Biometrie-Stichworte

Merkmale
- Kodierung mit Grossbuchstaben (X, Y), Merkmalswerte in Kleinbuchstaben (x, y).

- Qualitativ: Auswahl aus Kategorien.
- Quantativ: Numerische Werte (Ordinalskala)
  - § Diskret: Nur endliche Werte.
  - § Stetig: Alle reellen Zahlen.

Häufigkeit

- Absolut: Absolute Anzahl.
- Relativ: Häufigkeit / Anzahl Beobachtungen [%].

Quartilen: Obere (75% Quantile) und untere (25% Quantile) Quartile.
Median: x bei f(x)=0.5.

- Bei symmetrischen Verteilungen (symmetrische Abweichung = Schiefe [γ]) ist Mittelwert = Median, bei Rechtsschiefe ist Mittelwert > Median.
- Durch Logarithmierung der Werte rechtsschiefer Verteilungen kann die Schiefe teilweise ausgeglichen werden.

Mittelwert (<u>kleinste</u> quadratische Abweichung): Messwertsumme / Anzahl Messungen: $\bar{x}=\frac{\Sigma_n}{n}$.

Varianz: <u>Mittlere</u> quadratische Abweichung: $s_x^2=\frac{1}{n-1}\Sigma(x_i-\bar{x})^2$

Streuung (Standardabweichug): √Varianz.
Ausreisser: Messwerte die >1.5 Quartilsabstände vom Median abweichen (oben und unten).

Kontingenztafeln (Zweidimensionale Verteilungen)

→ Zusammenhänge zwischen Untergruppen (z.B. Dosis – Nebenwirkung).

		Ereignis		
Exposition		+	-	
	+	**a**	**b**	(a+b)
	-	**c**	**d**	(c+d)
		(a+c)	(b+d)	n

Freiheitsgrade: (Spalten-1)x(Zeilen-1)

- **Erwartete Häufigkeit** (sollte >5 sein): $\frac{\text{Zeilensumme}\times\text{Spaltensumme}}{\text{Gesamtzahl}}$.
- **Relatives Risiko**: $\boxed{RR=\frac{a/(a+b)}{c/(c+d)}}$, abhängig von der Gruppengewichtung (Interpretation!) bezogen auf die <u>Gesamtzahl</u> (nicht in Fall-Kontroll-Studien verwendbar).
- **Odds-Ratio**: $\boxed{OR=\frac{a/b}{c/d}=\frac{a g d}{b g c}}$, bezogen auf die <u>Gegenzahl</u>, verwendbar für Fall-Kontroll-Studien (unbekannte Kontrollfall-Anzahl). Bei 1 sind die Chancen gleich. Entspricht P/(1-P).
- **Abweichungsmass** χ^2 (chi^2): Abweichung <u>erwarteter</u> von <u>beobachteter</u> Häufigkeit (analytische Statistik) unter der Bedingung, das kein Zusammenhang erwartet wird.
  Ist χ^2>kritischer Wert (Tabelle S.73) ist H_0 widerlegt.

$$\chi^2=\Sigma\frac{(\text{Beobachtet-Erwartet})^2}{\text{Erwartet}}=\frac{(a\times d-b\times c)^2\times\text{Gesamtzahl}}{(a+b)\times(c+d)\times(a+c)\times(b+d)}$$

Streudiagramme (Vergleich quantitativer Variablen)

→ Nur Tendenzen können angegeben werden, keine Gültigkeit für Einzelfälle.

- **Kovarianz s_{xy}**: Enge von Zusaammenhängen durch Bildung des Druchschnitts der **Kreuzprodukte** = Abweichung der Wertepaare vom <u>Mittelwert</u>.

$$s_{xy}=\frac{1}{n-1}\Sigma(x_i-\bar{x})\cdot(y_i-\bar{y})$$

- § Gross, positiv: Starke Abweichung.
- § Klein: Geringe Abweichung.

- **Korrelationsprodukt r_{xy}**: Linearer Zusammenhang zwischen x,y mit Wertebereich (-1) –(+1).

$$\boxed{r_{x,y}=\frac{s_{x,y}}{s_x \times s_y}}\text{ bzw. }\frac{s_{x,y}}{\sqrt{s_x^2 \times s_y^2}}$$

  - § Bei +/- 1liegen <u>alle</u> Werte auf einer Gerade.
  - § Bei $r_{xy}=0$ liegt kein Zusammenhang vor.
  - § (-) = negative Steigung, (+) = positive Steigung (vgl. Regressionskoeffizient).

- **Bestimmtheitsmass r^2**: Durch x bedingter Anteil (0.5 = 50%) der Varianz von y.

$$r^2=\frac{s_y^2-s_{y|x}^2}{s_y^2}$$

- **Regressionsgerade**: Erwarteter Mittelwert (kleinster quadratischer Abstand) für y bei bekanntem x (y=a+b*x).

$$s^2_{y|x}=\frac{1}{n-1}\Sigma(y_i-(a+bx_i))^2 \quad \text{mit a=Achsenabschnitt, b=Steigung.}$$

  - § Bestimmung über „Methode der kleinsten Quadrate“, die Abweichungsquadrate von der Regressionsgeraden sind geringer als die vom Mittelwert.
  - § Verläuft durch den Schwerpunkt ($\bar{x},\bar{y}$), Steigung = Regressionskoeffizient.
  - § Bei Änderung von x um eine Einheit ändert sich y dementsprechend.

Wahrscheinlichkeitsrechnung

→ Die in einem Zufallsexperiment auftretende relative Häufigkeit des Ereignisses A [**h_n(A)**] konvergiert bei hoher Messwiederholung unter gleichen Bedingungen gegen den Grenzwert **P(A)** = Wahrscheinlichkeit (Propability, 0-1).

- $A \cup B$ **(Vereinigung)**: A oder B tritt ein. $\boxed{P(A)+P(B)-P(P(A \cap B))}$
- $A \cap B$ **(Durchschnitt)**: A+B tritt ein.
- $\bar{A}$ **(Komplement)**: A tritt nicht ein [P($\bar{A}$)=1-P(A)].

  - § „**Disjunkte**“ Ereignisse (1 Merkmal) schliessen sich gegenseitig aus (z.B. Hypertonie + Hypotonie), die Wahrscheinlichkeiten addieren sich: $P(A \cup B)=P(A)+P(B)$.
  - § „**Non-Disjunkte**“ Ereignisse (2 Merkmale, z.B. mehrere Risikofaktoren) können sich überschneiden.

Beispiel: Würfeln, Summe aller Ereignisse (**Ω**): A (1), B (2), D (3), D (4), E (5), F (6).
- Alle Ereignisse sind disjunkt → Wahrscheinlichkeiten addieren sich: P(Ω) = 1 = P(A)+...+P(F).
- Modellannahme: Alle Elementarereigneisse sind identisch (Experimentalbedingungen)

→ 6*P = 1 → P = $\frac{1}{6}$ oder allgemein:

- **Laplace-Regel** $P(A)=\frac{n_A}{N}$ mit n_A = Elementarereigniss von A; N = Anzahl Elementarereignisse.

Bedingte Wahrscheinlichkeit (Abhängigkeit von Ereignissen voneinander)

- Abhängigkeit **P(B|A)**: Das Auftreten eines Ereignisses beeinflusst ein anderes Ereignis (Therapiemethode – Therapierfolg).

$$\boxed{P(B|A)=\frac{P(A \cap B)}{P(A)}}\text{ mit B = relative Häufigkeit; A = Bedingung.}$$

  (Lies: „*Bedingte Wahrscheinlichkeit von B gegeben A*“)
- **Unabhängigkeit**: Die Wahrscheinlichkeit für B ist nicht von A abhägig

$$P(B|A)=P(B) \quad > \quad \frac{P(A \cap B)}{P(A)} \quad > \quad P(A \cap B)=P(A) \times P(B)$$

[Durchschnitt=Produkt der Einzelwahrscheinlichkeiten]

Beispiel: 2-facher Münzwurf (wenn A,B unabhängig): $P(A \cap B)=P(A)\times P(B)=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$

Satz von Bayes (Zerlegung von Wahrscheinlichkeiten)
Gewichtung nach Grösse und Untergruppen.

Allgemeine Zerlegung: $P(B)=P(B|A)\times P(A)+P(B|\bar{A})\times P(\bar{A})$ mit P(A): Prävalenz von A.

Bestimmung der Prädiktheit aus Prävalenz, Sensitivität und Spezifität (S. 55):

$$P(A|B)=\frac{P(B|A)\times P(A)}{P(B|A)\times P(A)+P(B|\bar{A})\times P(\bar{A})} \quad \text{bzw.} \quad P(K+|T+)=\frac{P(K+\cap T+)}{P(T+)}=\frac{a/n}{(a+b)/n}$$

$$\text{Positiver prädiktiver Wert}=\frac{\text{Sensitivität}\times\text{Prävalenz}}{\text{Sensitivität}\times\text{Prävalenz}+(1\text{-Spezifität})\times(1\text{-Prävalenz})}$$

$$\text{Negativer prädiktiver Wert}=\frac{\text{Spezifität}\times(1\text{-Prävalenz})}{\text{Spezifität}\times(1\text{-Prävalenz})+(1\text{-Sensitivität})\times\text{Prävalenz}}$$

Testgüte

- **Prävalenz:** Vorhandene Ereignisse [P(Krankheit+)]; Sensitivität / n.
- **Sensitivität:** Richtige Erkennung vorhandener Ereignisse [P(Test+|Krankheit+)] = a/(a+c).
- **Spezifität:** Keine Erkennung nicht eintretender Ereignisse [P(Test-|Krankheit-)] = d/(d+b).
- **Positive Prädiktheit:** P(Krankheit+|Krankheit+).
- **Negative Prädiktheit**: P(Krankheit-|Krankheit-).
  - Bei geringer Prävalenz ist der prädiktive Wert (auch bei hoher Sensitivität und Spezifität) sehr klein.
- **ROC-Kurve** (Sensitivität ↔ 1-Spezifität): Fähigkeit der Trennung von Untergruppen durch Parameter → Festlegung von Schwellenwerten (möglichst hohe Spez. und Sens.).
- **Youden-Index** (Sensitivität + Spezifität-1): Bewertung von Sensitivität und Spezifität.
- **AUC**: Diagnostische Güte durch Fläche unter ROC-Kurve, bei AUC=0.5 sind beide Untergruppen gleichmässig verteilt.

	Tatsächlich			
Diagnose		+	-	
	+	**a**	**b**	a/(a+b)=Positive Prädiktheit
	-	**c**	**d**	d/(c+d)=Negative Prädiktheit
		a/(a+c)=Sensitivität	d/(d+b)=Spezifität	

Vierfelderschema nur für Querschnittuntersuchungen gültig!

Mittelwert und Varianz von Wahrschinlichkeitsverteilungen

- μ: Erwarteter Wert für die Verteilung von Y, entspricht Mittelwert (Addition der gewichteten Häufigkeiten).
  $\mu=\sum w_i \times p_i$ mit ω_i=mögliche Werte von Y, p_i=Wahrscheinlichkeit vonP(Y= ω_i).
- σ^2: Erwartete Varianz
  $$\sigma^2=\sum (w_i \times \mu)^2 \times p_i$$

Bernoulli/Binominal-Verteilung

Experiment mit 2 möglichen Ausgängen (Erfolg = p, Kein Erfolg = 1-p) nach n unabhängigen Wiederholungen. Bei gleicher Erfolgswahrscheinlichkeit für alle Versuche ist p = P(x_i=1).

- T_n: Anzahl Erfolge (p) nach n Versuchen.

Normalverteilung (Gauss-Verteilung)

- Symmetrische Verteilung (Glockenkurve) um den Mittelwert **μ** = $\bar{x}$.
- Summe der Einzelverteilungen = Integral der Kurve (Gesamtfläche = 1).
- Durch die Dichtefunktion lässt sich die Wahrscheinlicheit von Intervallen (a-b) bestimmen.

- Je grösser die Varianz, desto flacher der Kurvenverlauf.
- Eine Normalverteilung kann durch **Z-Transformation** (s) in eine **standardisierte Verteilung** mit N ~ 0 (Erwartungswert μ), 1 (Varianz σ) umgewandelt werden:

$$\boxed{Z=\frac{x-\bar{x}}{s}=\frac{x-\mu}{\sigma}}$$

Kumulative Verteilungsfunktion ($F_{(y)}$: Fläche bis y)

Einteilung anhand der **Standardabweichung σ** (Tabelle S. 73):

- **$^{+}/_{-}$ 1σ ≈ 68%.**
- **$^{+}/_{-}$ 2σ ≈ 95%** (-1.96 - +1.96)**.**
- **$^{+}/_{-}$ 3σ ≈ 99%.**

Zentraler Grenzwertsatz

Bei grossen n sind die Wahrscheinlichkeiten für Mittelwerte symmetrisch verteilt, nach Z-Transformation ergibt sich eine Normalverteilung.

Schätzen

- Gesetz der grossen Zahl: Der Mittelwert strebt gegen den Erwartungswert μ.
- Erwartungswert des Mittelwertes $(\bar{Y}_n)$: $\boxed{E(\bar{Y}_n)=\mu}$.
- Erwartungswert der Varianz (s^2_y): $\boxed{E(s_y^2)=s^2}$.

Standardabweichung von Mittelwerten

- Die Mittelwerte liegen enger beieinander als Einzelwerte; die Varianz sinkt dabei um den Faktor $\frac{s^2}{n}$.
- Abschätzung aus nur 1 Stichprobe: Nach $s_x=\frac{s}{\sqrt{n}}$ ergibt sich $\boxed{s_{\bar{Y}}=\frac{s}{\sqrt{n}}}$.
- Standardfehler (Streuung) des Mittelwertes: $\boxed{SEM=\frac{s_y}{\sqrt{n}}}$.
- Verteilung des Mittelwertes = **T-Verteilung**: $\boxed{Z_n=\frac{\bar{Y}_n-\mu}{SEM}=T_n}$; Geschätzte Verteilung (s^2_y) mit n-1 Freiheitsgraden.
  - § $t_{n-1,\,97.5\%}$: 97.5% Quantil der T-Verteilung.
  - § α: Ausserhalb der T-Verteilung liegender Bereich ($t_{n-1,\,1-\frac{\alpha}{2}(\hat{=}97.5\%)}$).

Konfidenzintervall (CI)

Verhältnis empirischer Daten ($\bar{x}_n$) ↔ Wahrscheinlichkeit (μ).

- 95% Konfidenzintervall → max. 5% Abweichung von der Beobachtung.
- Beim Einsetzen in die Formel der T-Verteilung muss T innerhalb des Erwartungsintervalls (für n-1) liegen.
- Liegt 0 nicht im Konfidenzintervall wird H_0: μ=0 zurückgewiesen.
- Mit aus Tabellen gegebenem **t** berechnet sich das Konfidenzintervall nach:

$$\boxed{\bar{y}\pm\frac{s_y}{\sqrt{n}}\times(t_{n-1,\,97.5\%})}\quad\boxed{y=\mu\pm(1.96\times s)}$$

Statistisches Testen

Signifikanztest mit **Nullhypothese** H_0: Nachweis von Zusammenhängen durch <u>zurückweisen</u> der H_0 wenn μ_0 ausserhalb des Konfidenzintervalls (bzw. innerhalb α) liegt. Ein signifikantes Ergebnis bedeutet jedoch keinen Beweis!

- Fehler 1. Art (**α**): Zurückweisen einer richtigen H_0, α wird <u>vorher festgelegt</u> (i.d.R. 0.05 = 95% Konfidenz).
- Fehler 2. Art (**β**): Beibehalten einer falschen H_0, sinkt mit steigendem *Stichprobenumfang* und der *Testgüte* (Power, 1-P); steigt mit der *Varianz* σ^2 .
- Power (Güte): Signifikanz gegebener Zusammenhänge (1-β).

- Wenn T>t (T-Test) gilt die H_0 als widerlegt.

		Tatsächlich	
		H_0	H_1
Test	H_0	**1-α**	**β**
	H_1	**α**	**1-β (Power)**

Normalverteilung: N(0,1) ~ $\boxed{\frac{\overline{Y}_n-\mu}{s}\times\sqrt{n}}$.

T-Verteilung für **verbundene** (normalverteilte) Stichproben [97]

$$t_{n-1} \sim \boxed{\frac{\overline{Y}_n-\mu}{s}\times\sqrt{n}}\text{; wenn } \mathbf{H_0: \mu=0} \rightarrow \boxed{T=\frac{\overline{Y}_n}{s}\times\sqrt{n}} \text{ mit } (s=\sum\frac{(x_i-\overline{x})^2}{n-1}).$$

- Voraussetzung für T-Verteilungen: Normalverteilung N(0,1), gleicher Varianz **σ**.
- 2x Merkmalskontrolle in einer Gruppe.
- Wenn $|T|>t_{n-1,\,97.5\%}$ (t aus Tabellen) wird die H_0 abgelent.
- Die Wahrscheinlichkeit der Ablehnung soll (bei richtiger H_0) max. 5% (**α**=0.05) betragen.
- Bei <u>einseitigen</u> Tests liegt der gesamte α-Bereich über <u>oder</u> unter dem μ, durch den verringerten kritischen Wert λ steigt die Testgüte (Power).
  - § Rechtsseitig: $T \geq t_{n-1,\,1-\alpha}$.
  - § Linksseitig: $T \leq t_{n-1,\,\alpha}$.

T-Verteilung für **unverbundene** (normalverteilte) Stichproben [100]

- Verteilung <u>einer</u> Variable in 2 Stichproben mit <u>verschiedenen</u> Probanden.
- Hypothese $\mathbf{H_0: \mu_x=\mu_y}$ (keine Unterscheidung der Gruppen), Nachweis durch Differenz der Mittelwerte $(\overline{X}_n-\overline{Y}_n=0)$ bei unbekannter Varianz mit 2 Freiheitsgraden (n-2 da 2 Stichproben):

$$\boxed{T=\frac{\overline{x}_n-\overline{y}_n}{s_{(\overline{x}_n-\overline{y}_n)}}} \text{ mit s (Standardfehler) } = \sqrt{(\frac{1}{n_x}+\frac{1}{n_y})\times\frac{(n_x-1)\times s_x^2+(n_y-1)s_y^2}{n-2}} \;\; \mathbf{A}\; SEM=\frac{s}{\sqrt{n}}$$

P-Wert (Signifikanz) [94]

- Wenn P<5% ist eine Signifikanz gegeben → die Differenz ist grösser als per Zufall zu erwarten!
- Beobachtet (nicht tatsächlich = kein Beweis!) wird die Differenz zur H_0.
- Unter der H_0 ist die Abweichung = aktueller Wert oder grösser (aber nie negativ!).
- Bei grosser Fallzahl ist auch P hoch, was aber keine klinische Relevanz bedeutet!

χ^2 (chi²) [109]

- Merkmalszusammenhänge aus 4-Felder-Tafel (bei gegebener Randsumme) → Anweichung von der Annahme unabhängiger Merkmale (Erwartungshäufigkeit in Zellen).
- Zurückweisung der H_0 wenn χ^2 <u>gross</u> → in wenigstens einer Zelle sind Unterschiede vorhanden.

$$\boxed{\chi^2=\frac{n\times(a\times b)-(b\times c)}{(a+b)\times(c+d)\times(a+c)\times(b+d)}}$$

- Bei einer Gesamtzahl <20 und einer 2x2 Tabelle kann der exakte Wert durch den Fischer-Test bestimmt werden [111].
- H_0: $P(x_i, y_j) = P(x_i, x_j)$

Überlebenszeit-Analyse

Dauer bis Ereigniseintritt (i.d.R. Tod), tritt das Ereignis zum Beobachtungsende nicht ein wird der Proband zensiert.

- Besser geeignet ist die 5-Jahres-Überlebensrate [116] mit geringerer Inzidenz zensierter Fälle, da mehr vom Endpunkt (5a) unabhängige Informationen verwendet werden können.
- **Lifetable-Methode** [118]: Überlebensrate in mehreren Intervallen.

  Sterberate (1. Intervall): $\frac{\text{Verstorbene (1. Intervall)}}{\text{Patienten (Beginn)-Zensierte Fälle (1. Intervall)}}\times 2$

  Überlebensrate (1. Intervall): 1-Sterberate.

  Überlebende nach n Intervallen: In das Intervall eintretende Überlebende x Überlebenswahrscheinlichkeit.

- **Kaplan-Meier Methode** [121]: Bildung von sehr kleinen Intervallen („Hazards“) in der Lifetable-Methode.
- **Logrank-Test** [124]: χ^2 (mit 1 Freiheitsgrad: *df*1) aus Überlebenskurve.
  H_0: Wahrscheinlichkeit für jedes Intervall ist für alle Gruppen gleich.