Arithmetischer Mittelwert:

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

In Abhängigkeit der Symmetrie der Verteilung:

- Linksschief: $\bar{x} < m$
- (Grob) symmetrisch: $\bar{x} \approx m$
- Rechtsschief: $\bar{x} > m$

Standardabweichung:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Inter-Quartils-Abstand:

$$IQA = Q_3 - Q_1$$

1.5 * IQA Regel (Ausreisser bestimmen):

(a) kleiner als $Q_1 - 1.5 \times IQA$

(b) grösser als $Q_3 + 1.5 \times IQA$

Wahrscheinlichkeits-Axiome & -Regeln:

Jedem Ereignis A wird ein Wert $P(A)$ zugeordnet, so dass gilt:

(1) $P(A) \geq 0$

(2) $P(S) = 1$

(3) $A \cap B = \emptyset \Longrightarrow P(A \cup B) = P(A) + P(B)$

Die folgenden Regeln können daraus abgeleitet werden:

- $0 \leq P(A) \leq 1$
- $P(\emptyset) = 0$
- $P(A^c) = 1 - P(A)$
- $P(A) \leq P(B)$, falls $A \subseteq B$
- $P(A) = \sum_{s \in A} P(s)$
- $P(A) = P(A \cap B) + P(A \cap B^c)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Unabhängigkeit:

Zwei Ereignisse A und B sind **unabhängig**, wenn gilt:

- $P(A \cap B) = P(A) \times P(B)$

Bedingte Wahrscheinlichkeit:

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

1. Multipliziere auf beiden Seiten mit $P(B)$:
   $P(A \cap B) = P(A|B)P(B)$

Dann natürlich aber auch: $P(A \cap B) = P(B|A)P(A)$

2. Daraus folgt der Satz von Bayes:

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

3. Der Nenner kann zerlegt werden (Partition):
   - $P(B) = P(B \cap A) + P(B \cap A^c) = P(B|A)P(A) + P(B|A^c)P(A^c)$

Erwartungswert und Varianz:

- $\mu_X = \sum_i x_i p_i$
- $\sigma_X^2 = \sum_i (x_i - \mu_X)^2 p_i$

Regel 1:

$E(X + Y) = E(X) + E(Y)$

Falls X und Y unabhängig sind, $Var(X + Y) = Var(X) + Var(Y)$

Regel 2:

$E(a + bX) = a + bE(X)$

$Var(a + bX) = b^2 Var(X)$

$SA(a + bX) = |b| SA(X)$

Umfassende Regel (n Zufallsvariablen):

Betrachte n Zufallsvariablen $X_1, X_2, \ldots, X_n$ und eine beliebige Linearkombination

$$Z = a + \sum_{i=1}^{n} b_i X_i$$

Dann gilt:

- $E(Z) = a + \sum_{i=1}^{n} b_i E(X_i)$
- Falls die X_i unabhängig sind, $Var(Z) = \sum_{i=1}^{n} b_i^2 Var(X_i)$

Bernoulli-Verteilung:

- Die Wahrscheinlichkeit eines Erfolges ist p
- X ist die Anzahl der Erfolge

Abkürzende Notation: $X \sim Bern(p)$

Man erhält die folgende Verteilung:

Ergebnis von X	0	1
Wahrscheinlichkeit	$1 - p$	p

Daraus berechnen sich leicht:

- $\mu_X = p$
- $\sigma_X^2 = p(1 - p)$ und $\sigma_X = \sqrt{p(1 - p)}$

Binomial-Verteilung:

- X ist die Anzahl der Erfolge in den n 'Versuchen'

Abkürzende Notation: $X \sim Bin(n, p)$

Man erhält die folgende Verteilung:

Ergebnis von X	0	1	...	n
Wahrscheinlichkeit	?	?	...	?

Welches sind nun die Wahrscheinlichkeiten $p_k = P(X = k)$?

- Mit Hilfe der Multiplikations-Regel finden wir leicht $P(X = 0) = (1 - p)^n$ und $P(X = n) = p^n$

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Dazu müssen wir nur realisieren, dass $X = \sum_{i=1}^{n} X_i$, wobei X_i eine Bernoulli-Variable für den i-ten Versuch ist. Daher:

- $E(X) = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np$
- $Var(X) = Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} p(1 - p) = np(1 - p)$

Poisson-Verteilung:

Wenn die zugrundeliegenden 'Versuche' nicht beobachtbar sind, ist oftmals eine Poisson-Verteilung adäquat:

- Mögliche Ergebnisse: $0, 1, 2, \ldots$
- Wahrscheinlichkeiten: $P(X = k) = e^{-\mu} \cdot \frac{\mu^k}{k!}$
  (Diese summieren sich zu 1 auf da $\sum_{k=0}^{\infty} \frac{\mu^k}{k!} = e^{\mu}$)
- Hierbei ist μ eine positive Konstante, die dem Erwartungswert von X entspricht. Offensichtlich hängt dieser von dem gewählten Zeitintervall (Tag, Woche, Jahr, ...) ab

Abkürzende Notation: $X \sim Po(\mu)$

Kennzahlen (ohne Beweis):

- $E(X) = \mu$
- $Var(X) = \mu$

$X \sim Po(7.5)$

$P(X \geq 5) = \sum_{k=5}^{\infty} e^{-7.5} \cdot \frac{7.5^k}{k!}$ = mühsam!

$P(X \geq 5) = 1 - P(X \leq 4) = 1 - \sum_{k=0}^{4} e^{-7.5} \cdot \frac{7.5^k}{k!} = 1 - e^{-7.5}(1 + 7.5 + 28.125 + 70.312 + 131.896) = 1 - 0.132 = 0.868$

Stetige Zufallsvariablen:

- Verteilung wird durch eine Dichte(funktion) $f(\cdot)$ beschrieben
- Diese muss (nur) erfüllen: $f(x) \geq 0$ und $\int_{-\infty}^{\infty} f(x)dx = 1$
- Wir können die W-keit von Ereignissen als die entsprechende Fläche unter der Dichtekurve berechnen (mittles Integration)
- Beispiel: $P(a \leq X \leq b) = \int_a^b f(x)dx$
- Es folgt: $P(X = a) = \int_a^a f(x)dx = 0$
- Und daher auch: $P(a \leq X \leq b) = P(a < X < b)$ etc. (im Gegensatz zu diskreten Zufallsvariablen!)

$\mu_X = \int_{-\infty}^{\infty} x f(x)dx$

$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)dx$

Das $p \times 100$te Perzentil, x_p, einer stetigen Verteilung erfüllt die folgende Bedingung: $P(X \leq x_p) = p = p \times 100\%$ für ein $p \in (0, 1)$.

Das heisst, es löst die folgende Gleichung:

$$\int_{-\infty}^{x_p} f(x)dx = p$$

Alternative Berechnung der Varianz:

Somit haben wir $Var(X) = E[(X - \mu_X)^2]$, da:

(a) $Var(X) = \sum_i (x_i - \mu_X)^2 p_i$ falls X diskret ist

(b) $Var(X) = \int (x - \mu_X)^2 f(x)dx$ falls X stetig ist

Dies kann wie folgt vereinfacht werden:

$$E[(X - \mu_X)^2] = E[X^2 - 2\mu_X X + \mu_X^2] = E(X^2) - 2\mu_X E(X) + \mu_X^2 = E(X^2) - 2\mu_X^2 + \mu_X^2 = E(X^2) - \mu_X^2$$

Falls wir μ_X schon haben, brauchen wir dann nur noch $E(X^2)$ als

(a) $E(X^2) = \sum_i x_i^2 p_i$ falls X diskret ist

(b) $E(X^2) = \int x^2 f(x)dx$ falls X stetig ist

Gleichverteilung:

Die Dichte ist konstant zwischen zwei Zahlen $a < b$ (und sonst null):

- $f(x) = \begin{cases} \frac{1}{b-a} & \text{falls } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$
- Wahrscheinlichkeiten super einfach: "Höhe × Breite"

Abkürzende Notation: $X \sim U[a, b]$

Die Kennzahlen sind:

- $\mu_X = (a + b)/2$
- $\sigma_X^2 = (b - a)^2/12$

Normalverteilung:

Abkürzende Notation: $X \sim N(\mu, \sigma)$

Die Kennzahlen sind:

- $\mu_X = \mu$
- $\sigma_X^2 = \sigma^2$

- Gegeben: $X \sim N(\mu, \sigma)$
- Gesucht: $P(X < a)$

$$X \sim N(\mu,\sigma) \Longrightarrow \frac{X-\mu}{\sigma} \sim N(0,1)$$

[Siehe Übung . . .]

Daher:

$$P(X<a) = P\left(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right) = P\left(Z < \frac{a-\mu}{\sigma}\right) \quad \text{wobei } Z \sim N(0,1)$$

Fazit: Es reicht eine Tabelle für die Standardnormalverteilung $N(0,1)$

Beispiel: $X \sim N(750, 5)$ [Inhalt einer Flasche Wein . . .]

$$P(X < 744) = P\left(\frac{X-750}{5} < \frac{744-750}{5}\right) = P(Z < -1.2) \overset{\text{Tab.}}{=} 0.115$$

- Das $p$te Perzentil von $N(0,1)$, z_p, findet sich in Tabelle. Daher:

$$p = P(Z < z_p) = P\left(\frac{X-\mu}{\sigma} < z_p\right) = P(X < \mu + \sigma z_p) \Longrightarrow \boxed{x_p = \mu + \sigma z_p}$$

Die wichtigsten Perzentile von $N(0,1)$:

p	0.90	0.95	0.975	0.99	0.995
z_p	1.282	1.645	1.960	2.326	2.576

Stichproben-Verteilung:

(1) Der Erwartungswert von $\bar{X}$:

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \overset{R2}{=} \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) \overset{R1}{=} \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu_X = \mu_X$$

(2) Die Varianz von $\bar{X}$:

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \overset{R2}{=} \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) \overset{R1}{=} \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \sigma_X^2 = \frac{\sigma_X^2}{n}$$

$$\Longrightarrow SA(\bar{X}) = \sqrt{VAR(\bar{X})} = \frac{\sigma_X}{\sqrt{n}}$$

Stichproben-Verteilung (2):

- Falls der Stichprobenumfang n gross ist, ähnelt die Verteilung von $\bar{X}$ einer Normalverteilung (d.h. einer Glockenkurve)
- Da die Kennzahlen von $\bar{X}$ bekannt sind, wissen wir dann direkt um welche Normalverteilung es sich handeln muss:

$$\bar{X} \mathrel{\dot\sim} N\left(\mu_X, \frac{\sigma_X}{\sqrt{n}}\right)$$

Wir benötigen $n \geq 50$ für die Normal-Approximation.

Stichproben-Anteil:

- Das zugrundeliegende Zufallsexperiment ist Bernoulli
- Daher $\mu_X = p$ und $\sigma_X = \sqrt{p(1-p)}$
- $\bar{X}$ ist der Anteil (oder Prozentsatz) der 'Erfolge' in den n Wiederholungen des Experimentes
- Gebräuchlichere Bezeichnung: $\hat{p}$ (Stichprobenanteil)

Der allgemeine Zentrale Grenzwertsatz besagt nun:

$$\hat{p} \mathrel{\dot\sim} N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Regel:

- Wir benötigen $\mathrm{Min}\{np, n(1-p)\} \geq 10$ für die N-Approximation

Konfidenz-Intervalle:

Parameter: Mittelwert μ

- Schätzer: Stichprobenmittelwert $\bar{X}$
- $E(\bar{X}) = \mu \Longrightarrow$ unverzerrt
- $Var(\bar{X}) = \sigma^2/n$

Parameter: Anteil p

- Schätzer: Stichprobenanteil $\hat{p}$
- $E(\hat{p}) = p \Longrightarrow$ unverzerrt
- $Var(\hat{p}) = p(1-p)/n$

Konfidenz-Intervall für Mittelwert:

(1) Zentraler Grenzwertsatz: $\bar{X} \mathrel{\dot\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

(2) Standardnormalverteilung:
$Z \sim N(0,1) \Longrightarrow P(-1.96 \leq Z \leq 1.96) = 0.95$

$\bar{X} \pm 1.96 \times \frac{s}{\sqrt{n}}$ ist ein 95% Konfidenz-Intervall (KI) für μ.

Wir benötigen daher $n \geq 50$, um dem KI zu vertrauen

Konfidenz-Intervall = Schätzer $\pm\, z_{1-\frac{\alpha}{2}} \times$ Standard-Fehler (SF)
= Schätzer $\pm$ Fehler-Marge (FM)

- Betrachte FM* $\geq z_{1-\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}}$ und löse nach n auf
- Wir erhalten dann: $n \geq \left(\frac{z_{1-\frac{\alpha}{2}} \times s}{FM^*}\right)^2$

Konfidenzniveau $1-\alpha$	90%	95%	99%
Konstante $z_{1-\frac{\alpha}{2}}$	1.645	1.96	2.576

Konfidenz-Intervall für Anteil p:

$$\hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

wir benötigen $\mathrm{Min}\{n\hat{p}, n(1-\hat{p})\} \geq 10$.

$$n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{FM^*}\right)^2 \hat{p}(1-\hat{p})$$

- Also setzt man eine Schätzung für $\hat{p}$ ein oder, ganz auf Nummer sicher, nimmt den 'worst case' $\hat{p} = 0.5$ an

KI für Differenz zweier Mittelwerte:

a) Stichproben unabhängig:

$$\bar{X} - \bar{Y} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad n_1 \geq 50,\; n_2 \geq 50$$

b) Stichproben abhängig (Datenpaare):

$$D_i = X_i - Y_i \qquad \bar{D} \pm z_{1-\frac{\alpha}{2}} \times \frac{s_D}{\sqrt{n}}$$

KI für Differenz zweier Anteile (Stichproben unabhängig!):

$$\hat{p}_1 - \hat{p}_2 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$\mathrm{Min}\{n_1\hat{p}_1, n_1(1-\hat{p}_1)\} \geq 10$
$\mathrm{Min}\{n_2\hat{p}_2, n_2(1-\hat{p}_2)\} \geq 10$

Hypothesen-Test für Mittelwert:

1.
- H_0: $\mu = \mu_0$ (die Null-Hypothese)
- H_1: $\begin{cases} \text{(a) } \mu > \mu_0 & \text{(einseitig)} \\ \text{(b) } \mu < \mu_0 & \text{(einseitig)} \\ \text{(c) } \mu \neq \mu_0 & \text{(zweiseitig)} \end{cases}$

(die Alternativ-Hypothese)

2. $z = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ n $n \geq 50$ wie immer.

3.
$$PW = \begin{cases} \text{(a) } P(Z \geq z) & \text{(einseitig)} \\ \text{(b) } P(Z \leq z) & \text{(einseitig)} \\ \text{(c) } 2P(Z \geq |z|) & \text{(zweiseitig)} \end{cases}$$

- Der p-Wert misst die Evidenz gegen H_0 (oder für H_1)

$$PW \begin{cases} \leq 0.1 & \text{: milde Evidenz} \\ \leq 0.05 & \text{: anständige Evidenz} \\ \leq 0.01 & \text{: starke Evidenz} \end{cases}$$

Wir können $1 - PW$ konfident sein, dass H_0 falsch ist

Signifikanz-Niveaus:

- Man wählt dann eine kleine Zahl α im Voraus aus
- Diese Zahl heisst das Signifikanzniveau des Hypothesen-Tests
- Falls $PW \begin{cases} \leq \alpha & \text{: man verwirft } H_0 \\ > \alpha & \text{: man verwirft } H_0 \text{ nicht} \end{cases}$

(Falls $PW \leq \alpha$, dann ist der Test signifikant zum Niveau α)

- Schritte 1 und 2 sind wie zuvor
- Im Schritt 3 vergleicht man die Test-Statistik mit einem kritischen Wert und verwirft H_0 falls

$$\begin{cases} z \geq z_{1-\alpha} & \text{für (a)} \\ z \leq -z_{1-\alpha} & \text{für (b)} \\ |z| \geq z_{1-\frac{\alpha}{2}} & \text{für (c)} \end{cases}$$

- Ein zweiseitiger Test verwirft $H_0 : \mu = \mu_0$ zum Signifikanzniveau α genau dann, wenn das Konfidenzintervall mit Niveau $1-\alpha$ den Wert μ_0 nicht enthält

Güte eines Hypothesentests:

Szenario:

- Teste $H_0 : \mu = 10$ gegen $H_1 : \mu > 10$
- $\alpha = 0.05, \mu = 11, \sigma/s = 4$ und $n = 60$

Dann gilt:

$$\text{Güte} = P(z > z_{1-\alpha}) = P\left(\frac{\bar{X} - 10}{4/\sqrt{60}} > 1.645\right) = P(\bar{X} > 10.85)$$

Nun benutze den Zentralen Grenzwertsatz:

$$\bar{X} \mathrel{\dot\sim} N\left(11, \frac{4}{\sqrt{60}}\right) \quad \Longrightarrow \quad P(\bar{X} > 10.85) \approx 0.61$$

Hypothesen-Test für Anteil p:

2. $z = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ $\quad \mathrm{Min}\{np_0, n(1-p_0)\} \geq 10.$

HT für Differenz 2er Mittelwerte:

$\Delta = \mu$ $\quad \Delta = \Delta_0$ (die Null-Hypothese)

2. a, $z = \dfrac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ $\quad \mathrm{Min}\{n_1, n_2\} \geq 50$

b, $z = \dfrac{\bar{D} - \Delta_0}{s_D/\sqrt{n}}$ $\quad n \geq 50$

HT für Differenz zweier Anteile:

$\Delta = p_1 - p_2$

2. $z = \dfrac{\hat{p}_1 - \hat{p}_2 - \Delta_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$

$\mathrm{Min}\{n_1\hat{p}_1, n_1(1-\hat{p}_1), n_2\hat{p}_2, n_2(1-\hat{p}_2)\} \geq 10$