Wörterbuch Statistik von ANA LOGO

Was	Bedeutung
a	Merkmalsausprägung.
a	Summe der Abstandsquadrate.
A	Merkmalsraum = Menge aller möglichen Beobachtungswerte.
$\lvert A \rvert$	# der Elemente der Menge von Ereignis A.
α	Signifikanzzahl, Irrtumswahrscheinlichkeit. 5%. Wenn α kleiner ist, ist Konfidenzzahl γ größer. Hat man viele Daten, lieber poolen.
α-Fehler	Es stellt sich heraus, dass H0 korrekt ist. Sie wird aber verworfen anstatt angenommen zu werden.
α-Fehler-Kumulation	Wenn H0 zwanzigmal signifikant ist, kann in der Summe Signifikanz auch zufällig zustande kommen.
Abbildung I	$X : A \to B$. Also Merkmal X ist eine Abbildung von der Menge A in die Menge B (... nach der Menge B).
Abbildung II	$X : \Omega \to B$. Also Merkmal X ist eine Abbildung von Ω nach B.
Abbildung III	$a \to X(a)$. a wird abgebildet auf X (a).
Abhängige Variable	Merkmal, das in einem Quasi-Experiment erfasst wird, um zu überprüfen, wie sich systematisch variierte unabhängige Variablen auf abhängige V. auswirken.
Abhängigkeit von Ereignissen	P(B|A) = P(B|A kompl) = P(B). Ereignis B ist stochastisch (un-)abhängig von Ereignis A ...mit P(A) > 0 und P(B) > 0. Unabhängigkeit siehe dort.
Absolute Häufigkeit	Anzahl.
Abzählregel	Alle Teilereignisse zusammenzählen, z. B. 6 Ereignisse in einer Menge. Dann günstige durch mögliche, z. B. 6 / 9.
AD-Streuung	AD = Σ i=1 bis n ($\lvert xi-x^{quer} \rvert$) / n, average deviation, negative Vorzeichen der Abweichung lassen sich gut durch Beträge betrachten.
Additionspfadregel	Mehrstufiger Zufallsversuch mit mehreren Pfaden zum gesuchten Ereignis. Alle Wahrscheinlichkeiten werden addiert.
Additionssatz I	Kolmogorov III. P(A ∪ B). Für sich ausschließende Ereignisse: P(1) + P(2) + ... + P(n). Wenn sich Ereignismengen also nicht überschneiden.
Additionssatz II	Wahrscheinlichkeit dafür, dass Ereignis A oder Ereignis B oder beide eintreten (Vereinigung → Additionssatz).
Additionssatz III - allgemeiner AS	Für beliebige Ereignisse: P(A ∪ B) = P(A) + P(B) - P(A ∩ B). P(A ∪ B ∪ C) = P(A) + P(B) + P(C) - P (A ∩ B) - P (A ∩ C) - P(B ∩ C) + P(A ∩ B ∩ C).
Additionssatz III - spezieller AS - I	Für disjunkte Ereignisse: P(A ∪ B) = P(A) + P(B). P(A U B U C) = P(A) + P(B) + P(C). P(A U A)=P(S)=1. P(A)=1-P(A).
Additionssatz III - spezieller AS - II	Für disjunkte Ereignisse: P(A U A kompl)= P(S) =1. P(A kompl) = 1-P(A). A kompl (komplementär) dann, wenn A nicht eintritt.
Additionstheorem	Wenn P eines Ereignisses bestimmt werden muss, welches sich aus der Vereinigung zweier anderer Ereignisse ergibt, Bortz S.52
Ähnlichkeitsmaße	Werden im Rahmen der Clusteranalyse benötigt, um die Ähnlichkeit der zu gruppierenden Objekte zu ermitteln.
Allgemeines lineares Modell	ALM. Verfahren, das die Varianzanalyse sowie die lineare Regressionsrechnung integriert.

Was	Bedeutung
Alternativhypothese	H_1. Formulierung vor Durchführung eines statistischen Tests. Logisches Gegenteil der Nullhypothese: Zusammen ergeben beide alle möglichen Ereignisse.
ANOVA I	Analysis of variance, Varianzanalyse: Einfluss eines/mehrerer unabhängiger Merkmale (deren Ausprägung) auf ein/mehrere abhängige Merkmale.
ANOVA II	Für Gruppenvergleiche, z. B. Vergleich der Fledermauskontakte zwischen verschiedenen Hecken und Weinbergen.
a posteriori I	Vorsicht: Hinterher ist man immer schlauer. Unterschied zwischen zwei Gruppen wird im Nachhinein auf Signifikanz geprüft (Varianzanalyse).
a posteriori II	(A|B). Bei a priori P(A).
a posteriori III	Dieser Ansatz hilft bei der Frage, wie wir unsere Annahmen, Überzeugungen und Theorien im Licht von neuen Ergebnissen ändern sollen.
a priori I	Über den Unterschied zwischen zwei Gruppen besteht bereits vor der Untersuchung eine (meist gerichtete) Hypothese.
a priori II	P(A). Bei a posteriori P (A|B).
a priori III	Dieser Ansatz hilft bei der Frage, wie wir unsere Annahmen, Überzeugungen und Theorien im Licht von neuen Ergebnissen ändern sollen.
Arithmetischer Mittelwert I	Nur sinnvoll bei symmetrischer Verteilung, z. B. Normalverteilung.
Arithmetischer Mittelwert II	Ist nicht Erwartungswert; Summe aller Werte geteilt durch die Anzahl der Werte, x Oberstrich. Methode ist zufällig: Neuer Wert = neue Zahlen
Arithmetischer Mittelwert III	Excel: **MITTELWERT()**
Aufleitung von e-Funktion	Siehe e-Funktion aufleiten.
Axiome	Aussagen, die nicht bewiesen werden, sondern deren Gültigkeit vorausgesetzt wird.
Axiome der Wahrscheinlichkeit I	Kolmogoroff I: Für P eines zufälligen Ereignisses A: **$0 \leq P(A) \leq 1$**
Axiome der Wahrscheinlichkeit II	Kolmogoroff II: P eines **sicheren** Ereignisses ist gleich **1**
Axiome der Wahrscheinlichkeit III	Kolmogoroff I: Sind zwei Ereignisse **disjunkt**: **$P(A \cup B) = P(A) + P(B)$**
b	Regressionskoeffizient.
Balken- und Kreisdiagramm	Gut für nominale Merkmale.
Bayes	Thomas Bayes (1702-1763) unterschied erstmals A-priori und A-posteriori-Wahrscheinlichkeiten → Ergebnis: Die sehr umstrittene Bayessche Formel
Bayes-Satz I	Für zwei Ereignisse: $P(A\|B) = P(A) \cdot P(B\|A) / P(B)$. Die gesuchte Größe ist die erste, also abhängige Größe.
Bayes-Satz II	30% kommen mit Auto, 80% sind pünktlich, 95% Autofahrer pünktlich. 0,3·0,95/0,8= 35% aller Pünktlichen sind Autofahrer.
Bayes-Statistik I	Variante der statistischen Entscheidungstheorie. Rückschlüsse von a-priori-Wahrscheinlichkeiten auf a-posteriori-Wahrscheinlichkeiten
Bayes-Statistik II	Berechnung von Wahrscheinlichkeiten für verschiedene Hypothesen unter Voraussetzung eines empirisch ermittelten Untersuchungsergebnisses
Bedingte relative Häufigkeit	r (a|b) = h (a,b) / h (b): Bedingte relative Häufigkeit von a unter der Bedingung b. Werte aus Kontingenztabelle nehmen.
Bedingte Wahrscheinlichkeit I	$P(B\|A) = P(B \cap A) / P(A)$ oder mit Quer-Version: P(A|B quer) = P(A∩B quer) / P(B quer).

Was	Bedeutung
Bedingte Wahrscheinlichkeit II	P(B|A) = P(B∩A) / P(A). P(B|A) als bedingte P. P dafür, dass B eintritt, unter der Bedingung dass A mit P(A) ≠ 0 bereits eingetreten ist.
Bedingte Wahrscheinlichkeit III	P(B|A) = P(B∩A) / P(A). Baumdiagramme (spcl: Anwendung auf Testverfahren mit Fehlern)
Bedingte Wahrscheinlichkeit IV	Totale Wahrscheinlichkeit und Satz von Bayes.
Bernoulli-Experiment	Zwei Möglichkeiten im Experiment, z. B. bei der Münze Kopf oder Zahl. Mit Zurücklegen! Baumdiagramm anwenden.
Bernoulli-Kette	Mehrere Bernoulli-Experimente hintereinander. Baumdiagramm.
Bimodalität	Zwei Modalwerte, also Verteilung mit zwei Gipfeln.
Binominalkoeffizient I	**(n über k)** = **n! / k! · (n - k)!**
Binomialkoeffizient II	Ausrechnen mit 1. Baumdiagramm oder 2. Pascal'schen Dreieck
Binominalkoeffizient - **in R**	choose. Für Beispiel: n<-9 und k<-7: choose (n, k). Ergebnis: [1] 36.
Binomialverteilung I	Trefferwahrscheinlichkeit ist immer gleich. Mit Zurücklegen.
Binomialverteilung II	**n! / k! · (n-k)!**. P-Verteilung: Wie wahrscheinlich ist x Erfolg bei n Wiederholungen eines Zufallsexperiments ?
Binomialverteilung IV	Erfolg tritt in jedem Versuch mit **p = π** ein (Misserfolg p: 1-π). z. B. beim Münzwurf: Erfolg = Zahl und Misserfolg = Kopf.
Binomialverteilung V	In Klammern: (oben 8, unten 5): Wenn untere Zahl größer als obere, ist Binominalkoeffizient definiert als 0.
Binomialverteilung VI	Diskrete Wahrscheinlichkeitsverteilung für ein Merkmal, das z. B. aus zwei Ausprägungen, z. B. funktionsfähig/defekt, Wappen oder Zahl besteht
Bivariate Normalverteilung I	Werden zwei Merkmale und gemeinsam erhoben, verteilen sie sich bivariat normal...
Bivariate Normalverteilung II	... wenn nicht nur die Verteilung jedes Merkmals für sich allein, sondern auch deren gemeinsame Verteilung normal ist...
Bivariate Normalverteilung III	... In diesem Fall ergibt grafische Darstellung der gemeinsamen Verteilung eine dreidimensionale Glockenform.
Bivariate Verteilung	Werden zwei Merkmale und gemeinsam erhoben. Punktwolke oder dreidimensional.
Blindwert I	Zur Überwachung von Blindwerten, um Messwerte der Messproben (Analysenprobe) um einen unspezifischen Anteil korrigieren zu können
Blindwert II	Reagenzien-BP: Beinhalten alle in Prüfmethode beschriebene Reagenzien, aber null Anteil der Laboratoriumsprobe.
Blindwert III	Proben-BP: Beinhalten Reagenzien lt. Prüfmethode + Probenmatrix ohne zu bestimmende(n) Komponente(n). Messmatrix ohne Probe. Optimale Blindprobe.
Blindwertursachen I	Verwendete Reagenzien sind von zu bestimmender Komponente nicht völlig frei, systematische Verunreinigung der Prüfgeräte
Blindwertursachen II	Matrixbestandteile und Reagenzien zeigen ähnliche Reaktionen wie Reagenzien mit zu bestimmenden Komponenten (mangelnde Selektivität)
Blindwertursachen III	Unterschiede in den Prüfgeräten (Küvetten in der Photometrie), Reagenzienzersetzung (Lagerung, Alterung)
Bootstrapping I	Der Gag: Mit Zurücklegen. Gibt es erst seit 1985. Konfidenzschätzung. Münchhausen-Prinzip. Am eigenen Schnürsenkel aus dem Sumpf ziehen.
Bootstrapping II	Anwendung z. B. wenn man nicht weiß, wie die fünf Gewichte von Fischen verteilt sind. Es gibt Δ Intervalle.

Was	Bedeutung
Boxplot I	Whiskerlänge maximal ausrechnen: **Q1 - 1,5 · IQR und Q3 + 1,5 · IQR**, alles darüber sind Ausreißer. **In R boxplot(a)**.
Boxplot II	Für ökologische Daten sind Boxplots am besten geeignet.
β-Fehler	Es stellt sich heraus, dass H0 nicht korrekt ist. Sie wird aber angenommen anstatt abgelehnt zu werden.
c	Kritischer Wert oder Vergleichswert.
Chi^2-Test I	Signifikanztests zur Analyse von Häufigkeiten.
Chi^2-Test II	Drei Hypothesentests: Verteilungstest (Anpassungstest), Unabhängigkeits- und Homogenitätstest. Beginnt bei 0. Nicht symmetrische Graphen.
Chi^2-Verteilung	Nicht symmetrisch. Stammt aus der Feder Karl Pearsons.
Cluster	Einzelne gebündelte Gruppen von gepoolten Daten. Ein Pool ist also die größere Einheit.
Clusteranalyse	Heuristisches Verfahren (wenig Wissen + wenig Zeit: gute Lösung) zur systematischen Klassifizierung der Objekte einer gegebenen Objektmenge
dμ	Ableitung vom Mittelwert = Messunsicherheit vom MW.
dσ	Ableitung von STA = Messunsicherheit von STA.
δ	Mittlere lineare Abweichung.
Dach	$\hat{y}$ oder y-Dach. Ausgehend von vorhandenen Daten der erwartete Wert (*Regressionsgerade*)
Datentyp	z. B. kardinalskaliert, kategorial.
dbinom	Genau eine Zahl.
density	Relative Häufigkeit (vs. frequency für Häufigkeit).
Dendrogramm	Grafische Darstellung der Ergebnisse einer hierarchischen Clusteranalyse, die über die Anzahl der bedeutsamen Cluster informiert.
Deskriptive Statistik	Beschreibende Statistik. Beschreibt Daten einer Stichprobe z.B. durch Grafiken oder Kennwerte (Mittelwert, Varianz, etc.).
Deterministisch	Modell des Festlegens, was man **nicht** weiß.
Deterministische Wahrscheinlichkeit	Zuordnung eines Wertes für p, z. B. bei Glücksrad für Winkel des Kreissektors αω / 360°, also α-grau = 80° bedeutet p = 80/360 = 2/9.
Diagnostischer Test	ROC-Analyse und Likelihood-Quotient. Ferner siehe Prävalenz, Spezifität, Sensitivität, Negativer und positiver Voraussagewert.
Dichte der Gleichverteilung	Ist eine Konstante.
Dichtefunktion I	Wahrscheinlichkeitsdichtefunktion: Wahrscheinlichkeitsverteilung einer stetigen ZV. Empirische Dichtefunktion = relative Häufigkeit der Daten.
Dichtefunktion II	Ableitung der Verteilungsfunktion. Wenn man einmal Dichtefunktion hat, kann man in WS-Rechnung gehen.
Dichtefunktion III	Verteilungsfunktion kann man nur kennen, wenn man Dichtefunktion hat.
Dimension	Eine: x. Zwei: x + y ... z. B. in Formel (x·y) + (x-y).

Was	Bedeutung
Disjunkte Ereignisse	Unvereinbare Ereignisse, die nicht gleichzeitig auftreten können.
Diskretes Merkmal	Kann nur bestimmte Werte annehmen, z.B. # Freunde einer Personen nur als ganze Zahlen.
Diskrete Häufigkeitsverteilung I	Endliche bzw. abzählbare Menge: 2 Würfel à 6 ergibt diskrete Werte von 2 - 12. Variabilität begrenzt. Binomial- und Hypergeometrische Verteilung
Diskrete Häufigkeitsverteilung II	Grafik: **Stabdiagramm** verwenden.
ε	Tolerierter Fehler.
η	Korrelationskoeffizient, der die non- und linearen Zusammenhänge zwischen unabhängiger und abhängiger Variable erfasst (Varianzanalyse).
η	Substituent.
e-Funktion ableiten I	Aus $f(x) = 4\ e^{2x}$ wird $f'(x) = 8\ e^{2x}$. Nach $f(x) = e^{x+4}$ bleibt $f'(x) = e^{x+4}$. Aus $f(x) = 2\ e^{2-4x}$ wird $f'(x) = -8\ e^{2-4x}$. Aus $f(x) = (x+2)\ e^{x}$ wird $f'(x) = (x+3)\ e^{x}$.
e-Funktion ableiten II	Aus $F(x) = 8\ e^{t/4}$ wird $f(x) = 2\ e^{t/4}$.
e-Funktion aufleiten I	Aus f(x) = e^x wird F(x) = e^x , z. B. bei g(x) = e^(x²-2) muss man noch durch innere Ableitung dividieren, also G(x) = e^(x² - 2) / 2x.
e-Funktion aufleiten II	Aus f(x) = -3e^-x wird F(x) = 3e^-x : Es ändert sich nur das Vorzeichen, die Ableitung der Hochzahl wird mit dem Faktor vor e multipliziert.
Uneigentliche Integrale	Haben Grenzen -∞ und/oder ∞. Berechnung mit Hilfe von Grenzwerten. Funktion liegt im Integrationsbereich [a; b] und hat Defintionslücke.
Elementarereignisse	Ereignisse mit nur genau 1 Ergebnis. Merkmalsausprägungen, z.B. {2,4,6} ∪ {1,4,} = {1,2,4,6}.
Empirische arithmetischer MW	Hat z. B. die Dimension Länge mit der Einheit [cm].
Empirische Kovarianz	Siehe Kovarianz.
Empirische Messungen	Empirische Verteilungen.
Empirische Standardabweichung	s = √s² = √ (alle quadrierten Werte summiert) / n - STA². √ aus 1/n · Σ (xi-μ)², z. B. √ 18743/40 - 21,27². Excel: STABWN().
Empirische Verteilungen	Empirische Messungen.
Empirischer Korrelationskoeffizient	Siehe Korrelationskoeffizient. r.
Endlich mögliche Ergebnisse	Würfel hat maximal 6 Ergebnisse.
Endliche ZV	Nur endlich viele Werte x1, ..., xm möglich. Beispiel: Würfel hat nur 6.
Ereignis I	Teilmengen der Ergebnismenge. Sie können also eintreten. Jedem Ereignis werden Wahrscheinlichkeiten zugeordnet. ≠ Ergebnis.
Ereignis II	z. B. (X = x). Oder (X ≤ x) für alle Ereignisse, bei denen die Zufallsgröße alle Realisationen annimmt, die ≤ der reellen Zahl sind.
Ereignis III	Menge von Ergebnissen (auch { } möglich). Teilmenge der Ergebnismenge aller möglichen Elementarereignisse, z. B. die Zahlen 3-6 beim Würfeln.
Ereignisraum	Gesamtheit aller Elementarereignisse. P = 1 = 100% = {...}.
erf	Gaußsche Fehlerfunktion.

Was	Bedeutung
Ergebnis	≠ Ereignis. Können auch Mengen sein. z. B. bei Ω = {K,Z} und 6 Würfen Ergebnis {K,Z,Z,K,Z,Z}. Ereignis dann: 4 x K und 2 x Z.
Ergebnismenge I	Ω. Zahlen, Tupel oder Mengen. Am Beispiel von drei Würfen mit Würfel: $\{1,2,3,4,5,6\}^3$ mit $\Omega = 6^3 = 216$. Dabei ist ein mögliches Ergebnis (1,5,4).
Ergebnismenge II	Grundgesamtheit, Menge der möglichen Ergebnisse. **Das** spielt sich in der Realität ab. Gerechnet wird z.B. mit $\mathbb{R}$. $\mathbb{R}$ als Abbildung. Münze: {Z,K}.
Erwartungstreue	Gütekriterium für Schätzungen. Wenn EW dem wahren zu schätzenden Wert entspricht.
Erwartungswert I	μ. E(X) einer Zufallsvariable X. MW einer theoretischen (nicht empirischen) Verteilung einer ZV.
Erwartungswert II	Regeln: 1. Falls $X(\omega) = a$, dann $E(X) = a$. 2. $E(c \cdot X) = c \cdot E(X)$ 3. $E(X1 + X2) = E(X1) + E(X2)$.
Erwartungswert III	Formeln für EW gelten immer (abhängig/unabhängig). Bezüglich Zukunft.
Erwartungswert IV	Durchschnittlich zu erwartender Wert einer ZV ergibt sich aus der Verteilung. Wo ist also die Mitte der Verteilung?
Erwartungswert V	EW = MW der Normalverteilung = gewogener arithmetischer MW, STA mit μ-s und μ+s = 68%, μ-2s und μ+2s = 95,5%. Ist oft nicht ausrechenbar.
Erwartungswert VI	Arithmetischer MW nur wenn Stichprobe gut ist. Gibt gewichtete Wahrscheinlichkeit.
Erwartungswert VII	Für diskrete Zufallsvariablen: Σ xi · Pi. Also additiv → Würfel: (1/6 · 1 + 1/6 · 2 + 1/6 · 3 ...) = 3,5. EW zweier Würfel = 7.
Erwartungswert VIII	Für stetige Zufallsvariablen: $\int_{-\infty}^{+\infty} t \cdot f(t)dt$.
Erwartungswert IX	In dessen Nähe erwartet man mit größter WS die Realisierungen der Zufallsgröße.
Experiment	Stichprobe.
Exponentialfunktion I	Als allgemeine Funktion i. w. S. : $x \rightarrow a^x$.
Exponentialfunktion II	Als spezielle Funktion i. e. S. ist damit meist die natürliche Exponentialfunktion gemeint: $x \rightarrow e^x$. Schreibweise: $x \rightarrow \exp(x)$.
Exponentialverteilung I	Verteilungsfunktion ist die Stammfunktion. Durch Ableitung kommt man zu ihrer Dichtefunktion.
Exponentialverteilung II	$F_Z(x) = \int_{-\infty}^{x} f(t)dt$ = {0 falls x<0 und 1-e - λ•x, falls x ≥0}. E(Z) und $\sigma_Z = 1/\lambda$.
factor	Ordinale Faktoren = ordinale Skalen. **>factor**
Fakultät	0! = 1. (n+1)! = (n+1) · n!. Bei paarweise Δ Komponenten: n! / (n-k)!. Produkt aller ganzen Zahlen zwischen 1 und dieser Zahl.
Fallzahlbestimmung I	Bestimmung des erforderlichen (Mindest-) Stichprobenumfangs bei Schätzung eines (prozentualen) Anteils π: $(n \geq u^2_{\alpha/2} \cdot p \cdot (1- p)) / e^2$.
Fallzahlbestimmung II	n = Erforderliche Mindestfallzahl. $u^2_{\alpha/2}$ aus Tabelle von N(0,1). 1-α bzw. α Sicherheits- bzw. Irrtumswahrscheinlichkeit.
Fallzahlbestimmung III	p = „Erfahrungswert" für ungefähre Größenordnung des %-Anteils existiert kein Wert, setze p=0,5.
Fallzahlbestimmung IV	e = |π−p|. Exaktheit. Abstand zwischen Stichprobenschätzung p und wahrem π der Grundgesamtheit.
Fallzahlbestimmung V	Bestimmung der erforderlichen (Mindest-)Fallzahl n bei der Schätzung eines (unbekannten) Mittelwertes μ: $(n \geq u^2_{\alpha/2} \cdot \sigma\text{Dach}^2 / e^2$.
Fallzahlbestimmung VI	n = Erforderliche Mindestfallzahl. $u^2_{\alpha/2}$ aus Tabelle von N(0,1). 1-α bzw. α Sicherheits- bzw. Irrtumswahrscheinlichkeit.

Was	Bedeutung
Fallzahlbestimmung VII	**e = |μ−x quer|**. Exaktheit. Unterschied zwischen Stichprobenmittelwert x und wahrem μ der Grundgesamtheit.
Fallzahlbestimmung VIII	σDach² = Schätzung der Varianz des betrachteten stetigen Merkmals. Falls keine Schätzung existiert: s² = 1/n-1 • Σ I=1 bis n (xi - x quer)².
Fisher I	Sir Ronald A. Fisher (1890-1962) war dominierender Theoretiker der Statistik im 20. Jahrhundert. Auch heute noch dominierend in der Praxis.
Fisher II	Fisher-Produkte: Varianzanalyse, Planung von Experimenten, Zufallsanordnungen, Signifikanztests.
Fisher-Test	Prüft die Nullhypothese, dass sich zwei Anteile aus zwei kleinen Zufallsstichproben nicht signifikant unterscheiden.
Freiheitsgrade I	Nur bei Tests mit Populationsparameter-Schätzung (z. B. MW). Nicht immer benötigt, z. B. Fishers exakter Test.
Freiheitsgrade II	Aus Kontigenztabellen: (Zeilen minus 1) · (Spalten minus 1) = z. B. 3. Je symmetrischer, desto mehr FG.
Freiheitsgrade III	(n1 + n2 -2) / 2 (mit n = Anzahl der Messungen). Für Streuung: Man teilt bei Stichproben durch FG und nicht durch Anzahl der Messungen.
Freiheitsgrade IV	**df** (degrees of freedom), Stichprobenumfang n minus der # Parameter p, die aus diesem Stichprobenumfang geschätzt wurden: **df = f = n - p**
frequency	Häufigkeit (vs. density für Dichte = relative Häufigkeit).
Friedman-Test	Nicht-parametrischer Test. Vergleicht die Lage von drei oder mehr Gruppen abhängiger Stichproben.
Galton	Francis Galton (1822-1911) entwickelte Grundlagen der Regression und Korrelation. Er war ein Vetter von Charles Darwin.
GAM	General additive model.
Gammaverteilung	Kontinuierliche Wahrscheinlichkeitsverteilung über der Menge der positiven reellen Zahlen. NV = $a \cdot x^{b} \cdot e^{cx}$ für x>0.
Gauß	Karl Friedrich Gauß (1777-1855) unternahm grundlegende Arbeiten zur Normalverteilung.
Gaußverteilung	Normalverteilung. Glockenkurve. y-Achse: Häufigkeit der Ergebnisse. x-Achse: Ergebnisse. Bsp: Treppenstufen sind in der Mitte stärker abgenutzt
Gegenereignis I	> 3 ist gleich 1 - <4. !!! Also ausrechnen für Σ 1 bis 3. ≥ 4 ist gleich 1 - ≤ 3. !!!
Gegenereignis II	Bei bedingten Wahrscheinlichkeiten: P(Hquer|G) ist Gegenereignis zu 1- P(H|G).
Genauigkeit I	Wie nah liegen die Werte der Messung am wahren Wert? Wirklichkeitstreue. Accuracy.
Genauigkeit II	Wiederholte Messungen können sehr präzise, aber sehr ungenau sein.
Geometrischer Mittelwert	GM = **$\sqrt[n]{x1 \cdot x2 \cdot x3 \cdot \ldots xn}$** , größere Werte mit kleinerem Einfluss, für Ø Veränderung von Zuwachsraten, Excel: **GEOMITTEL()**
Gesamtabweichung	Differenz von Sollwert und Istwert, Systematische plus zufällige Abweichung.
Gewogener arithmetischer MW	μ, **(Σ xi · fi) / n**. Wie häufig kam die 2 vor? Antwort: 17x die 2 und 3x die 8 / 20.
Gleichverteilung	Verteilungsfunktion ist die Stammfunktion. Durch Ableitung kommt man zu ihrer Dichtefunktion.
Gleichverteilung (diskret)	Jedes mögliche Ereignis tritt mit der gleichen Wahrscheinlichkeit ein. Laplace: P(A) = Betrag von A / Betrag von Ω.
Gleichverteilung (stetig)	Die Dichte ist konstant. Wenn $\Omega \subseteq R^{n}$.

Was	Bedeutung
GLM I	Analyse-Methode aus dem Bereich der Regressionsanalyse, abhängige Variable Y (Zielgröße) hat diskreten Charakter, wie z. B. ja = 1.
GLM II	Generalized linear model: Bestehend aus Zufalls-, parametrischer Link- und systematischer Komponente.
GLM III	Verwendung bei kontinuierlichen erklärenden Variable, z. B. Betrachtung der Fledermaus-Kontakte entlang der Entfernung zur Hecke
Grenzabweichung	Untere Grenzabweichung oder obere Grenzabweichung.
Gütefunktion I	WS für einen α-Fehler. WS, dass ich H0 ablehne, obwohl H0 stimmt.
Gütefunktion II	Wie stimmt das, was ich gemessen habe? Durch einen Test wird die Güte der Anpassung geprüft. **In R manuell auszurechnen. r^2 ?**
Gütemaß R^2 I	Misst nur Stärke eines linearen Zusammenhangs. Wieviel erklärte Streuung an Gesamtstreuung: Güte der Anpassung der Regressionsgerade an Lage der Werte
Gütemaß R^2 II	Zwischen 0 und 1. Bei 1: Gesamte Streuung wird durch das Regressionsmodell aufgeklärt: Perfekter linearer Zusammenhang. **Gibt's nicht in R**.
h	Häufigkeit (vs. density für Dichte = relative Häufigkeit).
h_a	Absolute Häufigkeit.
h_r	Relative Häufigkeit (vs. frequency für Häufigkeit).
H-Test	Von Kruskal und Wallis, verteilungsfreies Testverfahren: Gehören Beobachtungen aus Δ unabhängigen Stichproben zu gleicher Grundgesamtheit?
Harmonischer Mittelwert	HM, bei Ermittlung von Indexzahlen wie Preis pro Liter, Einwohner pro km², km/h. Wenn Beobachtungen Reziprozitäten enthalten.
Häufigkeitsdichte	**h_r** (relative Häufigkeit) / Klassenbreite.
Häufigkeitstabellen I	Für diskrete Merkmale. Kreuztabellen.
Häufigkeitstabellen II	Da stetige Merkmale ganz unterschiedliche Merkmalswerte besitzen können, würde direkte Datenauswertung durch Tabellen wenig Sinn machen.
Häufigkeitsverteilung	Besonders geeignet für nominale, ordinale und diskrete Ausprägungen. Für stetige Merkmale müssen Klassen gebildet werden.
Hauptkomponentenanalyse	Aus vielen Variablen wenige machen.
Histogramm I	Kategorie (x-Achse) und Kategorienhäufigkeit (y-Achse). Für stetige klassifizierte Variablen. Welche Klassengrenze ist die ideale?
Histogramm II	Breite spielt eine Rolle. In Breite steckt die Information. Fläche eines Rechtecks entspricht: c · f (xj) . Wenn c 0 n: Fläche = h absolut.
Histogramm III	**In R: hist ()** oder mit Klassengrenzen: **hist(xxx, breaks=6)**.
Hypergeometrische Verteilung I	Ohne Zurücklegen.
Hypergeometrische Verteilung II	H. Lostrommel weiße + schwarze Kugeln (N-Objekte). Davon k Objekte mit bestimmter Eigenschaft. Ausgezeichnete Objekte = weiße Kugeln.
Hypergeometrische Verteilung III	Nicht ausgezeichnet: N - K. n Objekte gezogen (aus N) ohne zurücklegen. Dichotome Grundgesamtheit. Anwendung bei Qualitätskontrollen.
Hypothese	Annahme über die Verteilung einer Zufallsvariablen, z.B. betreffende Verteilung hat Mittelwert 35,8.
Hypothesentests	Dinge über zufällige Dinge rausfinden. 1000e Versuche: Was kann man über Wahrscheinlichkeit sagen?

Was	Bedeutung
Induktive Statistik	Schließende Statistik. Modelle aufstellen und validieren.
Integrale ausrechnen I	1. li Seite F(t) aufleiten aus 2. re Seite f(t) 3. f(t) einsetzen in Formel wie in ∫ (4t+5)dt 4. F(t) einsetzen wie in [F(t)] b über a = **F(b) - F(a)**.
Integrale ausrechnen II	Integral über Fläche der Ableitung rechnet man also aus mit Werten der Aufleitung.
Interpolation	1. Schritt: IDW (Inverse Distance Weighting) = Inverse Distanz-Gewichtung. Kann man in QGIS machen.
Intervallskalierung	Kardinalskalierung, Temperatur in Celsius: "doppelt so heiß" geht nicht, da Celsiusskala keinen Nullpunkt hat. Aussage bzgl. Kelvinskala OK.
Inverse Distanzgewichtung	Je weiter weg, desto unähnlicher. Nichtstatistisches Interpolationsverfahren der Geostatistik. Einfache Interpolation räumlicher Abhängigkeit von Geodaten.
Inverse Wahrscheinlichkeit	Wahrscheinlichkeit für den umgekehrten Fall. Zentrales Konzept von Bayes.
IQR	Quartilabstand.
Irrtumswahrscheinlichkeit	Signifikanzniveau. **α**. α-Risiko.
Ishikawa-Diagramm	Ursache-Wirkungs-Diagramm = Fischgräten-Diagramm.
Justieren	Messgerät so einstellen, dass Messabweichungen möglichst klein werden oder Messabweichungsbeträge die Fehlergrenze nicht überschreiten.
k	Klasse.
K	Trefferzahl ist zufällig, aber bekannt. Immer ganzzahlig.
Kalibrieren	Ermittlung/Festlegung v. funktionalem Zusammenhang 1. Zähl-/messbare Größe + 2. Bestimmende c (Objekteigenschaft) aus Daten mit zufälligen Abweichungen
Kardinalskalierung	Intervallskaliert oder rationalskaliert: Abstände zwischen zwei Ausprägungen können quantifiziert und verglichen werden.
Kardinalzahlen	Mengen, die als Repräsentanten von Mengen einer bestimmten Größe dienen.
Kategorisierung	Klassierung, Merkmalsausprägungen werden gruppiert. Man kann Kategorien nur zusammenzählen.
Kennwerte I	Mittelwerte und Streuung für Stichproben haben lateinische Buchstaben.
Kennwerte II	Mittelwerte und Streuung für Population = Grundgesamtheit haben griechische Buchstaben.
Klassenbildung I	Bei stetigen Merkmalen. Festgelegt werden muss: Wie viele Klassen lege ich fest und wo lege ich Klassengrenzen fest?
Klassenbildung II	Anzahl der Klassen wird bestimmt durch Größe der Stichprobe. Faustregel: Anzahl der Klassen = $\sqrt{n}$ (bei Dezimalstellen sinnvoll runden).
Klassenbildung III	Klassengrenzen dürfen sich nicht überschneiden (klare Definition). Klassengrenzen sollen aneinander Grenzen.
Klassenbildung IV	1. Klasse fängt bei dem minimalen Wert an. Letzte Klasse hört bei dem maximalen Wert auf. Glatte Zahlen zur besseren Lesbarkeit.
Klassenbildung V	Ideal: Prozentual gleich starke Klassen: Jede Klasse besitzt die gleiche Anzahl von n.
Klassenbreite	Spannweite / Anzahl der Klassen. Jede Klasse besteht aus gleich breiten Abschnitten.
Klassenhäufigkeit	Bei prozentual gleich starke Klassen: **n / Anzahl der Klassen**.

Was	Bedeutung
Kolmogorov-Smirnoff-Test	Anpassungstest zur Prüfung auf Normalverteilung. Test auf eine bestimmte Verteilungsfunktion.
Komplementärereignis	A^{quer}, Alle Ereignisse, die nicht zum Ereignis A gehören.
Konfidenzintervall	Streubereich für den Mittelwert in der Grundgesamtheit. Je mehr Daten, desto kleiner ist KI.
Konfidenzintervall I	**1-α**: Sicherheits- oder Vertrauenswahrscheinlichkeit üblich: 99%, 95%, 90%. **α**: Irrtumswahrscheinlichkeit üblich: 1%, 5%, 10%.
Konfidenzintervall II	$u_{\alpha/2}$: Werte aus der N(0,1)-Tafel. KI hängt ab von n, α und γ. Wenn α kleiner ist, ist γ größer.
Konfidenzintervall III	Intervall mit Vertrauensgrenzen, in Naturwissenschaften meist KI 95% verwendet mit z. B. 10.000 Läufen, in Lage und Breite zufälliges Intervall.
Konfidenzintervall IV	Für einen unbekannten Mittelwert μ einer Grundgesamtheit: P(x quer - $u_{\alpha/2}$ • s / √n ≤ μ ≤ x quer + $u_{\alpha/2}$ • s / √n) = 1-α.
Konfidenzintervall V	Beispiel für 95%-ige Sicherheitswahrscheinlichkeit zur Formel IV: P(x quer - 1,96 • s / √n ≤ μ ≤ x quer + 1,96 • s / √n) = 95%.
Konfidenzintervall VI	Sonderregel: Umfasst die Stichprobe > 5% der Grundgesamtheit (große Stichprobe), benutze statt s / √n lieber s / √n • √N-n/N-1.
Konfidenzintervall VII	Für eine unbekannte Proportion **π** einer Grundgesamtheit: P(p - $u_{\alpha/2}$ • √p(1-p)/n ≤ π ≤ p + $u_{\alpha/2}$ • √p(1-p)/n) = 1-α.
Konfidenzintervall VIII	Voraussetzung für VII : n • p ≥ 5 wenn p ≤ 0,5 und n • (1-p) ≥ 5 wenn p > 0,5.
Konfidenzintervall IX	Beispiel für 95%-ige Sicherheitswahrscheinlichkeit zur Formel VII: P(p - 1,96 • √p(1-p)/n ≤ π ≤ p + 1,96 • √p(1-p)/n) = 95%.
Konfidenzintervall X	Sonderregel: Umfasst die Stichprobe > 5% der Grundgesamtheit (große Stichprobe), benutze statt √p(1-p) / n lieber √p(1-p)/n • √N-n/N-1.
Kontingenztabelle I	Kreuztabelle für stetige Stichproben mit festgelegten Bereichen bzw. Intervallen.
Kontingenztabelle II	Für zwei verbundene Merkmale X und Y. Für qualitative Merkmale aus ordinal- und nominalskalierten Bereichen.
Kontingenztabelle III	Immer zwei Tabellen aufzeichnen: Eine für absolute Häufigkeit, andere für relative.
Korrelation	Zwei Δ Wertepaare (Messung Zufallsgrößen an einem Objekt), i.e.S. der lineare stochastische Zusammenhang. **$-1 \le r_{xy} \le +1$**. Ohne Einheit. **R :>cor**
Korrelationskoeffizient I	Das Maß der Abhängigkeit von zwei Variablen, z. B. Durchschnittseinkommen / Lebenserwartung. Wie sicher ist Rückschluss von a auf b?
Korrelationskoeffizient II	**r.** Kovarianz $s_{X,Y}$ / STA's $s_X \cdot s_Y$. [Σ (x·y) - 1/n · (Σx) · (Σy)] / [√ (Σx² - 1/n(Σx)²) · (Σy² - 1/n(Σy)²)]. $r_{X,Y}$ = 1/n · Σ i=1 bis n (X Dach (ωi) · Y Dach (ωi))
Korrelationskoeffizient III	Pearson. Zur Ausrechnung der Formel Tabellenschema (* siehe Reiter) erstellen. Maß für die lineare Abhängigkeit. Wird standardisiert. Folie 81.
Korrelationskoeffizient IV	Zwischen -1 und +1. -1 ist negative Korrelation (je größer x, desto kleiner y). < -0,6 und > 0,6 "statistisch erkennbarer Zusammenhang".
Korrelationskoeffizient V	Der empirische Korrelationskoeffizient bildet lediglich eine Schätzung für den (unbekannten) richtigen Korrelationskoeffizienten.
Korrelationskoeffizient VI	Liegt dieser nahe bei 1 oder -1 → Schätzung für die Werte b + a der Regressionsgleichung.
Kovarianz I	Summe der Produkte der Abweichungen der Werte der beiden Merkmalsträger von ihrem Mittelwert.
Kovarianz II	$s_{x,y}$: (X · Y)quer - X quer · Y quer. [Σ (xi - x quer) · (yi - y quer)] / [n - 1]. n-1 für empirische Kovarianz. In Excel: **KOVAR(B3:B7;C3:C7)**.
Kovarianz III	Beschreibung der Streuung der beobachteten Punktwolke durch Summe der Rechtecke Σ (xi - x quer) · (y i - y quer).

Was	Bedeutung
Kovarianz IV	Verallgemeinerung der Varianz. Maß für Grad des Miteinander-Variierens zweier Messwertreihen x und y. EW (Produkt A·B) - EW (A) · EW (B).
Kovarianz V	Erweiterung der Varianz: Korrelation · Varianz = Mittelwert (x · y) - Mittelwert (x) · Mittelwert (y), Ableitung der Semivarianz ?
Kovarianz VI	Positive K.: Viele Versuchspersonen haben bei einem hohen x-Wert auch einen hohen y-Wert. Wenn groß, dann gibt es Korrelation. **R :>cor**
Kovarianz VII	Negative K.: Viele Versuchspersonen haben bei einem hohen x-Wert einen niedrigen y-Wert.
Kovarianz VIII	K. von z-transformierten Variablen entspricht der Produkt-Moment-Korrelation.
Kreuztabelle	Für zweidimensionale Stichproben und viele Werte. Für /D,D)-Stichproben.
Kriging I	Geostatistisches Verfahren: Interpolation oder Annäherung von Werten an Orte, für die keine Stichprobe vorliegt durch umliegende Messwerte
Kriging II	Fortgeschrittene Statistik. Zeitliches Kriging interessant für Klimadaten.
Kriging III	Nicht so gut geeignet für Gewässer: Untere Datengrenze 50 Messungen.
Kruskal-Wallis-Test	Nicht-parametrischer Test. Vergleicht die Lage von drei oder mehr Gruppen unabhängiger Stichproben.
λ	Exponentialverteilung wird durch Parameter λ bestimmt mit Werten zwischen 0 (ausschließlich) und unendlich.
Lagemaß	Lageparameter. Kennwert. Sagt, wo z.B. das Maximum liegt.
Laplace-Bedingung	$\sigma > 3$. In einer Normalverteilung.
Laplace-Experiment	**P(A) = |A|/|Ω|**. Gleichwahrscheinlichkeit des Versuchsausgangs. Anwendbar bei vielen Zahlen mit kombinatorischen Hilfsmitteln.
Laplace-Regel	P für ein Laplace-Experiment. Siehe oben.
Latente Variable	Aus Ordinationsverfahren der multivariaten Statistik.
Lineare Regression I	Gerade der linearen Regression von Y auf X mit ihrer Steigung a0 als Regressionskoeffizient. Grundsätzliche Betrachtung der Normalverteilung.
Lineare Regression II	Anpassung der Abhängigkeit zwischen zwei Merkmalen (siehe auch multiple lineare Regression) eines Objektes einer Regressionsgleichung.
Lineare Regression III	**y = ax + b**, Parameter a und b werden aus Merkmalsdaten x und y nach Kleinst-Quadrate-Schätzung (KQ oder SQA) berechnet/geschätzt.
Lineare Regression IV	1. a = a0 = KOVAR $s_{X,Y}$ / s_X^2 *plus* b = b0 = Y quer - a0 · X quer, 2. Gerade berechnen mit y = a0·x + b0 mit Steigung a0 · D + Achsenabschnitt b0.
Linksschiefe Verteilung	Die meisten Werte liegen links vom EW. Fällt rechts steil ab.
Logarithmische Normalverteilung	e ^ μ + (σ² / 2). Median e ^ μ. Varianz: e ^ 2μ + σ² oder (e ^ σ²) - 1
Logarithmische Skalen	Richterskala für Erdbeben, pH-Wert, Tonleiter, Lautstärke (dB): Geometrische Mittelwertbildung.
Lorenzkurve	Darstellung von Merkmalen bzgl. ihrer c-Verteilung, z. B. Umsätze mehrerer Unternehmen (Merkmale) bezogen auf einen Markt.
m	Anzahl der Klassen oder Anzahl der möglichen Werte.
μ I	Arithmetische Mittel der Gesamtpopulation.

Was	Bedeutung
μ II	Erwartungswert bei **festgelegten** Verteilungen. Mit P's gewogene arithmetische Mittel der Realisationen.
μ III	Mittelwert in der Grundgesamtheit bei Verteilungsfunktionen. Griechischer Buchstabe: Zur Unterscheidung von der Stichprobe.
Marginalverteilung	Siehe Randverteilung.
Matrizen	Verwendung bei 1. allgemeinem Linearen Modell 2. Varianzanalyse 3. Regressionsanalyse 4. Faktorenanalyse.
Maximum-Likelihood-Methode I	ML. Keine Zufallsvariable. Elegante Methode, falls Verteilungstyp bekannt ist. Entspricht relativer Häufigkeit.
Maximum-Likelihood-Methode II	Beweist, dass man mit K, N, p etc. richtig liegt. Immer einen Graphen erstellen.
Maximum-Likelihood-Methode III	Schätzverfahren: Populationsparameter werden so geschätzt, dass **p** bzw. Likelihood des Auftretens der beobachteten Werte maximiert wird: IQ
Median I	Wird vorgezogen, wenn ein paar wenige Daten Ausreißer sind, also sehr viel höher oder niedriger als die Mehrzahl der Werte sind.
Median II	$x_{0,5}$, für Zwecke der Stoßzeiten usw. benötigt man aber einen weiteren Mittelwert, den harmonischen Mittelwert, robust gegen Ausreißer.
Median III	Zentralwert, größer oder kleiner als 50% der Stichprobe, unterschieldiche Formeln für gerade und ungerade Messreihen.
Median IV	Bei Ordinalskalen verpönt → Man darf keine neue Zwischenkategorie einführen. **R verweigert Berechnung.**
Merkmal	X. (Zufalls-) Variable. Interessierende Eigenschaften der Merkmalsträger, z. B. Augenzahl beim Würfel oder # der richtigen Zahlen beim Lotto.
Merkmalsausprägungen	Mögliche Meßwerte eines Merkmals. Elementarereignisse, z. B. die Zahlen der Menge (1,2,3,4,5,6).
Merkmalsträger	Untersuchungsobjekt, statistische Einheit, Element. Objekte, an denen die interessierenden Merkmale auftreten.
Messfehler	Tatsächlicher Wert minus Messwert.
Metrische Daten	Gemessene Daten führen zu einer Kardinalskalierung.
mindestens	**1-pbinom.** Mindestens 2 ist das Gegenereignis von "höchstens 3". Ω \ A.
Mittelwerte	Arithmetisch, Median, Modalwert, Harmonisch , Geometrisch, Quadratisch (rms). Bei NV sind Arithmetisch, Median und Modus dieselben Werte.
Mittelwerte im Vergleich I	Menge {10, 12, 14, 20}. Arithmetischer MW: (10 + 12 + 14 + 20) / 4 = **14**. Bei additiven Zusammenhängen.
Mittelwerte im Vergleich II	Menge {10, 12, 14, 20}. Median: 12 Δ 14 = **13**. Stabil gegen Ausreißer.
Mittelwerte im Vergleich III	Menge {10, 12, 14, 20}. Modalwert: **{ }**, da kein Wert am häufigsten vorkommt. Ist nicht die # (z.B. dreimal), sondern quantitativer Wert (z. B. 12)
Mittelwerte im Vergleich IV	Menge {10, 12, 14, 20}. Harmonischer MW: 4 / (1/10 + 1/12 + 1/14 + 1/20) = 4 / ((84+70+60+42)/840) = **13,125**, z.B. für Ø-Geschwindigkeiten. Bei **Quotienten**.
Mittelwerte im Vergleich V	Menge {10, 12, 14, 20}. Geometrischer MW: $\sqrt[4]{}$ (10 · 12 · 14 · 20) = $\sqrt[4]{}$ 33.600 = **13,54**. Dann √**Wachstum**sfaktor minus 1 = z. B. 0,89-1 = -0,11 (ØRate)
Mittelwerte im Vergleich VI	Menge {10, 12, 14, 20}. Geometrischer MW: $\sqrt[5]{}$ (1,02 · 1,04 · 1,06 · 1,08 · 1,10) = **1,0596**. 2% wäre kein Wert, sondern der Prozentsatz.
Mittelwerte im Vergleich VII	Teil 2, also z. B. für Ø Zinsen und für logarithmische Skalen, **GEOMITTEL**
Mittelwerte im Vergleich VIII	Menge {10, 12, 14, 20}. Quadratischer MW(rms): √ (10^2 · 12^2 · 14^2 · 20^2) / 4 = √ 840/4 = **14,49**. Selten verwendet.

Was	Bedeutung
Mittlere lineare Abweichung	δ, $1/n \cdot \Sigma \lvert x_i - \mu \rvert$, in Excel: **MITTELABW(B2:B13)**.
Modalwert I	Wert, der am häufigsten vorkommt.
Modalwert II	x_{mod}, größte absolute Merkmalsausprägung, für qualitative nicht-geordnete Daten, für Nominalskalen, wenig Aussagekraft.
Mode	Lokales Maximum.
Moment	Summe der Abweichungen der beobachteten Werte von einem fixen Wert (z. B. arithmetischem MW).
Multiple lineare Regression	Auf abhängiges Merkmal wirken mehrere (multiple) unabhängige Merkmale.
Multiplikations(pfad)regel I	Mehrstufiger Zufallsversuch bei voneinander unabhängigen Ereignissen mit nur einem Pfad. Alle Wahrscheinlichkeiten werden multipliziert.
Multiplikations(pfad)regel I	A + B sind voneinander unabhängig, wenn das Vorkommen von A keinen Einfluss auf p hat, dass B eintreffen wird. P(A und B) = P(A) * P(B)
Multiplikationssatz	$P(A \cap B) = P(A) \cdot P(B)$ oder schwieriger: $P(A \cap B) = P(A) \cdot P(B \mid A)$. P für Ereignis A und Ereignis B treten ein (Schnittmenge → Multiplikationssatz).
Multiplikationssatz - allgemeiner	$P(A \cap B) = P(A) \cdot P(B \mid A) = P(B) \cdot P(A \mid B)$. $P(A \cap B \cap C) = P(A) \cdot P(B \mid A) \cdot P(C \mid A \cap B)$. Stochastische Abhängigkeit und Unabhänggkeit.
Multiplikationssatz - spezieller	Für stochastisch unabhängige Ereignisse A, B und C gilt: $P(A \cap B) = P(A) \bullet P(B)$. $P(A \cap B \cap C) = P(A) \bullet P(B) \bullet P(C)$ für beliebig viele Ereignisse.
Multivariate Analysenmethoden I	Für mehrdimensionale Daten. Multivariat: Ein Objekt ist durch mehr als ein Merkmal ausgeprägt.
Multivariate Analysenmethoden II	Strukturprüfend: Regressions-, Varianz-, Diskriminanz- und Kontingenzanalyse, Logistische Regression, Strukturgleichungsmodelle.
Multivariate Analysenmethoden III	Strukturentdeckend: Faktoren-, Korrespondenz- und Clusteranalyse, Multidimensionale Skalierung, Neuronale Netze.
n	Stichprobenumfang, Anzahl der untersuchten Objekte, absolute Häufigkeit, Versuchszahl, Stoffmenge, Mächtigkeit von **ω**.
n - 1	Wenn man eigentlich EW einsetzen müsste, aber arithmetischen MW einsetzt, teilt man durch n-1.
N	Andere Variable für **Ω**.
Nachweisgrenze xi >= 3 sx (I)	Bestimmungsgrenze, unterer Bereich des Arbeitsbereiches einer Prüfmethode, siehe auch Blindwert. Im einfachsten Fall: xi >= 3 sx.
Nachweisgrenze xi >= 3 sx (II)	Wenn Prüfergebnis (Merkmalswert, xi) 3 mal größer ist als Standardabweichung sx, gilt Ergebnis als gesichert nachgewiesen.
Natürliche Einheit	Dimensionslos, z. B. Konstanten.
Negativer Voraussagewert	$P(K \mid T^-)$. Der Anteil der Gesunden unter den Personen mit positivem Testergebnis.
Nichtparametrischer Test	T-Test, keine Annahme wird über die genaue Form der Verteilung gemacht. Aussagen nur unter bestimmten Voraussetzungen gültig.
Nominalskalierung	Diskrete Ausprägungen ohne Ordnungsrelation. Daten mit geringster Informationsdichte. MW-Bildung nicht sinnvoll. Datentyp: Kategoriale Daten.
Normalverteilung I	Bei NV Daten liegen 95% der Fälle weniger als zwei Standardabweichungen vom arithmetischen Mittel entfernt. z. B. IQ 100 +/- 15.
Normalverteilung II	N. **∞** bis **-∞**. Gaußverteilung. Folgen beobachtete Werte (z. B. Messwerte) einer NV (NV = symmetrische, unimodale, glockenförmige Dichtekurve)?
Normalverteilung III	Der NV zugrundeliegende Zufallsvariable **x** besitzt den Erwartungswert **μ** und die Varianz σ^2. Dichtefunktion = Symmetrische Glocke um **μ**.

Was	Bedeutung
Normalverteilung IV	y-Achse ist die Höhe, auf x-Achse die Ergebnisse.
n-Tupel	Produkt der Mächtigkeiten der beteiligten Mengen. Produktmenge hilft, Tupel zu erstellen.
Nullhypothese I	H_0. Formulierung vor Durchführung eines statistischen Tests. Logisches Gegenteil der Alternativhypothese: Zusammen ergeben beide alle möglichen Ereignisse.
Nullhypothese II	Nachteil von Hypothesen: Sie müssen eine Aussage über zufällige Ereignisse machen.
$\omega_1 \rightarrow \omega_n$	Merkmalsträger.
Ω	Grundgesamtheit, Ergebnismenge.
oder	A oder B. Ergebnis aus {1,2,3} oder {2,3,4} ist {1,2,3,4}. oder bezeichnet also die vereinigende Menge: $A \cup B$.
Ordinalskalierung I	Für (Zufalls-)Variablen mit Ausprägungen, zwischen denen eine natürliche Rangordnung besteht.
Ordinalskalierung II	Abstände der Ausprägungen sind nicht interpretierbar. Nur Bildung des Medians sinnvoll: Streuungsmaße (σ oder σ^2) machen daher keinen Sinn.
Ordinalzahl	Menge, die den Ordnungstyp einer wohlgeordneten Menge repräsentiert.
p	Trefferwahrscheinlichkeit. Faas: "Wir müssen uns davon lösen, dass wir p kennen."
P	$P(\Omega)$, Serie von Versuchen (W bei...).
π	Prozentsatz in der Grundgesamtheit bei Verteilungsfunktionen. Griechischer Buchstabe: Zur Unterscheidung von der Stichprobe.
p-Wert I	Grad der Unwahrscheinlichkeit = Überschreitungs-Wahrscheinlichkeit. Große Werte sagen nur was aus, wenn H_0 gilt.
p-Wert II	Hypothesenentscheidung auf Basis des p-values in der Anwendung von Statistikprogrammen wie z. B. R.
Panelerhebung	Mehrfacherhebung, die sich auf eine repräsentative Teilauswahl (Panel) bezieht. Datenerhebung immer mit selbem Untersuchungsgegenstand.
Parameter I	In Ω: Zur Unterscheidung von der Stichprobe werden Parameter von Ω mit griechischen Buchstaben bezeichnet.
Parameter II	μ Mittelwert in der Grundgesamtheit, σ Streuung in der Grundgesamtheit, π Prozentsatz in der Grundgesamtheit.
Parameterschätzung	Dinge über zufällige Dinge rausfinden. 1000e Versuche: Was kann man über Wahrscheinlichkeit sagen?
Parametrischer Test	t- und Shapiro-Wilk-Test oder Schätzen. Voraussetzung: Vorliegen/Bekanntsein einer bestimmten Verteilung. Ausprägung der Beobachtung: Kardinalskalierung.
Pascal'sches Dreieck	5 über 2 ausrechnen: Zeile 5+1 = 6. Zeile und Stelle 2+1 = 3. Stelle: Wert ist 10.
pbinom	Höchstens. Eine Zahl.
PCA	Principal component analysis, Ergebnis z. B.: PCA1 + PCA2 = 30% --> 30% Einfluss von zwei Faktoren ist viel!
Pearson	Karl Pearson (1857-1936) arbeitete die Chi²-Methode aus.
Pearsonscher Korrelationskoeffizient	Siehe Korrelationskoeffizient.
Permutationen	Möglichkeiten der Reihenfolge.

Was	Bedeutung
Pfadregeln	Multiplikationspfadregel, Addititionspfadregel und Totalwahrscheinlichkeitsregel.
Φ	Standardnormalverteilung.
Plotten	Besser mit anderen Programmen als mit **R**.
Poisson-Verteilung I	Bestes theoretisches Modell für seltene Ereignisse.
Poisson-Verteilung II	3. Frage: $n \geq 100$ und $p \leq 0{,}05$? Setzung: $p=n \cdot p$. Näherungsformel für Verteilung der seltenen Ereignisse. Spezielle Wahrscheinlichkeitsverteilung.
Polygon	Obere Grenze + untere Grenze / 2, anstatt Kategoriengrenzen wie bei Histogrammen betrachtet man Kategorienmitten, Grafik: Lit. inf 508.2 19l S. 127
Poolen	Alle Messergebnisse zusammengenommen (vs. Cluster - siehe dort).
Positive Korrelation	$r_{xy} = +1 \rightarrow y = ax + b$ mit $a > 0$.
Positiver Voraussagewert	$P(K \mid T^{+})$. Der Anteil der Kranken unter den Personen mit positivem Testergebnis.
Prävalenz	(a + b) / n.
Präzision I	Reproduzierbarkeit einer Messung.
Präzision II	Eine Messung ist umso präziser, desto mehr signifikante Stellen im Ergebniswert sind. Wie nah liegen Messwerte beieinander (wahrer Wert egal)?
Präzision III	Ausmaß gegenseitiger Annäherung voneinander unabhängiger Merkmalsergebnisse bei mehrfacher Anwendung einer festgelegten Prüfmethode.
Präzision IV	Beschreibt den zufälligen Fehler (zufällige Abweichung, Varianz). Es gibt die Wiederholpräzision und die ergleichspräzision.
Primärstatistik	Daten werden für eine Untersuchung auf Basis von Erhebungen gesammelt: Field Research mit Beobachtung, Experiment und/oder Befragung.
Probe	Stichprobe.
Prozentualer Fehler	Messfehler / Tatsächlichen Wert x 100.
Quadratisches Mittel	RMS, root mean square, quadratischen Mittelung, im Gegensatz zum geometrischen Mittel: Größere Werte mit stärkerem Einfluss als kleinere.
Quadratische Relation	Wenn Intervalllänge 2, dann Versuchsanzahl x4.
QQ-Plot	Quantil-Quantil-Plot. Ersetzen keine statistischen Tests. Sagt nicht, aus welchem Verteilungstyp (exp, normal) sie kommen. **qqnorm**(ZufZahlen) und **qqline**.
Quantile	Lagemaß. Quantile der Ordnung q = 1 sind q-Quantile. Spezielle Quantile: Median, die Quartile, die Quintile, die Dezile und die Perzentile.
Qualitative Merkmale I	Ein Merkmal bezeichnet eine Eigenschaft, „Qualität“, der untersuchten Objekte. Ausprägungen sind keine echten Zahlen.
Qualitative Merkmale II	Man kann mit ihnen nicht rechnen!
Qualitative Merkmale III	Nominale Variablen sind nur namentlich unterscheidbar: Keine Zahlen, keine Ordnung, nur Namen.
Qualitative Merkmale IV	Ordinale Variablen sind sortierbar, es gibt eine „innere Ordnung“, es kann damit nicht gerechnet werden.
Quantitative Merkmale I	Merkmalsausprägungen sind echte Meßwerte (von Natur aus Zahlen), d.h. addier-, subtrahier- und multiplizierbar.

Was	Bedeutung
Quantitative Merkmale II	Stetige und/oder metrische Variablen können fast jeden Wert und Zwischenwert annehmen, d.h. sie sind beliebig fein meßbar.
Quantitative Merkmale III	Bei diskreten Variablen können nur endlich viele, ganz konkrete Werte angenommen werden (i.d.R. ganzzahlig, keine Brüche).
Quartile I	Für ordinalskalierte Merkmale geeignet und wenn diskrete/stetige Merkmale schief verteilt sind. Excel: **QUARTILE()**, Dimension hat z.B. die Einheit [cm]
Quartile II	Beispiel für 1. Quartil: Trenngrenze bei ≤ 25% der Werte und 75% oberhalb des 1. Quartils.
Quartil 0	Kleinster vorkommender Wert.
Quartilabstand	**QA = Q3 - Q1** oder **Q.75 - Q.25** (engl. Interquartilerange **IQR**). Spannweite der mittleren 50%.
r	Empirischer Korrelationskoeffizient. Komplizierte Formel. Mit a = y quer minus b xquer. Und mit a = Quotient aus ... (komplizierte Formel).
R^2 I	**B**, Bestimmtheitsmaß, Determinationskoeffizient, erklärter Variabilitätsanteil (Varianz) einer abhängigen Variablen Y durch ein statistisches Modell
R^2 II	Sollte bei >95% liegen für gute statistische Aussagekraft.
Randomisierung	Zufällige Zuordnung der Untersuchungsteilnehmer zu den Untersuchungsbedingungen.
Randverteilung	Ergebnisse in Randspalte von Kontingenztafeln. Diskrete und stetige. Absolute und relative Häufigkeiten.
Rationalskalierung	Kardinalskalierung.
Realisationen	Zufallswerte: Die Zahlen der Zufallsgröße, wenn man Elementarereignisse Zahlen zuordnen will.
Rechtschiefe Verteilung	Die meisten Werte liegen rechts vom EW. Fällt links steil ab.
Regression	"... von F auf J...".
Regressionsanalyse	Siehe Lineare Regression.
Regressionsgleichung	**Y = a + bX** mit Achsenabschnitt a und Steigung b. Regression zwischen X und Y. Parameter a und b werden aus Merkmalsdaten gewonnen.
Relative Häufigkeit	Anzahl in Bezug zur Gesamtzahl. **K / n** = Schätzung für die (Treffer-) Wahrscheinlichkeit **p**. **Anzahl mit der spezifischen Ausprägung / n · 100%**.
Residuen I	**e**, Abweichung zwischen Zielgröße y (Vorgabe einer Kalibrierung) und dem berechneten Zielgrößenwert $y^0 = a + bx$. $e_i = y_i - y^0_i$
Residuen II	Ist lineares Modell korrekt? Wenn ja, müssen sich die Werte von e_i im gesamten Bereich von y^0_i ohne erkennbare Struktur um 0 sammeln.
Responsefaktor **RF**	Korrekturfaktor für Messgeräte für falsche Messungen, Rückmeldung.
Richtigkeit I	trueness = accuracy of the mean, Annäherungsmaß vom MW der Merkmalsergebnisse an Bezugswert (kann richtiger oder festgelegter Wert sein)
Richtigkeit II	Beschreibt den systematischen Fehler (systematische Abweichung).
σ I	Standardabweichung.
σ II	Streuung in der Grundgesamtheit bei Verteilungsfunktionen. Griechischer Buchstabe: Zur Unterscheidung von der Stichprobe.
σ^2	Varianz.

Was	Bedeutung
Σ	Summenzeichen. Σ av = "Summe aller av für v von 1 bis ∞.
Satz der totalen Wahrscheinlichkeit	P(F) = P(A) · P(F|A) + P(B) · P(F|B) + P(C) · P(F|C).
Satz von Bayes	P(A|B) = P(A) · P(B|A) / P(BF).
Schätzer	Ist etwas Beobachtetes. Dann Rechnen mit dem, was man herausgefunden hat.
Sekundärstatistik	Wenn Primärstatistikdaten nicht für statistische Untersuchung erhoben wurden, aber z. B. in einer Verwaltung. Untersuchungsdaten sind also schon vorhanden
Semivarianz	Funktion der Entfernung zum Bezugspunkt: Σ quadrierter Punktpaar-Δ im Variogramm. Danach Teilung durch Menge der Punkte. Beim Kriging.
Semivariogramm	Variogramm. Errechnung von räumlicher Varianz.
Sensitivität	$P(T^{+}\| K)$. Der Anteil positiver Testergebnisse unter den Kranken.
Shapiro-Wilk-Test	Überprüfung der Hypothese, ob die Beobachtungen X normalverteilt sind. Signifikanztest.
Sichere Ereignisse	Beim Würfeln eine Augenzahl zwischen 1-6.
signifikant I	Nicht zufällig.
signifikant II - höchst	p ≤ 0,001 (entspricht 99,9 %)
signifikant III - hoch	p ≤ 0,01 (entspricht 99 %)
signifikant IV	p < 0,05 (entspricht 95 %)
Signifikanz I	"Ein in den Daten sichtbares Muster ist nur schwer durch Zufall zu erklären. Also, so der Umkehrschluss, steckt ein System dahinter."
Signifikanz II	d. h. Wenn wirklich nur der Zufall wirken würde, hätte das beobachtete Muster eine WS von höchstens 5% (Signifikanzniveau).
Signifikanz III	Wenn man ein Messergebnis auswertet, soll man nur jene Stellen berücksichtigen, deren Werte sicher sind. Diese Stellen sind signifikant.
Signifikanz IV	Jede Stelle ≠ 0 (6,42 sec hat 3 signifikante Stellen). Die 0 zwischen zwei Ziffern gilt auch als signifikante Stelle (3,07sec hat 3 signifikante Stellen).
Signifikanz V	Werte >1, dazu alle Stellen rechts vom Komma (1,760sec hat 4 signifikante Stellen). Letzte Null kann, muss aber nicht signifikant sein.
Signifikanz VI	Zahlen aus Zählungen (1 Kuh, 2 Kühe, 3 Kühe, ...) und Zahlen in wissenschaftlicher Schreibweise.
Signifikanzniveau I	Irrtumswahrscheinlichkeit **α**. Durch deren Vorgabe (z. B. **α**=0,05) wird WS bestimmt, einen **α**-Fehler zu begehen.
Signifikanzniveau II	WS = 95% für die korrekte Entscheidung, dass eine richtige Nullhypothese nicht zurückgewiesen wird.
Skalenniveaus	Nominal (gleich-ungleich), ordinal (größer-kleiner), Intervallskala (Gleichheit von Differenzen), Verhältnisskala (Gleichheit von Verhältnissen).
Skalenparameter	Regelt die Stauchung. Bestimmt, wie breit ... ist.
Spannweite	**R** = range. **x(n) - x(1)**. Maximalwert minus Minimalwert. Kann sehr leicht durch Extremwerte verzerrt werden.
Spearman-Korrelation	Für geordnete Datensätze = ordinale Skalen (nicht für numerische Skalen). $(\Sigma \text{ der Differenzen})^2 = (d1)^2 + ... + (dn)^2$. Werte: 0 bis 1.

Was	Bedeutung
Spezifität	P(T⁻ | K quer). Der Anteil negativer Testergebnisse unter den Gesunden.
Spline	Funktion mit Knoten, die stückweise aus Polynomen höchstens n-ten Grades zusammengesetzt ist. Für Interpolationen und Approximationen.
Stängel-Blatt-Diagramm	Mit geringem Arbeitsaufwand Skizzierung von vielen Rohdaten.
Standardabweichung I	σ. √ (alle quadrierten Werte summiert) / n - STA^2. √ 1/n · Σ $(X(\omega_i)$ - Xquer$)^2$, z. B. wenn 19 = xquer: √aus 1/40 · Σ $(25^2 - 19^2) + (18^2 - 19^2)$ + ...
Standardabweichung II	In Excel: **STABWN(B1:B12)**. **Empirische STA ist different.**
Standardabweichung III	Durchschnittliche Entfernung aller gemessenen Ausprägungen eines Merkmals vom Durchschnitt. Gleiche Maßeinheit wie untersuchtes Merkmal.
Standardabweichung IV	STA einer diskreten Zufallsvariablen: √ VAR einer diskreten Zufallsvariablen.
Standardabweichung V	Bei normalverteilten Merkmalen: Innerhalb 1 STA Entfernung vom MW liegen 68% aller Antwortwerte. 95% im Umkreis von 2 STA's. > 95%: Ausreißer.
Standardabweichung VI	Variabilität der einzelnen Messdaten.
Standardfehler I	Abweichung der Stichprobenmittel vom Mittelwert. Geht also nur, wenn man x quer und Ω kennt.
Standardfehler II	s / √n . Variabilität der Durchschnittswerte von Stichproben. Der Weg: 1. STA 2. KI 3. Voraussage MW in Ω.
Standardfehler III	Mittlerer Fehler des Mittelwerts. Je mehr Messdaten, desto kleiner ist Standardfehler. Variabilität Durchschnitswerte von Stichproben.
Standardfehler IV	Publikation: Entweder publiziert man STA oder Standardfehler.
Standardisiertes Merkmal	Gut für Vergleichbarkeit. Bei mehreren Werten muss jeder Wert standardisiert werden.
Standardisierung I	Varianz standardisieren, da bei steigender Anzahl Messungen auch steigende quadratische Abweichung.
Standardisierung II	XDach (ω) = Normalverteilter Messwert X(ω) - X quer / s_x: Wieviel STA liegt Schüler/Objekt entfernt vom arithmetischen Mittel ?
Standardisierung III	Für Vergleichbarkeit Δ verteilter ZV's und für Faktorenanalyse. z-Transformation einer ZV: Verwendung beim **z**entralen Grenzwertsatz.
Standardisierung IV	U = X - μ / σ_x. Mit U ~ N(0,1). Standardisierung einer beliebigen Zufallsvariablen X ~ N(μ,σ2). Resultierende ZV hat EW = 0, VAR = 1 und STA = 1.
Standardnormalverteilung I	**z = x - μ / σ** (Transformation). MW=0 und Var=1. Um Verteilungen vergleichen zu können, wird NV standardisiert um zur Standard-NV zu gelangen.
Standardnormalverteilung II	**Φ (z) = x - μ / σ** (Transformation): Dieser Term ist immer anwendbar für Grenzen der Standardnormalverteilung.
Stationarität	**μ + σ²** ist konstant. Ist ein Prozeß. Wichtige Eigenschaft in Zeitreihenanalysen.
Statistik	Benutzt Stichproben und leitet daraus Schlussfolgerung = Inferenz ab.
Stetige klassifizierte Variablen	Stetige Punkte, Zuordnung von P erfolgt über Intervalle: Histogramme.
Stetige kontinuierliche Verteilung	Größere Bereiche. Einzelne Punkte haben die Wahrscheinlichkeit 0. Zweckmäßig: Kumulierte Häufigkeiten grafisch darstellen. Grafik: **Polygone**.
Stetige Merkmale	Für Häufigkeitsverteilung müssen Klassen gebildet werden.
Stetige Zufallsvariable	Werden mit dieser Dichtefunktion beschrieben: ∫ -∞/∞ f(t)dt.

Was	Bedeutung
Stichprobe I	s. Experiment. Stichprobe *der Länge n* . Einzelne unabhängige Probe bei gleicher Situation. Eine Stichprobe beim Würfel wäre Quatsch.
Stichprobe II	Man geht immer davon aus, dass man eine gute Stichprobe hat. Liefert konkrete Konfidenzschätzer. SP muss repräsentativ sein. In R: sample.
Stichprobe III	Eine Stichprobe ist dann repräsentativ, wenn jede Untersuchungseinheit aus der Grundgesamtheit dieselbe Chance hat, ausgewählt zu werden.
Stichprobe Typ A	Mit genau 2 Merkmalswerten.
Stichprobe Typ D	Mit > 2 Merkmalswerten, aber noch in einer Liste angebbar.
Stichprobe Typ S	Sehr viele bis unendlich viele Merkmalswerten (stetig).
Stichproben-Minimum-Größe	Mindestens 30 Stichproben für repräsentative Aussage zum Verteilungstyp, für endliche Ω: Benutzung der Moßig-Formel.
Stichproben-Umfang	Laut WHO: **$0{,}4 \cdot \sqrt{N}$ bzw. $0{,}4 \cdot \sqrt{\Omega}$.**
Stochastik	Befasst sich mit zufälligen Ereignissen (Wahrscheinlichkeitsrechnung).
Stochastische Unabhängigkeit	Wenn P für ihr gleichzeitiges Eintreten, i. e. $\lvert A \cap B \rvert = P(A) \cdot P(B)$ ist, dann sind sie voneinander unabhängig.
Streuung	Wichtig für Signifikanz der Aussage.
Streudiagramm	Von der Urliste ausgehen, zwei unterschiedliche Untersuchungsvariablen. Excel Diagrammtyp: **PUNKT()**.
Streuungsmaß	Eng oder weit verteilt um den Mittelwert? Meist verwendet: Standardabweichung (die empirische).
Subjektive Wahrscheinlichkeiten	In der Summe immer 1, z. B. gibt man für ein Fussballspiel für einen Sieg $p3 = 0{,}2$, für ein Unentschieden $p1 = 0{,}45$ und Niederlage $p0 = 0{,}35$.
Summe der Wahrscheinlichkeiten	1 = 100%.
T	Zufallsverteilung für die Zahl...
t-Test	Für Differenzen, Sollwerte oder den Korrelationskoeffizienten.
T-Test	Parametrischer Test oder Nichtparametrischer Test.
t-Verteilung	Fast so wie Normalverteilung, aber schmäler.
Tests	Jeder Test hat Voraussetzungen, dass er klappt.
Toleranz	Höchstwert minus Mindestwert, obere Grenzabweichung minus untere Grenzabweichung.
Totale Wahrscheinlichkeit	Gewogenes Mittel der Likelihoods.
Transformation	Transformation der Standardnormalverteilung in eine beliebige Normalverteilung und v.v. $U \sim N(0,1) \rightarrow U \bullet \sigma + \mu \sim N(\mu,\sigma 2)$. ~ d.h. "verteilt wie".
Transformierte Messwerte	z. B. z-Transformation mit Φ (z) = 1 / $\sqrt{2\pi}$ ...
Trefferwahrscheinlichkeit	**p**
Tupel I	Tupel der Länge n: Aufzählung von n nicht notwendig voneinander verschiedenen Objekten in einer vorgegebenen Reihenfolge, z.B. (a,a,c,n,x,y)

Was	Bedeutung
Tupel II	u. a. für Produktmengen. Bei $\{0,1\}^4$ steht der Exponent für die Tupeleinträge = Tupelelemente: {(0,0,0,0), (0,0,0,1), ..., (1,1,1,1)1}.
Tupel III	Oft werden bei Laplace-Experimenten Ergebnisse als Tupel oder Mengen dargestellt. Wichtig: Reihenfolge der Einträge (0,1) vs. (1,0).
Unabhängigkeit von Ereignissen	P(A|B)= P(A|C)= P(A|B∩C)= P(A). P(B|A)= P(B|C)= P(B|A∩C)=P(B). P(C|A)= P(C|B)= P(C|A∩B)=P(C): Ereignisse A, B, C sind voneinander unabhängig.
und	A und B. Ergebnis aus {1,2,3} und {2,3,4} ist lediglich {2}. und bezeichnet also die eingrenzende Schnittmenge von beiden: A ∩ B.
Univariate Analysenmethoden	Objekt ist nur durch ein Merkmal ausgeprägt (z. B Wirkstoffgehalt einer Tablette): Eindimensional.
Unmögliche Ereignisse	Wie die Zahlen 7 oder 3,14 beim Würfeln.
Urliste	Enthält alle Werte = alle Informationen.
Varianz I	Mittlere Abweichung vom Mittelwert², mathematische Verzerrung der Abweichung, daher nicht in der Einheit der Merkmale: Aussage schwieriger
Varianz II	Das mit p gewogene arithmetische Mittel aus quadratischen Abweichungen der Realisationen vom EW. Excel: **VARIANZEN(B2:B12)**.
Varianz III	σ^2 (Grundgesamtheit) oder s^2 (Stichprobe = empirisch)
Varianz IV	$s^2 = \Sigma (X_i$ - xquer$)^2$ / n-1 oder $s^2 = (\Sigma X_i{}^2 - (\Sigma Xi)^2/n)$ / n-1. Zweite Formel Taschenrechner-freundlicher.
Varianz V	Multipliziert mit Häufigkeiten des Messwertes: $s^2 = \Sigma f_i (x_i$ - xquer$)^2$ / n-1. Mit Stichprobenumfang $n = \Sigma f_i$.
Varianz VI	Wenn ganzes Ω erfasst wird → ohne FG, also nur durch n teilen. Da fast nie gesamte Ω erfasst wird → erwartungstreu schätzen → n-1 für FG.
Varianz VI	Diskrete Zufallsvariablen: Σ P(Z=x) · (x-E (Z))². E (X²) - (E(X))², bei Würfel 1/6 · (1-3,5)² + 1/6 · (2-3,5)² + ... + (6-3,5)².
Varianz VII	Var (X+Y) = Var (X) + Var (Y). Var (aX + bY) = a² Var (X) + b² Var (Y) → wichtige Formeln. Ausreißer werden stark gewichtet.
Varianz VIII	Bei VAR einer Zufallsvariablen Verschiebungssatz verwenden: [p1·(Wert1)² + p2·(Wert2)² + ...] - EW.
Varianz IX	Die positive Zahl Var(X) = E((X-μ)²) = Σ i=1 bis k (xi-μ)² · P(X=xi) = E(X²) - (E(X))² heißt Varianz von X.
Varianz X	Regeln: 1. VAR (X+a) = Var (X) 2. Var (aX) = a² Var(X) 3. und noch eine weitere.
Varianzanalyse - einfache	Für > 2 nicht verbundene Stichproben.
Variationskoeffizient I	v. Vergleicht Δ STA's unterschiedlicher Stichproben miteinander. **s / x quer** bzw. **· 100%** (für Angaben in Prozent).
Variationskoeffizient II	CV = coefficient of variation. Wie viel % des arithmetischen Mittels ist STA?
Variogramm	Stellt die räumliche Beziehung eines Punktes (regionalisierte Variable) zu Nachbarpunkten dar. Entfernungsstufen ("lag") durch Punktpaare.
Verschiebungssatz	Für VAR von ZV's. Beim Ausrechnen Achtung: Wenn zweimaliges Ziehen, zuerst VAR ausrechnen, dann VAR1 verdoppeln, dann Wurzel ziehen (STA)
Verteilungsformen	Symmetrisch, asymmetrisch, unimodal, bimodal, schmalgipflig, breitgipflig, linkssteil, rechtssteil, u-förmig, abfallend.
Verteilungsfunktion I	Wertemenge jeder Funktion: [0; 1]. Jede Funktion ist monoton steigend. P's der Realisationen werden in Rtg. Zahlenachse kumulativ aufsummiert
Verteilungsfunktion II	Links von kleinster Realisation nimmt Funktion stets Wert 0 an, rechts von größter Realisation stets den Wert 1.

Was	Bedeutung
Verteilungsfunktion III	Σ der P's einer ZV. Die Werte dieser Funktion benennen keine Einzel-P's, sondern die P des Wertes selbst sowie aller kleineren Werte.
Verteilungsfunktion IV	$F_X(x)$ = VF einer Zufallsvariablen X an der Stelle x, bei Gleichverteilungen: Durch Integration (Aufleitung) kommt man zu ihr von der Dichtefunktion
Verteilungsfunktion V	$F_X(x)$ = VF einer Zufallsvariablen X an der Stelle x, bei Exponentialverteilungen: Durch Ableitung kommt man zu ihr von der Dichtefunktion.
Verteilungsfunktion VI	Errechnet sich bei stetigen ZV durch das Integral der Dichtefunktion: ∫ (f(t) dt = [F(t)]a bis b = F (b) - F (a).
Wahrer Wert	Tatsächlicher Merkmalswert, i.d.R. nur ideeller Wert, praktisch nicht ermittelbar, Vermeidung aller Ergebnisabweichungsfaktoren unmöglich.
Wahrscheinlichkeit	Hängt ab von n und wie Zufallsvariablen verteilt sind.
Wahrscheinlichkeiten	A-priori-Wahrscheinlichkeit, A-posteriori-Wahrscheinlichkeit, Vertrauens- und Irrtumwahrscheinlichkeit.
Wahrscheinlichkeit des Komplements	$P(A^{quer}) = 1 - P(A)$
Wahrscheinlichkeitsmaß	Jedem Ereignis wird eine Zahl P(A) zwischen 0 und 1 zugeordnet für Beschreibung einer P-Verteilung. Ist eine Funktion.
Wahrscheinlichkeitsraum	Modell zur zur Beschreibung von Zufallsexperimenten. Maßraum mit (Ω, Σ, P).
Wahrscheinlichkeitstheorie	Arbeitet gegenteilig wie Statistik: Vom allgemeinen (z.B. einer Population) wird auf die (Stich-)probe geschlossen.
Wahrscheinlichkeitsverteilung	
Wiegefehler	Gibt es immer, daher wiegt man öfter.
Wilcoxon-Mann-Whitney-Test	Nicht-parametrischer Test. Vergleicht die Lage zweier unabhängiger Stichproben. Uni Landau: Round-Up!
Wilcoxon-Vorzeichen-Rang-Test	Nicht-parametrischer Test. Vergleicht die Lage zweier abhängiger Stichproben (z.B. Paarvergleiche).
X	Merkmal (eindimensional).
X(ω)	Beobachtungswert.
z I	Normalverteilte Prüfgröße. z-Zahl steht immer für 2 Werte!!! Ergeben Flächenstück unter der Standardnormalverteilung (Dichte).
z II	z = -1,96 bis z = 1,96 hat 95% Fläche, d. h. Intervall wo 95% der Werte liegen lautet: [x quer -1,96 · s < X < xquer + 1,96 · s]
z III	Aus Symmetriegründen: Φ (-z) = 1 - Φ (z)
z-Tabelle	Bedeutung: Benutzung der Tabelle, da Berechnung sonst nur mit Computern möglich wäre.
z-Transformation I	Transformation: z = x - μ / σ., u.a. von σ² zu σ. Umrechnen des Rohwertes x in den z-Wert: z = 190 cm - 175 cm / 15 cm = 1, MW=0 und σ=1.
z-Transformation II	Transformation pro Wert: z = (x - x quer) / s
Zählprinzip für Mengen	Binomialkoeffizient.
Zentraler Grenzwertsatz I	**ZGS**. Die zentrale Aussage der Statistik. Erstaunlich ist: Man landet immer bei einer Normalverteilung, wenn man genug Messwerte hat.
Zentraler Grenzwertsatz II	Xi (i=1,2,...,n) unabhängige Zufallsgrößen mit EWen μi und Varianzen σi², dann Zufallsgröße X= ΣXi den EW = Σμi und VAR Σσi². P(X≤x)~Φ(x-μ/σ).

Was	Bedeutung
Zentraler Grenzwertsatz III	Summenverteilung kann beliebig sein. Ab n > 30 ergibt sich brauchbare Annäherung. Sind die Xi normal verteilt, dann ist X exakt normal verteilt.
Zentralwert	Median. 50% der Werte unter dem Median, 50% darüber.
Zufallsereignis	Steht im Gegensatz zum Elementarereignis. Oft in wissenschaftlichen biologischen Versuchen mit vielen Variablen und möglichen Resultaten.
Zufallsexperiment	Experimente mit ungewissem Ausgang (Würfel oder Lotto), beliebig oft wiederholbarer Vorgang nach bestimmter Vorschrift ausgeführt.
Zufallsgröße	P(X), Funktion, Zufallsvariable. Elementarereignissen werden Zahlen zugeordnet.
Zufallsvariable I	Jedem Ergebnis eines Zufallsexperiments wird eine reelle Zahl zugeordnet, z. B. beim Würfeln Zuordnung einer Zahl von 1-6 zu jedem Wurf.
Zufallsvariable II	Meist große Buchstaben. Realisierung im konkreten Versuch kleine Buchstaben. z.B. die Summe beider Augenpaare beim Würfeln.
Zufallsvariable III	Kopf oder Zahl. Grund: Ich will damit rechnen. Hat endlich oder unendlich verschiedene diskrete (abzählbar unendlich) Werte.
Zufallsvariable IV	Zahl der Augen unter den Gezogenen (A), also ZV A: 6 aus 49 vorher ausgewählt. 49 Tiere N, 6 kranke Tiere, n Tiere fangen --> wieviele davon sind krank?
Zufallsvariable V	Vorteile: Wenn unabhängig, kann man besser rechnen.
Zufallsvariable VI	Z gibt p an, wie oft rote Kugel gezogen wird, wenn von 10 Kugeln (3 rot, 7 grün) vier gezogen werden, davon zwei rote sein sollen.
Zufallsvariable (diskret)	Nimmt Zufallsvariable x diskrete Werte an, spricht man von Wahrscheinlichkeitsfunktion, hat endlich viele oder abzählbar unendlich viele Werte.
Zufallsvariable (konstant)	Nimmt nur einen Wert an. Spezialfall der diskreten Zufallsvariable.
Zufallsvariable (stetig)	Für stetige Zufallsvariablen wird die Verteilungsfunktion durch Integration über die Wahrscheinlichkeitsdichte berechnet: **F(x) = ∫+/-∞ f(x)dx**.