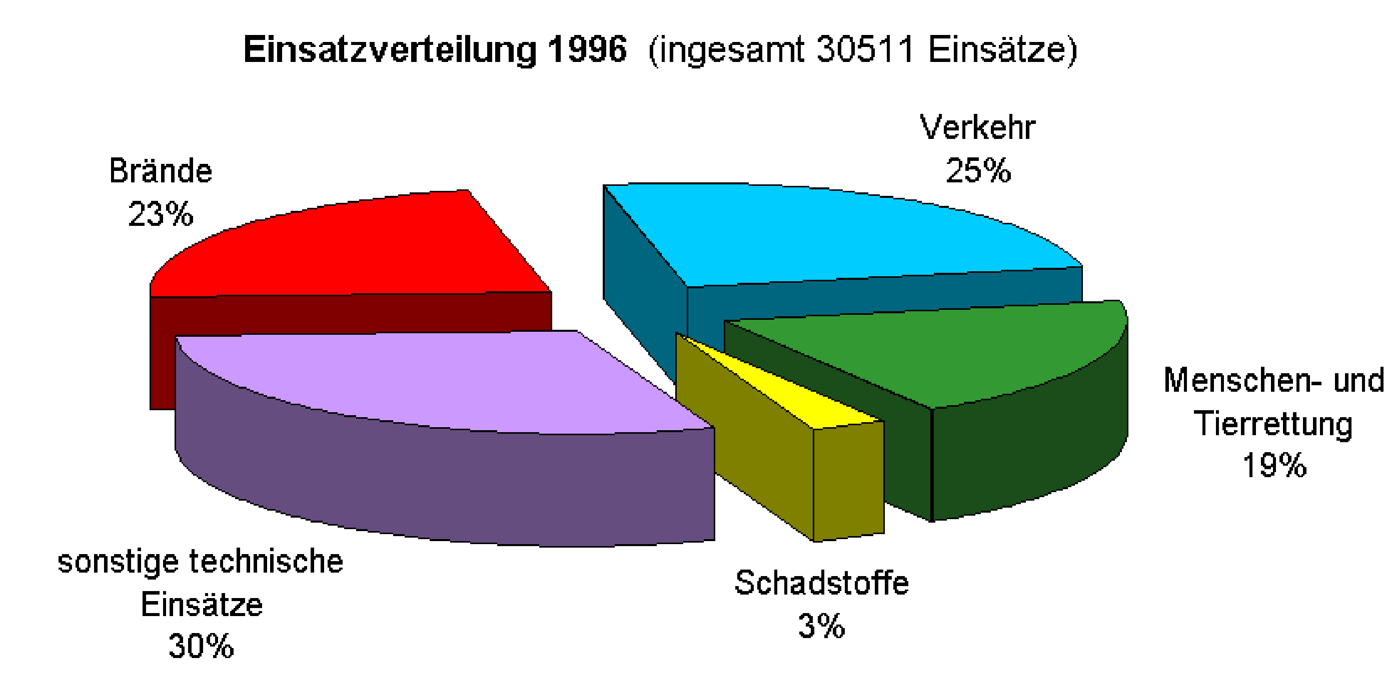
2. EIN KATEGORIALES MERKMAL – (univariate kategoriale Daten)



Einführendes Beispiel: Einsätze der Wiener Feuerwehr 1996

wieviele Einsätze in den einzelnen Kategorien ?

2 Möglichkeiten die Daten darzustellen (numerisch)

- **Anzahl** (*absolute Häufigkeiten*, d.h. wieviele Einsätze in einzelnen Kategorien)
- **Anteil** (Angaben in *Prozent* oder mittels *relativer Häufigkeiten**)

(* relative Häufigkeiten sind Prozentzahlen/100 - genaueres später)

Daten sind *aggregiert*
(in größeren Einheiten zusammengefasst)

dient zur übersichtlichen Darstellung (aber Informations-verlust)

Darstellung der Daten in Tabellenform: (detaillierter)

Art des Feuerwehreinsatzes	Anzahl
Brand	6.892
Brand- und Explosionsgefahr, Erhebung nach Brand und Explosion	907
Rettung, Bergung und in Sicherheit bringen	2.814
Unfall in versperrtem Raum	3.030
Verkehrsunfall	75
Verkehrshindernis nach Verkehrsunfall	3.860
Verkehrshindernis ohne vorhergehendem Verkehrsunfall	869
Falsch geparkte Kraftfahrzeuge, Ortsveränderung von Kfz	2.922
Schäden an Bauteilen, Einstürze von Objekten oder Teilen derselben	626
Austretende feste und flüssige Schadstoffe	893
Geruchs- und Rauchbelästigung, Gasausströmung ohne Explosionsgefahr	658
Technischer Einsatz mit Wasser	3.647
Technischer Einsatz mit Wasser in Form von Eis oder Schnee	346
Sonstiger technischer Einsatz	2.269
Hilfeleistung (Beistellung von Personal, Geräten oder Einsatzfahrzeugen)	90
Sonstige Einsätze	613
gesamt	30.511

(Feuerwehreinsätze in Wien, 1996; Quelle: Magistrat der Stadt Wien)

Definitionen in diesem Beispiel:

Beobachtungseinheiten:	einzelne Einsätze (der Feuerwehrteams)
Variable:	Art des Feuerwehreinsatzes
Ausprägung:	kategorial (nominal) für jeden Einsatz wird festgelegt, in welche Kategorie er fällt, z.B. 1 ... Brand 2 ... Brand und Explosionsgefahr usw.

Rohdaten: (Einzelbeobachtungen)

Datum	Uhrzeit	Art des Einsatzes	Standort der Feuerwache
1996-01-01	0:01	1	3
1996-01-01	0:03	1	1
1996-01-01	0:11	3	4
1996-01-01	0:21	11	1
1996-01-01	0:21	5	1
1996-01-01	0:25	5	5
1996-01-01	0:38	2	7
...	...	...	...
1996-12-31	23:59	3	6

so könnten Daten aussehen, wenn sie zur Verarbeitung im Computer gespeichert sind

die zusammenfassenden Tabellen erhält man durch Auszählen wie oft eine bestimmte Kategorie vorgekommen ist

(welches Skalenniveau haben die anderen Variablen ?)

KATEGORIALE DATEN

enstehen durch Abzählen und Zuordnung in Kategorien

- **wenn Daten in KLASSEN oder KATEGORIEN eingeteilt werden**

Bsp.: Auszählung, wieviele Patienten *männlich* bzw. *weiblich* sind

- **wenn Daten GRUPPIERT werden**

Bsp.: Patienten aus 9 Bundesländern werden in 3 Regionen eingeteilt:

Region WEST – Region MITTE – Region OST,
dann Auszählung, wieviele Patienten in einer Region

Bsp.: Anzahl von Behandlungen in 3 Gruppen:
wenig (bis 40) – mittel (41-100) - viel (mehr als100)
dann wieder Auszählung, wieviele Patienten In jeder Kategorie

man unterscheidet:

- UNGEORDNETE KATEGORIEN (nominal)
- GEORDNETE KATEGORIEN (ordinal)

Darstellung auf 2 Arten:

- **absolute HÄUFIGKEITEN**
- **relative HÄUFIGKEITEN bzw. Prozentsätze**

ABSOLUTE UND RELATIVE HÄUFIGKEITEN

Beispiel: Präsentatorenwerbung und normale Werbung

Marke	(absolute) Häufigkeit	relative Häufigkeit	Prozente
FL	45	0,186	18,6
BB	64	0,264	26,4
KH	52	0,215	21,5
CC	81	0,335	33,5
Gesamt	242	1,000	100,0

- **(absolute) Häufigkeiten:**

  durch Abzählen wie oft eine bestimmte Kategorie in den Daten vorkommt
  z.B.: 45 Kinder haben FL gewählt

- **relative Häufigkeiten** (Anteile):

$$\text{relative Häufigkeit} = \frac{\text{absolute Häufigkeit in einer Kategorie}}{\text{Gesamtanzahl}}$$

  z.B.: der Anteil (relative Häufigkeit) für FL = 45 / 242 = 0,186

- **Prozente**

$$\text{Prozent} = \text{relative Häufigkeit} \times 100$$

Anmerkung:

bei der Angabe von Prozenten oder relative Häufigkeiten sollte immer die Gesamtzahl der Beobachtungen mitangegeben werden (dies ist ein Indikator für die Wertigkeit und Glaubwürdigkeit z.B. einer Meinungsumfrage)

KUMULIERTE RELATIVE HÄUFIGKEITEN:

(nur bei geordneten Kategorien sinnvoll)

Beispiel: Canada Social Survey, 1991
insgesamt 11924 Personen zu verschiedensten Themen befragt

Frage nach subjektiv erlebter Stressbelastung:

Kategorie	absolute Häufigkeit	relative Häufigkeit	kumulierte relative Häufigkeit
keine	2310	0,20	0,20
gering	3783	0,33	0,53
hoch	4397	0,39	0,92
sehr hoch	844	0,07	1,00
gesamt	11334	1,00	-

kumulierte relative Häufigkeit:

Summe der relativen Häufigkeiten aller vorhergehenden Kategorien, einschliesslich der aktuellen

erlaubt Aussagen wie:

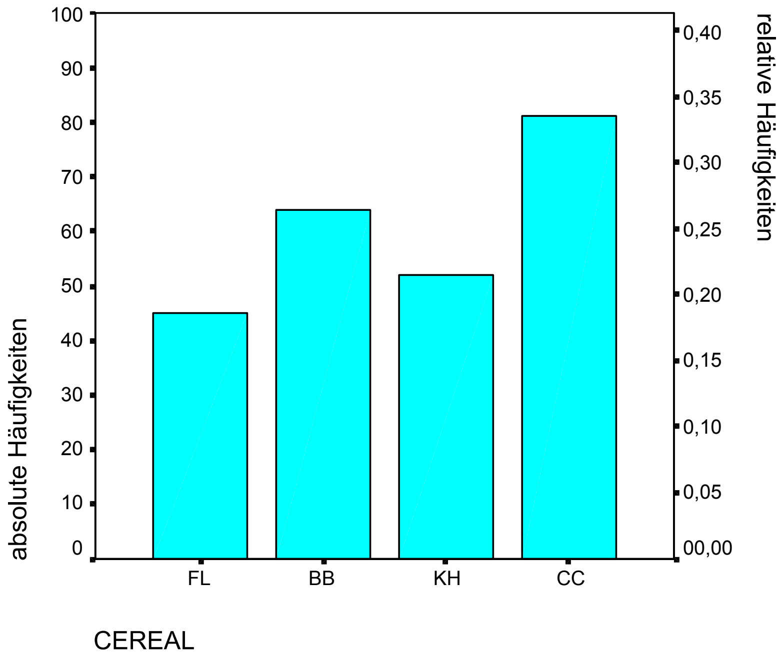
- 53% haben keine oder nur geringe Stressbelastung
- 80% haben zumindest geringe Stressbelastung (100% minus 20%)

Berechnung:

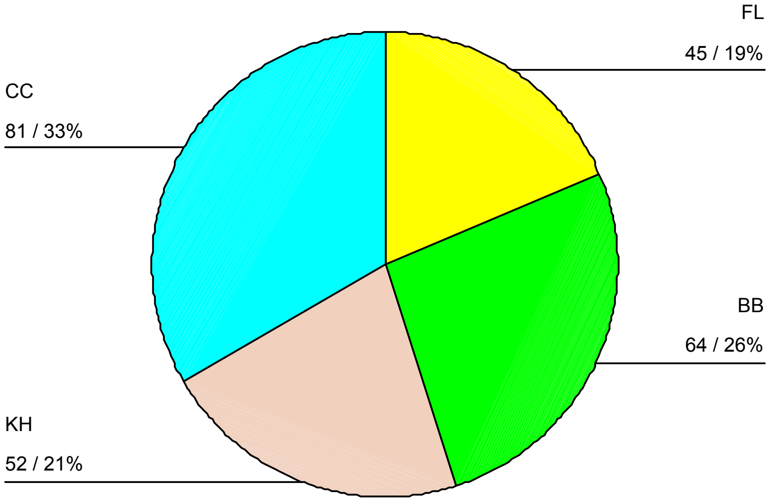
Kategorie	relative Häufigkeit	kumulierte relative Häufigkeit
keine	**0,20**	**0,20**
gering	**0,33**	0,53 = 0,20 + **0,33**
hoch	**0,39**	0,92 = 0,20 + 0,33 + **0,39**
sehr hoch	**0,07**	1,00 = 0,20 + 0,33 + 0,39 + **0,07**

GRAPHISCHE DARSTELLUNGEN FÜR EINFACHE KATEGORIALE DATEN

BAR-CHART (BALKENDIAGRAMM)



PIE-CHART (KREISDIAGRAMM)



WICHTIGE FRAGESTELLUNGEN
BEI EINEM KATEGORIALEN MERKMAL

Fragestellungen für mehrere Kategorien:

- Kommen alle Kategorien gleich häufig vor ?

Diese Frage stellt man vor allem dann, wenn man wissen möchte, ob eine oder mehrere Kategorien häufiger vorkommen als andere.

Beispiel: Showmaster-Werbung:
Wählen die Kinder, nachdem sie Werbung für CC gesehen haben, eher dieses Produkt ?

- Entsprechen die Häufigkeiten in den Kategorien einer bestimmten Vorgabe ?

Diese Frage stellt man, wenn man wissen möchte, ob die Häufigkeiten eine bestimmte vorspezifizierte Verteilung haben.

Beispiel: Sind die Marktanteile von Produkt A und B tatsächlich 65% bzw. 20% ?

Fragestellungen für nur eine einzelne Kategorie:

- Entspricht die Häufigkeit (Prozentsatz, Anteil) in nur einer Kategorie einer bestimmten Vorgabe ?

Diese Frage stellt man vor allem dann, wenn man wissen möchte, ob ein bestimmter Prozentsatz (Anteil), der aus einer Stichprobe errechnet wurde, mit einem vorgegebenen Wert übereinstimmt, oder diesen über- bzw. unterschreitet.

Beispiel: Gewinnt Kandidatin X die Bundespräsidentenwahl ?

- In welchem Bereich kann man den Anteil einer Kategorie in der Grundgesamtheit erwarten ?

Diese Frage stellt man, wenn man abschätzen möchte, wie groß die Schwankungen sein können, wenn man Daten nur aus Stichproben hat, aber über die Grundgesamtheit etwas aussagen möchte

Beispiel: Ausgehend von einer Meinungsumfrage, in welchem Bereich ist der tatsächliche Wähleranteil von Partei X zu erwarten ?

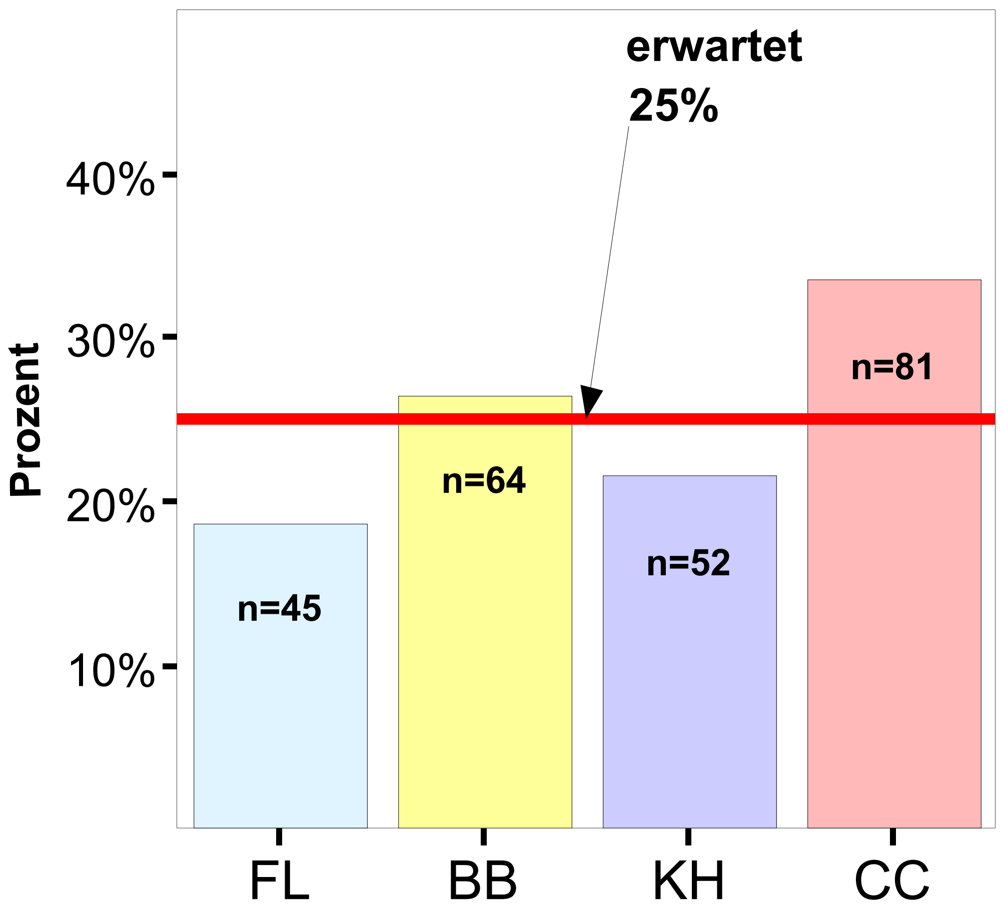
FRAGESTELLUNG 1:

Kommen alle Kategorien gleich häufig vor ?

Beispiel: Präsentatorenwerbung und normale Werbung

War die Werbemaßnahme wirksam ? (bzw.)

Wird das Produkt CC, das beworben wurde, häufiger ausgewählt, nachdem ein Werbespot gesehen wurde, oder sind die Unterschiede in den beobachteten Häufigkeiten nur zufällig ?



Exkurs:

Der „faire Würfel“ und der Zufall beim Würfeln

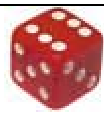
Wahrscheinlichkeit einen „6“er zu würfeln = 1/6 ??

nur wenn Würfel „fair“ ist
(unrealistische Annahme – setzt perfekte Geometrie voraus)

wie lässt sich das prüfen ? → ausprobieren !
d.h. oftmaliges Würfeln und auszählen

da Ergebnis zufällig ist, kann verschiedenes passieren

Beispiel: 18 Mal würfeln (ein Versuchsdurchgang)

							fair ?
Versuch 1	3	3	3	3	3	3	ok
Versuch 2	3	3	2	4	4	2	ok ?
Versuch 3	0	0	0	0	0	18	nicht ok
Versuch 4	0	2	0	1	0	15	nicht ok ?
Versuch 5	2	1	2	4	5	4	ok ??

ab wann glaubt man nicht mehr an fairen Würfel ?

um zu beurteilen, ob Würfel fair ist, benötigen wir Methode,
um das Ergebnis eines einzelnen Versuchs(durchgangs)
beurteilen zu können
diese Methode nennt man einen **statistischen Test**

Ergebnis des Tests, ob ein Würfels fair ist:

im Prinzip 2 Möglichkeiten:

„der Würfel ist fair“	oder	„der Würfel ist nicht fair“

diese beiden „Statements“ nennt man **Hypothesen**

die sog. **Nullhypothese (**H_0**)** besagt: **der Würfel ist fair**

- alle Zahlen kommen gleich wahrscheinlich vor, nämlich mit 1/6

die **Alternativhypothese (**H_A**): der Würfel ist nicht fair**

- zumindest eine Zahl ist wahrscheinlicher als die anderen

Entscheidung:

man kann eine **Wahrscheinlichkeit** angeben, **ob** der jeweilige Versuchsdurchgang dafür spricht, dass **der Würfel fair ist**

diese Wahrscheinlichkeit wird mit *p-Wert* (*p-value*) oder *Signifikanz(wert)* bezeichnet

- ist diese Wahrscheinlichkeit (p-Wert) **klein**, glaubt man **nicht** daran, dass der Würfel fair ist

- ist diese Wahrscheinlichkeit (p-Wert) **groß**, vertraut man darauf, dass der Würfel fair ist

was heißt groß bzw. klein ? (oder)
ab wann glaubt man das eine oder das andere ?

ab einem **p-Wert von weniger als** (üblicherweise) **5%** glaubt man **nicht** mehr an die „Fairness“ des Würfels
man sagt dann, man *verwirft* die Nullhypothese zugunsten der Alternativhypothese

zurück zum Beispiel „Wirksamkeit der Werbemaßnahme“:

was hat der „faire Würfel“ mit dem Beispiel zur Wirksamkeit einer Werbemaßnahme zu tun ?

- Würfel mit möglichen Ergebnissen „1“ – „6“ (6 mögliche Kategorien) entspricht dem ausgewählten Produkt mit Ergebnissen FL, BB, KH, CC (4 mögliche Kategorien)

- Anzahl des Auftretens einer bestimmten Zahl beim Würfeln (z.B. „6“) entspricht der Anzahl, wie oft ein bestimmtes Produkte gewählt wurde (z.B. „CC“)

- ein Versuchsdurchgang beim Würfeln (18 Mal den Würfel zu werfen) entspricht einer Stichprobe beim Werbungs-beispiel (242 untersuchte Kinder)

- „Fairness“ des Würfel entspricht „keine Präferenz für ein bestimmtes Produkt“ in der Population aller vergleichbaren Kinder

- die Frage im Werbungsbeispiel: „Wird ein bestimmtes Produkt häufiger gewählt ?“ entspricht der Frage „ist der Würfel unfair ?“

Anmerkung: Beantwortungsmöglichkeit der weiterführenden Frage *„war die Werbemaßnahme effektiv?“* beruht auf folgender Idee:

- wenn die Kinder kein Produkt kennen und alle gleich aussehen und auch für keines geworben wurde, dann werden sie zufällig eines auswählen (mit Wahrscheinlichkeit 1/4)
  → unterschiedliche Häufigkeiten sind Stichprobenschwankungen
- wenn aber Werbemaßnahme erfolgte, und trotzdem alle Produkte gleich häufig ausgewählt, dann gewisser Hinweis, dass die Werbemaßnahme nicht erfolgreich war
  → d.h. umgekehrt, wenn Produkte nicht gleich häufig ausgewählt, dann gewisser Hinweis auf Effektivität der Werbung

Idee zur Beantwortung der Frage ob Werbemaßnahme wirksam war ?

- sind die beobachteten **UNTERSCHIEDE** in den Häufigkeiten nur **GERING**, kann man annehmen, dass sie auf Grund von Stichprobenschwankungen (also zufällig) zustande gekommen sind:

  in „Wirklichkeit“ (in der Population) gibt es keine Unterschiede
  "die Werbung ist nicht wirksam"

- sind die **UNTERSCHIEDE GROSS** (größer als man dem Zufall zuschreiben würde):

  in „Wirklichkeit“ (nicht nur in der Stichprobe, auch in der Population) gibt es Unterschiede
  "die Werbung ist wirksam"

Maßzahl für die Größe der Unterschiede in den Häufigkeiten der einzelnen Kategorien

d.h. für die Abweichung von einer *Gleichverteilung*
Gleichverteilung heißt: gleiche Häufigkeiten in allen Kategorien

Pearson's X^2

je größer diese Maßzahl ist, umso eher gibt es tatsächlich (in der Population) Unterschiede in den Häufigkeiten

Konzept der beobachteten und der erwarteten Häufigkeiten:

beobachtete Häufigkeiten: $\mathbf{o_j}$ (o ... observed)

jene Häufigkeiten für die einzelnen Kategorien **j**, die beobachtet wurden

	FL	BB	KH	CC	*n*
Beobachtete Häufigkeit	45	64	52	81	242

erwartete Häufigkeiten: $\mathbf{e_j}$ (e ... expected)

welche Häufigkeiten, bei gegebener Gesamtanzahl *n* von Beobachtungen, würden wir in den einzelnen Kategorien erwarten, wenn es keine Unterschiede gäbe

	FL	BB	KH	CC	*n*
erwartete relative Häufigkeit	0,25	0,25	0,25	0,25	-
erwartete (absolute) Häufigkeit	60,5	60,5	60,5	60,5	242

(Berechnung der erwarteten Häufigkeit mittels: 0,25 x 242 = 60,5)

Pearson's X^2

die Größe der Maßzahl X^2 hängt von der Differenz zwischen beobachteten und erwarteten Häufigkeiten ab

je mehr die beobachteten und erwarteten Häufigkeiten voneinander abweichen, umso größer wird X^2 werden

Berechnung von X^2 :

$$X^2 = \sum_{j=1}^{J} \frac{(e_j - o_j)^2}{e_j}$$

J ... Anzahl der Kategorien,
hier 4

am Beispiel:

	FL	BB	KH	CC	*n*
Beobachtete Häufigkeit	45	64	52	81	242
erwartete (absolute) Häufigkeit	60,5	60,5	60,5	60,5	242

$$X^2 = \frac{(45-60{,}5)^2}{60{,}5} + \frac{(64-60{,}5)^2}{60{,}5} + \frac{(52-60{,}5)^2}{60{,}5} + \frac{(81-60{,}5)^2}{60{,}5}$$

$$= 3{,}97 + 0{,}20 + 1{,}19 + 6{,}95 = 12{,}31$$

was bedeutet nun ein bestimmter X^2 - Wert: also z.B. 12,31

man kann berechnen, wie wahrscheinlich es ist, einen bestimmten X^2 - Wert oder einen noch größeren zufällig zu erhalten, wenn die Annahme zutrifft, dass in der Population alle Kategorien gleich häufig sind (wenn also die Nullhypothese zutrifft)

diese Wahrscheinlichkeit heißt p-Wert oder Signifikanz

(die Bestimmung des p-Werts erfolgt hier unter Zuhilfenahme der sogenannten χ^2 -Verteilung (sprich: "chi-Quadrat Verteilung") daher nennt man diesen Test auch χ^2 –Test)

Vorgehensweise beim Durchführen des Tests:

1. **berechnen des p-Werts** (händisches Berechnen nur äußerst mühsam, daher erfolgt dies meistens per Computer)

2. **Vergleich des p-Werts mit Signifikanzniveau (wird vorher festgelegt, meist 5% bzw. 0.05)**

- ist **p-Wert < 0.05**, dann glaube ich nicht mehr an zufällige Abweichungen von der Gleichverteilung – mein konkretes Ergebnis kann natürlich durch Zufall zustande gekommen sein, aber das ist sehr unwahrscheinlich (< 5%) - Schlussfolgerung: *Nullhypothese (Ho) verwerfen*
$\rightarrow$ **<u>Unterschiede auch in Population</u>**

- ist **p-Wert** $\geq$ **0.05**, dann muss ich annehmen dass die Häufigkeiten für alle Kategorien in der Population gleich sind, und etwaige Unterschiede nur zufällig in meiner konkreten Stichprobe beobachtet wurden – Schlussfolgerung: *Nullhypothese (Ho) beibehalten*
$\rightarrow$ **<u>keine Unterschiede in Population</u>**

Berechnung mit Statistikprogrammen:

SPSS

CEREAL

	Beobachtetes N	Erwartete Anzahl	Residuum
FL	45	60,5	-15,5
BB	64	60,5	3,5
KH	52	60,5	-8,5
CC	81	60,5	20,5
Gesamt	242		

Statistik für Test

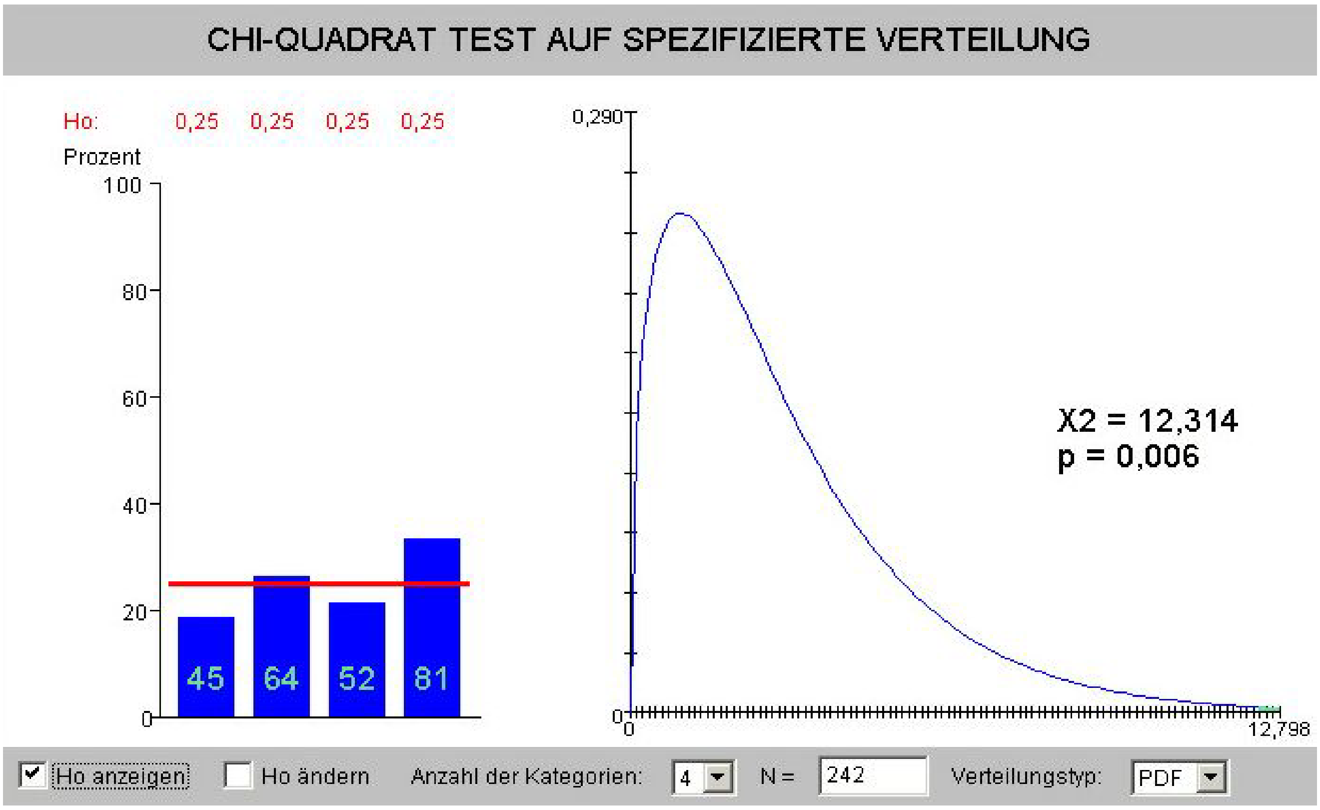
	CEREAL
Chi-Quadrat	12,314
df	3
Asymptotische Signifikanz	,006

- der p-Wert („Asymptotische Signifikanz“) ist 0.006
- dieser Wert ist wesentlich kleiner als 0.05, daher wird die Nullhypothese verworfen
- die Produkte werden nicht gleich häufig ausgewählt - demnach scheint Werbemaßnahme effektiv gewesen zu sein

R

```
                 BB   CC   FL   KH
beobachtete H  64.0 81.0 45.0 52.0
erwartete H    60.5 60.5 60.5 60.5


Chi-squared test for given probabilities

data:  table(CEREAL)
X-squared = 12.314, df = 3, p-value = 0.006381
```

grafische Darstellung des Ergebnisses zur Fragestellung 1:



alternative Vorgehensweise
(wenn kein Computer zur Verfügung)

dazu gibt es die sogenannte χ^2 -Tabelle

1. suchen des kritischen χ^2 - Werts

- bestimmen der Freiheitsgrade *df (degrees of freedom)*

$$\boxed{df = J - 1}$$

  bei Fragestellung ob Häufigkeiten gleich verteilt sind:
  in unserem Beispiel: df = 4 - 1 = 3

- nachsehen in Zeile mit Freiheitsgraden und in Spalte mit 0.950
  kritischer Wert ist in unserem Beispiel 7.815

2. Vergleich des X^2 - Werts mit dem kritischen χ^2 - Wert:

- ist berechneter $\mathbf{X^2}$ - Wert <u>größer</u> als χ^2 - Wert aus Tabelle, dann glaube ich nicht mehr an zufällige Abweichungen von der Gleichverteilung - <u>Unterschiede auch in Population</u>

- <u>sonst</u> muss ich annehmen, dass die Häufigkeiten für alle Kategorien in der Population gleich sind, und etwaige Unterschiede nur zufällig in der Stichprobe beobachtet wurden - <u>keine Unterschiede in Population</u>

FRAGESTELLUNG 2:

Entsprechen die Anteile von einzelnen Kategorien in einer Stichprobe den tatsächlichen Anteilen in der Grundgesamtheit ?

Bsp.: Repräsentativität einer Meinungsumfrage

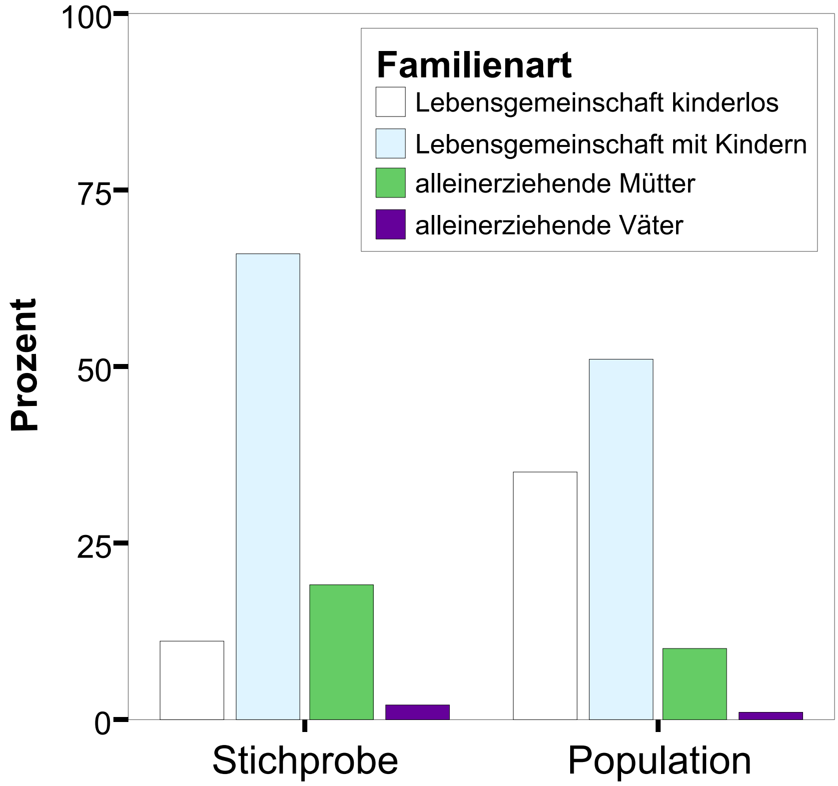
Telefonumfrage an 200 zufällig ausgewählten Personen
Thema: Gesetzesmaßnahme zur speziellen Förderung von Familien mit Kindern

---

Ist die Stichprobe repräsentativ bezüglich verschiedener Familienarten ?

Häufigkeiten von Familienarten in der Stichprobe und in Österreich (Angabe des Statistischen Zentralamts für 1996)

	Stichprobe		Österreich
Familie	Häufigkeit	Prozent	Prozent
Lebensgemeinschaft kinderlos	23	11,5	35,8
Lebensgem. mit Kindern	133	66,5	51,7
alleinerziehende Mütter	39	19,5	10,8
alleinerziehende Väter	5	2,5	1,7



Grundidee zur Beantwortung der Frage gleich wie bei Fragestellung 1:

- entsprechen die Stichproben-Häufigkeiten anteilsmäßig jenen in der Population ?
  (weichen sie nur zufällig von diesen ab ?)

- oder gibt es eine systematische Abweichung ?
  (überzufällige Abweichungen)

der Unterschied zu Fragestellung 1 ist nur:

hier sind die erwarteten Werte nicht alle gleich
sie haben eine spezielle Verteilung (jene der Grundgesamtheit)

Prüfen der Frage wieder mittels χ^2 -Test:

- Vergleich von beobachteten und erwarteten Häufigkeiten

Berechnung der erwarteten Häufigkeiten:

welche Häufigkeiten, bei gegebener Gesamtanzahl *n* von Beobachtungen, würden wir in den einzelnen Kategorien erwarten, wenn sie der Verteilung in der Population entsprechen würden

erwartete Häufigkeit =
relative Häufigkeit in Population (π) x Stichprobengröße (*n*)

$$e_j = \pi_j \cdot n$$

am Beispiel:

	LGo	LGm	Ma	Va	*gesamt*
erwartete relative Häufigkeit	0,358	0,517	0,108	0,017	1
erwartete (absolute) Häufigkeit	71,6	103,4	21,6	3,4	200

z.B.: für j=2: e_j = 0,517 x 200 = 103,4

wie vorher $X^2 = \sum_{j=1}^{J} X_j^2 = \sum_{j=1}^{J} \frac{(e_j - o_j)^2}{e_j}$

(LGo ... Lebensgemeinschaft ohne Kinder, etc.)

Berechnung von X^2:

	LGo	LGm	Ma	Va	*gesamt*
beobachtete Häufigkeit	23	133	39	5	200
erwartete Häufigkeit	71,6	103,4	21,6	3,4	200
Beitrag zum X^2-Wert *)	32,99	8,47	14,02	0,75	56,23

*) z.B.: für j=1: $X^2_1 = (71{,}6-23)^2 / 71{,}6 = 32{,}99$

Ergebnis: $X^2 = 56{,}23$

Prüfen des X^2 – Werts bei händischer Berechnung:

1. **suchen des kritischen χ^2 - Werts**
   - Freiheitsgrade df = 4 - 1 = 3
   - nachsehen in Tabelle: Zeile mit df, Spalte mit 0.950
     kritischer Wert ist: 7,82
2. **Vergleich des X^2 - Werts mit dem kritischen χ^2 - Wert:**
   - $X^2 = 56{,}23 > \chi^2 = 7{,}82$

     daher: sehr unplausibel, dass Anteile in der Stichprobe nur zufällig von jenen in der Population abweichen

ERGEBNIS:
die erhobene Stichprobe ist nicht repräsentativ
(die Abweichungen zwischen beobachteten und erwarteten Werten sind "signifikant")

Berechnung mit Statistikprogrammen:

SPSS

Familienart

	Beobachtetes N	Erwartete Anzahl	Residuum
Lebensgemeinschaft kinderlos	23	71,6	-48,6
Lebensgemeinschaft mit Kindern	133	103,4	29,6
alleinerziehende Mütter	39	21,6	17,4
alleinerziehende Väter	5	3,4	1,6
Gesamt	200		

Statistik für Test

	Familienart
Chi-Quadrat	56,231
df	3
Asymptotische Signifikanz	,000

- der p-Wert („Asymptotische Signifikanz") < 0.001
- dieser Wert ist wesentlich kleiner als 0.05, daher wird die Nullhypothese verworfen
- die in der Stichprobe beobachteten Prozentsätze bestimmter Familienarten entsprechen nicht den Anteilen in der Population –
- demnach scheint Stichprobe nicht repräsentativ zu sein

R

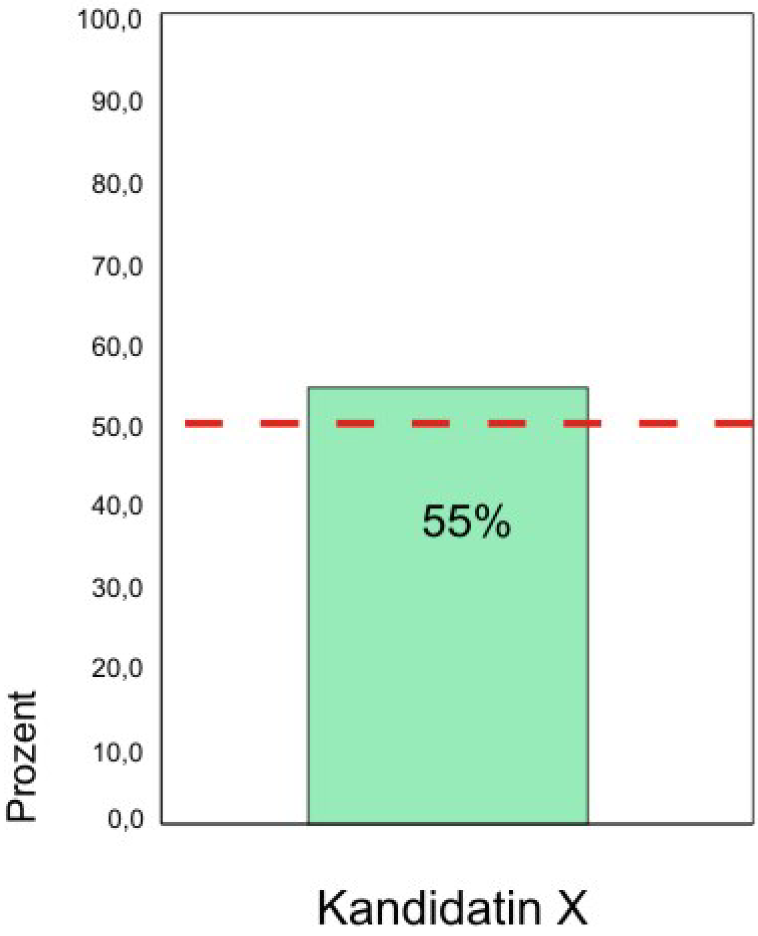
```
Chi-squared test for given probabilities

data: absH
X-squared = 56.2314, df = 3, p-value = 3.749e-12
```

FRAGESTELLUNG 3:

Hat ein Prozentsatz (Anteil) einen bestimmten Wert ?

Beispiel: Gewinnt Kandidatin X die Bundespräsidentenwahl ?

Bei einer Meinungsumfrage kurz vor der Bundespräsidentenwahl gaben 55 % (220 von 400 Befragten) an, Kandidatin X wählen zu wollen. Kann man auf Grund dieser Stichprobendaten darauf schließen, dass Kandidatin X über 50% kommen wird, oder ist die absolute Mehrheit in der Stichprobe nur zufällig ?



Exkurs:

Die „faire Münze“ und der Zufall

(Idee analog zum „fairen“ Würfel)

beim Werfen eines Euro:  

Wahrscheinlichkeit für Kopf oder Zahl: 1/2 ??

nur wenn Euromünze „fair“ ist
(unrealistische Annahme – setzt perfekte Geometrie voraus)

wie lässt sich das prüfen ? → ausprobieren !
d.h. oftmaliges Werfen und auszählen

da Ergebnis zufällig ist, kann verschiedenes passieren

Beispiel: 10 Mal werfen (ein Versuchsdurchgang)

	Zahl	Kopf	%	fair
● ● ● ● ● ● ● ● ● ●	10	0	1,0	✗
● ○ ● ○ ○ ○ ● ○ ● ●	5	5	0,5	✓
● ○ ○ ● ● ● ○ ● ○ ●	6	4	0,6	?
● ○ ● ○ ○ ○ ○ ● ○ ○	3	7	0,7	??

ab wann glaubt man nicht mehr an „faire“ Münze ?

zur Bewertung, ob Münze fair ist, benötigen wir wieder eine Methode, um das Ergebnis eines einzelnen Versuchs (Versuchsdurchgangs) beurteilen zu können
→ **statistischen Test**

Ergebnis des Tests, ob eine Münze fair ist:

wieder 2 Möglichkeiten:

„die Münze ist fair“	oder	„die Münze ist nicht fair“

mit entsprechenden **Hypothesen**

die **Nullhypothese (**H_0**)** besagt: **die Münze ist fair**

- **„Kopf“ und „Zahl“ kommen gleich wahrscheinlich vor (mit 1/2)**

die **Alternativhypothese (**H_A**): die Münze ist nicht fair**

- **eine Seite ist wahrscheinlicher als die andere**

Entscheidung:

wieder mittels p-Wert, der Aufschluss darüber gibt, für welche der beiden Alternativen man sich entscheiden soll

- p-Wert **klein:** $\rightarrow$ **Münze ist** (eher) **nicht fair**
  („H_0 zugunsten von H_A verwerfen“)
- p-Wert **groß**, $\rightarrow$ **Münze ist** (eher) **fair**
  („H_0 beibehalten“)

die Grenze (groß / klein) war bisher 5%, man nennt sie *Signifikanzniveau* oder *Irrtumswahrscheinlichkeit*

man kann sie selbst wählen (andere übliche Wahl ist 1%)
Wahl hängt von der Fragestellung ab:
wenn man strenger zur Alternativhypothese sein will, wählt man kleineren Wert (es müssen bei 1%-Signifikanzniveau größere Abweichungen von der H_0 in der Stichprobe auftreten als bei 5%, bevor man H_0 verwirft)

Arten von Alternativhypothesen

manchmal kann man die Alternativhypothese noch genauer spezifizieren:

3 mögliche Formulierungen der Alternativhypothese:

ohne *vorherige* Vermutung welche Seite häufiger kommt:
Möglichkeit 1:

- eine Seite ist wahrscheinlicher als die andere

wenn man schon *vorher* eine bestimmte Vermutung hat:
Möglichkeit 2:

- Kopf ist wahrscheinlicher als Zahl

Möglichkeit 3:

- Zahl ist wahrscheinlicher als Kopf (bzw.)
  Kopf ist weniger wahrscheinlicher als Zahl

Möglichkeit 1 nennt man **zweiseitige Alternativhypothese**

Möglichkeit 2 und 3 heißen **einseitige Alternativhypothese**

Anmerkung:

man verwirft bei einseitig formulierten Alternativhypothesen die Nullhypothese leichter, d.h. es genügen kleinere Abweichungen von der Nullhypothese, um sie zu verwerfen, als bei zweiseitigen Alternativhypothesen
das bedeutet, man sollte eine Alternativhypothese nur dann einseitig formulieren, wenn man konkrete Vermutungen hat (also im Vorhinein) – nachträgliche Umdefinition ist Selbstbetrug bzw. verfälschend

zurück zum Beispiel:

was hat die „faire Münze“ mit dem Beispiel der Stimmanteile für Kandidatin X zu tun ?

- Münze mit möglichen Ergebnissen „Kopf“ – „Zahl“ (2 mögliche Kategorien) entsprechen den Präferenzen FÜR bzw. NICHT FÜR die Kandidatin

- Anzahl des Werfen von „Kopf“ entspricht der, wie oft die Kandidatin X bevorzugt wurde

- ein Versuchsdurchgang beim Werfen der Münze (10 Mal) entspricht einer Stichprobe bei einer Meinungsumfrage

- „Fairness“ der Münze entspricht einem Verhältnis von 50:50 FÜR bzw. NICHT FÜR die Kandidatin in der Population aller wahlberechtigten Österreicher

- die Frage Beispiel: „Erzielt Kandidatin X abolute Stimmenmehrheit entspricht der Frage „ist die Münze unfair ?“

wieder **Maßzahl** notwendig um Abweichung von Nullhypothese (gleiche Häufigkeiten für bzw. gegen Kandidatin in Population) beurteilen zu können

z-Wert
(hier kein spezieller Name wie vorher Pearson's X^2 , allgemein nennt man solch eine Maßzahl ***Teststatistik***)

je größer diese Maßzahl ist, umso eher weichen die tatsächlichen Häufigkeiten (in der Population) von dem vorgegebenen Wert (hier 50%) ab

z – Wert zum Testen von Anteilen

$$\boxed{z = \frac{r_j - \pi_0}{\sigma}}$$

wobei:

r_j... relative Häufigkeit für Kategorie j, d.h. FÜR Kandidatin X

n ... Stichprobengröße

π_0...relative Häufigkeit in Population (durch Nullhypothese festgelegt, in unserem Beispiel 0.5)

σ... wie sehr kann Ergebnis in Stichprobe schwanken (sprich „sigma“)

$$\sigma = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

Berechnung für unser Beispiel:

$$r_j = \frac{220}{400} = 0,55$$

$$\sigma = \sqrt{\frac{0,5(1-0,5)}{400}} = 0,025$$

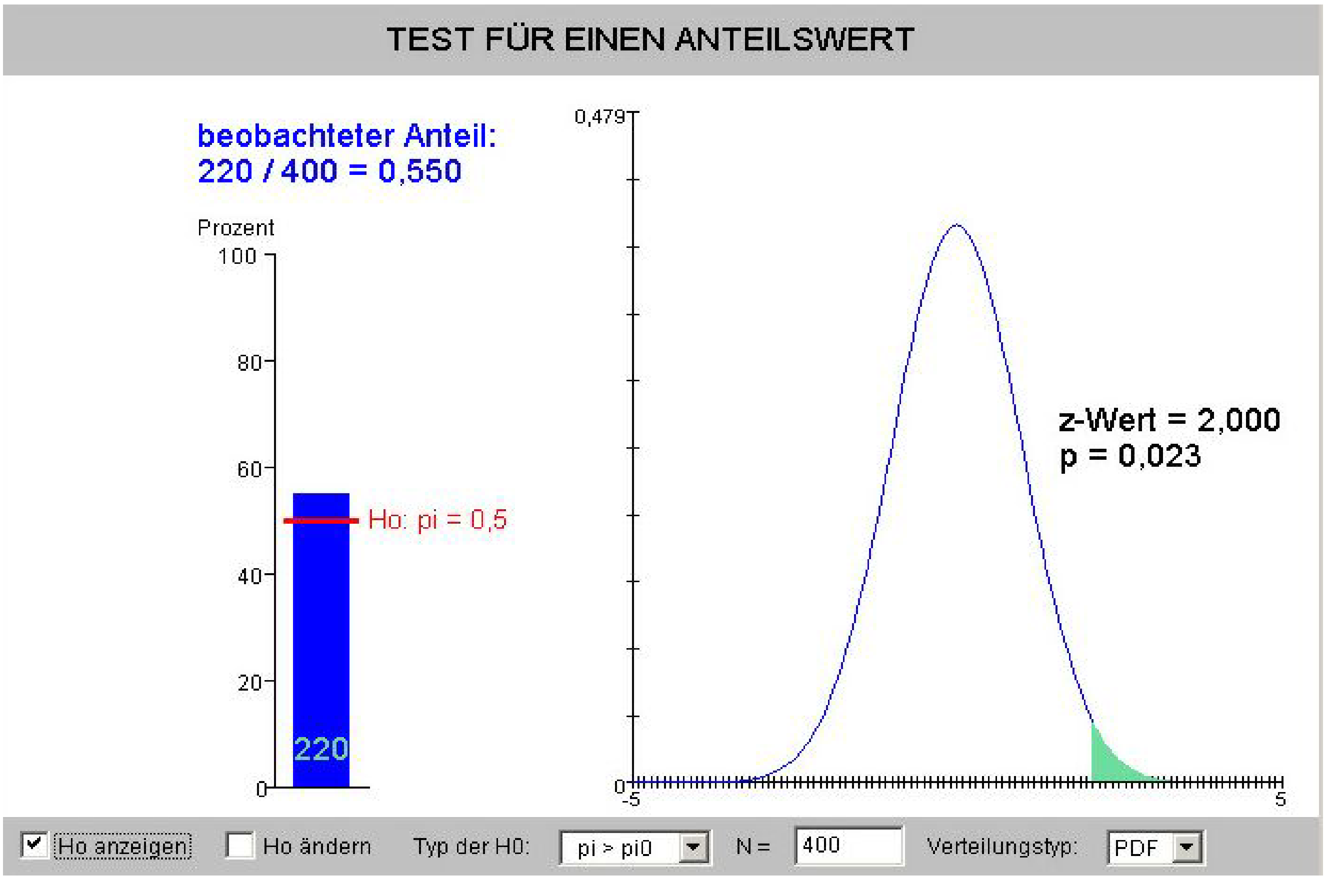
einsetzen

$$z = \frac{r_j - \pi_0}{\sigma} = \frac{0,55 - 0,50}{0,025} = 2$$

was bedeutet nun ein bestimmter z-Wert ?

hier war der Wert z = 2,0

unter Zuhilfenahme der sog. *Standardnormalverteilung* kann man den zugehörigen *p-Wert* ausrechnen



Ergebnis:

- der p-Wert von 0,023 ist kleiner als 0.05, das Ergebnis ist *signifikant*
- die Nullhypothese („Kandidatin X erreicht 50% der Stimmen“) wird zugunsten der einseitigen Alternativhypothese („Kandidatin X erreicht mehr 50% der Stimmen“) verworfen
- wir ziehen den Schluss, dass die Kandidatin eine absolute Stimmenmehrheit erreichen wird

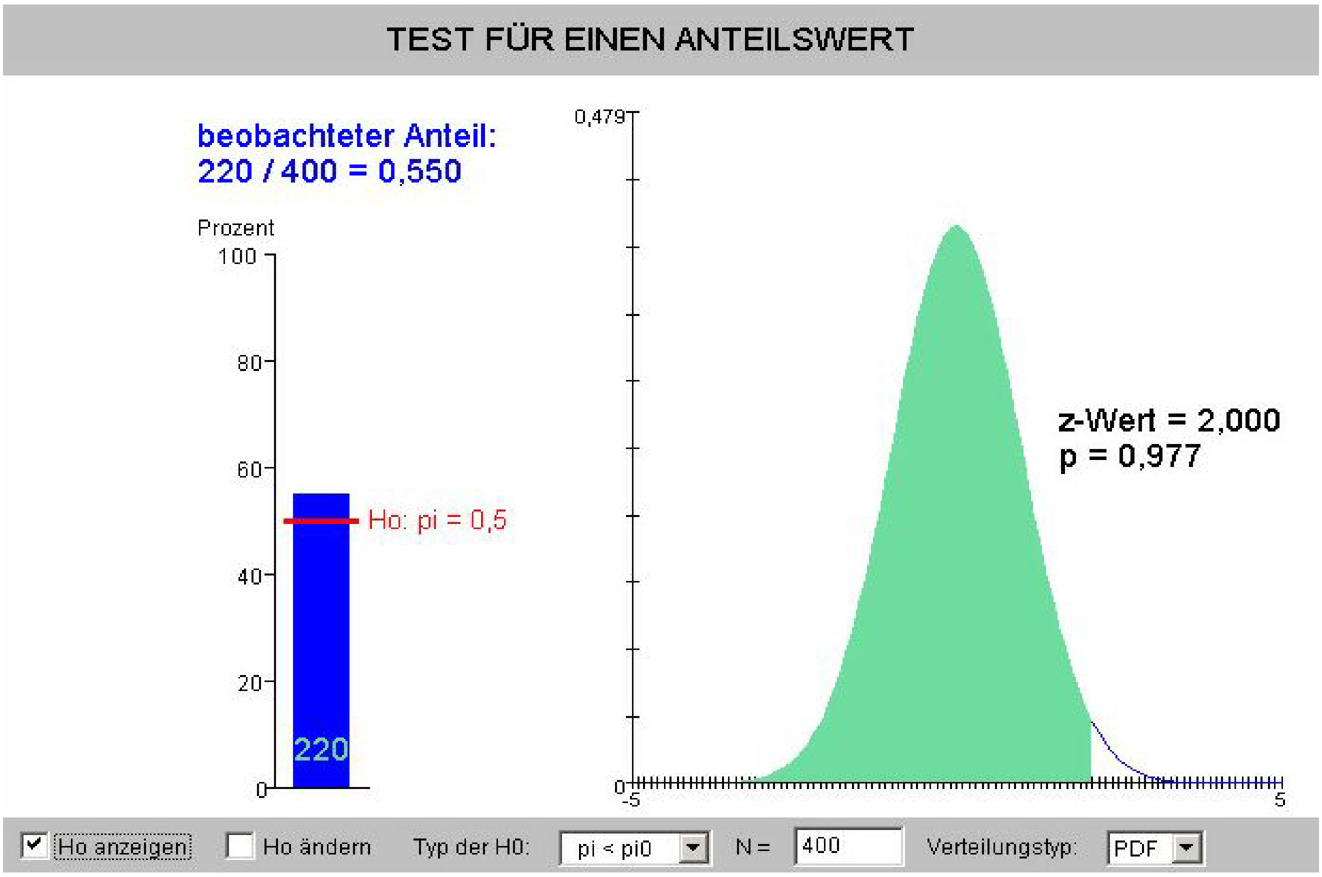
der p-Wert bei verschiedenen Formulierungen der Alternativhypothese

die Beziehung zwischen dem z-Wert und seiner Bedeutung (ausgedrückt durch den p-Wert) hängt immer von der jeweils gewählten Alternativhypothese ab

im obigen Beispiel: Alternativhypothese einseitig formuliert: „Kandidatin X erreicht <u>mehr</u> als 50% der Stimmen“

wie hätte das Ergebnis ausgesehen, wenn die Frage umgekehrt gewesen wäre, nämlich:

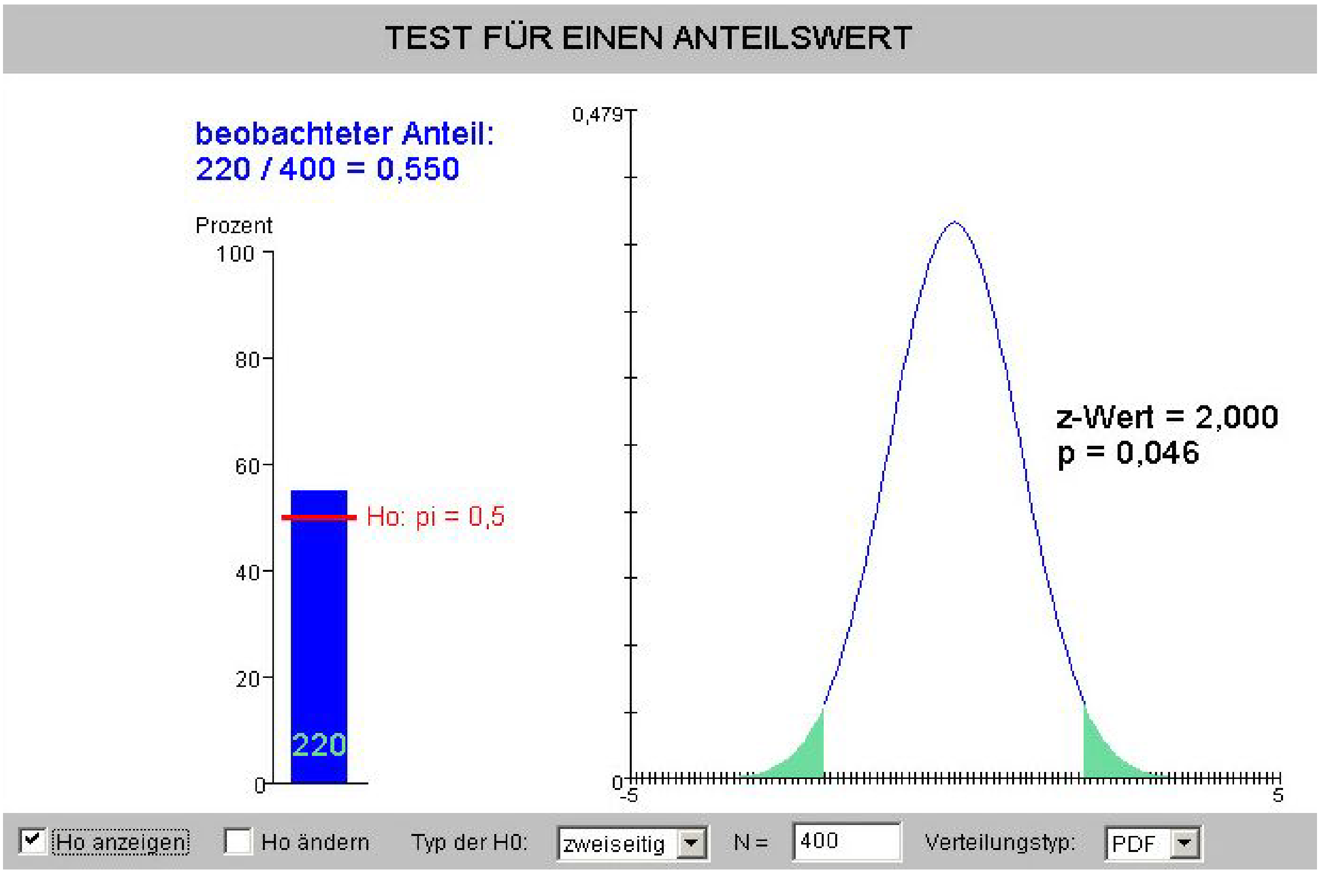
„Kandidatin X erreicht <u>weniger</u> als 50% der Stimmen“



der p-Wert ist jetzt 0,977, das ist sehr viel größer als 0,05 und wir würden die Nullhypothese beibehalten – 220 Befürworter von 400 Befragten sprechen natürlich nicht für die Alternativhypothese, dass die Kandidatin WENIGER als 50 % erreicht wie sieht das bei einer **zweiseitigen Alternativhypothese** aus ?

sie würde lauten:
„Kandidatin X erreicht einen Stimmanteil von ungleich 50%"
(es können also sowohl mehr als auch weniger als 50% sein)

bei gleichem z-Wert von 2,0 wäre der p-Wert jetzt 0,046



Ergebnis:

- der p-Wert von 0,046 ist kleiner als 0,05
- das Ergebnis ist signifikant
- die Nullhypothese wird zugunsten der zweiseitigen Alternativhypothese („Kandidatin X erreicht einen Stimmanteil von ungleich 50%") verworfen

Schlussfolgerungen:

- bei zweiseitigen Tests müssen wir sowohl nach oben als auch nach unten schauen

- der p-Wert beim zweiseitigen Test hat sich gegenüber dem Test mit einseitiger Alternative verdoppelt
  (bzw.)
- bzw. der p-Wert des einseitigen Tests ist nur halb so groß als jener beim zweiseitigen Test

- Achtung: bei der Umrechnung von zweiseitigen und einseitigen p-Werten muss man immer die Richtung, in der die einseitige Alternativhypothese formuliert ist, beachten

Berechnung mit Statistikprogrammen:

sowohl in SPSS als auch in R funktioniert das Berechnen des Tests für einen Anteilswert über die χ^2-Verteilung

R

```
1-sample proportions test without continuity
correction

data:  220 out of 400, null probability 0.5
X-squared = 4, df = 1, p-value = 0.02275
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval:
 0.5088852 1.0000000
sample estimates:
   p
0.55
```

das Ergebnis ist das Gleiche wie schon vorher beschrieben:

- der p-Wert von 0,023 ist kleiner als 0.05, das Ergebnis ist *signifikant*
- die Nullhypothese („Kandidatin X erreicht 50% der Stimmen“) wird zugunsten der einseitigen Alternativhypothese („Kandidatin X erreicht mehr 50% der Stimmen“) verworfen
- wir ziehen den Schluss, dass die Kandidatin eine absolute Stimmenmehrheit erreichen wird

Anmerkung: die Chi-Quadrat Verteilung und die Standardnormalverteilung stehen in einer engen mathematischen Beziehung, daher kann man (wie in R) auch die Chi-Quadrat Verteilung zur Beurteilung solch einer Fragestellung heranziehen

SPSS

in SPSS ist keine eigene Möglichkeit vorgesehen Tests auf Anteile für nur eine einzelne Kategorie zu berechnen

man behilft sich damit, dass man eine zweite Kategorie mitberücksichtigt, und dann einen Chi-Quadrat - Test analog zu Fragestellungen 1 und 2 berechnet

Idee:

- ein kategoriales Merkmal muss immer mindestens 2 Ausprägungen (Kategorien) haben
- es gibt neben der Kategorie „Zustimmung zu Kandidatin X" („Kopf" beim Münzenwerfen) natürlich auch die Kategorie „Ablehnung von Kandidatin X" („Zahl" bzw. „nicht Kopf")

Kandidatin X

	Beobachtetes N	Erwartete Anzahl	Residuum
Zustimmung	220	200,0	20,0
Ablehnung	180	200,0	-20,0
Gesamt	400		

Statistik für Test

	Kandidatin X
Chi-Quadrat	4,000
df	1
Asymptotische Signifikanz	,046

Ergebnis wie oben, aber:
der Chi-Quadrat Test kann nur zweiseitig formuliert werden, daher muss hier der p-Wert halbiert werden, um die Null- bzw einseitige Alternativhypothese richtig beurteilen zu können

händischer Berechnung von FRAGESTELLUNG 3:

1.man berechnet den z-Wert wie oben dargestellt

2.bestimmen des kritischen Werts

(aus Normalverteilungstabelle):

die wichtigsten kritischen Werte sind:

Alternativhypothese	kritischer Wert 95% Sicherheit	kritischer Wert 99% Sicherheit
(a) zweiseitig: $\pi \neq \pi_0$	1.96	2.576
(b) einseitig: $\pi > \pi_0$	1.645	2.326
(c) einseitig: $\pi < \pi_0$	-1.645	-2.326

- Freiheitsgrade gibt es bei der Normalverteilung keine

3. Vergleich des z - Werts mit dem kritischen Wert

man verwerft die Nullyhypothese wenn:

- (a) Absolutbetrag von z-Wert > kritischer Wert
- (b) z-Wert > kritischer Wert
- (c) z-Wert < kritischer Wert

ERGEBNIS für Beispiel:

z-Wert von 2,0 > 1,645,
daher verwirft man die Nullhypothese („der wirkliche Anteil an Stimmen für Kandidatin X ist 50%“) zugunsten der einseitigen Alternativhypothese („der wirkliche Anteil ist größer als 50%“) mit 95% Sicherheit

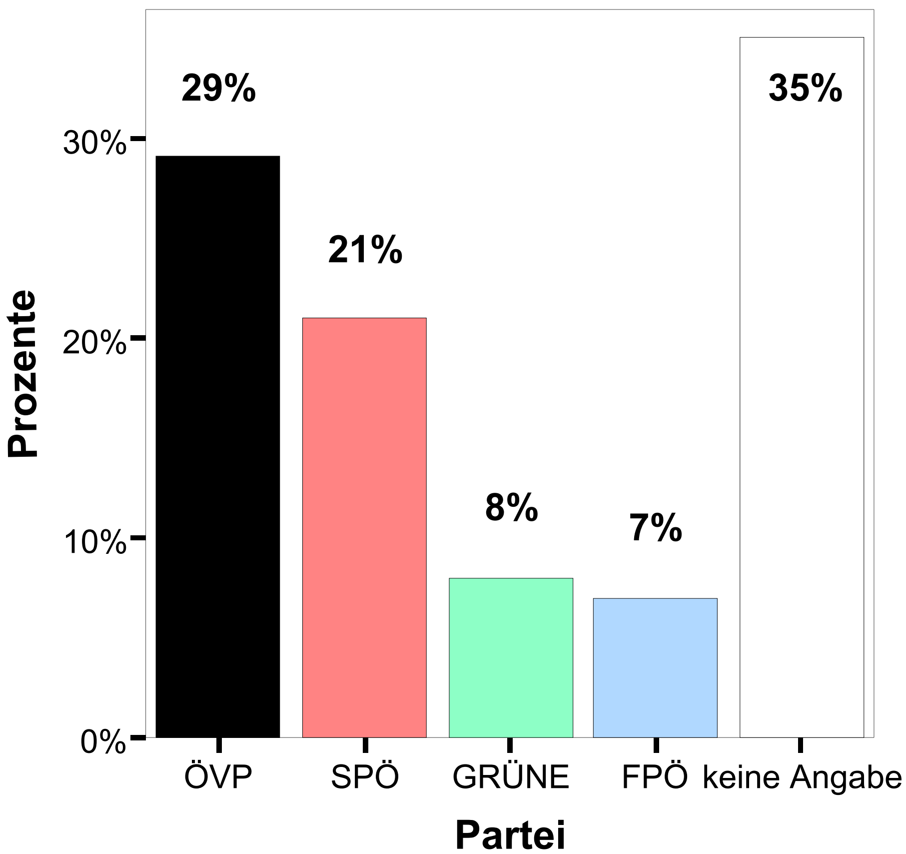
FRAGESTELLUNG 3:

In welchem Bereich kann man den Anteil einer Kategorie in der Grundgesamtheit erwarten ?

Bsp.: Meinungsumfrage zur Parteipräferenz

Telefonumfrage an 500 Personen zu den Nationalratswahlen 2002 (3 Wochen vor der Wahl)

"Sonntagsfrage": Falls am nächsten Sonntag Nationalratswahlen wären, welcher Partei würden Sie am ehesten Ihre Stimme geben ?



Mit welchen Stimmanteilen können die einzelnen Parteien rechnen ?

---

beobachtete Zahlen werden nicht genau mit Wahlergebnis übereinstimmen

weil:
1. noch 3 Wochen bis zur Wahl
2. noch Unentschlossene, "keine" oder "falsche" Antwort, Nicht- und Ungültig-Wähler,...
3. Daten beruhen nur auf Stichprobe:
   selbst wenn Stichprobe repräsentativ, alle Antworten ehrlich und keine Änderungen,
   nicht exakt dieses Ergebnis bei Wahl zu erwarten

bei anderen Stichproben - andere Ergebnisse

Daten aus Stichprobe schwanken zufällig um tatsächliche Werte in Population (gleiche Idee wie bei Chi-Quadrat Tests)

wie groß können diese Schwankungen sein ?

zunächst zur Berechnung Ausschluss jener, die keine Angabe gemacht haben

	Anzahl	Prozent	Prozent für Partei
ÖVP	145	29	44,6
SPÖ	105	21	32,3
GRÜNE	40	8	12,3
FPÖ	35	7	10,8
keine Angabe	175	35	-

hier: Berechnung der „Prozent für Partei“ (also unter Weglassen der Kategorie „keine Angabe“) simples Dividieren durch Anzahl „gültiger Angaben“
Meinungsforschungsinstitute verwenden ausgefeiltere Methoden

Bestimmung von "Schwankungsbreiten":

Größe der Schwankungsbreite c

$$c = 1{,}96 \cdot \sqrt{\frac{r_j(1-r_j)}{n}}$$

r_j ... relative Häufigkeit für Kategorie j
n ... Stichprobengröße
1,96 ... Faktor zur Berechnung von c mit 95%-Sicherheit

Konfidenzintervalle (KI)

Berechung von zwei Grenzen so, dass es sehr plausibel ist, dass tatsächlicher Wert innerhalb dieser Grenzen liegt

$$\text{KI: } \left[r_j - c \;\; ; \;\; r_j + c \right]$$

Beispiel: Konfidenzintervall für SPÖ

$$c = 1{,}96 \cdot \sqrt{\frac{0{,}32 \cdot (1-0{,}32)}{365}} = 1{,}96 \cdot 0{,}026 = 0{,}0508$$

95%-Konfidenzintervall: [0,272 ; 0,374]

es war zu erwarten, dass der Anteil der SPÖ zwischen 27,2 und 37,4 Prozent liegen würde

Nationalratswahl 1999, Gesamtergebnis

	Stimmen	Mandate	%	95% Konfidenzintervall	
ÖVP	2.076.831	79	42,3	39,2	50,0
SPÖ	1.792.499	69	36,51	27,2	37,4
FPÖ	491.328	18	10,01	8,7	16,0
Grüne	465.021	17	9,47	7,3	14,2
LIF	48.085	0	0,98		
KPÖ	27.567	0	0,56		
sonstige	8.354	0	0,17		

Interpretation von Konfidenzintervallen:

- Ziehen sehr vieler Stichproben
- jeweils Berechnen der relativen Häufigkeit und eines Konfidenzintervalles für die relative Häufigkeit der Kategorie j
- dann: 95% dieser Konfidenzintervalle würden die tatsächliche relative Häufigkeit in der Population (die ja unbekannt ist) einschließen
- für eine einzelne Stichprobe kann man keine exakten Angaben machen

Beispiel:

wenn man 15 verschiedene Stichproben erheben würde, könnte man die folgenden 15 Konfidenzintervalle für SPÖ erhalten (die strichlierte Linie zeigt den tatsächlichen Anteil für SPÖ, 36.51%)