Begriffe		Qualitative
Statistik	Entwicklung und Anwendung von Methoden zur Erhebung,	Faktoren
	Aufbereitung, Analyse und Interpretation von Daten	Faktorstufe
Beschreibende	Vollständige Kenntnis über das Untersuchungsobjekt	
Statistik		
Schliessende	Für Untersuchung liegend die Daten des zu untersuchenden Objekts	
Statistik	nur zum Teil vor.	
Hypothese	Eine Hypothese ist eine Aussage deren Gültigkeit man für möglich	Komplexitä
	hält, die aber nicht bewiesen oder verifiziert ist.	Kompliziert
Nullhypothese	Die Nullhypothese H0 ist eine Aussage von der angenommen wird,	Symbole
	dass sie stimmt.	h_i
Alternativhypothe	Die Alternativhypothese H1 beschreibt eine Annahme, sie ist also das	f_i
se Fehler 1. Art	Gegenteil der Nullhypothese. Fehlerhaftes Verwerfen einer Hypothese	II————
(alpha)	rememates verwerien einer nypothese	H_i
Fehler 2. Art	Fehlerhaftes Annehmen einer Hypothese	
(beta)	rememates Amemmen enter hypothese	$ F_i $
Zufällige Fehler	Nicht reproduzierbar	l
Systematische	Reproduzierbar (können vermieden werden, unterliegt keinen	μ
Fehler	grossen Schwankungen)	9
Validierung	Mache ich das Richtige (Überprüfung des Modells)	σ^2
Verifikation	Mache ich es richtig (Verifiziertes Modell kann nicht valide sein)	⊩
Merkmalsträger	Der Gegenstand der statistischen Untersuchung	σ
Abgrenzungsmerk	Sachlich: wer/was ist unter Merkmalsträger zu verstehen	\bar{x}
mal (sachlich,	z.B. Wer gilt als "Mitarbeiter" eines Unternehmens	$\boldsymbol{\iota}$
räumlich, zeitlich	Räumlich: Räumliche Grenzen, in denen der Merkmalsträger liegen	H_0
,	muss	
	z.B. ein Bürogebäude eines Konzerns	H_1
	Zeitlich: Zeitpunkt oder Zeitraum, an der ein Merkmalsträger	
	"existieren" muss, um Teil der Grundgesamtheit zu sein ≠ zum	ω
	Zeitpunkt der Messung/Erhebung!	Ω
Grundgesamtheit	Die Menge aller Merkmalsträger die für eine Untersuchung in Frage	
	kommen	σ -Alge
Merkmal	Eigenschaften der Merkmalsträger die von Interesse sind	4
Merkmalswert	Der Wert der Beobachtung / Messung	\mathbb{A}
Primärstatistik	Die Daten wurden genau für diesen Zweck erhoben (teuer)	
Sekundärstatistik	Existierende Daten wobei es ungewiss ist, wie die Daten erhoben	n
V II I I	wurden. (günstig)	N
Vollerhebung	Befragung aller Merkmalsträger (Kosten und Umfang meist zu gross)	IV
Teilerhebung	Befragung der essentiellen Merkmalsträger (wird meist gemacht)	Δ
Diskrete Funktion	Mit Lücken	\overline{R}
Stetige Funktion	Ohne Lücken	
Formale	Zahlenmässig begründete Abhängigkeit	Ableitungs
Abhängigkeit	Let den West eines Manhards bevordt 1910 (* 1. M.)	x^a
Sachliche	Ist der Wert eines Merkmals kausal/ursachlich für den Wert eines	
Abhängigkeit	zweiten Merkmals abhängig	\boldsymbol{x}
Menge Tupel	Ungeordnet, ohne Redundanzen Geordnet, mit Redundanzen	
Tupei Zufallsexperiment	Ein Experiment welches beliebig oft durchgeführt werden kann und	x^2
zuiaiisexperiment	das Ergebnis komplett vom Zufall abhängig ist	x^{-}
Disjunkt	Keine gemeinsame Teilmenge	-
Zielgrösse	Beschreiben die Grösse, die man optimieren möchte	$\frac{1}{x}$
Einflussgrösse	Sind Grössen welche die Zielgrösse beeinflussen. Es wird zwischen	\boldsymbol{x}
L143361 033C	Streu,- und Störgrössen unterschieden. Man unterscheidet zwischen	
	Steuergrössen und Störgrössen	\sqrt{x}
Steuergrössen	Eine einstellbare Grösse (die man auch für eine gewisse Zeit halten	_
<u> </u>	kann)	e^{x}
Störgrössen	Eine Grösse deren Wert man nicht beeinflussen kann	C
Faktoren	Aus allen Einflussgrössen werden die wesentlichen/relevant Faktoren	Linearitäts
	genannt. Es wird zwischen Quantitativen und Qualitativen Faktoren	
	unterschieden:	Produktreg
Quantitative	Quantitative Faktoren: Die Werte sind auf einer Ordinalskala	1 TOGUNUE
	beschreiben	

Qualitative			tive Faktoren: Die Werte sind auf einer Nominalskala							
Faktoren		beschri								
Faktorstufen			rte die ein Faktor in einem Versuch annehmen soll, werden							
			cufen genannt. De Faktor nicht genau gemessen werden, so sollte der Abstand							
			n Faktor nicht genau gemessen werden, so sollte der Abstand torstufen mindestens							
			/arianz sein	310113						
Komplexität			nzahl an Faktore	en						
Komplizierth	eit			ierig zu beschrei	bende Faktoren					
Symbole										
h_i		Absolut	lute Häufigkeit (Anzahl)							
f_i		Relative	e Häufigkeit (Ant	teil)						
$\overline{H_i}$		Kumulie	erte absolute Hä	iufigkeit						
$\overline{F_i}$		Kumulie	erte relative Häu	ıfigkeit						
μ		Mittelw	vert							
σ^2		Varianz								
σ		Standar	rdabweichung							
\bar{x}			etisches Mittel (Durchschnitt)						
H_0			pothese							
H_1			ativhypothese							
ω			ntarereignis Teilmenge der Ergebnismenge ismenge / Ergebnisraum (Menge aller möglicher Ausgänge							
Ω		eines Zı	Lufallsexperiments) Nahrscheinlichkeiten							
σ -Alge	bra		t aus allen möglichen Ergebniskombinationen (Potenzmenge							
\mathcal{A}			gebnismenge)							
n		Anzahl	Messungen / Stichprobenumfang							
N		Grösse	der Grundgesamtheit							
Δ		Mittlere	e absolute Abweichung							
R		Spannw	veite							
Ableitungsre	geln					1				
x^a	a ·	x^{a-1}	1	0	tan(x)	$\frac{1}{\cos^2(x)}$				
x	1		a^x	$\ln(a) \cdot a^x$	tan(x)	$1 + \tan^2(x)$				
x^2	22	c	ln(x)	$\frac{1}{x}$	arcsin(x)	$\frac{1}{\sqrt{1-x^2}}$				
$\frac{1}{x}$	_	$\frac{1}{x^2}$	$\log_b(x)$	$\frac{1}{\ln(b)\cdot x}$	arccos(x)	$-\frac{1}{\sqrt{1-x^2}}$				
\sqrt{x}	$\frac{1}{2}$	$\frac{1}{\sqrt{x}}$	sin(x)	$\cos(x)$	$\operatorname{arctan}(x)$	$\frac{1}{1+x^2}$				
e^{x}	e^a	_	cos(x)	$-\sin(x)$						
Linearitätsre	gel		$\frac{d}{dx}(f(x))$	(x) + g(x)	= f'(x)	+ q'(x)				
			$dx \left(\int \left(\omega \right) \right)$) ' 3 (~)	, , (~)	· 3 (~)				

 $\frac{d}{dx}(f(x) \cdot g(x)) = f'(x) \cdot g(x) + f(x) \cdot g'(x)$

Qualitative Eaktoren: Die Werte sind auf einer Nominalskala

	Produktregel mit Konstante c	$\frac{d}{dx}\left(c\cdot f\left(x\right)\right) = c\cdot f'(x)$
nd	Kettenregel	$\frac{d}{dx}\left(f\left(g\left(x\right)\right)\right) = f'(g(x)) \cdot g'(x)$
	Quotientenregel	$d f(x) $ $f'(x) \cdot g(x) - f(x) \cdot g'(x)$
		$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{f(x)g(x) - f(x)g(x)}{(g(x))^2}$

Experimente

Zyklischer Prozess des Experimentierens nach Shewhard (Plan \rightarrow Do \rightarrow Check \rightarrow Act)

- 1. Hypothese aufstellen
- 2. Experiment durchführen
- 3. Hypothese überprüfen
- 4. Hypothese/Modell gegebenenfalls anpassen

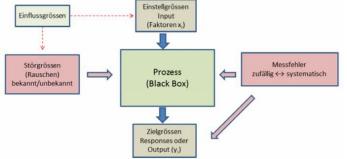
Hindernisse für den Erkenntnisgewinn:

- 1. Komplexität: Hohe Anzahl an Faktoren
- 2. Kompliziertheit: Unbekannte oder schwierig zu beschreibende Faktoren
- 3. Rauschen/Dynamik: Unterschiedliche Ergebnisse bei gleichen Faktoren

Experimente werden immer nach einem bestimmen Schema durchgeführt:

- Ausgangssituation beschreiben
- . Untersuchungsziele festlegen / Zielgrössen definieren
- 3. Faktoren auswählen und gewichten
- 4. Versuchsplanung erstellen
- Versuche durchführen
- . Ergebnisse auswerten und Vertrauensintervalle bestimmen
- . Ergebnisse interpretieren und Massnahmen ableiten
- 8. Überprüfen der «Verbesserungen»





DoE: Design of Experiment

Wie sind Experimente zu planen, damit mit möglichst wenigen Einzelexperimenten der Zusammenhang zwischen Einflussfaktoren und Zielgrössen möglichst genau ermittelt werden können.

Vorgehen

- 1. Ausgangssituation spezifizieren / Problem beschreiben / Ziel definieren
 - a. Kunde und dessen Bedürfnisse definieren
 - b. Liegen bereits Daten vor
 - c. Welche Probleme müssen gelöst werden
 - Welche Ressourcen (Zeit und Geld) stehen zur Verfügung (Kosten/Nutzen Analyse)
 - e. Betroffene Gruppen und deren Beziehung untereinander listen (Wiederstände, Supporter, Wissensträger)
- Zielgrösse beschreiben: Dabei möglichst alle Grössen sammeln und diese dann auf die wichtigen Reduzieren

Einfluss-Zielgrössen-Matrix

- . Für jede Zielgrösse eine Spalte anlegen
- In der ersten Spalte alle Einflussgrössen sammeln und in Einflussgrössen und Steuergrössen unterteilen
- Für jede Einflussgrösse das vorhandenen Wissen über Grösse und Einfluss auf jede Zielgrösse sammeln (z.B. stark, schwach, kein, linear, nicht linear)

Fehlerrechnung:								
-	l benötigt um den Bereich	abzuschätzen, in der	nen der tatsächliche					
-	Wahrscheinlichkeit liegt.							
Zufällige Fehler	<u> </u>	Nicht reproduzierbar						
Systematische Fehler	Reproduzierbar							
Absoluter Fehler Δt	Bei Summen und Diffe		die absoluten Fehler					
	(gleiche Einheit wie Me							
		Der absolute Fehler kann mithilfe der relativen Fehlers						
	berechnet werden →	Δt = relativer Fehle	er * Wert					
Relativer Fehler	$\frac{\Delta t}{t}$ wobei Δt = absolute Fehler und t = Messwert Bei Produkten und Quotienten addieren sich die relativen Fehler (einheitenlos \rightarrow %)							
	Faktor verbessert werd Fehleranteil)	Mit Hilfe des relativen Fehler lässt sich gut Abschätzen, welcher Faktor verbessert werden sollte. (der mit dem grösseren Fehleranteil)						
	multipliziert werden. Z	Bei Potenzen kann der relative Fehler mit dem Exponenten multipliziert werden. Z.B $r^{2*}\pi \rightarrow 2*f_r$ da sich der relative Fehler bei Multiplikationen addiert ($r^*r^*\pi \rightarrow f_r + f_r = 2*f_r$)						
Nennwert die Fehlerang								
Mindestens	Letzte Stelle des Messy	vertes + 1 Stelle (auf	halbe gerundet)					
Höchstens	Letzte Stelle des Messy	vertes (auf 0.3/0.4 g	erundet)					
Beispiel	Gemessener Wert (t)	∆t von	∆t bis					
	15.32s	± 0.005s	± 0.04s					
	15.3s	± 0.05s	± 0.4s					
	15.320s	± 0.0005s	± 0.004s					
Beispiele (Masseinheiten	beachten)							
Schätzung eines Rechted	cks							
Länge wird abgelesen:		28.15 - 22.35 cm = 5	.8 cm					
Fehlerschätzung beim A	blesen:	± 0.05 cm						
**	Fehler addieren sich (links i	und rechts)						
Länge des Rechtecks:		5.8 ± 0.1 cm						
Berechnung der Fläche								
Breite des Rechtecks geg	geben mit	0.9 ± 0.1 cm						

 $\Delta B = 0.1/0.9 = 11.1\%$

 $\Delta L = 0.1/5.8$ cm = 1.7%

 $A = 0.9 * 5.8 = 5.2 \pm 0.7 \text{ cm} 2$

0.128 * 5.22 cm2 = 0.668 cm2

Breite des Rechtecks gegeben mit

Berechnung des relativen Fehlers Relativer Fehler der Länge:

//Multiplikation: relative Fehler addieren sich $\Delta B + \Delta L = 12.8\%$

Fläche A = L * B absoluter Fehler der Fläche Bei Messgeräten ist der relative Fehler nicht auf den gemessenen Wert, sondern auf

Messbereich bezogen. Tipp für Rechnungen mit Kombinationen von +/- und */:

Resultat berechnen ohne beachten der Fehlerangaben

2. Resultat berechnen unter Nutzen der Maximalwerte 3. Fehler Δ ergibt sich durch die Differenz von 1. und 2.

Diagramme	
Balkendiagramm	Y-Achse: Häufigkeit und X-Achse: Balken pro Klasse
Histogramm	Der Balken geht über die gesamte Klassenbreite
Polygonzug	Verbinden der Balken mit einer Linie, wobei jeweils der rechte Ecken verbunden wird. Beim Balkendiagramm wird die Mitte
	genommen.

Skalen		Dichte:
Wir arbeiten hauptsächli	ch mit metrischen Skala (Intervall und Verhältnis)	Wenn die Klassenbreiten
Nominalskala	Sind zwei Einheiten gleich oder ungleich? = / ≠	Ist die Klassenbreite gleic
(qualitativ)	Enthält Namen die gleichgewertet werden	Häufigkeiten berechnen
	Geschlecht: {Feminin, Maskulin} Ortsname: {Berlin, Rom, Bern, Paris} Familienstand: {verheiratet, ledig, geschieden, verwitwet}	Häufigkeit für einen bestimmten Wert (z.B. 45)
Ordinalskala / Rangskala (qualitativ)	Es lässt sich zusätzlich eine Ordnung herstellen $=/\neq \mathrm{und} >/<$ Die Werte sind nicht mehr gleichgewichtet, sondern intensitätsmässig geordnet (in Klassen)	Arithmetisches Mittel der Gesamtheit
Intervaliskala	Schulnote: {sehr gut, gut, genügend, schlecht} Umfragen: {Trifft zu, Trifft eher zu, Trifft eher nicht zu, Trifft nicht zu} Qualitätsstufe: {Standard, Business, First Class} Es lässt sich zusätzlich eine Aussage über die Abstände machen	
(metrische Skala / Kardinalskala) (quantitativ)	= / ≠ und > / < und +/− Es kann der einfache Abstand (Intervall) gemessen werden. Hat keinen absoluten Nullpunkt Temperatur: {-12,, 0,, 42}	Modus (häufigster Wert)
Verhältnisskala (metrische Skala / Kardinalskala) (quantitativ)	Uhrzeit: {20:00, 0:00, 10:00} Es lässt sich zusätzlich eine Aussage über das Verhältnis machen = / ≠ und > / < und +/- und ·/: Hat einen absoluten Nullpunkt, deshalb Vergleich Aussagen möglich. Negative Werte sind nicht möglich. Besitzt das höchste Informationsniveau! Umsatz: {OM, 1M, 2M, 3M,}	Median, 1. Quartil, etc.
	Alter: {0,1,,40,, gut, gut, genügend, schlecht} Gewicht: {0kg, 50kg, 60kg,,80kg,120kg }	Beispiel für klassifizierte I

Häufigkeitsverteilung

n Gesamtzahl der Merkmalsträger (z.B Glühbirnen) v Anzahl verschiedene Merkmalsträger (Klassen) $H_{i} = \sum_{a=1}^{i} h_{a}$ $F_{i} = \sum_{a=1}^{i} f_{a}$ $N = \sum_{i=1}^{v} h_{i}$ h_i absolute Häufigkeit H_i kumulierte absolute Häufigkeit f_i relative Häufigkeit F_i kumulierte relative Häufigkeit d_i Klassendichte D_i kumulierte Klassendichte x Bestimmter Wert innerhalb der Klasse

 \acute{x} Klassenmittelwert

 x^u Untere Klassengrenze

 x^{o} Obere Klassengrenze ${\cal F}_{i-1}/h_{m-1}$ Häufigkeit der vorherigen Klasse

 F_{i+1}/h_{m+1} Häufigkeit der nächsten Klasse

hte	:
nn	dia

Wenn die **Klassenbreiten unterschiedlich gross** sind muss mit Dichte d_i gerechnet werden. Ist die Klassenbreite gleich gross bzw. die Häufigkeit unklassifiziert, so ist $d_i = h_i$

 $F(x) = F_{i-1} + \frac{x - x^u}{x^o - x^u} \cdot (F_i - F_{i-1})$ bestimmten Wert (z.B. 45) Unklassifiziert: Arithmetisches Mittel $\bar{x} = \frac{1}{n} \sum_{i=1}^{c} x_i \cdot h_i = \sum_{i=1}^{c} x_i \cdot f_i$ der Gesamtheit

Der Modus ist immer in der Klasse mit der höchsten Dichte. Modus $Mo = x^{u} + \frac{d_{i} - d_{i-1}}{(d_{i} - d_{i-1}) + (d_{i} - d_{i+1})} \cdot (x^{o} - x^{u})$ (häufigster Wert) Median.

 $Me/Q = x^u + \frac{\frac{n}{t} - \overline{H_{i-1}}}{h} \cdot (x^o - x^u)$ 1. Quartil etc. 1. Bestimmen von t a. t = In wie viele Teile die Gesamtheit unterteilt ist (2=Median, 4=1,Quartil, etc.)

> 2. Klasse finden, in welcher der Median/Quartil liegt a. $(n/t) < H_i \rightarrow i = Klasse$

Beispiel für klassifizierte Häufigkeit

Welcher Anteil der Mitarbeiter ist < 45 Jahre alt? 0 bis 40 10 10 0.2 0.2 40 bis 50 15 25 0.3 0.5 25 50 bis 65 50 0.5

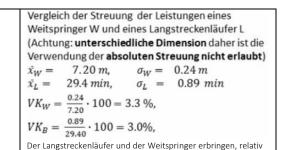
 $f = \frac{45 - 40}{50 - 40} * (0.5 - 0.2) = 0.15$ $F(x < 45) = 0.2 + 0.15 = 0.35 \, bzw. 35\%$ Anteil < 45 Jahre = 35% Anteil > 45 Jahre = 100% - 35% = 75%

Rechnung mit Dichte:

d1 = 10/40 = 0.25d2 = 15/10 = 1.5d3 = 25/15 = 1.7Modusklasse ist also Klasse 3, da sie die grösste Dichte hat.

		11
Lagemasse / Lageparamet		I
Arithmetisches Mittel /	Der Klassische Durchschnitt: Man addiert alle Messwerte und	I
Mittelwert	dividiert durch die Anzahl Messwerte	l
	$_{1}$ n	l
	$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$	l
	$x = \frac{1}{n} \sum_{i=1}^{n} x_i$	l
	i=1	l
Harmonisches Mittel	Ist zur Berechnung des Durchschnitts einzusetzen wenn das	1
	Merkmal aus einem Bruch hervorgeht.	l
	n	
	$\sum h_i$	
	- $i=1$	
	$\bar{x_h} = \frac{\sum\limits_{i=1}^{n} h_i}{\sum\limits_{i=1}^{n} \frac{h_i}{x_i}}$	
	$\sum \frac{h_i}{}$	
	$\sum_{i=1}^{\infty} x_i$	
		ł
	Bsp. Auf einer Strecke von 2 Kilomer benötigt ein Fahrzeug auf	
	der Hinfahrt 10km/h und auf der Rückfahrt 30km/h	
	$\overline{MH} = \frac{\sum_{i=1}^{v} h_i}{\sum_{i=1}^{v} \frac{h_i}{x_{i}}} = \frac{(2+2) \ km}{\frac{2 \ km}{10 \ km/_{h}} + \frac{2 \ km}{30 \ km/_{h}}} = 15 \ \frac{km}{h}$	I
	$MH = \frac{1}{2 km} = \frac{15}{2 km} = 15 \frac{1}{h}$	I
	$\sum_{i=1}^{i} \frac{1}{x_{i}} \frac{10 km_{/b}}{10 km_{/b}} \frac{1}{30 km_{/b}}$	
Geometrisches Mittel	Ist die n-te Wurzel aus dem Produkt aller beobachteten	1
ocometrisches Mittel	Merkmalswerte	I
	ivierkillaiswerte	
	$\bar{x_g} = \sqrt[n]{\prod_{i=1}^n x_i}$	
	$ \bar{x_a} = n \prod x_i$	
	\ i=1	I
	Verwendete man immer dann, wenn man Mittelwerte aus	
	aufeinander aufbauenden Wachstumsfaktoren	I
	berechnen will. Wichtig beim Geometrischen Mittel ist, dass	I
	man nicht den Prozentsatz selbst sondern die einzelnen	I
	x :	I
	Faktoren (Brüche) $\overline{100}$ einsetzt.	J
Modus	Gibt den Wert an, der am häufigsten vorkommt	J
Median	Die Mitte in einem geordneten Datensatz. Gibt es eine gerade	I
	Anzahl Elemente wird einfach der Schnitt der beiden in der	
	Mitte liegenden Werte genommen.	ı
Quantil	Unterteilt die Gesamtheit in 2gleich grosse Teile	J
Quartil	Unterteilt die Gesamtheit in 4 gleich grosse Teile	l
Dezil	Unterteilt die Gesamtheit in 10 gleich grosse Teile	J
Perzentil	Unterteilt die Gesamtheit in 100 gleich grosse Teile	l
Streumasse / Streuparam	eter	l
Die Varianz und Standard	abweichung werden in der Praxis für die Streuung eingesetzt.	1
Spannweite	Die Differenz zwischen dem grössten und kleinsten	1
	beobachteten Merkmal	
Zentraler	Die Differenz zwischen dem ersten und dritten Quartil	1
Quartilsabstand /	$Q_3 - Q_1$	I
Interquartilsabstand	$\mathbf{q}_3 - \mathbf{q}_1$	I
		l
	T 1,00 -	l
	25%	I
	0,75	I
		I
	50% 0,50	I
		ľ
	0,25	١
	173% / V	1
	25%	ı
	23% 0,00 Q ₁ Q ₂ Q ₃ x	
	0,00	

Mittlere absolute	Der Durchschnitt der Summe aller Differenzen zum Mittelwert
Abweichung	15
	1 10
	$\bar{d} = \frac{1}{n} \sum_{i=1}^n x_i - \bar{x} \cdot h_i$ Die Betragsstriche der mittleren absoluten Abweichung ist
	unvorteilhaft (Fallunterscheidung). Deshalb arbeitet man viel
	öfter mit der Varianz
Varianz	Durch das Quadrieren wird der Varianzwert sehr gross, weshalb man eher mit der Standardabweichung rechnet.
	$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot h_i$
	t=1 Rsp. vereinfacht: n
	$\sigma^2 = \sum_{i=1}^n x_i^2 \cdot h_i - \bar{x}^2$
	i=1 oder n
	$\sigma^2 = \sum_{i=1}^n (x_i - \bar{x})^2 * f_i$
Standardabweichung	Arithmetisches Mittel der Abweichung vom Mittelwert der
	Gesamtheit \bar{x}
	$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot h_i}$
	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$
-	57.5 126.56 3164.06
	$\bar{x} = (20 * 10 + 45 * 15 + 57.5 * 25) * \frac{1}{50} = 46.25$
	$\sigma^2 = \frac{1}{50} * (6890.63 + 23.44 + 3164.06) = 201.56 \rightarrow \sigma = 14.2$
Variationskoeffizient (%)	Die Standardabweichung im Verhältnis zum arithmetischen Mittel
	$v = \frac{\sigma}{\bar{x}} \cdot 100$
	1. Standardabweichung: CHF 0.85
	2. Durchschnittlicher Preis für einen Espresso: CHF 4.25
	$v = \frac{\sigma}{\bar{x}} = \frac{0.85}{4.25} = 0.2 \cdot 100 = 20\%$



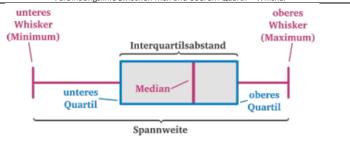
Boxplot

Der Boxplot vermittelt einen schnellen Eindruck, in welchem Bereich die Daten liegen und wie sie sich in diesem Bereich aufteilen.

 Der geordnete Datensatz wird in 4 Abschnitte aufgeteilt, die etwa gleich viele Werte umfassen

gesehen eine nahezu gleichmässige Leistung

- a. Minimum
- b. Maximum
- c. Median
- d. Beim unteren Quartil (Min 25% aller Messwerte kleiner/gleich und Max 75% aller Messwerte (grösser/gleich)
- Beim oberen Quartil Max 25% aller Messwerte kleiner/gleich und Min 75% aller Messwerte (grösser / gleich)
- 2. Die oberen und unteren Enden der Quartile mit Strichen verbinden = B
- Verbindungslinie zwischen Min und unterem Quartil sowie eine Verbindungslinie zwischen Max und oberem Quartil = Whisker



Zeitreihen

X-Achse = Zeit / Y-Achse = Merkmalswerte → Punktdiagramm

Gleitender Mittelwert

x: 1 2 3

Ziel: Glättung der Zeitreihe/Kurve, in dem die hohen und niedrigen Werte gegeneinander Abgeglichen werden.

Man berechnet immer das arithmetische Mittel über eine Auswahl aller Messwerte und verschiebt diese Auswahl kontinuierlich nach vorne. Aus den neuen Messwerten (arithemtische Mittel) wird anschliessend eine neue Zeitreieh erstellt.

- 1	^1				-	-			-		
	yi		5		8	7		6	9	11	9
	\bar{y}_i		-		6,67	7,0	0	7,33	8,67	9,67	-
	5 8	8	7	6	9	11	9	\rightarrow	$\overline{y}_2 =$	$\frac{5+8+7}{3}$ =	6,67
5	8		7	6	9	11	9	\rightarrow	$\bar{y}_3 =$	$\frac{8+7+6}{3}$ =	7,00
	y ↑ 111 – 9 –								\		
	7-		/	<u> </u>	>		1				
	5										

22 21 20 19		•	z u	wisc						t das				
20			(0	ohne		e Ar	nalys			ntige				en)
13			d		eide					veich lein	_			
								Ach		Zusa inea		enh	ang	3
Regressionsfunktion für line	eare Zusamm	enhän	ge											
Der gesuchte Wert sollte a	uf der Y-Achs	e, der	geg	gebe	n au	ıf de	r X-	Achs	<u>e</u> lie	egen				
Regressionsgerade $\widehat{\mathcal{Y}}$	$\hat{y} = a_1$ Beschreibt of Merkmal X u	den Zu:	sam	nmei		-				n una	abhá	ingi	gen	
Regressionsparameter	$a_1 = \bar{y}$ Gibt den ter des Merkma	ndenzi	elle	n W				kmal	sΥä	an, v	venr	n de	r W	'ert
Regressionsparameter	$b_1 = \sum_{i=1}^{\infty}$ Gibt als Stein des Merkmals X	gungsr als Y te	nas nde	s an enzie	, um ell är	wie nder	vie , we	nn c					· W	ert
	$\bar{x} = \frac{\sum_{j} \bar{y}}{\bar{y}} = \frac{\sum_{j} \bar{y}}{\bar{y}}$	$\frac{x_i}{\frac{y_i}{n}}$												
Beispiel	12 Studenten gingen im letzten Semester neben dem Studium einer Erwerbstätigkeit nach. In der nachfolgenden Tabelle sind der zeitliche Aufwand (Std./Woche) für die Erwerbstätigkeit X und der zeitliche Aufwand (Std./Woche) für das Studium Y angegeben.													
	Student		A	В	C	D	E	F	G	Н	I	J	K	L
	Erwerbstätig	keit	1	2	2	3	3	4	5	6	8	12	15	23
	Studium		39	37	36	40	36	37	34	36	33	33	32	27
	Ein Student anhand der sein Studiun Zusammenh Regressions	vorlieg n aufbi nang zv	geno ring visc	den gen k	Date Cann	en e . Es	rmit best	teln eht	, wie	eviel linea	Zeit arer	er	für	ne
	Arbeitstabelle für die Regressionsgeraden $\widehat{\mathcal{Y}}$ und $\widehat{m{\mathcal{X}}}$													

	Student	xi	yi	x _i y _i	x _i ²	y _i ²					
	A	1	39	39	1	1.521					
	В	2	37	74	4	1.369					
	C	2	36	72	4	1.296					
	D	3	40	120	9	1.600					
	E	3	36	108	9	1.296					
	F	4	37	148	16	1.369					
	G	5	34	170	25	1.156					
	H	6	36	216 264	36 64	1.296 1.089					
	J	8 12	33	396	144	1.089					
	K	15	32	480	225	1.024					
	L	23	27	621	529	729					
	Summe	84	420	2.708	1.066	14.834	W				
	Σν.	9.4		Σν.	420		T				
	$\bar{x} = \frac{\Delta x_i}{2}$	= = = :	\overline{y}	$=\frac{\sum y_i}{n}=$	$\frac{420}{10} = 35$	5	Zi				
	n	12		n	12		2				
	. 1 Σ	$x_i y_i - n\bar{x}$	\bar{y} 2'	708-12*7*	35	_					
	$p_1 = \frac{1}{2}$	$\sum x_i^2 - n\bar{x}^2$	= -1/	708-12*7* 066-12*7*	$\frac{33}{7} = -0.4$.9	EI				
				5 – (-0.49			<u> </u>				
				5 - (-0.4	9) 7 – 3	0.43	E				
Resultat für $\widehat{\mathcal{Y}}$	\hat{y} : -0.49	x + 38.4	.3								
-	Auf den Fall	6 Studen	Erwerbst	tätigkeit ang	ewendet fo	lgt:	Eı				
	-0.49*6	+ 38.43	= 35.4	49 → der St	udent kanr	ı	Sy				
				d von 35.49							
				r tatsächlich			١.				
		U		och weitere			U				
				auf die Höh			0				
	hahen	_					(
	$h2 = \frac{\Sigma}{2}$	$b2 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum y_i^2 - n\bar{y}^2} = \frac{2'708 - 12*7*35}{14'834 - 12*35*35} = -1.73$									
							Lá				
				- (-1.73))*35 = 6	7.55] [
Resultat für $\widehat{\mathcal{X}}$	-1.73y -						(g				
	_	-		eibt die Tend							
		-		n Zeitaufwar		udium					
				rwerbstätig							
				anfallende Ze	eitaufwand	fur die					
D : 1 C 1	Erwerbstäti	_					4				
Beispiel für den				e 200 Besch		m					
Zusammenhang mit				gruppen aus							
Häufigkeitsverteilung				ie Verteilung	-		U				
			n Bescha	iftigten auf d	iie Tarifgru	ppen	Er				
	ersehen we		1 10								
	Tarifgruppe	43 32 36		124							
	Weiblich Männlich	19 18 23		76							
	Summe	62 50 59		200							
				nhang zwisc	han dan Mi	arkmalan					
				enzugehörigl		cikillaicii					
Resultat				tlich mehr F		länner.	1				
				figkeitsverte		-					
				enen Tarifgri		achen,					
				iten mit folge							
	berechnet v	verden.	_	_							
	_	$H_i(Ta)$	rif grup	$pe) * H_i(Ge)$	eschlecht)						
	J J	i =		npe) * H _i (Ge							
	Bsp.										
	$f_i = \frac{62*124}{200}$	= <u>38.43</u> →	relative	Häufigkeit,	weiblich in	G1					
	200										

Weiblich 38.44 (43) 31 (32) 36.58 (36) 17.98 (13) 124 Männlich 23.56 (19) 19 (18) 22.42 (23) 11.02 (16) 76 Summe 62 50 59 29 200			
Summe 62 50 59 29 200			
Alk also reliable as 118 of also become a locus as a configuration.			
Mit der relativen Häufigkeitswerten kann man nun feststellen,			

Wahrscheinlichkeitsrechnu	ung / Stochastik		
Menge	Ungeordnet, ohne Redundanzen		
Tupel	Geordnet, mit Redundanzen		
Zufallsexperiment	Ein Experiment welches beliebig oft durchgeführt werden kann und das Ergebnis komplett vom Zufall abhängig ist (z.B Werfen eines Würfels)		
Elementarereignis ω	Ist ein möglicher Ausgang des Zufallsexperiments, wobei zwei Elementarereignisse sich immer gegenseitig ausschliessen.		
Ergebnismenge Ω	Umfasst alle möglichen Elementarereignisse eines Zufallsexperiments. Z.B {1,2,3,4,5,6}		
Ereignis	Eine Teilmenge der Ergebnismenge. Z.B {2,4,6}		
System der Ereignisse ${\cal A}$	Bei einem Zufallsvorgang gemessene Ereignisse, bilden zusammen ein System von Ereignissen. Dieses weist Eigenschaften auf, welche es ermöglichen Relation (Durchschnitt, Vereinigung, etc.) zu bilden.		
Unmögliches Ereignis	Die leere Menge		
Disjunkte Ereignisse	A und B sind disjunkt, wenn sie keine gemeinsame Teilmenge besitzen.		
Laplace Experiment (gut Fälle / alle Fälle)	Ein Experiment bei dem jedes Ergebnis <u>dieselbe</u> Wahrscheinlichkeit hat und die Ergebnismenge endlich/abzählbar ist. $P(A) = \frac{ A }{ \Omega } = \frac{\text{Anzahl der für das Ereignis A günstigen Ergebnisse}}{\text{Anzahl aller möglichten Ergebnisse}}$		
	Wie gross ist die Wahrscheinlichkeit im Lotto (6 aus 49) genau drei Richtige anzukreuzen? $P\{3 \ richtige\} = \frac{\binom{6}{3}*\binom{43}{3}}{\binom{49}{6}} = 0.0176$		
Unabhängige Ereignisse	Zwei Ereignisse A und B sind voneinander unabhängig wenn gilt: $W(A) = W(A B) \qquad bzw.$		
	$W(A) = W(A \overline{B})$ bzw. $W(A B) = W(A \overline{B})$ Beispiel: Es wurden folgende Wahrscheinlichkeiten berechnet: P(A) = 0.65; $P(A B) = 0.75$; \Rightarrow Da $P(A) \neq P(A B)$ sind die beiden Ereignisse A und B abhängig.		

Additionssatz	Die Wahrscheinlichkeit das A oder B eintritt
	A und B sind vereinbar
	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
	$ (A \cup B) = I(A) + I(B) = I(A \cap B) $ 2. A und B sind unvereinbar
	$P(A \cup B) = P(A) + P(B)$
Multiplikationssatz	Die Wahrscheinlichkeit, dass A und B eintritt
	1. Sind A und B abhängig
	$P(A \cap B) = P(A) \cdot P(B A)$
	oder
	$P(A \cap B) = P(B) \cdot P(A B)$
	2. Sind A und B unabhängig
	$P(A \cap B) = P(A) \cdot P(B)$
	Es gibt 5 Kugel, 4 rote und 1 weise. Wie gross ist die
	Wahrscheinlichkeit, dass man ohne zurücklegen zwei rote zieht? → Da man nicht zurücklegt, ist die Wahrscheinlichkeit der
	weiteren Kugeln <u>abhängig</u> von der von den vorherig gezogenen.
	1. Zug 2. Zug Ereignis W(Ereignis)
	$\begin{array}{ccc} R2 R1 & R1 \cap R2 & \frac{12}{20} \end{array}$
	W_{2/R_1} RIOW2 $\frac{4}{20}$
	W_1 $W_1 \cap R_2$ $\frac{4}{20}$
	All
	W_2/W_1 $W_1 \cap W_2$ $\frac{0}{20}$
Bedingte	Die Wahrscheinlichkeit von A unter der Voraussetzung von B
Wahrscheinlichkeit	$P(A B) = \frac{P(A \cap B)}{P(B)}$
	Verliebt verliebt 0.4 0.6
	P(S L) = 0.8 $P(-S L) = 0.2$ $P(S -L) = 0.7$
	versalzen O.K versalzen O.K
	$P(S \cap L) = P(-S \cap L) = P(-S \cap L) = P(-S \cap L) = P(-S \cap L) = 0.32$ 0.08 0.18 0.42
Komplementäre	Die Wahrscheinlichkeit, dass A nicht eintritt
Wahrscheinlichkeit	$P(\bar{A}) = 1 - P(A)$

Korrelation		Permutationen	Bei identischen Elementen werden diese in Klassen
Die Korrelation ist eine Kennzahl für den Zusammenhang zwischen mehreren Streudiagrammen		mit Wiederholung/Zurückle	zusammengefasst. Es gibt dabei k Klassen mit jeweils n _k identischen Elementen
Formale Abhängigkeit Zahlenmässig begründete Abhängigkeit		gan	
Sachliche Abhängigkeit	Ist der Wert eines Merkmals kausal/ursachlich für den Wert		$p(n_k) = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$
	eines zweiten Merkmals abhängig		$n_1! \cdot n_2! \cdot \ldots \cdot n_k!$
Kovarianz	$1\sum_{i=1}^{n}$	Beispiel:	Von einem 6-stelligen Zahlenschloss weiss man, dass es sich mit
	$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$		einer bestimmten Folge der Ziffern 1, 1, 4, 4, 4 und 8 öffnen lässt. Wie viele Versuche sind maximal notwendig um das
	n = 1		Zahlenschloss zu öffnen?
	Merkmalswertkombinationen (x_i, y_i)		Gegeben sind n=6 Ziffern, die in k=3 Klassen von untereinander
Kovarianz bei	Bei Stichproben verwendet man eine korrigierte Varianz, wobei		gleich Ziffern zerfallen. Die Klasse «1» enthält n1 = 2 Elemente,
Stichproben	man nicht nur n sondern durch n $-$ 1 teilt $Kovarianz$ σ_{xy}		die Klasse «4» n2 = 3 und die Klasse «8» n3 = 1 Element.
Korrelationskoeffizient Lineare Abhängigkeit	$r_{xy} = \frac{Kovarianz}{Standardabweichunq_x \cdot Standardabweichunq_y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$		$p_{2,3,1}(6) = \frac{6!}{2! \cdot 3! \cdot 1!} = \frac{720}{12} = 60 \text{ Permutationen}$
	Es resultiert immer ein Wert r zwischen -1 und 1:		2. 0. 1. 12
	a) Je mehr der Wert bei -1 liegt, desto mehr ähneln die	Bei der Kombinatorik geht e eine Ordnung zu bringen.	es darum, aus n Elementen, k auszuwählen und anschliessend in
	Punkte im Streudiagramm einer Gerade mit negativer	Binomialkoeffizient	Aus n Optionen, k auswählen
	Steigung (stark linear abhängig) b) Je mehr der Wert bei O liegt, desto grösser ist die		
	Streuung der Punkte (linear unabhängig)		$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$
	c) Je mehr der Wert bei +1 liegt, desto mehr ähneln die		
	Punkte im Streudiagramm einer Gerade mit positiver	Kombinationen ohne	Im Rechner ist das die Funktion nCr(n,k)
Permutationen und Komb	Steigung (stark linear abhängig)	Wiederholung/Zurücklegen	Anzahl Möglichkeiten = $\frac{n!}{(n-k)!}$
T CITICALIONEIT AND ROTTE			Alizam Mognerikerten $-\frac{(n-k)!}{(n-k)!}$
	Ist jedes vorgegebene Element genau einmal anzuordnen	(Mit Beachtung der Anordnung)	Im Rechner ist das die nPr(n,k) Funktion
	anzuoranen	Anordinang)	Aus 5 Bewerber soll eine Rangliste der ersten 3 Plätze
	JA NEIN		gemacht werden. Wie viele verschiedene Listen sind
J	Text		möglich?
Sind die vorgegeb	Darf ein vorgegebenes		$V_3(5) = \frac{5!}{(5-3)!} = \frac{5!}{2!} = 60$
Elemente alle verschieden	Element wiederholt ausgewählt werden	Kombinationen ohne	Ist gleich dem Binomialkoeffizienten
		Wiederholung/Zurücklegen	
JA N	EINŢ		Anzahl Möglichkeiten = $\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$
		(ohne Beachtung der Anordnung)	Anzani Mognenkerten = $\binom{k}{k} = \frac{(n-k)! \cdot k!}{(n-k)! \cdot k!}$
	mutationen mit judgment judgme	Anordinang	
			Im Rechner ist das die Funktion nCr(n,k)
			Bsp. Lotto Beim Lotto müssen aus 49 Zahlen 6 Zahlen ausgewählt
			werden. Wie viel Tipps sind möglich?
lst die Anordnung der Ist die Anordnung der Elemente von Elemente von Bedeutung Bedeutung			
			$\binom{49}{6} = \frac{49!}{(49-6)!*6!} = 13'983'816$
	JA NEIN JA NEIN	Kombinationen mit Wiederholung/Zurücklegen	Anzahl Möglichkeiten = n^k
Kombination mit Kombination ohne Kombination ohne		(14)	In einem Einkaufsladen gibt es unterschiedlich bemalte
Wiederholung Beachtung	und mit Wiederholung und Wiederholung und mit Wiederholung und	(Mit Beachtung der Anordnung)	Vasen zu kaufen. Der Kunde möchte 3 Vasen für seinen
Anordnu	ng Anordnung Anordnung Anordnung	Anordinang)	Garten kaufen. Wie viele Möglichkeiten hat er, die Vasen in seinem Garten anzuordnent.
Permutationen			$V_3^W(4) = 4^3 = 64$
Permutationen	Anzahl Möglichkeiten n Objekte anzuordnen = Fakultät	Kombinationen mit	(n+k-1)! $(k+n-1)!$
ohne	$p(n) = n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 2 \cdot 1$	Wiederholung/Zurücklegen	Anzahl Möglichkeiten = $\frac{(n+k-1)!}{k! \cdot (n-1)!} = \binom{k+n-1}{k}$
Wiederholung/Zurückle gen (Merke: 0!=1)	Bei einem Ziffernschloss muss man eine 5-stellige Zahl		In einem Rat werden 3 Sitze neu vergeben, es bewerben
Ben (Merke, 0:-1)	einstellen, die aus den Ziffern 0-9 gebildet wird. Wie viele	(ohne Beachtung der Anordnung)	sich 6 Verbände darauf. Die wiederholte Auswahl eines
	Kombinationen gibt es?	/ inor unung)	Verbandes ist möglich. Wie viele mögliche Sitzverteilungen
	$10^5 = 100'000$		gibt es? k= 3
	Eine Maschine muss vier Aufträge A, B, C, D nacheinander		
	abarbeiten. Wie viel Anordnungen sind möglich: 4!=24		$K_3^W(6) = {6+3-1 \choose 3} = {8 \choose 3} = {8! \over (8-3)! * 3!} = 56$
	→ für den ersten Platz gibt es 4 Möglichkeiten, für den zweiten 3, usw. → n!		
	5) 45 m 2 m		

Zufallsvariablen

Zusammenhang Zufallsvariable und Merkmal

Zufallsvariable X	Merkmal X
Realisation x	Merkmalswert x
Wahrscheinlichkeit	relative Häufigkeit
Wahrscheinlichkeitsfunktion	einfache relative Häufigkeitsverteilung
Verteilungsfunktion	kumulierte relative Häufigkeitsverteilung
Erwartungswert	arithmetisches Mittel
Varianz	Varianz

Realisation

Wert der Zufallsvariable für ein Ereignis

z.B. Im Monopoly ist die Summe der Augenzahlen zweier Würfel entscheidend, wie weit ein Spieler vorrücken darf:

Zufallsvariable = Augensumme Realisationen = {2,3,4, ..., 12}

Eine Zufallsvariable hat für jedes Ereignis eine bestimmte Wahrscheinlichkeit, dass das Ereignis eintreffen kann.

$$f(x) = P(X = x) = \begin{cases} Wahrscheinlichkeit_1 & \text{für x=Ereignis} \\ Wahrscheinlichkeit_2 & \text{für x=Ereignis} \end{cases}$$

Diskrete Massenfunktion	Kann mit einem Stabdiagramm veranschaulicht werden (Ordinate (Y) = Wahrscheinlichkeit, Abszisse (X) = Ereigniswerte) Wahrscheinlichkeit kann direkt abgelesen werden Hat Lücken und nur positive Werte Die Summe aller Einzelwahrscheinlichkeiten entspricht 1 = Fläche unter dem Graphen	
Stetige Dichtefunktion /	- Auf der X-Achse sind unendliche viele Werte	
Kontinuierliche	 Die Summe aller Einzelwahrscheinlichkeiten 	
Verteilungsfunktion	entspricht 1 = Fläche unter dem Graphen	
Erwartungswert	$E(X) = \sum_{i=1}^{n} x_i \cdot P(x_i)$	
Varianz	$VAR(X) = \sum_{i=1}^{n} (x_i - E(X))^2 \cdot P(x_i)$	

Beispiele:

Ein Zufallsvorgang besteht im dreimaligen Werfen einer Münze. Entscheidend ist die Anzahl an Wappen.

Geben Sie die Wahrscheinlichkeits- und Verteilungsfunktion an.

Zufallsvariable: Ar Realisation: 0.

Anzahl Wappen

	0,1,2,3	
хi	f(xi)	F(xi)
0	0.125	0.125
1	0.375	0.500
2	0.375	0.875
3	0.125	1.000

a) Berechnen Sie den Erwartungswert.

0 * 0.125 + 1 * 0.375 + 2 * 0.375 + 3 * 0.12 = 1.5 Wappen

b) Berechnen Sie die Varianz und die Standardabweichung $\sigma^2 = (0-1.5)^2*0.125 + \underbrace{(1-1.5)^2*0.375 + (2-1.5)^2*0.375 + (3-1.5)^2*0.125}$

 $\sigma = \sqrt{0.75} = 0.866$

Stichproben	
Zufallsstichprobe	Aus der Grundgesamtheit werden Elemente zufällig
	ausgewählt
Einfache Stichprobe	Die Elemente der Stichprobe haben alle die gleiche
	Wahrscheinlichkeit

Geschichtete Stichprobe	Ist es möglich, Elemente mit gleichen Eigenschaften in
	Gruppen einzuteilen ist es sinnvoller, Teilstichproben pro
	Gruppe/Schicht zu nehmen, um genauer Aussagen über
	die Gesamtheit zu machen

Schätzverfahren

Ist der Mittelwert, Standardabweichung und die Verteilungsfunktion nicht bekannt müssen diese mit Hilfe von Schätzfunktionen geschätzt werden. Ziel dabei ist es, von einer Stichprobe auf die Grundgesamtheit zu schliessen und dabei den Fehler einer falschen Schätzung zu minimieren.

Stichprobe auf die Grundgesamtheit zu schliessen und dabei den Fehler einer falschen			
Schätzung zu minimieren.			
Punktschätzung			
Schätzfunktion für den	1 _ n		
Mittelwert	$\bar{x} = -\sum x_i$		
	$x - \frac{1}{n} \sum_{i=1}^{n} x_i$		
	i=1		
Schätzfunktion für Varianz	$\frac{1}{n}$		
und Standardabweichung	$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$		
	$n-1 \stackrel{\longrightarrow}{\underset{i=1}{\longleftarrow}}$		
	1 n		
	$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$		
	$n-1 \stackrel{\frown}{=} (w_i w)$		
	V 1=1		

Varianz σ^2 Stichprobe	bekannt	unbekannt
mit Zurücklegen	$\sigma_{\overline{X}}^2 = \frac{\sigma^2}{n}$	$\hat{\sigma} \frac{2}{X} = \frac{s^2}{n}$
$\frac{n}{N} < 0.05$	$\sigma_{\overline{X}}^2 \approx \frac{\sigma^2}{n}$	$\hat{\sigma}^2_{\overline{X}} \approx \frac{s^2}{n}$
ohne Zurücklegen $\frac{n}{N} \ge 0,05$	$\sigma \frac{2}{X} = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$	$\hat{\sigma} \frac{2}{X} = \frac{s^2}{n} \cdot \frac{N-n}{N}$

ACHTUNG: Evtl. Wurzel ziehen! Wir arbeiten meist mit der Standardabweichung

Varianz σ^2 Stichprobe	bekannt	unbekannt
mit Zurücklegen	$\sigma_{p}^{2} = \frac{\Theta \cdot (1 - \Theta)}{n}$	$\hat{\sigma}_P^2 = \frac{P \cdot (1 - P)}{n}$
$\frac{n}{N}$ < 0,05	$\sigma_p^2 \approx \frac{\Theta \cdot (1 - \Theta)}{n}$	$\hat{\sigma}_{P}^{2} \approx \frac{P \cdot (1-P)}{n}$
	$\sigma_{p}^{2} = \frac{\Theta \cdot (1 - \Theta)}{n} \cdot \frac{N - n}{N - 1}$	$\hat{\sigma}_{P}^{2} = \frac{P \cdot (1 - P)}{n} \cdot \frac{N - n}{N}$

Varianz σ^2 Verteilung des Merkmals X	bekannt	unbekannt		
bekannt und normalverteilt	\overline{X} ist normalverteilt	\overline{X} ist t-verteilt mit $k = n - 1$ Freiheitsgraden Wenn $n > 30$: \overline{X} ist approximativ normalverteilt		
bekannt und nicht normalverteilt (n > 30) unbekannt (n > 30)	X ist approx	\overline{X} ist approximativ normalverteilt		

P = Wahrscheinlichkeit (Anteilswerte)

Intervallschätzung

Konfidenzintervall für den Mittelwert

Beispiel: Bekanntheitsgrad (unbekannte Varianz)

Ein Chemieunternehmen möchte den Bekanntheitsgrad eines von ihm hergestellten Waschmittels in Erfahrung bringen. Dazu werden 400 Personen zufällig ausgewählt und befragt. Das Waschmittel war 30 % der Befragten zumindest namentlich bekannt. Erstellung des zentralen 95%-Konfidenzintervalls für O.

Schritt 1: Festlegung der Verteilungsform von P

Wie: $n * P * (1 - P) > 9 = 400 * 0.3 * 0.7 = 84 > 9 \rightarrow wahr$, also approximativ normalverteilt **Schritt 2**: Festlegung der Varianz / Standartabweichung von P

Resultat:
$$\hat{\sigma}_P = \sqrt{\frac{P*(1-P)}{n}} = \sqrt{\frac{0.3*0.7}{400}} = 0.02$$

Schritt 3: Ermittlung des Quantilswertes z → Gemäss geg. Konfidenzintervall (Unterscheidung einseitig/Beidseitig(zentral))

Desultat aus Taballa

Resultat aus Tabelle

Schritt 4: Berechnung des maximalen Schätzfehlers

Resultat: $z * \hat{\sigma}_P = 1.96 * 0.02 = 0.04$

Schritt 5: Ermittlung der Konfidenzgrenze

Resultat: $W(0.30 - 0.04 \le \Theta \le 0.30 + 0.04) = 0.95$

 $W(0.26 \le \Theta \le 0.34) = 0.95$

Der Bekanntheitsgrad in der Grundgesamtheit wird mit einer Wahrscheinlichkeit von 95% vom Intervall [26%; 34%] überdeckt.



Genauigkeit erhöhen:

- Konfidenzgrenze behalten, Umfang n erhöhen
- Umfang n behalten, Konzidenzniveau senken

Konfidenzintervall für beidseitig begrenzt:

$$W(\mu \ \text{-}\ z \cdot \sigma_{\overline{X}} \, \leq \, \overline{X} \, \leq \, \mu \, + \, z \cdot \sigma_{\overline{X}}) \ = \, 1 \, \text{-} \, \alpha$$

$\sigma_{\bar{v}} =$ findet sich mit Tabelle links!

 $1-\alpha$ gibt die Wahrscheinlichkeit an, dass sich die Zufallsvariable/Stichprobenfunktion innerhalb des Intervalls befindet. (Konfidenzintervall)

Der tägliche Kaffeekonsum in einem Bürg

Der tagliche kaffeekonsum in einem Buro:								
١	xi	f(xi)	Wie hoch ist die Wahrscheinlichkeit, dass der Stichprobenmittelwert					
ı	1	20	bei n=100 im Intervall (2.3;2.5) liegt?					
	2	30	$\bar{x} = \frac{1 * 20 + 2 * 30 + 3 * 40 + 4 * 10}{2 \times 30 + 3 * 40 + 4 * 10} = 2.4$					
١	3	40	100 Varianz-Berechnung siehe "Streuparameter" → 0.84					
١	4	10						
			$\sigma_{\overline{X}} = \sqrt{\frac{0.84}{100}} = 0.0917$					
1	-	0.5	2.4					

$$z = \frac{\overline{x} - \mu}{\sigma_{\overline{X}}} = \frac{2.5 - 2.4}{0.0917} = 1,0905$$

Intervall von z=-1,09 bis +1,09 => $0.8621-0.1379 = 0.7242 \rightarrow 72.42\%$

notwendiger Stichprobenumfang n beim Konfidenzintervall für das arithmetische Mittel

In diesem Fall wird gefordert, dass die Schätzung ein vorgegebenes Mindestmass an Genauigkeit e besitzt und dass diese Mindestgenauigkeit mit einer vorgegebenen Konfidenz bzw. Sicherheit erzielt wird.

gegeben: Konfidenz, Genauigkeit (e) → e = \overline{x} - μ

gesucht: Stichprobenumfang n

Beispiel: Zuckerabfüllung

gegeben: Konfidenz z = 1.96 Genauigkeit e = 0.2 g Standartabweichung σ = 1.2 $n \ge \frac{z^2 * \sigma^2}{e^2} = \frac{1.96^2 * 1.2^2}{0.2^2} = 138.3$

Es müssen 139 Packungen entnommen werden, um die gewünschte Genauigkeit zu erzielen (wegen n > 30 ist im Falle einer beliebig verteilten Grundgesamtheit die Approximation durch die Normalverteilung zulässig).

<u>Beispiel</u>: Eine Molkerei liefert an eine Lebensmittelkette 40'000 Flaschen Milch mit 1000ml Soll-Füllmenge. Die Stichproben haben eine durchschnittliche Füllmenge von 1000.25ml. Aufgrund von zahlreichen Kontrollen weiss man, dass die Ist-Füllmenge normalverteilt mit einer Streuung von σ = 1.2ml ist. Wie viele Flaschen Milch müssen der Lieferung

entnommen werden, wenn folgende Dinge gegeben sind.

gegeben: z= 1.96 (=95% Konf. Int beidseitig) e=0.25ml ($\bar{x} - \mu$)

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{e}{\frac{\sigma}{\sqrt{n}}} \Rightarrow$$
 nach n auflösen = 88.51 \Rightarrow 89 Stichproben

Konfidenzintervall für die Varianz

Es wird die folgende Schätzfunktion verwendet: $s^2 = \frac{1}{x-1} * \sum (x_i - \bar{x})^2$

Voraussetzungen für eine erwartungstreue Schätzung: das Merkmal X ist in der Grundgesamtheit normalverteilt und die Entnahme erfolgt mit Zurücklegen.

-Zweiseitiges Konfidenzintervall

Formel

$$W\left(\frac{(n-1)*s^2}{y_{_1-\frac{\alpha}{2}}} \le \sigma^2 \le \frac{(n-1)*s^2}{y_{\frac{\alpha}{2}}}\right) = 1 - \alpha$$
 (Konfidenzintervall)

 $\alpha = Irrtumswahrscheinlichkeit$

n =Stichprobenumfang/Freiheitsgrade

y = Mit k = n-1 und α z.B. 95% Konf. Int $\rightarrow \alpha = 0.975$ $(y_{1-\frac{\alpha}{2}})$ und $\alpha = 0.025$ $(y_{\frac{\alpha}{2}})$, in der

Chi² Tabelle → y herausfinden

-Einseitiges Konfidenzintervall (nach oben begrenzt)

Formel:

$$W\left(\sigma^2 \le \frac{(n-1)*s^2}{y_{\alpha, k=n-1}}\right) = 1 - \alpha$$
 (Konfidenzintervall)

 $\alpha = Irrtumswahrscheinlichkeit \\$

n = Stichprobenumfang/Freiheitsgrade

v = Mit

1							
Testverfahren							
Signifikanzniveau	Das Signifikanzniveau wird meist bei 5% angesetzt. Ist der Wert						
	kleiner wie 5% wird angenommen, dass ein Ergebnis signifikant ist.						
Hypothese /	Die Nullhypothese H0 ist eine Aussage von der angenommen wird,						
Nullhypothese H0	dass sie stimmt.						
Alternativhypothese	Die Alternativhypothese H1 beschreibt eine Annahme, sie ist also						
H1	das Gegenteil der Nullhypothese.						
Fehler 1. Art (alpha)	Fehlerhaftes Verwerfen einer Hypothese						
Fehler 2. Art (beta)	Fehlerhaftes Annehmen einer Hypothese						
Parametertest	Man möchte wissen, ob der angegebene Benzinverbrauch eines						
	Autos eingehalten wird. μ=10l/100km, σ=1l/100km						
	Es werden 25 Autos getestet. Dabei kommt der Mittelwert						
	10.2L/100km zustande. Liegt das Ergebnis im Bereich statistischer						
	Schwankungen, wenn 1-α=0.95						
	Annahme: Normalverteilter						
	Strichprobenmittelwert $\mu_0 - z \frac{\sigma}{\ln z} \le \bar{x} \le \mu_0 + z \frac{\sigma}{\ln n}$						
	z für zweiseitige Tests: 1.96						
	Intervall:						
	$10 - 1.96 \frac{1}{\sqrt{25}} \le \overline{x} \le 10 - 1.96 \frac{1}{\sqrt{25}}$						
	9.61≤ \overline{x} ≤ 10.39 → Nullhypothese annehmen!						
Anteilswert	Schritt: Wähle die Signifikanzzahl α und bestimme daraus die Werte für Z aus (
(unbekannte	Tabell (1-α)						
Wahrscheinlichkeit)	Schritt: Berechne die Annahmegrenzen zu						
wani schemiichkeit)	$1. \ \ c_u = p_0 - z \sqrt{\frac{p_0(1-p_0)}{n}} \text{ oder } c_u = p_0 - z \sqrt{\frac{p_0(1-p_0)}{n}} \sqrt{\frac{N-n}{N-1}} \text{ und } \\$						
	$2. c_o = p_0 + z \sqrt{\frac{p_0(1-p_0)}{n}} \qquad \text{oder } c_o = p_0 + z \sqrt{\frac{p_0(1-p_0)}{n}} \sqrt{\frac{N-n}{N-1}}$						
	3. Man berechne den Anteil $\overline{p} = \frac{k}{n}$						
	4. Fällt \overline{p}° in den Annahmebereich: $c_{u} \leq \overline{p}^{\circ} \leq c_{o}$						
	wird die Hypothese angenommen, sonst						
	abgelehnt						

Allgemeine Aussagen

Je grösser das Konvidenzniveau, desto breiter das Konvidenzintervall

Zum Testen einer Hypothese ist eine neue Stichprobe nötig

Man begeht keine Fehler, wenn man die Nullhypothese verwirft, wenn sie falsch ist

Der Beta-Fehler wird grösser, je knäpper die Nullhypothese falsch ist

Man kann bei einem Hypothesentest nur dann eine relativ sichere Aussage machen, wenn man die Hypothese verwirft

Je grösser der Stichprobenumfang, desto <u>schmaler</u> wird die Dichtefunktion des Stichprobenmittelwertes

Die Länge der Konfidenzintervalls variiert mit der Stichprobe

Berechnet man für 100 verschiedene Stichproben Konfidenzintervalle für μ zum Niveau 0.9, so sollen etwa 90 Konfidenzintervalle den wahren Wert von μ enthalten

Ist die Varianz einer Grundgesamtheit bekannt, so verwendet man bei der Berechnung von Konfidenzintervallen für den Erwartungswert μ Quantile der Standartnormalverteilung

Die Konfidenzintervalle für den Erwartungswert einer Zufallsvariable sind <u>nicht</u> nur für normalverteilte Grundgesamtheiten berechenbar

Bei einer endlichen Ergebnismenge kann man alle möglichen Wahrscheinlichkeiten berechnen, wenn man die Wahrscheinlichkeiten aller Elementarereignisse kennt

Bei einem klassischen Hypothesentest wählt man im Allgemeinen nur die Wahrscheinlichkeit des alpha-Fehlers klein

Wenn die Nullhypothese nur knapp falsch ist, ist allgemein die Wahrscheinlichkeit sehr gross eine falsche Entscheidung zu treffen

Man begeht keinen Fehler wenn man die Nullhypothese verwirft wenn sie falsch ist

Beim klassischen Signifikanztest muss man damit rechne, dass der beta-Fehler sehr gross wird, wenn die Nullhypothse nur knapp falsch ist

Statistische Aussagen über eine Grundgesamtheit sind auch dann möglich, wenn $\underline{\text{nicht}}$ alle Elemente der Grundgesamtheit beobachtet werden.

Die Brenndauer einer Glühbirne ist eine Zufallsvariable.

Falls X eine Zufallsvariable ist, so hat X keinen einzelnen Wert, sondern X kann mehrere mögliche Werte annehmen.

Realisation einer Zufallsvariable bedeutet etwas anderes als Zufallsvariable.

Ein diskretes Merkmal kann <u>nicht</u> nur endlich viele Werte annehmen.

Median, Mittelwert und Spannweite sind alle <u>keine</u> Lageparameter zur Bestimmung der Verteilung eines Merkmals.

Die Summe der Abweichungen zum Mittelwert ist stets Null.

Verteilungen				Berechnen der Maximalen	Schätzfehlers:	1
Zentraler Grenzwertsatz	wertsatz Der zentrale Grenzwertsatz besagt, dass sich mit grösserem			$t*\sigma_{\bar{x}} = 2.060 * 0.34 = 0.70 g$		
	Stichprobenumfang n, alle Verteilungen	der Normalverteilung		6. Berechnen der Konfidenzgrenzen		
	approximieren. Als Faustregel gilt, dass b	ei einem n über 30		W(124.58 -0.70 ≤ µ ≤ 124.5	58 + 0.70)=0.95	
Stichproben die Normalverteilung genommen werden kann.			W(123.88 ≤ µ ≤ 125.28)=0.	·		
Stetige Verteilungen		Exponentialverteilung	Anwendungen:			
Stetige Verteilungen sind überabzählbar. Das heisst, sie beinhalten so viele Werte, dass diese nicht einfach gezählt werden können.			- Zeitspanne zwischen zwei	Anrufen in einer		
			Telefonzentrale			
Normalverteilung	1	· · · · · · · · · · · · · · · · · · ·		 Dauer eines Telefongesprägen 	ichs.	
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot (\frac{x-\mu}{\sigma})^2}$			- Lebensdauer eines Geräts, wenn Defekte durch äussere			
	$\sigma \sqrt{2\pi}$				Verschleiß verursacht werden.	Poissonverteilung
Chan danda annali iantali.				Wahrscheinlichkeitsdichte:	λ = Durchschn. Eintritt	
Standardnormalverteilu ng	Die Standardnormalverteilung ist der einfachste Fall der Normalverteilung, wenn der Mittelwert = 0 und die Varianz = 1			$f(t) = \lambda * e^{-\lambda t}$		
=	ist.			Verteilungsfunktion: (entspricht	der aufsummierten Wahrsch.)	
$\mu = 0$, $\sigma = 1$				$F(t) = 1 - e^{-\lambda t}$		
	$f(x) = \phi = \frac{1}{\sqrt{2\pi}} \cdot e^{-}$	$\frac{1}{2} \cdot x^2$		Erwartungswert	Varianz	
	$\int (\omega) = \varphi = \sqrt{2\pi}$				**	
	V 2 N			$E(x) = \frac{1}{4}$	$VAR(x) = \frac{1}{\lambda^2}$	
7 \\/ant banach:	77			Λ	n 10 00 und 11 00 Ub	
Z-Wert berechnen	$Z = \frac{X - \mu}{}$			Ein Geschäft wird täglich zwische		
	$Z = \frac{r}{r}$			durchschn. 3,5 Kunden besucht. Wahrscheinlichkeit, dass der Abs		
	σ		4	zweier Kunden höchstens 0.2 Stu		
Kleiner: Z Wert direkt ablesen Grösser: 1 – Z Wert aus der Tabelle			$F_F(0,2 3,5) = 1 - e^{-3,5}$			
			L	<u> </u>		
	Beidseitig: Min/Max Z-Werte herauslese		Weibull Verteilung	Beschreibt die Lebensdauer von	Geräten oder Materialen mit	
	In einer Frabrik wird Zucker abgefüllt. Der	•	211	Abnutzungserscheinung		
	Tüte soll 1000g beinhalten. Gegeben: μ=1002g, σ=1,5g		Diskrete Verteilungen Bernoulli Experiment	Let also 7. fall and a sign and a sign and		
	Sollgewicht unterschreitet?	e hoch ist die Wahrscheinlichkeit, dass eine Tüte das		Ist ein Zufallsexperiment mit gen (Treffer oder Niete).	au zwei moglichen Ergebnissen	
	$z = \frac{1000 - 1002}{1.5} = -1,33$ > Tabelle > 0.0	040 >0 400/	Bernoulli Kette			
		918 7 9.18 <u>%</u>	(binomial Verteilt)	$P(Ereignis) = \binom{n}{k} \cdot p$	$a^k \cdot a^{n-k}$	
Chi-Quadrat-Verteilung	Voraussetzungen		(binomial vertency	$\begin{pmatrix} 1 & (Ereignis) - \begin{pmatrix} k \end{pmatrix} & p & q \end{pmatrix}$		
(Varianz!) - Zufallsvariablen sind unabhängig und			Bsp: Wie gross ist die Wahrscheinlichkeit mit 5 Würfel, 2x eine Sechs zu würfeln			
	Wird aus der Normalverteilung abgeleitet					
Anwendung:			n = 5			
	Schätzung von Verteilungsparametern (z.B. Varianz) $\chi_n^2 = Z_1^2 + \ldots + Z_n^2$			k = 2		
				p = 1/6		
				q = 5/6		
				$\binom{5}{2} (\frac{1}{6}^2)(1 - \frac{1}{6})^{5-2} = 0.161 = 16\%$		
r-Verteilung Voraussetzungen			(2) 0 0			
ū	Freiheitsgrade: k=n-1			20% der von einer Maschine produzierten Bolzen sind unbrauchbar. Wie hoch ist die Wahrscheinlichkeit, dass von 4		
	Das Füllgewicht von Leberwürsten ist no	rmalverteilt. Das Soll				
	Mindestgewicht ist 125g.Aus den täglich produzierten 600 Würsten werden 26 gewogen:			zufällig ausgewählten Bolzen <u>höchstens</u> zwei unbrauchbar sind: Genau zwei Möglichiche Ausgänge: defekt / i.O $P(h "ochstens" 2) = \sum_{x=0}^{2} {4 \choose x} * 0.2^x * 0.8^{4-x}$		
						Rechteckverteilung
						Neonteekvertenang
				P(nochstens 2) = 2	(x) * 0.2" * 0.8" "	
				= Summe aus P(0 unbrauchbar) -		
			Binomialverteilung	Voraussetzungen	, , , , ,	1
				- Die Experimente sind vone	einander unabhängig	1
				- Es gibt nur <u>zwei</u> Ausgangsı		
				- Anzahl der Versuche ist fix		
				- Das Experiment wird imm	9	
	X normalverteilt und σ² unbekannt	> c verteint iiii c k ii 1		1	1 15 65 5 1 1 1 1	
	X normalverteilt und σ^2 unbekannt Freiheitsgraden	7 C VETCENCIMEN II I		- Mit grösserem n, nähert si		
				ähnlich der Dichtefunktior	einer Normalverteilung an	
	Freiheitsgraden	₃ von x		_	einer Normalverteilung an	
	Freiheitsgraden 3. Festlegen der Standardabweichung Varianz unbekannt, ohne Zurückleg	₃ von x		ähnlich der Dichtefunktion Wahrscheinlichkeitsfunktion (Ber	n einer Normalverteilung an noulli)	
	Freiheitsgraden 3. Festlegen der Standardabweichung Varianz unbekannt, ohne Zurückleg $rac{s}{\sqrt{n}}=0.34g$	₃ von x		ähnlich der Dichtefunktior	n einer Normalverteilung an noulli)	
	Freiheitsgraden 3. Festlegen der Standardabweichung Varianz unbekannt, ohne Zurückleg $\frac{s}{\sqrt{n}}=0.34g$ 4. Festlegen von t	g von ⊼ gen n/N<0.05 → $\hat{\sigma}$ =		ähnlich der Dichtefunktion Wahrscheinlichkeitsfunktion (Ber	einer Normalverteilung an noulli) noulli)	
	Freiheitsgraden 3. Festlegen der Standardabweichung Varianz unbekannt, ohne Zurückleg $rac{s}{\sqrt{n}}=0.34g$	g von ⊼ gen n/N<0.05 → $\hat{\sigma}$ =		$\begin{aligned} & \text{\"{a}hnlich der Dichtefunktion} \\ & \textbf{Wahrscheinlichkeitsfunktion (Ber} \\ & P(Ereignis) = \binom{n}{k} \cdot p^k \cdot q \end{aligned}$	einer Normalverteilung an noulli) noulli)	

$\sigma^2 = n \cdot p \cdot (1 - p)$ Gegen eine Krankheit wurde ein neues Medikament entwickelt. Die Heilungschance liegt bei 90%. Wie gross ist die Wahrscheinlichkeit, dass bei 5 zufällig gewählten Patienten mindestens 4 geheilt werden? $P(4) = {5 \choose 4} * 0.9^4 * (1 - 0.9)^{5-4} = 0.328$ $P(5) = 0.590 \quad P(4+5) = 0.9185$ sonverteilung Voraussetzungen Wie hoch ist die Wahrscheinlichkeit, dass das Ereignis in einem Intervall genau oder höchstens x-Mal eintritt, wenn bekannt ist, dass in diesem Intervall das Ereignis im Mittel μ -Mal auftritt. Hängt stark vom Mittelwert (μ) ab typische Beispiele: Druckfehler/Seite, Arbeitsunfälle/Tag Die Ereignisse treten unabhängig voneinander auf. (z.B Telefonanrufe) Seltene Ereignisse häufen sich (z.B. Bitfehler bei Flugzeugabstürzen) → Verteilung der seltenen Ereignisse Ankunftsrate: Eintreffende Ereignisse / Zeit z.B. 24 Kunden in 8 Stunden → 24/8=3 BEACHTE: Die Werte in der Verteilung Tabelle sind auf kumuliert → Ist ein einzelner Wert gesucht, muss die Differenz zum vorherigen Wert berechnet werden vorherige Tabellenwert abgezogen oder einfach die Formel verwendet werden. Wahrscheinlichkeitsfunktion Erwartungswert: $E(X) = \sigma^2 = \mu$ Durchschnittlich 1 Telefonanruf/Minute. Wie gross ist die Wahrscheinlichkeit, dass 2 Anrufe pro Minute eingehen? $f_P(2|1) = \frac{1^2 * e^{-1}}{2!} = 0.18$ Achtung: Falls bis zu 2 Anrufe gefragt sind, müssen diese auf kumuliert werden. $\sum_{k=0}^{2} \frac{1^k * e^{-1}}{k!}$ Bei mehr als 2:

Alle Realisationen in einem bestimmten Intervall [a, b] sind

gleich wahrscheinlich