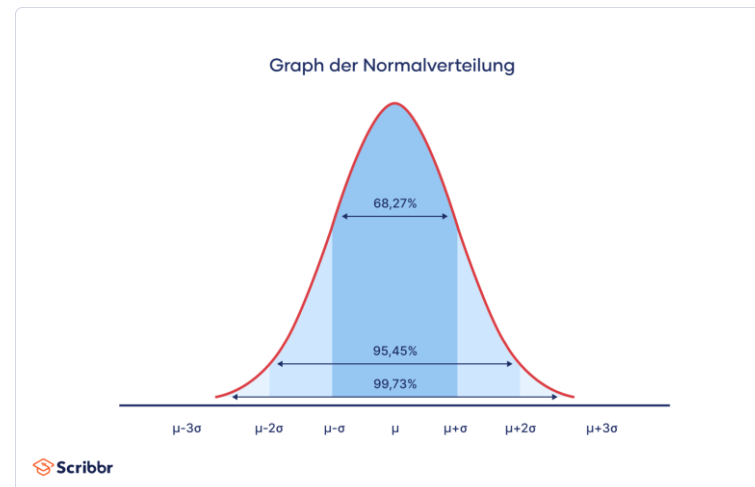
Voraussetzung:

- Metrische Daten sind **normalverteilt**
- Alternative für 1 Stichprobe: Chi-Quadrat-Anpassungstest

Keine Normalverteilung
notwendig!

Normalverteilung:

- Verteilung von Datenpunkten um einen Mittelwert – Daten symmetrisch um den Mittelwert angeordnet
- Meisten Datenpunkte nahe am Mittelwert, Abnahme mit Entfernung



Quelle:

www.scribbr.de/statistik/normalverteilung/ (2024)

- **Einfacher t-Test:** **Stichprobentest**, ob unbekannter Mittelwert (μ) eines **metrischen, normalverteilten** Merkmals der Grundgesamtheit (**GG**) mit vorgegeben Testwert (μ_0) **übereinstimmt**

Übungsaufgabe 11:

- Der Berliner Mieterbund und die Stadtverwaltung sehen die durchschnittliche Miete pro qm bei derzeit 6,- €. Überprüfen Sie, ob die zufällig ausgewählten Wohnungen im Mietpreisspiegel als durchschnittlich eingestuft werden können oder nicht.
- Nutzen Sie die **Übungsdatei „Mietspiegel“** für Ihre Analyse und unterstellen Sie ein Signifikanzniveau von 0.05!

Starthypothese:

Es wird vermutet, dass die zufällig ausgewählten Wohnungen in unserem Datensatz dem Mittelwert von 6,00 € entsprechen.

Hypothesen:

- H_0 = Die durchschnittliche Miete pro qm für die zufällig ausgewählten Wohnungen **entspricht** 6,00 €.
- H_1 = Die durchschnittliche Miete pro qm für die zufällig ausgewählten Wohnungen **entspricht nicht** 6,00 €.

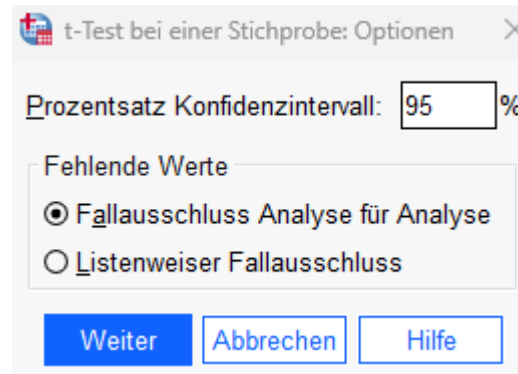
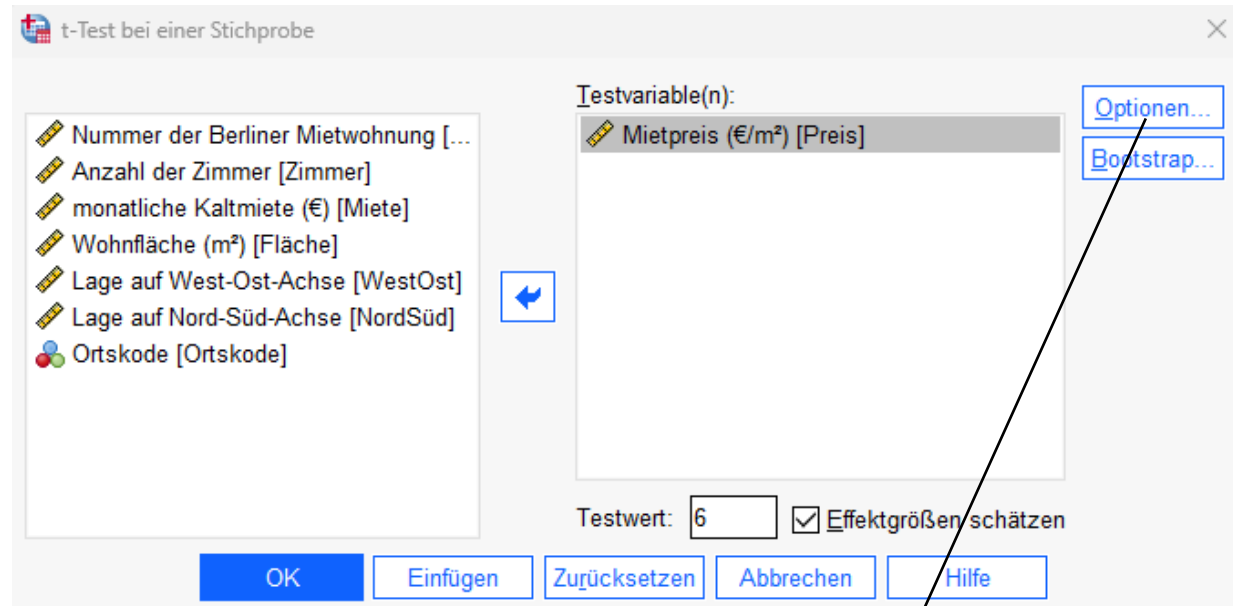
$$H_0 : \mu = \mu_0 = 6,0$$

$$H_1 : \mu \neq \mu_0 \neq 6,0$$

$$[\text{alternativ } H_0 : \mu - \mu_0 = 0,0]$$

$$H_1 : \mu - \mu_0 \neq 0,0]$$

1. Analysieren
2. Mittelwert und Proportionen vergleichen
3. t-Test bei einer Stichprobe



Ergebnis:

Statistik bei einer Stichprobe

	N	Mittelwert	Std.- Abweichung	Standardfehler des Mittelwertes
Mietpreis (€/m²)	7518	6,0203	1,90528	,02197

Test bei einer Stichprobe

				Testwert = 6			
		Signifikanz		Mittlere Differenz	95% Konfidenzintervall der Differenz	Unterer Wert	Oberer Wert
T	df	Einseitiges p	Zweiseitiges p				
Mietpreis (€/m²)	,924	7517	,178	,356	,02030	-,0228	,0634

Bewertung:

Da $\alpha^* = 0.356 > 0.05$ wird die Nullhypothese nicht abgelehnt und die Nullhypothese wird beibehalten.

D.h. es gibt **keine ausreichende Evidenz**, um zu zeigen, dass die durchschnittliche Miete pro qm **signifikant von 6,00 € abweicht**. Demnach entspricht die durchschnittliche Miete pro qm wahrscheinlich 6,00 €.

Es gibt keine statistischen signifikante Unterschiede zwischen der durchschnittlichen Miete pro qm aus unserem Datensatz und dem angenommen Wert von 6,00 €.

- **Zwei-Stichproben-Test**, bei dem mit einem vorgegebenen Signifikanzniveau geprüft wird, ob **unbekannte** Mittelwerte μ_1 und μ_2 eines **metrischen** und $N(\mu_1, \sigma_1)$ - verteilten bzw. $N(\mu_2, \sigma_2)$ -verteilten Merkmals aus **zwei disjunkten statistischen Grundgesamtheiten** $\Gamma_1 = \{y_{i1}, i = 1, 2, \dots, N_1\}$ vom Umfang N_1 und $\Gamma_2 = \{y_{i2}, i = 1, 2, \dots, N_2\}$ vom Umfang N_2 **übereinstimmen**.

Vereinfacht:

- Test zur **Überprüfung**, ob der **Mittelwert zweier Gruppen**, μ_1 und μ_2 , die jeweils **aus unterschiedlichen Grundgesamtheiten** stammen und **normalverteilt** sind, **übereinstimmen**. Der Test wird mit einem festgelegten **Signifikanzniveau** durchgeführt.

Unterstellung:

- **Standardabweichung** $\sigma_1, \sigma_2 > 0$ bzw. die **Varianzen** $\sigma_1^2, \sigma_2^2 > 0$ in beiden disjunkten Grundgesamtheiten gleichsam unbekannt sind.

Übungsaufgabe 12:

Bei gleicher Ausstattung (z.B. Zimmeranzahl) unterscheidet sich der Mietpreis pro qm signifikant im Hinblick auf den Standort. Überprüfen Sie dies anhand der zwei Berliner Randbezirke Zehlendorf und Köpenick z.B. für Drei-Zimmerwohnungen!

- a) Definieren Sie die Datenauswahlbedingung!
- b) Überprüfen Sie, ob eine Normalverteilung vorliegt!
- c) Stellen Sie die Nullhypothese auf!
- d) Überprüfen Sie auf dem Signifikanzniveau von 0.05, ob die Nullhypothese widerlegt werden kann!
- e) Gibt es weitere Anwendungsvoraussetzungen für den (doppelten) t-Test?

Übungsaufgabe 12: (a) Filterung der Daten

Mittelwert- und Varianzanalyse

Ansicht **Daten** Transformieren Analysieren Grafik Extras Erweiterungen Fenster Hilfe

Koo Fälle auswählen

Auswählen

☐ Alle Fälle
☒ Falls Bedingung zutrifft

Falls... Zimmer = 3 & (Bezirk="Zeh" ...

☐ Zufallsstichprobe

Stichprobe...

☐ Nach Zeit- oder Fallbereich

Spannweite...

☐ Filtervariable verwenden:

Ausgabe

☒ Nicht ausgewählte Fälle filtern
☐ Ausgewählte Fälle in neues Dataset kopieren
 Datasetname:
☐ Nicht ausgewählte Fälle löschen

Aktueller Status: Fälle filtern anhand der Werte von filter_\$

OK Einfügen Zurücksetzen Abbrechen Hilfe

Fälle auswählen: Falls

Zimmer = 3 & (Bezirk="Zeh" | Bezirk="Köp")

Funktionsgruppe:

Alle
 Arithmetisch
 Verteilungsfunktionen
 Umwandlung
 Aktuelles Datum/aktuelle Uhr
 Datumsarithmetik

Funktionen und Sondervariablen:

Weiter Abbrechen Hilfe

t-Test bei unabhängigen Stichproben

Testvariable(n):

Mietpreis (€/m²) [Preis]

Gruppierungsvariable:

Bezirk("Zeh" "Köp")

Gruppen definieren...

☒ Effektgrößen schätzen

OK Einfügen Zurücksetzen Abbrechen Hilfe

Gruppenstatistiken					
Stadtteil, drei Anfangsbuchstaben		N	Mittelwert	Std.- Abweichung	Standardfehler des Mittelwertes
Mietpreis (€/m²)	Zeh	32	9,1862	1,79339	,31703
	Köp	77	6,3216	1,54561	,17614

- **KS-Test oder Shapiro-Wilk-Test:** Überprüfung, ob **Daten normalverteilt** sind
 - **Normalverteilung Voraussetzung** für t-test oder einfache lineare Regression
- 1. Analysieren, 2. deskriptive Statistiken 3. explorative Datenanalyse

Tests auf Normalverteilung

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Mietpreis (€/m²)	,096	32	,200 [*]	,971	32	,522

*. Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors

Tests auf Normalverteilung

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Mietpreis (€/m²)	,086	77	,200 [*]	,977	77	,176

*. Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors

Hypothesen für KS- bzw. SW-Test:

- H_0 = Variablen sind normalverteilt.
- H_1 = Variablen sind nicht normalverteilt.
- Bei Signifikanz von $\alpha^* = 0,2$ bzw. $0,522$ und $0,2$ bzw. $0,176$ muss die H_0 **angenommen** werden. Die Daten sind somit normalverteilt und die Voraussetzung erfüllt.

Beispielhypothesen:

H_0 = Es gibt **keinen signifikanten Unterschied** in den Mietpreisen pro qm zwischen den Stadtteilen Zehlendorf und Köpenick für 3-Zimmer-Wohnungen.

H_1 = Es gibt **einen signifikanten Unterschied** in den Mietpreisen pro qm zwischen den Stadtteilen Zehlendorf und Köpenick für 3-Zimmer-Wohnungen.

Doppelter t-Test mit Überprüfung auf Varianzhomogenität nach Levene:

Levene-Test:

- Prüfung, ob Varianzen zwischen Gruppen signifikant unterschiedlich sind.
- H_0 : „Varianzen zwischen den Gruppen gleich sind (Homogenität der Varianzen).“ -> Alternativer Test
- H_1 : „Varianzen zwischen den Gruppen sind nicht signifikant unterschiedlich und die Annahme der Varianzhomogenität ist erfüllt.“
- Niedriger p-Wert deutet darauf hin, dass die Varianzen zwischen den Gruppen signifikant unterschiedlich sind
- Hoher p-Wert deutet darauf hin dass die Varianzen zwischen den Gruppen nicht signifikant unterschiedlich sind

Übungsaufgabe 12: (d) Test der Hypothesen

Doppelter t-Test mit Überprüfung auf Varianzhomogenität nach Levene:

Test bei unabhängigen Stichproben										
		Levene-Test der Varianzgleichheit		t-Test für die Mittelwertgleichheit						
		F	Sig.	T	df	Signifikanz		Mittlere Differenz	Differenz für Standardfehler	95% Konfidenzintervall der Differenz
						Einseitiges p	Zweiseitiges p			Unterer Wert Oberer Wert
Mietpreis (€/m ²)	Varianzen sind gleich	,359	,550	8,401	107	<,001	<,001	2,86469	,34100	2,18870 3,54069
	Varianzen sind nicht gleich			7,899	51,106	<,001	<,001	2,86469	,36267	2,13663 3,59276

$\alpha^*=0.550>0.05$
 H_0 anzunehmen

Die beobachtete Mittelwertdifferenz (2.86 €/qm) ist signifikant von null verschieden ($H_0 : \mu_1=\mu_2$ widerlegt)

1. Varianzen sind nicht signifikant unterschiedlich
2. d.h. die 3-Zimmer-Mietwohnungspreise unterscheiden sich signifikant
3. d.h. der Standort der Wohnung hat signifikante Bedeutung für die Miete

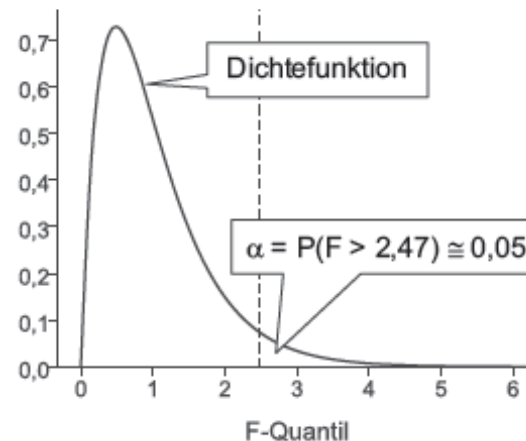
t-Test für zwei Stichproben: Hinweise

- Im Falle einer Verletzung der Normalverteilungsannahme als auch der Varianzhomogenitätsbedingung sind zwei Fälle zu unterscheiden:
- **$N \leq 50$: t-Test nicht anwendbar**
- **$N > 50$: t-Test anwendbar** („aufgeweichte Normalverteilungsannahme“) über Welch-Test
- Die Ergebnisse finden sich im „normalen“ t-Testergebnis:

Hier kein Unterschied!

Test bei unabhängigen Stichproben											
Levene-Test der Varianzgleichheit				t-Test für die Mittelwertgleichheit							
		F	Sig.	T	df	Signifikanz		Mittlere Differenz	Differenz für Standardfehler	95% Konfidenzintervall der Differenz	
						Einseitiges p	Zweiseitiges p			Unterer Wert	Oberer Wert
Mietpreis (€/m ²)	Varianzen sind gleich	,359	,550	8,401	107	<,001	<,001	2,86469	,34100	2,18870	3,54069
	Varianzen sind nicht gleich			7,899	51,106	<,001	<,001	2,86469	,36267	2,13663	3,59276

- Weiteres Testverfahren: **F-Test** (ebenfalls Überprüfung auf Varianz-Homogenität aber unter der Prämisse einer F-Verteilung = asymmetrische Verteilungsannahme)
- **t-Test** auch anwendbar bei **verbundenen Stichproben** (z.B. vor/nach Sanierung)



Keine Normalverteilung
notwendig!

Testvoraussetzung:

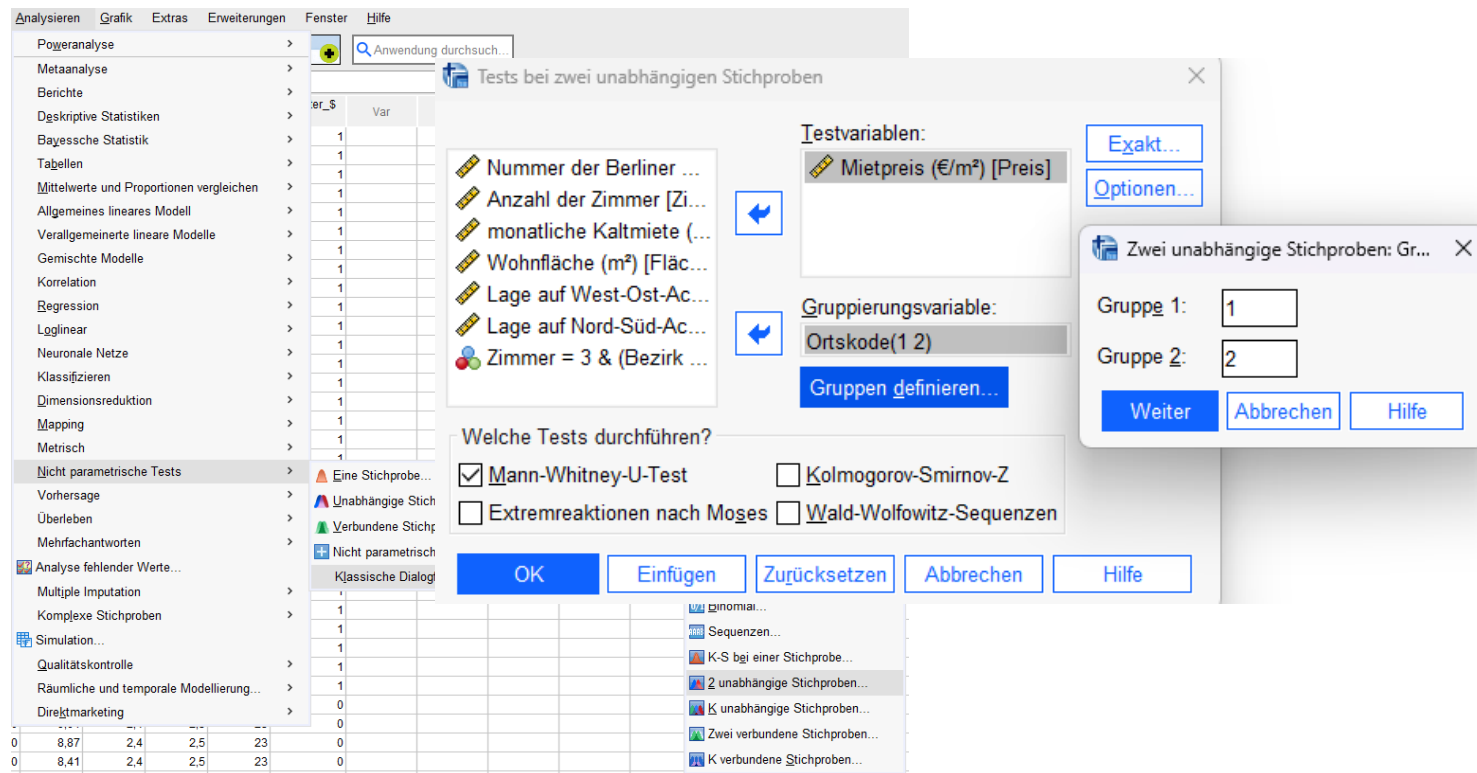
- Zwei unabhängige Stichproben mit beliebiger Verteilung
- Vergleich der Stichprobe anhand Ränge der abhängige Variablen
- Test auf Signifikanz
- Verletzung der Normalverteilung oder ordinalskalierte Daten
- nicht parametrischer Test

Ziel:

- Überprüfung, ob statistische GG in ihrer mittleren Lage übereinstimmt (Median) mittels Rängen
- Verwerfung von H_0 = kein Unterschied bzw. abhängige Variablen sind gleich
- Gleichverteilungstest, oft bei kleinen Stichproben

Beispiel:

- Überprüfung, ob die Mietpreise pro qm in den Bezirken Charlottenburg und Friedrichshain gleich verteilt sind.
- 1. nicht-parametrische Tests, 2. klassische Dialogfelder, 3. 2 unabhängige Stichproben



The screenshot displays the SPSS software interface for performing a Mann-Whitney-U-Test. The 'Analyse' menu is open, and the path 'Nicht parametrische Tests' > 'Zwei unabhängige Stichproben...' is selected. The 'Tests bei zwei unabhängigen Stichproben' dialog box is shown, with 'Mietpreis (€/m²) [Preis]' entered as the 'Testvariablen:' and 'Ortskode(1 2)' as the 'Gruppierungsvariable:'. The 'Welche Tests durchführen?' section has 'Mann-Whitney-U-Test' checked. A secondary dialog box 'Zwei unabhängige Stichproben: Gr...' is also visible, showing 'Gruppe 1' as 1 and 'Gruppe 2' as 2.

		Ränge		
	Ortskode	N	Mittlerer Rang	Rangsumme
Mietpreis (€/m²)	Charlottenburg	414	468,05	193774,50
	Friedrichshain	397	341,29	135491,50
	Gesamt	811		

Vorgehen:

- Bildet anhand der Variable „Mietpreis“ Ränge innerhalb N
- Sortierung vom niedrigsten zum höchsten, anschließend Addition

Teststatistiken ^a	
	Mietpreis (€/m²)
Mann-Whitney-U-Test	56488,500
Wilcoxon-W	135491,500
Z	-7,704
Asymp. Sig. (2-seitig)	<,001

a. Gruppenvariable: Ortskode

Interpretation:

- $\alpha = 0.001 < 0.05$, demnach wird die H_0 verworfen und H_1 angenommen
- Es gibt somit einen signifikanten Unterschied in den Mietpreisen
- Wie groß ist der Effekt?

Teststatistiken^a

	Mietpreis (€/m²)
Mann-Whitney-U-Test	56488,500
Wilcoxon-W	135491,500
Z	-7,704
Asymp. Sig. (2-seitig)	<,001

a. Gruppenvariable: Ortskode

Berechnung der Effektstärke:

$$r = \left| \frac{z}{\sqrt{n}} \right|$$

r – Wert	Effektstärke
ab 0,1 bis 0,299	Schwach
ab 0,3 bis 0,499	Mittel
ab 0,5	Stark

- Beispiel: $r = \left| \frac{7,704}{\sqrt{811}} \right| \approx 0,2705$
- Demnach besteht eine signifikante schwache Effektstärke
- MW-U-Test zeigt, dass Mittelwertunterschiede zwischen den verschiedenen Testsubjekten in den unabhängigen Stichprobe existieren.

Frage:

- Lässt sich das auch gleichzeitig für alle Stadtbezirke überprüfen?

Kruskal-Wallis-Test (H-Test)

Keine Normalverteilung
notwendig!

Testvoraussetzung:

- K unabhängige Stichproben mit beliebiger Verteilung
- Nicht-parametrischer Test mit größerem Anwendungsbereich
- Daten können ordinalskaliert sein

Ziel:

- Vergleicht > 2 unabhängige Stichproben unter der Ermittlung der Ränge der abhängigen Variablen
- Ermittlung signifikanter Unterschiede

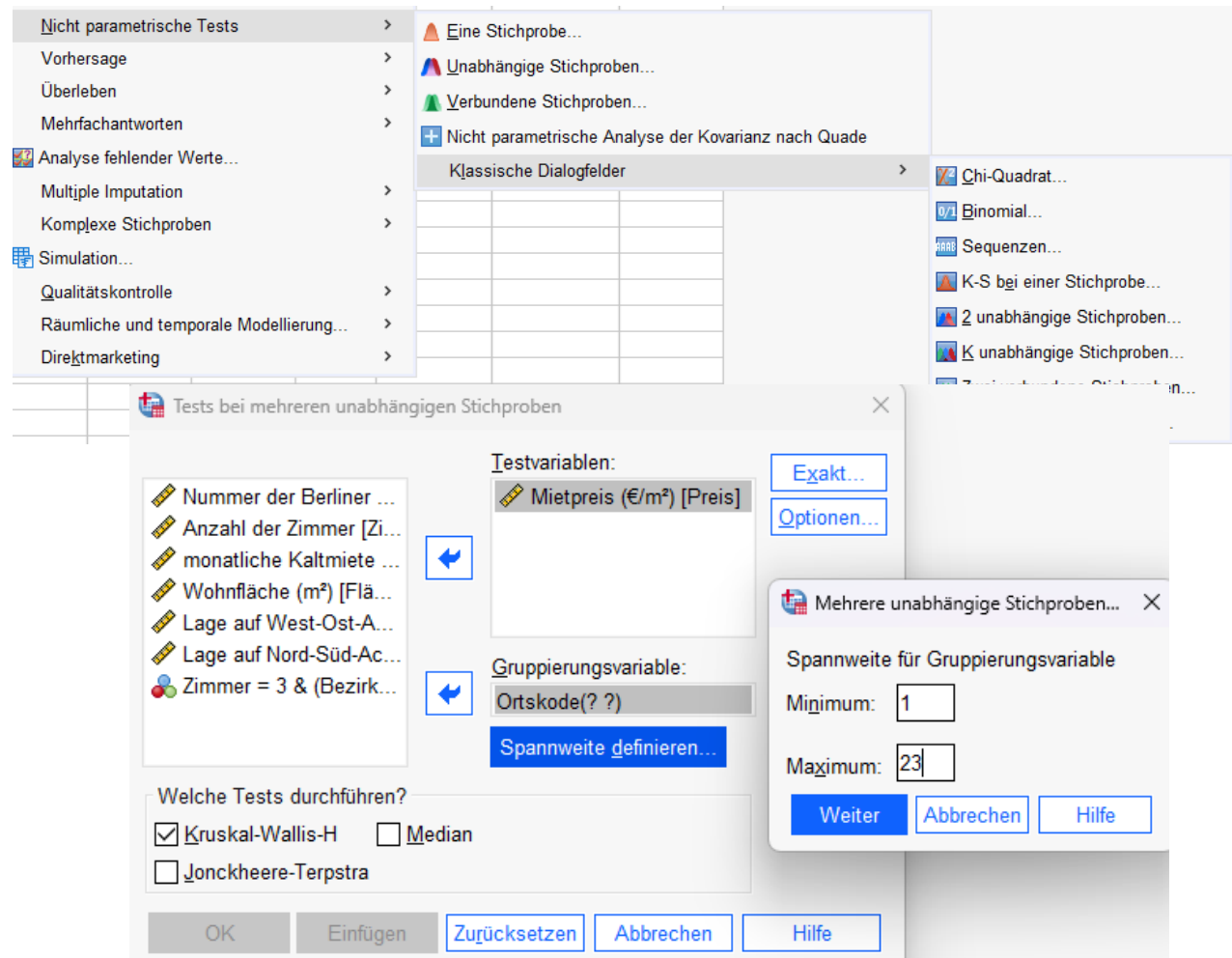
Hypothesen:

- H_0 = Es gibt keinen Unterschied. Die abhängigen Variablen sind gleich.
- H_1 = Es gibt Unterschied. Die abhängigen Variablen sind unterschiedlich.

Kruskal-Wallis-Test (H-Test)

Testvoraussetzung:

- K unabhängige Stichproben mit beliebiger Verteilung
- Nicht-parametrischer Test mit größerem Anwendungsbereich



The screenshot shows the SPSS software interface for performing a Kruskal-Wallis Test. The 'Nicht parametrische Tests' menu is open, and the 'Tests bei mehreren unabhängigen Stichproben' dialog box is active. In this dialog, the 'Testvariablen:' field contains 'Mietpreis (€/m²) [Preis]', and the 'Gruppierungsvariable:' field contains 'Ortskode(? ?)'. The 'Welche Tests durchführen?' section has 'Kruskal-Wallis-H' checked. A sub-dialog box 'Mehrere unabhängige Stichproben...' is also open, showing the 'Spannweite für Gruppierungsvariable' with 'Minimum: 1' and 'Maximum: 23'.

Überprüfung: Gleichverteilung der Mietpreise über alle Bezirke

Teststatistiken^{a,b}

	Mietpreis (€/m²)
Kruskal-Wallis-H	2132,654
df	22
Asymp. Sig.	<,001

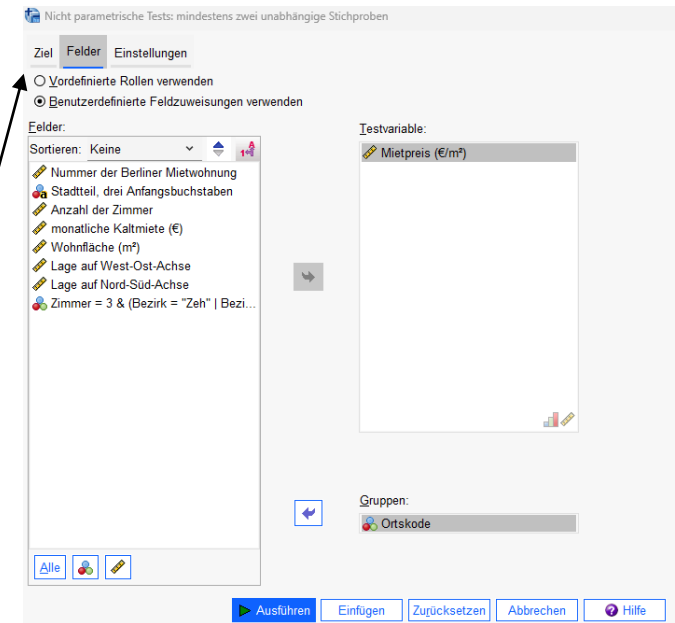
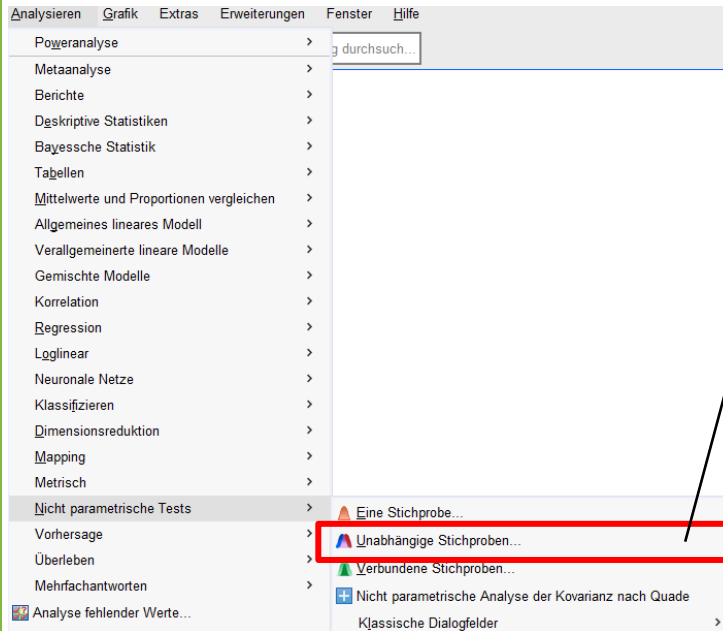
a. Kruskal-Wallis-Test

b. Gruppenvariable: Ortskode

- Mittelwerthomogenität (H_0) für die 23 Stadtbezirke verworfen, Annahme von H_1 es liegen Unterschiede in den Gruppen vor.
- Lediglich signifikante Unterschiede – **Paarweiser Vergleich** um zu testen welche Gruppen sich unterscheiden.

Kruskal-Wallis-Test (H-Test): Paarweiser Vergleich

Mittelwert- und Varianzanalyse



Automatische Erkennung, dass KW-Test durchzuführen ist

Hypothesentestübersicht

	Nullhypothese	Test	Sig. ^{a, b}	Entscheidung
1	Die Verteilung von Mietpreis (€/m²) ist über die Kategorien von Ortskode identisch.	Kruskal-Wallis-Test bei unabhängigen Stichproben	< ,001	Nullhypothese ablehnen

a. Das Signifikanzniveau ist ,050.

b. Asymptotische Signifikanz wird angezeigt.

Weitere Informationen zum PV im Ausgabedokument (Vergleich der Signifikanzen)

Weitere Testansätze:

- Jonckheere-Terpstra-Test (Reihenfolge des Hypothesentest änderbar)
- Wilcoxon-Test (für verbundene/gepaarte Stichproben)

Frage:

- Gibt es einen vergleichbaren Testansatz für k -Stichproben bei Erfüllung der Normalverteilungsannahme?

- Gehört wie die Regressionsanalyse zu den **strukturprüfenden Verfahren** und dient der **Feststellung von Mittelwertunterschieden** zwischen **zwei oder mehr Gruppen von Merkmalsträgern**.

Mathematisches Prinzip:

- Testung, ob Varianz zwischen den Gruppen größer als innerhalb Gruppen
Ergebnis = Aussagen, **ob Gruppen** bzgl. der (abhängigen) Variablen **signifikant voneinander unterscheiden**, bzw. ob die Einteilung in Gruppen anhand der (unabhängigen) Variablen gerechtfertigt ist

Durchführbarkeit:

- Im Vorfeld, welche abhängige und welche unabhängige Variablen im Modell sind

Unterscheidung:

- ANOVA (Analysis of Variance – 1 unabhängige Variable)
- MANOVA (Multivariate Analysis of Variance – minds. 2 unabhängige)

Testvoraussetzung:

- k unabhängige **normalverteilte** Stichproben

Einfaktorielle Varianzanalyse:

- Eine endlich statistische GG $\Gamma = \{Y_i, i = 1, 2, \dots, N\}$ vom Umfang N wird durch Festlegung eines kategorialen Gruppierungsmerkmals in k disjunkte Teilgesamtheiten aufgeteilt und hinsichtlich eines metrischen Erhebungsmerkmals X beschrieben. Sofern X in den k-Teilgesamtheiten normalverteilt mit einer homogenen Standardabweichung σ ist, gilt:
- $X_j \sim N(\mu_j, \sigma_j)$, $\sigma_j = \sigma > 0$ ($j = 1, 2, \dots, k$, $k \geq 2$), dann heißt das parametrische Verfahren zum Prüfen der Homogenitätshypothese $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ über die k unbekannten Mittelwerte μ_j auf der Grundlage von k unabhängigen Zufallsstichproben vom Umfang n_j einfaktorielle oder einfache Varianzanalyse

Vereinfacht:

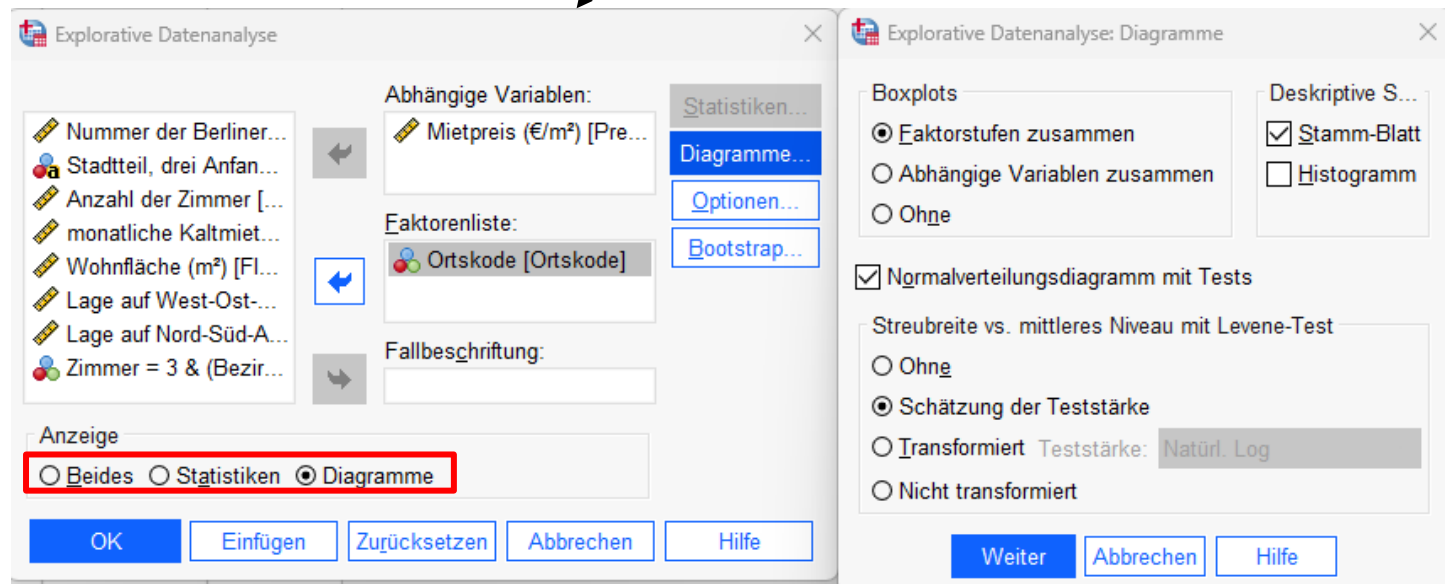
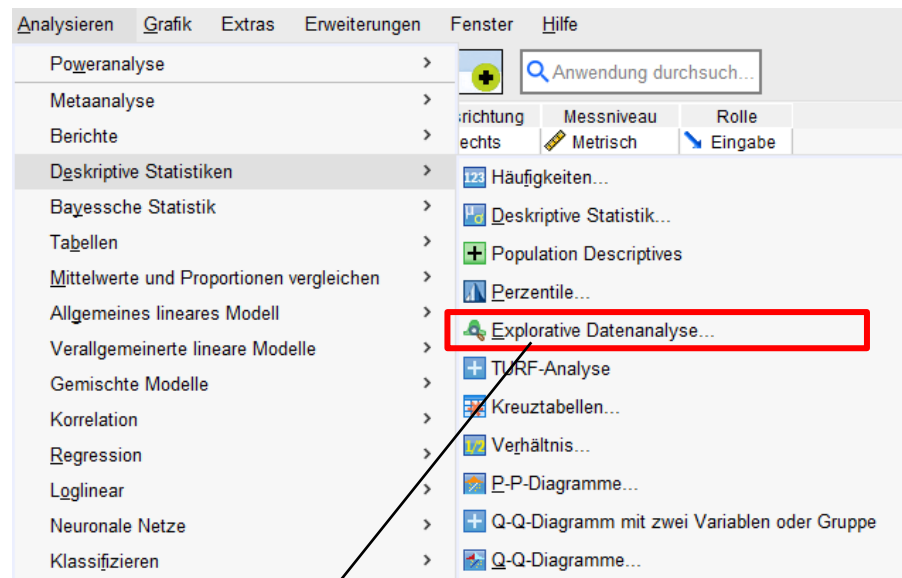
- Eine Gruppe von N-Beobachtungen wird basierend auf **ein** bestimmte Eigenschaft in k verschiedene Gruppen aufgeteilt
- Die Gruppen werden durch Mittelwert und Standardabweichung beschrieben
- Wenn betrachtendes Merkmal in jeder Gruppe normalverteilt ist sowie Standardabweichung (in allen Gruppen) gleich sind -> Einfaktorielle Varianzanalyse anwendbar

Ziel:

- Überprüfung, ob Mittelwerte der Gruppe gleich sind oder nicht bei zwei oder mehr Gruppen

- Varianzanalyse = Verallgemeinerung des t-Tests für zwei oder mehr unabhängige Stichproben
- Auch ANOVA genannt
- Überprüfung der **Normalverteilung** über **KS-Test** und der **Varianzhomogenität** über **Levene-Test**, dann erst Anwendung der **ANOVA**
- Kategoriale Merkmale (nominale oder ordinale Skala) = FAKTOR

- Beispiel: Mietspiegel Berlin**



Beispiel: Test auf Normalverteilung + Homogenität (explorativ)

- **Beispiel:** Monatliche Kaltmiete (€/qm) für Eigentumswohnungen in Berlin in Abhängigkeit zum Standort (Ortskode)

Tests auf Normalverteilung

Mietpreis (€/m²)	Ortskode	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistik	df	Signifikanz	Statistik	df	Signifikanz
	Charlottenburg	,104	414	<,001	,947	414	<,001
	Friedrichshain	,089	397	<,001	,908	397	<,001
	Hellersdorf	,134	56	,014	,971	56	,189
	Hohenschönhausen	,094	803	<,001	,971	803	<,001
	Köpenick	,074	215	,007	,980	215	,003
	Kreuzberg	,063	213	,042	,940	213	<,001
	Lichtenberg	,098	665	<,001	,944	665	<,001
	Marzahn	,118	288	<,001	,896	288	<,001
	Mitte	,079	321	<,001	,956	321	<,001
	Neukölln	,106	459	<,001	,938	459	<,001
	Pankow	,046	462	,023	,977	462	<,001
	Prenzlauer Berg	,078	89	,200 [*]	,985	89	,387
	Reinickendorf	,109	352	<,001	,960	352	<,001
	Schöneberg	,043	185	,200 [*]	,987	185	,085
	Spandau	,141	268	<,001	,873	268	<,001
	Steglitz	,041	365	,200 [*]	,991	365	,024
	Tempelhof	,068	151	,088	,990	151	,402
	Tiergarten	,145	181	<,001	,875	181	<,001
	Treptow	,071	179	,026	,986	179	,079
	Wedding	,097	593	<,001	,974	593	<,001
	Weißensee	,076	360	<,001	,930	360	<,001
	Wilmerdorf	,104	391	<,001	,898	391	<,001
	Zehlendorf						

Test auf Normalverteilung über explorative Datenanalyse

*. Dies ist eine untere Grenze der ech

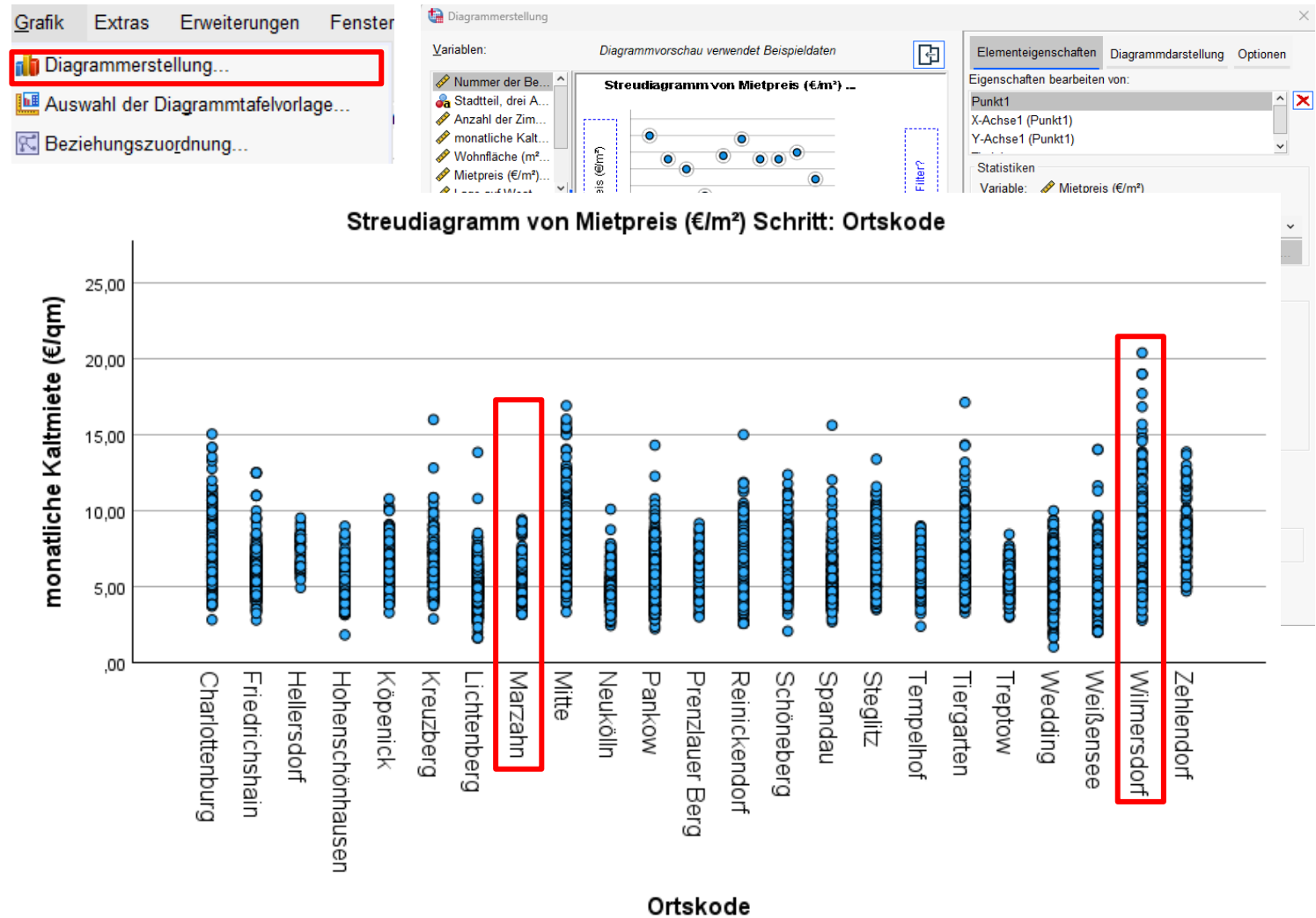
a. Signifikanzkorrektur nach Lilliefors

Test auf Homogenität der Varianz

		Levene-Statistik	df1	df2	Signifikanz
Mietpreis (€/m²)	Basiert auf dem Mittelwert	46,586	22	7495	<,001
	Basiert auf dem Median	39,513	22	7495	<,001
	Basierend auf dem Median und mit angepaßten df	39,513	22	5095,880	<,001
	Basiert auf dem getrimmten Mittel	44,001	22	7495	<,001

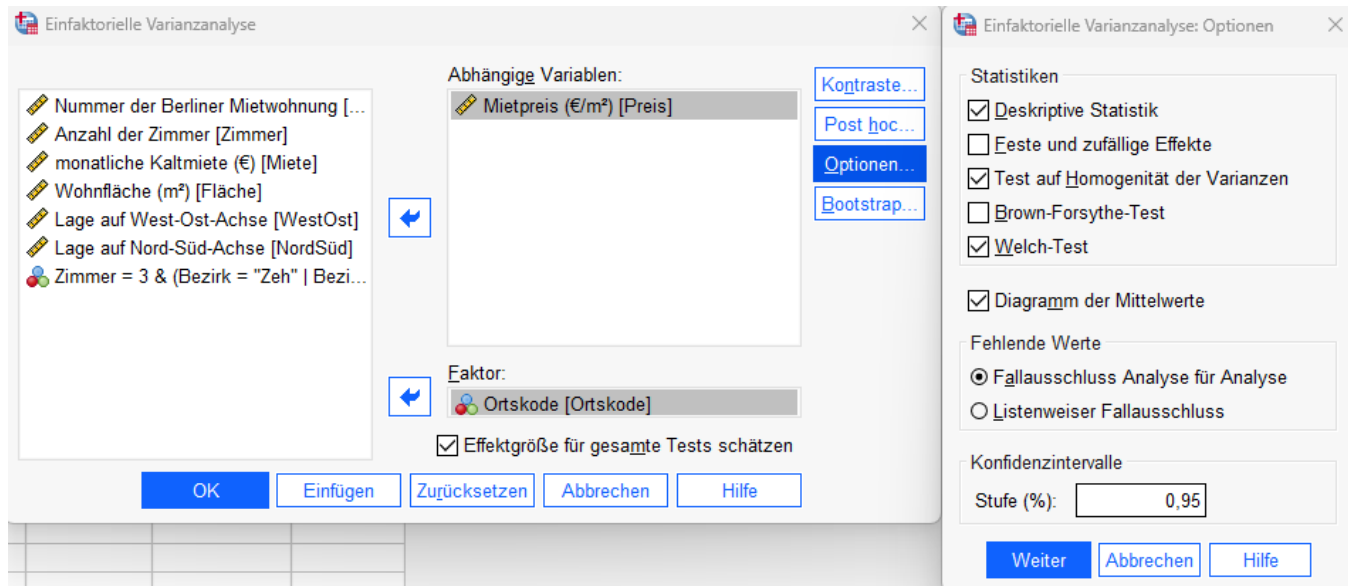
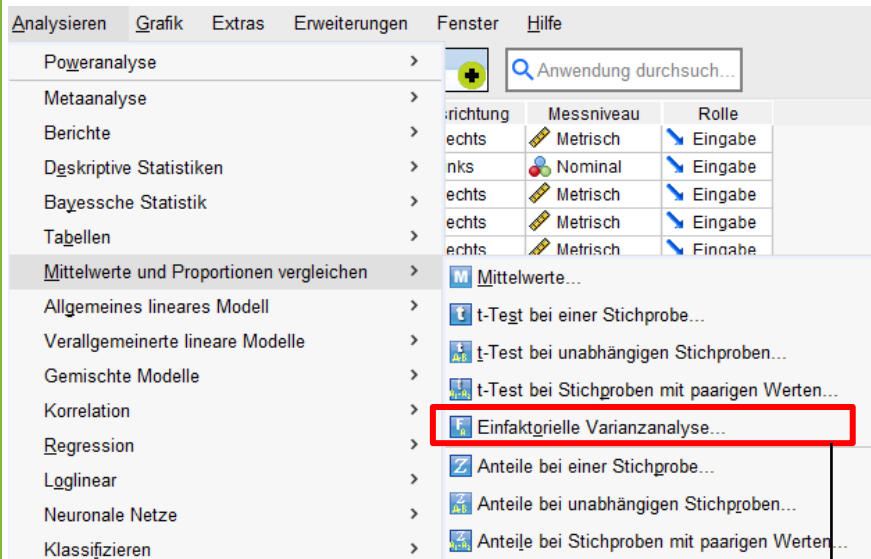
Veranschaulichung:

- Erstellung eines Streudiagramms



Beispiel: Einfaktorielle Varianzanalyse

Mittelwert- und Varianzanalyse



Tests der Varianzhomogenität

		Levene- Statistik	df1	df2	Sig.
Mietpreis (€/m ²)	Basiert auf dem Mittelwert	46,586	22	7495	<,001
	Basiert auf dem Median	39,513	22	7495	<,001
	Basierend auf dem Median und mit angepaßten df	39,513	22	5095,880	<,001
	Basiert auf dem getrimmten Mittel	44,001	22	7495	<,001

- **Identisch zur explorativen Datenanalyse**

ANOVA

Mietpreis (€/m²)

	Quadratsumme	df	Mittel der Quadrate	F	Sig.
Zwischen den Gruppen	8022,952	22	364,680	141,882	<,001
Innerhalb der Gruppen	19264,461	7495	2,570		
Gesamt	27287,414	7517			

Robuste Testverfahren zur Prüfung auf Gleichheit der Mittelwerte

Mietpreis (€/m²)

	Statistik ^a	df1	df2	Sig.
Welch	132,318	22	1741,490	<,001

a. Asymptotisch F-verteilt

Übungsaufgabe 13:

- Verwenden Sie zur Lösung der folgenden Problemstellungen die SPSS-Datendatei „**Mietspiegel**“.
- a. Es soll analysiert werden, ob ein signifikanter Unterschied in der monatlichen Kaltmiete aufgrund des Ortskodes zu beobachten sind. Welche Voraussetzungen sind zu prüfen?
- b. Sind die Voraussetzungen erfüllt?
- c. Überprüfen Sie die Mittelwerthomogenität auf der Basis eines Signifikanzniveaus von 0.05!
- d. Welche Aussagen können Sie ableiten?

Hypothese:

H_0 = Es bestehen keine signifikanten Unterschiede zwischen dem monatlichen Kaltmieten und dem Ortskode.

Mittelwert- und Varianzanalyse

Tests auf Normalverteilung

		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistik	df	Signifikanz	Statistik	df	Signifikanz
monatliche Kaltmiete (€)	Charlottenburg	,138	414	<,001	,833	414	<,001
	Friedrichshain	,178	397	<,001	,788	397	<,001
	Hellersdorf	,105	56	,190	,969	56	,156
	Hohenschönhausen	,124	803	<,001	,925	803	<,001
	Köpenick	,114	215	<,001	,837	215	<,001
	Kreuzberg	,165	213	<,001	,656	213	<,001
	Lichtenberg	,044	665	,004	,978	665	<,001
	Marzahn	,086	288	<,001	,953	288	<,001
	Mitte	,145	321	<,001	,788	321	<,001
	Neukölln	,130	459	<,001	,872	459	<,001
	Pankow	,111	462	<,001	,913	462	<,001
	Prenzlauer Berg	,129	89	<,001	,938	89	<,001
	Reinickendorf	,105	352	<,001	,915	352	<,001
	Schöneberg	,163	185	<,001	,849	185	<,001
	Spandau	,145	268	<,001	,795	268	<,001
	Steglitz	,095	365	<,001	,906	365	<,001
	Tempelhof	,167	151	<,001	,868	151	<,001
	Tiergarten	,223	181	<,001	,525	181	<,001
	Treptow	,095	179	<,001	,955	179	<,001
	Wedding	,121	593	<,001	,938	593	<,001
	Weißensee	,113	360	<,001	,837	360	<,001
	Wilmerdorf	,156	391	<,001	,827	391	<,001
	Zehlendorf	,130	111	<,001	,934	111	<,001

a. Signifikanzkorrektur nach Lilliefors

Test auf Homogenität der Varianz

		Levene- Statistik	df1	df2	Signifikanz
monatliche Kaltmiete (€)	Basiert auf dem Mittelwert	101,286	22	7495	<,001
	Basiert auf dem Median	68,860	22	7495	<,001
	Basierend auf dem Median und mit angepaßten df	68,860	22	2702,044	<,001
	Basiert auf dem getrimmten Mittel	87,097	22	7495	<,001

ANOVA

monatliche Kaltmiete (€)

	Quadratsumme	df	Mittel der Quadrate	F	Sig.
Zwischen den Gruppen	140604269,64	22	6391103,165	87,900	<,001
Innerhalb der Gruppen	544951827,68	7495	72708,716		
Gesamt	685556097,32	7517			

H_0 verwerfen

**signifikant
verschieden!**

ABER:

Tests der Varianzhomogenität

		Levene-Statistik	df1	df2	Sig.
monatliche Kaltmiete (€)	Basiert auf dem Mittelwert	101,286	22	7495	<,001
	Basiert auf dem Median	68,860	22	7495	<,001
	Basierend auf dem Median und mit angepaßten df	68,860	22	2702,044	<,001
	Basiert auf dem getrimmten Mittel	87,097	22	7495	<,001

[Homogenitätshypothese ist zu verwerfen]

Deswegen: **Welch- oder Brown-Forsythe-Test:**

Robuste Testverfahren zur Prüfung auf Gleichheit der Mittelwerte

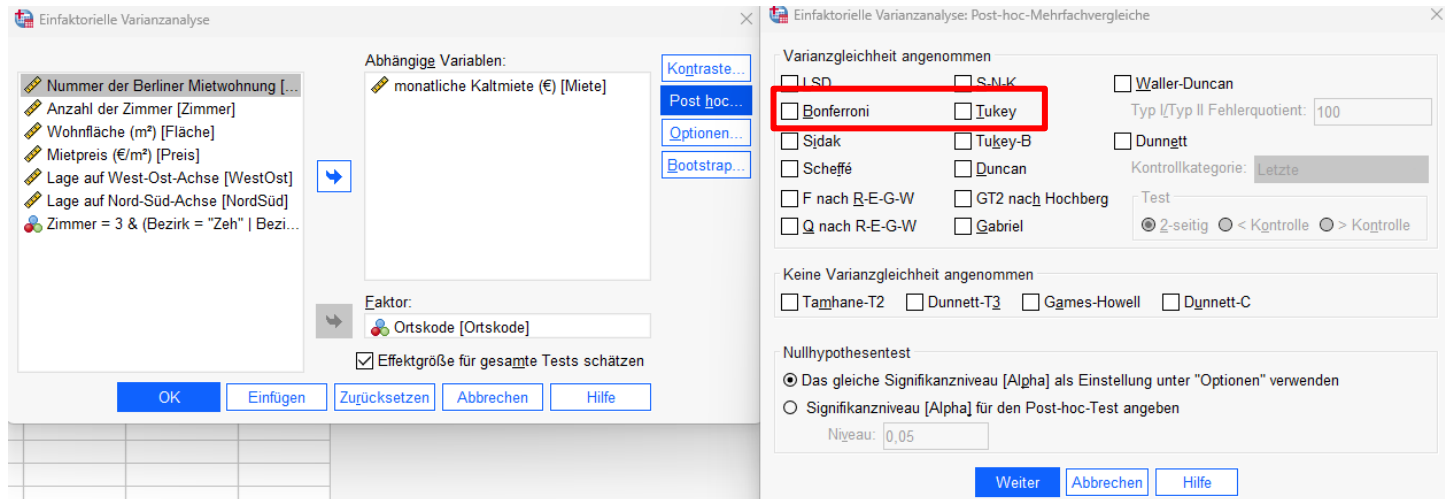
monatliche Kaltmiete (€)

	Statistik ^a	df1	df2	Sig.
Welch	71,274	22	1730,446	<,001

a. Asymptotisch F-verteilt

Ergebnis: Signifikante Mittelwertunterschiede -> Mittelwerthomogenität ist zu verwerfen!

- Nicht-parametrisches Gegenstück zu ANOVA = Kruskal-Wallis-H-Test
- Bei signifikanten Unterschieden zwischen Stichproben
 - Merkmalsunterschiede im paarweisen Mittelwertvergleich
 - Sogn. Post-hoc Tests z.B. nach Scheffé



- Grundsätzlich Varianzanalyse auch **mit mehr als einem Faktor** (d.h. mehr kategorialen Merkmalen) realisierbar mit **multifaktorielle Varianzanalyse**

Voraussetzung:

- Unabhängigkeit + Normalität + Varianzhomogenität der Daten

Zweifaktorielle Varianzanalyse

Testvoraussetzung:

- k unabhängige normalverteilte Stichproben (= ANOVA)

Ziel: Untersuchung **zwei** kategorialer Merkmale

- Zählt zu GLMs (General Linear Models)
- Nicht nur kategoriale (Feste Faktoren), sondern auch metrische Einflussfaktoren (Kovariaten = Zusatzvariablen ohne Einfluss auf abhängige Variablen) überprüfbar

Unterscheidung:

- Univariat (1 abhängiges Merkmal), bivariat (2 abhängige Merkmale), multivariat (> 2 abhängige Merkmale)

Analysieren Grafik Extras Erweiterungen Fenster Hilfe

Poweranalyse >  Anwendung durchsuch...

Richtung	Messniveau	Rolle
rechts	 Metrisch	 Eingabe
rechts	 Metrisch	 Eingabe
rechts	 Metrisch	 Eingabe
rechts	 Metrisch	 Eingabe
nks	 Nominal	 Eingabe
rechts	 Ordinal	 Eingabe

Allgemeines lineares Modell >  Univariat...
 Verallgemeinerte lineare Modelle >  Multivariat...
 Gemischte Modelle >  Messwiederholung...
 Korrelation >  Varianzkomponenten...

Nummer der Eigen...
 Wohnfläche (m²) [...]
 Zimmeranzahl [Zi...]
 Erhaltungszustand...
 logPreis
 logFläche
 Quadratmeterpreis
 logquadratmeterpr...

Abhängige Variable:
 Verkaufswert (1000...
 Feste Faktoren:
 Zustandscode (1=...
 Ortscode [Ortscode]
 Zufallsfaktoren:
 Kovariate(n):
 WLS-Gewichtung:

Modell...
 Kontraste...
 Diagramme...
 Post hoc...
 Geschätzte Randmittel...
 Speichern...
 Optionen...
 Bootstrap...

OK Einfügen Zurücksetzen Abbrechen Hilfe

Übungsaufgabe 14:

- Verwenden Sie zur Lösung der folgenden Problemstellungen die SPSS-Datendatei Eigentum_neu.sav. Die Datei beinhaltet Daten von zufällig ausgewählten Eigentumswohnungen, die 2010 verkauft wurden. Überprüfen Sie, ob sich die Verkaufswerte signifikant in Bezug auf den Standort und den Wohnungszustand unterscheiden (Signifikanzniveaus von 0.05).
- Welche Aussagen können Sie ableiten?

Tests der Zwischensubjekteffekte

Abhängige Variable: Verkaufswert (1000 €)

Quelle	Typ III Quadratsumme	df	Mittel der Quadrate	F	Sig.
Korrigiertes Modell	174779,890 ^a	7	24968,556	16,613	<,001
Konstanter Term	1018197,998	1	1018197,998	677,452	<,001
Zustandscode	10923,369	4	2730,842	1,817	,138
Ortscode	25779,443	2	12889,721	8,576	<,001
Zustandscode * Ortscode	739,249	1	739,249	,492	,486
Fehler	87172,973	58	1502,982		
Gesamt	1908629,000	66			
Korrigierte Gesamtvariation	261952,864	65			

a. R-Quadrat = ,667 (korrigiertes R-Quadrat = ,627)

- **Signifikante Abweichung** der Mittelwerte gegeben
- Kein Resultat des Wohnungszustandes ($\alpha^* = 0,138 > 0,005$), sondern vielmehr des **Standortes** ($\alpha^* < 0,001$ somit Mittelwerthomogenität des Zustandes ist nicht signifikant widerlegt!)
- Es besteht **keine Interaktion!**
- Fehlerterm ist relativ klein, somit hat das Gesamtmodell eine relativ hohe Bestimmtheit (mehr im folgenden...)

Literatur

- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2016¹⁴): Multivariate Analysemethoden – Eine anwendungsorientierte Einführung. Springer-Verlag, Berlin Heidelberg. DOI:10.1007/978-3-662-46076-4
- Baur, N. (2012): Mittelwertvergleiche und Varianzanalyse. In: Fromme, S. (Hrsg.): Datenanalyse mit SPSS für Fortgeschrittene 2: Multivariate Verfahren für Querschnittsdaten. Springer Fachmedien, Wiesbaden. S. 12 – 52. DOI: 10.1007/978-3-531-18794-5_2
- Eckstein, P. (2000³): Angewandte Statistik mit SPSS – Praktische Einführung für Wirtschaftswissenschaftler. Gabler Verlag, Wiesbaden. DOI: 10.1007/978-3-322-93157-3
- Eckstein, P. (2017⁶): Datenanalyse mit SPSS – Realdatenbasierte Übungs- und Klausuraufgaben mit vollständigen Lösungen. Springer Fachmedien, Wiesbaden. 10.1007/978-3-658-18039-3
- Kohn, W. & Öztürk, R. (2013²): Statistik für Ökonomen – Datenanalyse mit R und SPSS. Springer-Verlag, Berlin Heidelberg. DOI: 10.1007/978-3-642-37352-7
- Schnell, R. (2019²): Survey-Interviews – Methoden standardisierter Befragungen. Springer Fachmedien, Wiesbaden. DOI: 10.1007/978-3-531-19901-6

Internetlinks/SPSS:

- SPSS (2024): t-Tests. Verfügbar unter: <https://www.ibm.com/docs/de/spss-statistics/29.0.0?topic=features-t-tests> [zugegriffen am 15.05.2024]
- SPSS (2024): Einfaktorielle ANOVA. Verfügbar unter: <https://www.ibm.com/docs/de/spss-statistics/29.0.0?topic=features-one-way-anova> [zugegriffen am 15.05.2024]
- SPSS (2024): GLM – Univariat. Verfügbar unter: <https://www.ibm.com/docs/de/spss-statistics/29.0.0?topic=features-glm-univariate-analysis> [zugegriffen am 15.05.2024]

Empfehlung YouTube:

- „Statistik am PC“



Kontakt:

Lehrstuhl Immobilienentwicklung

M.Sc. Alexander Schuhmacher (GB I, Raum 323)

E-Mail: alexander.schuhmacher@tu-dortmund.de