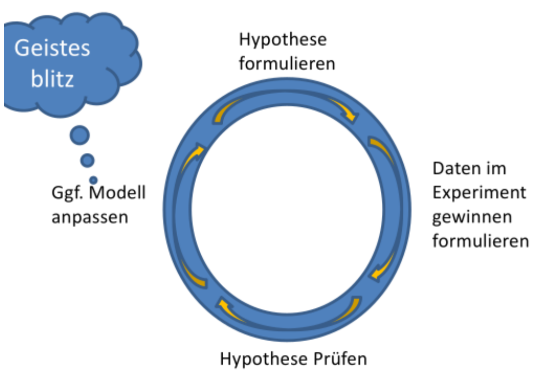
1 TODO Versuchsplanung

1.1 TODO Grundbegriffe

- **Testverteilung**: Können benutzt werden um Ergebnisse von Experimente miteinander zu vergleichen
- **Hypothese**: Eine Annahme
- **Zielgrösse**: Ergebnis eines Versuchs
- **Einflussgrössen**:
  - **Störgrössen (Rauschen)**: können nicht eingestellt werden
  - **Steuergrösse**: können eingestellt und gehalten werden
- **Faktoren**: für den Verusch wesentlichen Einflussgrössen
- **Faktorarten**:
  - **Quantitativ**: beschreiben Zahlenwerte auf der Messskala
  - **Qualitativ**: Name, Beschreibung, Bezeichnung

1.2 Wissenserwerbs Shewhart



1.3 DoE

1. Ausgangssituation und Problem beschreiben: Problem wirklich verstanden?
2. Untersuchungsziel festlegen: Was wollen wir erreichen?
3. Zielgrössen und Faktoren festlegen
4. Entscheidung: Soll das Problem analytisch / experimentell gelöst werden? Nein => Ende
5. Planung Versuchplan aufstellen, Blockbildung, Aufwandsabschätzung
6. Entscheidung: Experiment am realen Object oder Model durchführen?
7. Nachbearbeitung Versuch durchführen, Ergebnisse auswerten, intepretieren und Massnahmen ableiten, Absicherung, Dokumentation und weiteres Vorgehen

1.4 Daten sammeln

Was ist der Anlass, Daten zu sammeln?

- Reine Sammelwut (Big Data, Data Minig) um im Nachgang Zusammenhänge zu finden
- bewusstes erfassen von Daten (Data-Farming) um Normen einhalten zu könne, Schwachstellen zu erkennen

1.5 Statitische Grundbegriffe

Merkmalsträger (Mensch, Maschine, etc.). **Grundgesamtheit** ist die Menge aller Merkmalsträger. **Merkmal** (Augenfarbe, Ausfallrate) sind die Eingeschaften des Merkmalträgers. **Merkmalswert** (Blaue Auge, 5mal pro Stunde) sind die Ausprägung der Merkmale.

1.6 Skalen

- **Nominalskala** (qualitative Merkmale wie z.B. verheiratet, Feminin / Maskulin, etc.)
- **Ordinalskala** (Klassen, Rangfolge, sehr gut, gut, ...)
- **Metrische Skala** (reelen Zahlen, quantitative Merkmale)
  - **Intervallskala** (willkürlicher Nullpunkt Temperatur in Grad Celsius
  - **Verhältnisskala** (absoulter Nullpunkt, Gewicht in kg)

1.7 Ablauf

Datenerhebung, **Datenaufbereitung** und **-darstellung**, **Datenanalyse** und **-intpretation**

2 Fehlerfortpflanzung

Absoulter Fehler (Addiert sich bei Summer und Differenz). **Relativer Fehler** (Addiert sich bei Produkt und Quotiente). Berechnung mittels Partiele Differnatiation
Fehlerabschätzung ohne Angaben:

- min. halben Einheit von letzten signifikante Stelle
- max. 4 Einheiten der letzten signifikante Stelle

Relativer Fehler

$$\Delta x / x$$

Partiele Differnetiation

$$E = \frac{xy^2}{z}$$
$$\Delta E = |\frac{\delta E}{\delta x}\Delta x| + |\frac{\delta E}{\delta y}\Delta y| + |\frac{\delta E}{\delta z}\Delta z|$$

3 Häufigkeitsverteilung

Nach Erhebung liegen die Daten in Form einer **Urliste** vor (geordnet nach Erfassung).

3.1 Einfach Häufigkeisverteilung

Gibt an wie häufig ein Merkmalswert x_i aufgetreten ist (absolut oder relativ). h_i entspricht der **absoluten** Häufigkeits, f_i der **relativen Häufigkeit**

$$f_i = \frac{h_i}{n}$$

3.2 Kumulierte Häufigkeitsverteilung

Die kumulierte *absoulte* Häufigkeitsverteilung gibt die Anzahl Objekte kleiner oder gleich x_i

$$H_i = \sum_{i=1}^{k} h_i$$

Die kumulierte *relative* Häufigkeitsverteilung gibt den Anteil der Messungen an mit kleiner oder gleich x_i

$$F_i = \sum_{i=1}^{k} f_i$$

3.3 Klassifizierte Häufigkeitsvereilung

Bei mehr als 15 verschieden Merkmalswerten sollte die Werten in Klassen gruppiert werden, um eine überschaubare Darstellung zu erreichen. Die Anzahl Klassen sollte nicht $\sqrt{n}$ überschreiten.

- **Klassenbreite**: $b_m = x_m^o - x_m^u$
- **Klassenmitte**: $\frac{x_m^o + x_m^u}{2}$
- **Klassendichte**: $d_m = \frac{h_m}{b_j}$

4 Prognosen

Formale Abhängikeit (Zahlenmässiger Zusammenhang, muss nicht zwingend ein echter **Zusammenhang** haben). Sachliche Abhängikeit (Ist Anstieg von a Ursache für den Anstieg von b). Korrelation (Stäre des linearen rechnerischen Zusammenhangs)
Kovarianz:

$$\sigma_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$
$$\sigma_{XY} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x}\ \overline{y}$$

Variante 2 Analog zu der Varianz
Korrelationskoeffizient (-1 - 1):

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad mit\ |\sigma_{XY}| \leq \sigma_X \sigma_Y$$
$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{(\sum_{i=1}^{n}(x_i - \overline{x})^2)(\sum_{i=1}^{n}(y_i - \overline{y})^2)}}$$

gleitender Mittelwert

$$M_t = \frac{1}{T+t}(\sum_{T=1}^{t-1} x_T + x_t)$$

Lineare Regression funktioniert nur mit linearen zusammenhang. Bei einem exponentiellen kann dieser aber logrimiert werden. Bei einem logrimischen Zusammenhang kann potenziert werden um einen linearen Zusammenhang zu erhalten.

$$y = a_1 + b_1 x_i$$
$$a_1 = \overline{y} - b_1 \overline{x}$$
$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\ \overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}$$

Newton - Algorithmus (TI-Nspire):

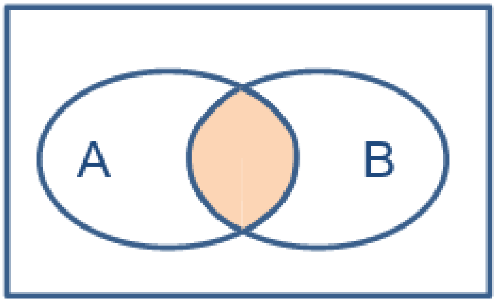
$$f(x) = a_0 + a_1(x - x_1) + a_2(x - x_1)(x - x_2) + ... + a_n(x -$$

5 Wahrscheinlichkeit

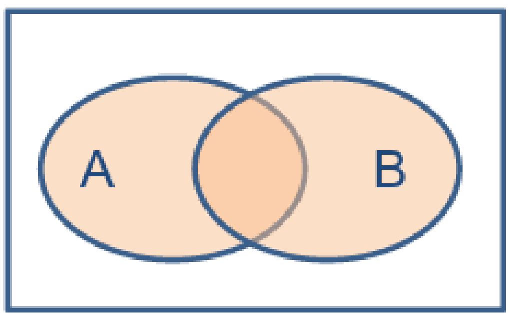
- Zufallsexperiment
- Ergebnismenge / Ergebnisraum Ω (Alle mögliche Ergebnisse eines Zufalsexperiment)
- (System A der) Ereignisse (Teilmenge der Ergebnismenge)
- Wahrscheinlichkeitsraum
  - Beschrieben durch Ereignismenge, System der Erignisse und Wahrscheinlichkeitsmass P

Logisch	Menge
A oder B treffen ein	A $\cup B$
A und B treffen ein	A $\cap B$
Nicht A trifft ein	A^C

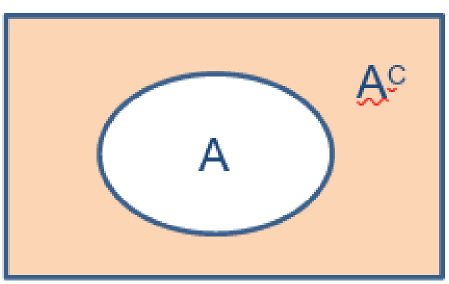
- Bedingte Wahrscheinlichkeit
  - $P(A|B)$



A und B

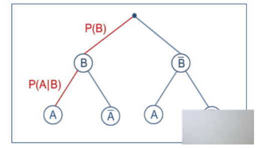


A oder B



Das zu A komplementäre Ereignis

$P(A \cap B) = P(B) \cdot P(A|B)$



- Laplace-Experiment
- stochastisch unabhängig

Wahrscheinlichkeitsmass-Regeln:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cap B) = P(A) \cdot P(B)$$

$$P(A \cap B) = P(B \cap A)$$
$$P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$$

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

Baysschen Regel

$$P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$$

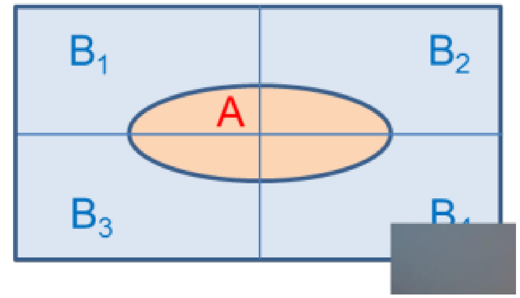
Laplace-Experiment

$$P(A) = \frac{|A|}{|\Omega|}$$

Vollständige Wahrscheinlichkeit:

$$\bigcup_i^n B_i = \Omega \text{ und } P(B_i) > 0 \forall i$$

$$P(A) = \sum_i^n P(A|B_i) \cdot P(B_i)$$
$$= \sum_i^n \frac{P(A \cap B) \cdot P(B)}{P(B)}$$
$$= \sum_i^n P(A \cap B)$$



6 Kombinatorik

- Mengen (Reihenfolge unwichtig)
- Tupel (Reihenfolge **wichtig**)
- Urnenexperiment
  - Kugel zurücklegen (wiederholen)
  - Reihenfolge beachten (geordnet)

- Geordnete Proben mit Wiederholung
  - Pin Code iPhone (4 Stellen, 10 Ziffern entsprechen 10^4 Möglichkeiten)
- Geordnete Proben ohne Wiederholung
  - Anzahl Möglichkeiten: $A = n(n-1)(n-2)...(n-k+)$
  - Formel: $A = \frac{n!}{(n-k)!}$

Anzahl Permutation einer n-Menge:

$$n!$$

Binominal Koeffizient:

$$\binom{n}{k} = \frac{n(n-1)..(n-k+)}{k!}$$

Reihenfolge relevant, mit Zurücklegen:

$$A = n^k$$

Reihenfolge nicht releveant, mit Zurücklegen:

$$A = \binom{n+k-1}{k}$$

Reihenfolge relevant, ohne Zurücklegen:

$$A = \frac{n!}{(n-k)!} = \binom{n}{k} \cdot k!$$
$$A = \prod_{n}^{(n-k+1)} n$$

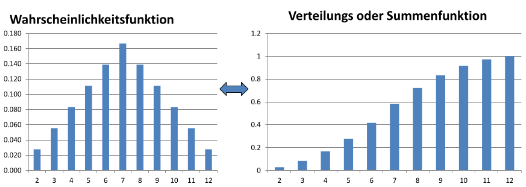
Reihenfolge nicht relevant, ohne Zurücklegen:

$$A = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

7 Zufallsvariable und die wichtigsten theoretischen Verteilungen

- Zufallsvariable, Abbildung / Funktion die den Elementen der Ergebnismenge eine Zahle zuordnet

$$X : \{Wappen, Zahl\} \to \mathbb{R},$$
$$X(\omega) = \begin{cases} 1, & \text{falls} \omega = \text{Wappen} \\ -1, & \text{falls} \omega = \text{Zahl} \end{cases}$$



7.0.1 Diskrete Wahrscheinlichkeitsfunktion

- Summer aller Einzelwahrscheinlichkeit ist 1 (Fläche)

$$P(x_i \leq X \leq x_n) = \sum_{j=i}^{n} f(x_j)$$

1. Wichtige diskrete Verteilungen **Bernoulli**:
   - Binär (ja / nein)
   - Erfolgswahrscheinlichkeit p bleiben gleich

   $$E(X) = \overline{x} = p$$
   $$\sigma^2 = pq$$

   Binominal:
   - Binär
   - wird n-mal wiederholt

   $$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

   $$E(X) = np$$

   $$\sigma^2 = np(1-p)$$

   Poisonverteilung:
   - Näherung der Binominalverteilung wenn $np \leq 10$ und $n \geq 1500p$
   - Wie hoch die Wahrscheinlichkeit, dass Ereignis E im Intervall I genau / höchstens x-mal auftritt, wenn bekannt ist, dass E μ -mal auftritt
   - Anzahl Anrufe an der Hotline zwischen 10:00 und 12:00

   $$f(x) = P(X = x) = \frac{\mu^x}{x!} e^{-\mu}$$

   $$E(X) = \sigma^2 = \mu$$

7.0.2 Kontinuierliche Verteilungsdichtefunktion

- Summer aller Wahrscheinlichkeit ist 1 (Fläche)
- Die Berechnung der Wahrscheinlichkeit nur in Bereichen möglich

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

7.0.3 Grundlegende Begriffe der Statistik

- Erwartungs- oder Mittelwert: $E[X] = \mu = \sum_i x_i f(x_i)$

8 Stetige Verteilungen

8.1 Verteilungen

Normalverteilung:

$$z = \frac{x-\mu}{\sigma}$$

Normalverteilung - Stichprobe:

$$Z_{stichprobe} = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

t-Verteilung:

$$t = \frac{\overline{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

$$r = n - 1$$

$$E(X) = 0 (\text{für} r > 1)$$

$$\sigma^2 = \frac{r}{r-2} (\text{für} r > 2)$$

Chi-Quadrat:

$$\chi^2_{(p,k)}$$

$$E(X) = r$$

$$\sigma^2 = 2r$$

Chi-Quadrat - untere Grenze:

$$S_u = S\sqrt{\frac{n-1}{\chi^2_{(1-\frac{\alpha}{2},n-1)}}}$$

Chi-Quadrat - obere Grenze:

$$S_o = S\sqrt{\frac{n-1}{\chi^2_{(\frac{\alpha}{2},n-1)}}}$$

- echte Zufallszahlen schlecht für Experimente
- besser Pseudo-Zufallszahlen

9 TODO Schliessende Statistik

1. Normalverteilung, wenn $n \geq 30$
2. t-Verteilung, wenn n $n < 30$
3. Chi Quadrat Verteilung

9.0.1 Stichproben und Experimente

1. Isolierende Abstraktion
    - What are the objectives?
2. Generalisierende Abstraktion
    - Modellverifikation
    - Modellvalidierung

9.0.2 Chi-Quadrat-Verteilung

- Zufallsvariablen, die unabhängig und Normalverteilt sind:

$$Y = X_1^2 + X_2^2 + X_3^2 + \cdots + X_r^2$$

"Echte" Grösse	Stichproben Grösse
μ	$\overline{X} \to \sigma_{\overline{x}}$
σ	$\hat{\sigma}$

- Konfidenzintervall

10 TODO Schätzverfahren

- Nicht blind raten:
    - Schätzverfahren
    - Schätzfunktionen
    - Gütekriterien für Schätzfunktionen
- Schätzverfahren:
    - die unbekannten Parameter anhand einer Stichprobe zu schätzen
    - durch Angabe eines Punktes: Punktschätzung
    - durch Angabe eines Intervalls: Intervallschätzung
- Schätzfunktion:
    - Punktschätzung: Hier werden die Werte für μ, σ^2 und für unbekannte Wahrscheinlichkeiten p geschätzt
    - Intervallschätzung: Hier geht es um die Bestimmung eines Vertrauensintervall für die oben genannten Parametern

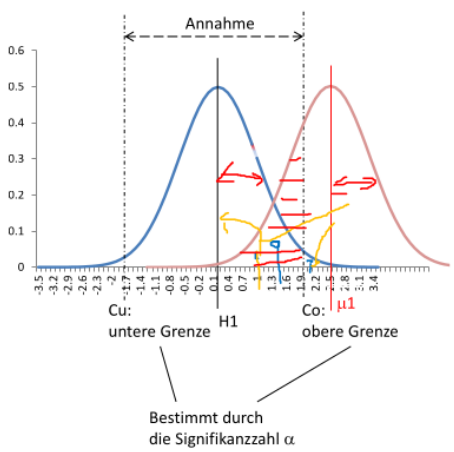
10.0.1 Erstellen eines Konfidenzintervall

1. Feststellung der Verteilungsform von $\overline{X}$ bzw. P
2. Festellung der Varianz von $\overline{X}$ / P ggf. schätzen.
3. Ermittlung der Quantilwertes z oder t aus Tabelle / Rechner
4. Berechnung des maximalen Schätzfehlers (Produkt aus Quantilswert und Standardabweichung von X)
5. Ermittlung der Konfidenzgrenzen (Schätzfehler $\pm$ Stichprobenmittel $\overline{X}$

11 TODO Testverfahren

- Testverfahren
  - Parametertest
  - Verteilungs- bzw. Anpassungstest
  - Unabhängigkeitstest
- Fehler
  - Fehler erster Art (fehlerhaftes Ablehnen einer Hypothese)
  - Fehler zweiter Art (fehlerhaftes Annehmen einer Hypothese)

1. Fehler
   - Wenn der ermittelte Werte ausserhalb des Annahmebereichs, wird die Hypothese verworfen
2. Fehler (siehe Bild)
   - $\mu 1$ echter Wert, gemäss Hypothese ist H1 der Mittelwert
   - Nun wird die Hypothese angenommen; es wird ein Fehler gemacht



11.0.1 Hypothese

1. Nullhypothese H_0 und der Alternativhypothese H_1 (im Allgemeinen die Verneinung der Nullhypothese)
2. Festlegen der Signifikanzzahl α, bzw. das Vertrauensintervall $1 - \alpha$
3. Bestimmen der Testgrössen und Annahme- und Ablehungsbereiches
4. Berechnen der Werten

Um die Hypothese zu beweisen, wird oft die Verneinung als Nullhypothese formuliert und versucht diese durch einen statistischen Test zu widerlegen.

11.0.2 Testverfahren

- Parametertest
- Anteilswert
- Differenztests

1. Abhängige Stichprobe
   (a) Bilder der Nullhypothese $H_0 : \mu_1 = \mu_2$, folglich $\overline{x} = 0$
   (b) Die Differenzen der Stichprobe $d_i = x_i - y_i$ für alle i erfassen
   (c) Die Nullhypothese H_0 gilt, wenn der Mittelwert von d_i im Bereich von $1 - \alpha$ liegt
   (d) Signifikanzzahl α wählen
   (e) Grenzen festlegen (mit Tabellen)
2. Unabhängige Stichprobe
   (a) Lege Signifikanzzahl α fest
   (b) Mit Hilfe der Tabelle den Z-Wert festlegen
   (c) Berechne den Annahmebereich:
   $$c = \pm z \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$
   (d) Berechne die Mittelwerte der Stichproben μ_1 und μ_2
   (e) Man nehme die Nullhypothese $H_0 : \mu_1 = \mu_2$ an, wenn die Differenz $d_i = \mu_1 - \mu_2$ in den Annahmenbereich fällt

12 TI-Nspire

12.1 Shortcuts

Summenzeichen Σ: `menu, 4, 5`
Produktzeichen Π: `menu, 4, 6`
Fakultät !: `menu, 5, 1`
nCr: `menu, 5, 3`
solve: `menu, 3, 1`
Numerisch Lösen: `menu, 3, 6`
Ableitung: `menu, 4, 1`
Integral: `menu, 4, 3`

12.2 Funktionen

Allgemein: `menu, 6, 3, <Zahl>`
Datensatzlänge: `count(x)`
Datensatzlänge mit Wert y: `count(x)`
Median: `median(x)`
Maximum: `max(x)`
Minimum: `min(x)`
Absoluter Wert: `abs(x)` oder `|x|`
Arithm. Mittelwert $\overline{x}$: `mean(x)`
Empirishe-Varianz σ_x^2: `varPop(x)`
Standardabweichung σ_x: `stDevPop(x)`

Stichproben-Varianz: `varSamp(x)`
Summe Σ: `sum(x)`
Approx. Zahlenwert $\approx$: `approx(x)`
Kombinationen $\binom{n}{k}$: `nCr(n,k)`

12.3 Verteilungen

Verteilungen die mit *Pdf* (probability density function) enden sind die Dichtefunktionen f, solche die mit *Cdf* (cumulative distribution function) enden sind die Verteilungsfunktionen F. Im Ti-Nspire zugänglich über `menu, 6, 5, <Zahl>`.

12.3.1 Normalverteilungs-

Dichte: *normPdf(x, μ, σ)*
Summe: *normCdf(a, b, μ, σ)*
Inverses: *invNorm(%, μ, σ)*
z-Value: *invNorm(%, 0,1)*

12.3.2 Student-T Verteilungs-

Dichte: `tPdf(x, \nu)`
Summe: `tCdf(a, b, \nu)`
Inverses: `invT(%, \nu)`

12.3.3 χ^2-Verteilungs-

Summe: `menu, 6, 5, 8, (a, b, \nu)`
Inverse: `menu, 6, 5, 9, (%, \nu)`

12.3.4 Zentralle Konfidenzintervalle

1. `menu 6, 6, 1 (2 falls t)`
2. Dateneingabe: Statistik
3. σ = Standardabweichung (nicht Stichprobenwert)
4. μ eintragen und n eintragen
5. Konfidenzniveau korrekt eintragen

12.4 Liste

Lineare Regression:

1. x-liste und y-liste eintragen
2. `menu, 6, 1, 4`

Statistische Berechnungen:

1. x-liste (und y-liste) eintragen
2. `menu, 6, 1, <Zahl>`

12.5 Spreadsheet

Datensatz hinzufügen:

1. Name in erste Zelle einfügen
2. Daten eintragen

Spalte link einfügen: `menu, 2, 3`
Spalte verschieben: `menu, 1, 1`
Zellenname: SpalteIndex (e.g. a1)
Bereichsreferenz: Startzelle:Endzelle
Datensatz kopieren:

1. Shift + Pfeile (selektieren)
2. ctrl, c
3. Zu obersten Punkt neuer Spalte navigieren
4. ctrl, v

Datensatz sortieren:

1. Spalte selektiren
2. `menu, 1, 6`

Kumulierte Spalte hinzufügen:

1. In neuer Spalte ins Ausdruckszelle (Zweite Zellen von oben) gehen
2. `menu, 3, 7, 1` oder `cumultativeSum(<var>)`

Gleitender Mittelwert:

1. Werte eintragen
2. In Zielspalte, Zelle n den Wert eintragen in der Form: (a1+a2+...+an) / n
3. Mit selektierter Zelle: `menu, 3, 3`
4. Mit Pfeilen neue Zellen selektieren

Newton-Alorithmus:

1. X-Werte in Spalte a, Y-Werte in Spalte b
2. In Zelle c2, den Wert `=approx((b2-b1)/(a2-a1))` eintragen
3. Mit selektierte Zelle: `menu, 3, 3`
4. Mit Pfeilen ganze Spalte ausfüllen
5. Ab Schritt 2 in Zelle d3 wiederholen: `=approx((c3-c2)/(a3-a1))`

13 Formelsammlung

13.1 Lageparameter

- **Klassenbreite**: $b_m = x_m^o - x_m^u$
- **Klassenmitte**: $\frac{x_m^o + x_m^u}{2}$
- **Klassendichte**: $d_m = \frac{h_m}{b_j}$

Modus: Bei unterschiedlicher Klassenbreite anstellen von h_m Klassendichte nutzen.

$$Mo = x_m^u + \frac{h_m - h_{m-1}}{(h_m - h_{m-1}) + (h_m - h_{m+1})} \cdot (x_m^o - x_m^u)$$

Median:

$$Me = x_m^u + \frac{\frac{n}{2} - H_{m-1}}{H_m - H_{m-1}} \cdot (x_m^o - x_m^u)$$

Quantile, Quartile, Dezile, Perzentile analog zu Median, einfach $\frac{n}{2}$ passend ersetzen

arithmetisches Mittel (mean, average):

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{v} x_i \cdot h_i$$

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{v} x_i \cdot f_i$$

arithmetisches Mittel - Klassifiziert ($\grave{x}_i$ Klassenmittelwert):

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{v} \grave{x}_i \cdot h_i$$

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{v} \grave{x}_i \cdot f_i$$

Harmonisches Mittel: Merkmal muss verhältnisskaliert sein

$$\overline{MH} = \frac{\sum_{i=1}^{v} h_i}{\sum_{i=1}^{v} \frac{h_i}{x_i}}$$

Geometrisches Mittel:

$$MG = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{\frac{endwert}{anfangswert}}$$

Spannweite:

$$R = x_{max} - x_{min}$$

Zentraller Quartilsabstand:

$$ZQA = Q_3 - Q_1$$

Mittlere absoulte Abweichung (h_i absoulte Häufigkeit):

$$\delta = \frac{1}{n} \sum_{i=1}^{v} |x_i - \overline{x}| h_i$$

Varianz (h_i absoulte Häufigkeit):

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{v} (x_i - \overline{x})^2 h_i$$

Stichprobenvarianz (h_i absoulte Häufigkeit):

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{v} (x_i - \overline{x})^2 h_i = \frac{n}{n-1} \sigma^2$$

$$\sigma_{\overline{x}}^2 = \frac{\sigma^2}{n}$$

Standardabweichung (Wurzerl der Varianz):

$$\sigma = \sqrt{\sigma^2}$$

Variantionskoeffizienten: Misst die relative Streuung

$$VK = \frac{\sigma}{\overline{x}} \cdot 100$$

Kovarianz:

$$\sigma_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$
$$\sigma_{XY} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x}\ \overline{y}$$

Variante 2 Analog zu der Varianz
Korrelationskoeffizient (-1 - 1):

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad mit\ |\sigma_{XY}| \leq \sigma_X \sigma_Y$$
$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{(\sum_{i=1}^{n}(x_i - \overline{x})^2)(\sum_{i=1}^{n}(y_i - \overline{y})^2)}}$$

13.2 Prognossen

gleitender Mittelwert

$$M_t = \frac{1}{T+t}(\sum_{T=1}^{t-1} x_T + x_t)$$

Lineare Regression funktioniert nur mit linearen zusammenhang. Bei einem exponentiellen kann dieser aber logrimiert werden. Bei einem logrimischen Zusammenhang kann potenziert werden um einen linearen Zusammenhang zu erhalten.

$$y = a_1 + b_1 x_i$$
$$a_1 = \overline{y} - b_1 \overline{x}$$
$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\ \overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}$$

Newton - Algorithmus (TI-Nspire):

$$f(x) = a_0 + a_1(x - x_1) + a_2(x - x_1)(x - x_2) + ... + a_n(x - x_1)...(x - x_n)$$

13.3 Anteilswert / Wahrscheinlichkeit (p)

k = Anzahl gute-Fälle, n = Stichprobenumfang

$$\overline{p} = \frac{k}{n}$$

13.4 Fehler

Fehler:

$$e = z\frac{\sigma}{\sqrt{n}}$$

Maximaler Fehler - mit zurücklegen (grosse Grundgesamtheit):

$$n \geq \frac{z^2\sigma^2}{e^2}$$

Maximaler Fehler - ohne zurücklegen:

$$n \geq \frac{z^2 N \sigma^2}{e^2(N-1) + z^2\sigma^2}$$

Vertrauensintervall - $\overline{x}$:

$$\mu_0 - z\frac{\sigma}{\sqrt{n}} \leq \overline{x} \leq \mu_0 + z\frac{\sigma}{\sqrt{n}}$$

Vertrauensintervall - Anteilswert:

$$c_u = p_0 - z\sqrt{\frac{p_0(1-p0)}{n}}$$
$$c_u = p_0 - z\sqrt{\frac{p_0(1-p0)}{n}}\sqrt{\frac{N-n}{N-1}} \quad |\frac{n}{N} \geq 0.05$$

$$c_o = p_0 + z\sqrt{\frac{p_0(1-p0)}{n}}$$
$$c_o = p_0 + z\sqrt{\frac{p_0(1-p0)}{n}}\sqrt{\frac{N-n}{N-1}} \quad |\frac{n}{N} \geq 0.05$$

13.5 Varianz für das Stichprobenmittel $\bar(X)$:

Stichprobe	σ^2 bekannt ($\sigma^2_{\overline{X}}$)	σ^2 unbekannt ($\hat{\sigma}^2_{\overline{X}}$)
Unendliche Grundgesamtheit	$\frac{\sigma^2}{n}$	$\frac{s^2}{n}$
$\frac{n}{N} < 0.05$	$\approx \frac{\sigma^2}{n}$	$\approx \frac{s^2}{n}$
$\frac{n}{N} \geq 0.05$	$\frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$	$\frac{s^2}{n} \cdot \frac{N-n}{N}$

13.6 Varianz für den Anteilswert (P):

für $np(1-P) > 9$ aprox. normalverteilt

Stichprobe	σ^2 bekannt (σ^2_P)	σ^2 unbekannt ($\hat{\sigma}^2_P$)
Unendliche Grundgesamtheit	$\frac{P(1-P)}{n}$	$\frac{P(1-P)}{n}$
$\frac{n}{N} < 0.05$	$\approx \frac{P(1-P)}{n}$	$\approx \frac{P(1-P)}{n}$
$\frac{n}{N} \geq 0.05$	$\frac{P(1-P)}{n} \cdot \frac{N-n}{N-1}$	$\frac{P(1-p)}{n} \cdot \frac{N-n}{N}$