Wahrscheinlichkeits-Verteilungen

Raphael Hartmann

SUMMER SCHOOL KOGNITIVE MODELLIERUNG 2022

Übersicht

- Zufallsvariablen
- Verteilungen
  - Normalverteilung
  - Bernoulliverteilung
  - Binomialverteilung
  - Poissonverteilung
  - Exponentialverteilung
- Beschreibende Modelle
- Erklärende Modelle
- Likelihood Funktion

Zufallsvariablen

Zufallsvariablen

- **Was ist eine Zufallsvariable?:**
  - Bezeichnung: $\boldsymbol{X}$
  - auch Zufallsgröße, zufällige Größe oder auf Englisch *random variable*
  - ist eine Funktion/Zuordnungsvorschrift, die jedem Ergebnis einen Größe zuordnet
- **Beispiele:**
  - Münzwurf: Jedem Ergebnis eines Münzwurfs, also Kopf oder Zahl, wird eine Größe/Zahl zugeordnet, also 0 oder 1.

  - Reaktionszeit: Jedem Ergebnis bei einem Reaktionszeit-Experiment, also der Zeit bis zur Reaktion auf einen Stimulus, wird eine Größe/Zahl zugeordnet, also Zeit-Werte in Sekunden, Millisekunden oder andere Zeiteinheiten.

Eigenschaften von Zufallsvariablen

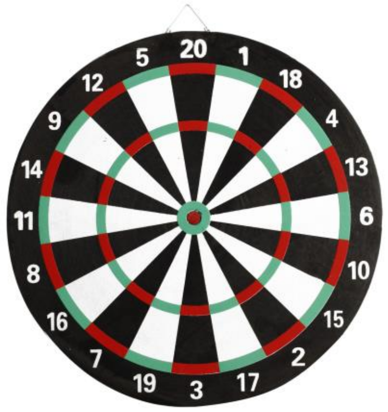
- **Abzählbarkeit:**
  - Können wir die möglichen Ergebnisse abzählen (bspw. Münzwurf mit zwei möglichen Ergebnissen), so ist die Zufallsvariable diskret.
  - Können wir die möglichen Ergebnisse nicht abzählen (weil auch Komma-Zahlen möglich sind bspw. Zeit, Alter, Körpergröße etc.), so ist die Zufallsvariable stetig/kontinuierlich.
- **Konstant:**
  - Gibt es nur ein mögliches Ergebnis (und somit nur ein zugeordnete Größe), so sagt man die Zufallsvariable ist konstant. In dem Fall spricht man eigentlich nicht mehr von einer Zufallsvariable
- **Realisierung:**
  - Die zugeordnete Größe eines zufälligen Ergebnisses nennt man Realisierung.
  - Eine Realisierung ist also eine tatsächliche Beobachtung (bspw. 0 bei Kopf-Wurf).

Kleine Aufgabe

- Was sind die Ergebnisse und die entsprechenden Größen bei folgenden Experimenten/Szenarien:
  - Basketball in Korb werfen
  - Lösung:
    - Ergebnisse:
    - Größen:
  - Bei einem Corona-Schnelltest
  - Lösung:
    - Ergebnisse:
    - Größen:

Kleine Aufgabe

- Was sind die Ergebnisse und die entsprechenden Größen bei folgenden Experimenten/Szenarien:
  - Eine Person versucht mit einen Dart eine Dartscheibe (45cm-breite Durchmesser) zu treffen und wir interessieren uns für den **Abstand** (nicht die Felder) zum Mittelpunkt
  - Lösung:
    - Ergebnisse:
    - Größen:
  - Anmerkung:



Verteilung einer Zufallsvariable

- **Wahrscheinlichkeitsverteilung:**
  - Verteilungen erlauben es, den Ereignissen eine Wahrscheinlichkeit zuzuordnen.
  - Eine Wahrscheinlichkeitsverteilung kann durch unterschiedliche Funktionen repräsentiert werden:
    - **Wahrscheinlichkeitsfunktion** (für diskrete Zufallsvariablen),
    - **Dichtefunktion** (für stetige/kontinuierliche Zufallsvariablen) oder
    - **Verteilungsfunktion** (allgemein für beide)
    - Quantilfunktion (allgemein für beide)
    - ...
  - Link: https://de.wikipedia.org/wiki/Liste_univariater_Wahrscheinlichkeitsverteilungen

Wahrscheinlichkeitsfunktion (PMF)

- **Bezeichnung:**
  - Auch: Zähldichte genannt; Englisch: *probability mass function* (PMF)
  - Man spricht oft auch von „Daten-generierende Funktion“
- **Definition:**
  - Für eine diskrete Zufallsvariable X wird mit $f_X(x)$ die W‘keitsfunktion bezeichnet.
  - Hier gilt: $f_X(x) = P(X = x_i)$, wobei x_i eine bestimmte Größe ist von den möglichen Größen und P die Wahrscheinlichkeit angibt, dass wir diese Größe beobachten.
- **Beispiele:**
  - Fairer Münzwurf: $f_X(x) = P(X = x_i) = \frac{1}{2}$, wobei $x_1 = 0$ (Kopf) und $x_2 = 1$ (Zahl)
  - Fairer Würfelwurf: $f_X(x) = P(X = x_i) = \frac{1}{6}$, wobei x_i in $\{1, \ldots, 6\}$ die Augenzahl des Würfels ist

Dichtefunktion (PDF)

- **Bezeichnung:**
  - Englisch: *probability density function* (PDF)
  - Man spricht oft auch von „Daten-generierende Funktion"
- **Definition:**
  - Für eine kontinuierliche Zufallsvariable X wird mit $f_X(x)$ die Dichtefunktion $f_X(x)$ bezeichnet.
  - Hier gibt es keine einfache Interpretation wie die Wahrscheinlichkeit. Der Wert einer Dichtefunktion ist keine Wahrscheinlichkeit!
- **Beispiel:**
  - Normalverteilung: $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma})$, wobei x in $(-\infty, \infty)$

Verteilungsfunktion (CDF)

- **Bezeichnung:**
  - Englisch: *cumulative distribution function* (CDF)
- **Definition:**
  - Für eine beliebige Zufallsvariable X wird mit $F_X(x)$ die Verteilungsfunktion bezeichnet.
  - Hier gilt: $F_X(x) = P(X \leq x)$, wobei x eine bestimmte Größe und P die Wahrscheinlichkeit angibt, dass wir diese oder eine kleinere Größe beobachten.
- **Beispiel:**
  - Fairer Würfel: $F_X(x) = P(X \leq x_4) = \sum_{i=1}^{4} P(X = x_i) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{4}{6}$

  - Std.-Normalverteilung: $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt = \frac{1}{2}$ für $x = 0$

Quantilfunktion

- **Bezeichnung:**
  - Englisch: *quantile function*
- **Beschreibung:**
  - Für eine beliebige Zufallsvariable X mit Verteilungsfunktion $F_X(x)$ ist $F_X^{-1}(p)$ die Quantilfunktion.
  - Die Quantilfunktion ist also die Umkehrfunktion (inverse Funktion) von der Verteilungsfunktion.
  - Bspw. für $p = .1$ (=10%) gibt die Funktion den x–Wert an, unter dem 10% der Daten sind.
- **Unterform:**
  - Quartilfunktion: gibt x-Werte für 0%, 25%, 50%, 75% und 100% der Daten.

Liste zentraler Verteilungen

- Normalverteilung (kontinuierlich)
- Bernoulliverteilung (diskret)
- Binomialverteilung (diskret)
- Poissonverteilung (diskret)
- Exponentialverteilung (kontinuierlich)

Normalverteilung

kontinuierlich

„Word recognition task“ data

subj	cat	Log(rt)
0	0	6,30991827822652
0	0	6,37672694789863
0	0	6,44413125670044
0	0	6,38012253689976
0	1	6,75925527066369
0	0	6,39526159811545
0	0	6,41673228251233
0	0	6,28226674689601
0	0	6,41999492814714
0	1	7,2211050981825
0	0	6,50578406012823
0	0	6,64378973314767
0	0	6,16120732169508
0	0	6,44730586254121
0	1	6,44730586254121
0	0	6,42324696353352
0	0	6,5510803350434
0	0	6,26149168432104
…	…	…

Beschreibung

- **Andere Bezeichnung:**
  - Gauss-Verteilung
- **Verwendung:**
  - IQ
  - Oft auch für andere Variablen in der Psychologie angenommen
  - Modellierung: in linearer Regression für die Residuen angenommen
- **Schreibweise:**
  - $X \sim \mathcal{N}(\mu, \sigma)$
  - X folgt einer Normalverteilung mit Mittelwert μ und Standardabweichung $\sigma > 0$

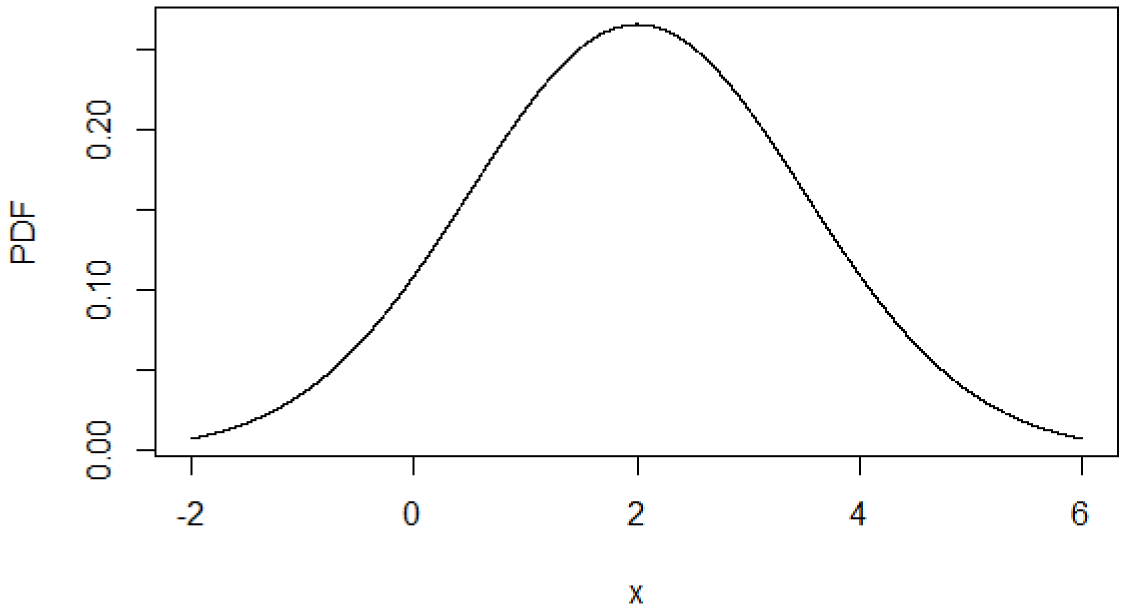
Dichtefunktion (PDF)

- **Formel:**
  - $f_X(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$
- **R Funktion:**

```
> dnorm()
```

- **R Code:**

```
> x <- seq(-2, 6, by=0.001)
> PDF <- dnorm(x, mean=2, sd=1.5)
> plot(x, PDF, type="l")
```



Kurze Aufgabe

- Zeichnet eine Dichtefunktion der Normalverteilung mit
  - Mittelwert 3 und
  - Standardabweichung 1.2
  - Passt dabei die Werte für x so an, dass die Dichtefunktion gut sichtbar ist (wie in der vorherigen Folie)

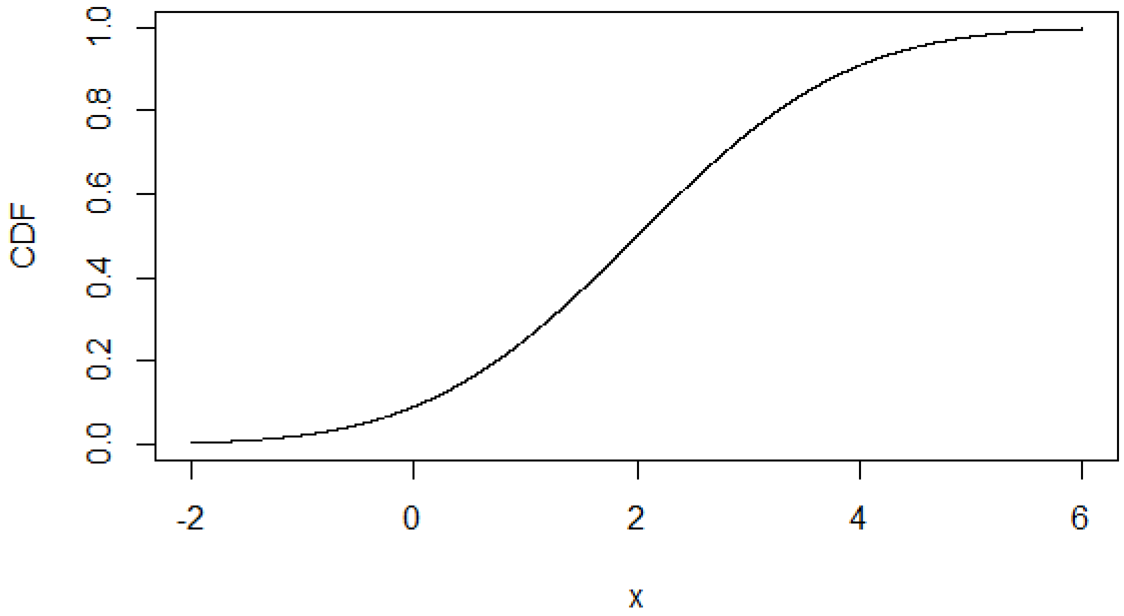
Verteilungsfunktion (CDF)

- **Formel:**
  - $F_X(x|\mu,\sigma) = \int_{-\infty}^{x} f_X(t)\ dt$
- **Besonderheit:**
  - Für diese CDF gibt es keine einfache Funktion
- **R Funktion:**

```
> pnorm()
```

- **R Code:**

```
> x <- seq(-2, 6, by=0.001)
> CDF <- pnorm(x, mean=2, sd=1.5)
> plot(x, CDF, type="l")
```



Kurze Aufgabe

- Zeichnet eine Verteilungsfunktion der Normalverteilung mit
  - Mittelwert 3 und
  - Standardabweichung 1.2
  - Passt dabei die Werte für x so an, dass die Verteilungsfunktion gut sichtbar ist (wie in der vorherigen Folie)

- Wie viel Prozent der Daten liegen unterhalb von $x = 1$?
  - Benutzt hierfür die `pnorm()` Funktion mit dem Mittelwert und die Standardabweichung von oben

Bernoulliverteilung

diskret

„Word recognition task“ data (Teildatensatz)

subj	cat	rt
0	0	550
1	0	824
2	0	589
3	0	766
4	1	1118
5	0	912
6	0	487
7	1	610
8	0	686
9	0	1022
10	0	624
11	1	1072
12	0	592
13	1	835
14	0	828
15	0	662
16	0	967
17	0	942
…	…	…

Beschreibung

- **Verwendung:**
  - Eine Entscheidung/Aufgabe mit zwei Optionen (Anzahl Trials ist 1)
  - Eine Diagnose (positiv vs. negativ)
  - Ein Münzwurf (Kopf vs. Zahl)
- **Schreibweise:**
  - $X \sim Bernoulli(\theta)$
  - X folgt einer Bernoulliverteilung mit Wahrscheinlichkeitsparameter θ, wobei θ die Wahrscheinlichkeit für die eine von den beiden Optionen ist (bspw. W'keit für positive Diagnose)
- **Sonstige Eigenschaften:**
  - Erwartungswert: θ; Varianz: $\theta(1-\theta)$

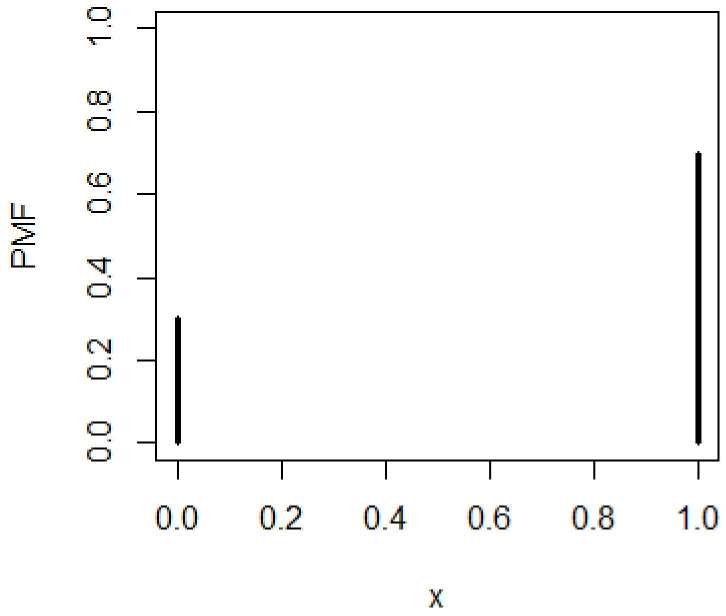
Wahrscheinlichkeitsfunktion (PMF)

- **Formel:**
  - $f_X(x|\theta) = \theta^x(1-\theta)^{1-x}$, wobei x entweder 0 oder 1, und θ in [0,1]
  - Im vorherigen Bsp. steht 1 für positive Diagnose
- **R Funktion:**

```
> dbinom(size=1)
```

- **R Code:**

```
> x <- c(0, 1)
> PMF <- dbinom(x, size=1, prob=.7)
> plot(x, PMF, type="h", lwd = 3, ylim=c(0,1))
```



Kurze Aufgabe

- Zeichnet eine Wahrscheinlichkeitsfunktion der Bernoulliverteilung mit
  - Wahrscheinlichkeitsparameter $\theta = .2$
  - Wenn 0 für die „falsche“ Entscheidung und 1 für die „richtige“ Entscheidung steht, was ist dann die Interpretation von θ?

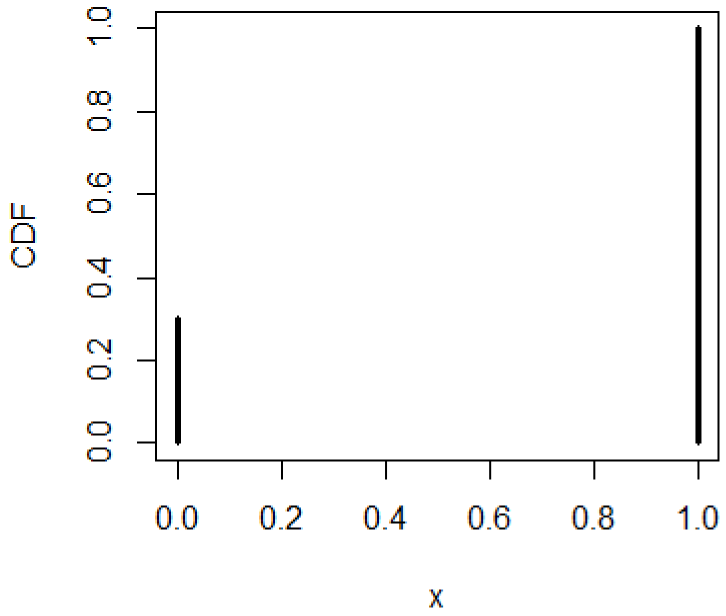
Verteilungsfunktion (CDF)

- **Formel:**
  - $F_X(x|\theta) = (1-\theta)^{1-x} = \begin{cases} 1-\theta \text{ falls } x = 0 \\ \quad 1 \quad \text{ falls } x = 1 \end{cases}$
- **R Funktion:**

```
> pbinom(size=1)
```

- **R Code:**

```
> x <- c(0, 1)
> CDF <- pbinom(x, size=1, prob=.7)
> plot(x, CDF, type="h", lwd = 3, ylim=c(0,1))
```



Kurze Aufgabe

- Zeichnet eine Verteilungsfunktion der Bernoulliverteilung mit
  - Wahrscheinlichkeitsparameter $p = .2$

Binomialverteilung

Erweiterung der Bernoulliverteilung für mehrere Trials

„Word recognition task“ data (Teildatensatz)

subj	cat	rt
0	0	550
0	0	588
0	0	629
…	…	…
1	0	824
1	0	663
1	0	548
…	…	…
2	0	589
2	0	521
2	1	904
…	…	…
3	0	766
3	1	1272
3	0	918
…	…	…

→

subj	cat = 0	cat = 1
0	26	4
1	24	6
2	20	10
3	17	13
4	19	11
5	25	5
6	22	8
7	21	9
8	25	5
9	23	7
10	22	8
11	22	8
12	22	8
13	19	11
14	18	12
…	…	…

Beschreibung

- **Verwendung:**
  - Mehrere Entscheidungen/Aufgaben mit zwei Optionen (Anzahl Trials $n > 1$)
  - Mehrere Münzwürfe (Kopf vs. Zahl)
- **Schreibweise:**
  - $X \sim Binom(n, \theta)$
  - X folgt einer Binomialverteilung mit n Trials und Wahrscheinlichkeitsparameter θ, wobei θ die Wahrscheinlichkeit für die eine von den beiden Optionen ist (bspw. W'keit für „richtige" Entscheidung)
- **Sonstige Eigenschaften:**
  - Erwartungswert: $n\theta$; Varianz: $n\theta(1-\theta)$

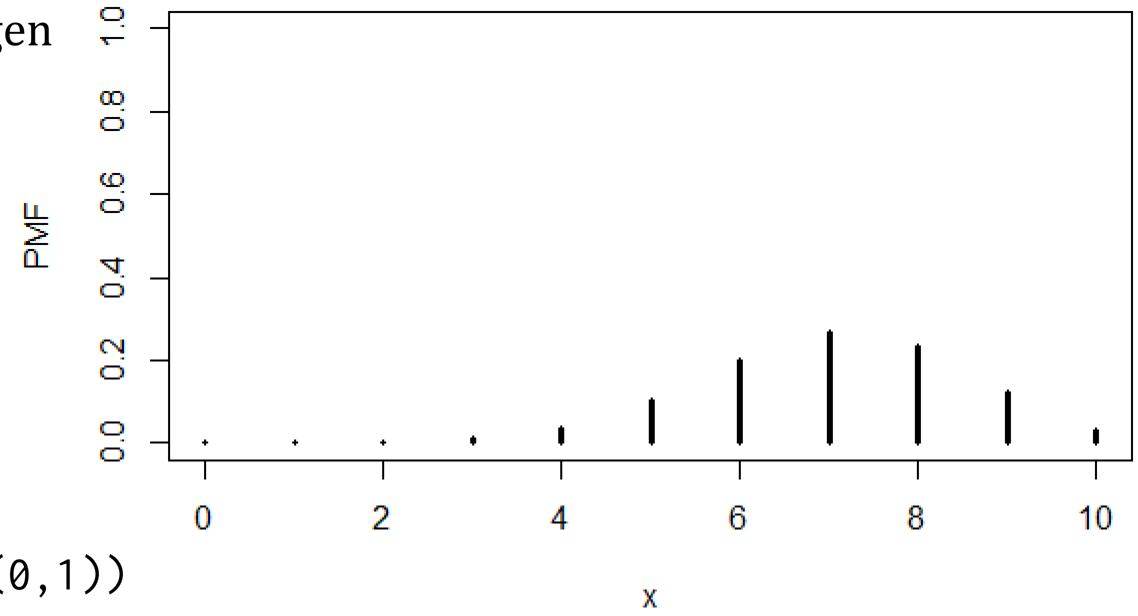
Wahrscheinlichkeitsfunktion (PMF)

- **Formel:**
  - $f_X(x|\theta) = \binom{n}{x} \cdot \theta^x (1-\theta)^{n-x}$, wobei x in $\{1,\dots, n\}$, θ in $[0,1]$
  - Bsp.: x ist die Anzahl „richtiger“ Entscheidungen
- **R Funktion:**

```
> dbinom(size=n)
```

- **R Code:**

```
> n <- 10
> x <- seq(0, n, by=1)
> PMF <- dbinom(x, size=n, prob=.7)
> plot(x, PMF, type="h", lwd = 3, ylim=c(0,1))
```



Kurze Aufgabe

- Angenommen ihr habt $n = 50$ Trials und beobachtet $x = 40$.
  - Wie würdet ihr intuitiv die Wahrscheinlichkeit θ schätzen bei $x = 40$ „richtigen“ Entscheidungen von $n = 50$ Trials/Entscheidungen?

- Zeichnet eine Wahrscheinlichkeitsfunktion der Binomialverteilung mit
  - Wahrscheinlichkeitsparameter $\theta = .8$ und
  - Anzahl Trials $n = 50$

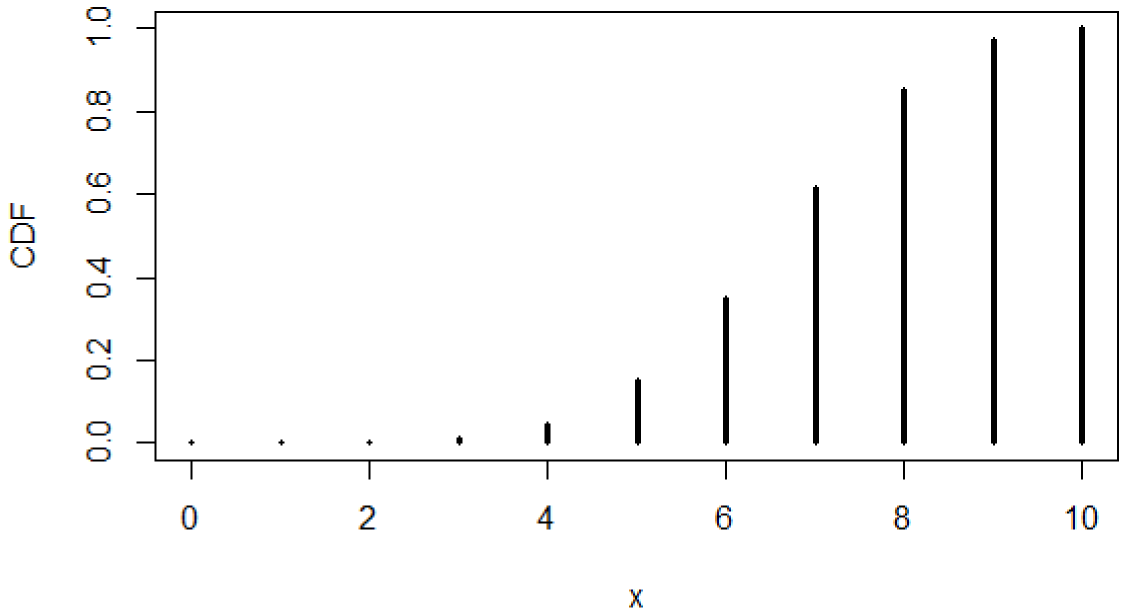
Verteilungsfunktion (CDF)

- **Formel:**
  - $F_X(x|\theta) = \sum_{j=1}^{x} \binom{n}{i} \cdot \theta^j (1-\theta)^{n-j}$
- **R Funktion:**

```
> pbinom(size=n)
```

- **R Code:**

```
> n <- 10
> x <- seq(0, n, by=1)
> CDF <- pbinom(x, size=n, prob=.7)
> plot(x, CDF, type="h", lwd = 3)
```



Kurze Aufgabe

- Zeichnet eine Verteilungsfunktion der Binomialverteilung mit
  - Wahrscheinlichkeitsparameter $\theta = .8$ und
  - Anzahl Trials $n = 50$

Poissonverteilung

diskret

Datensatz in R: gala (gala.RData)

	Species	Endemics	Area	Elevation	Nearest	Scruz	Adjacent
Baltra	**58**	23	25,09	346	0,6	0,6	1,84
Bartolome	**31**	21	1,24	109	0,6	26,3	572,33
Caldwell	**3**	3	0,21	114	2,8	58,7	0,78
Champion	**25**	9	0,1	46	1,9	47,4	0,18
Coamano	**2**	1	0,05	77	1,9	1,9	903,82
Daphne.Major	**18**	11	0,34	119	8	8	1,84
Daphne.Minor	**24**	0	0,08	93	6	12	0,34
Darwin	**10**	7	2,33	168	34,1	290,2	2,85
Eden	**8**	4	0,03	71	0,4	0,4	17,95
Enderby	**2**	2	0,18	112	2,6	50,2	0,1
Espanola	**97**	26	58,27	198	1,1	88,3	0,57
Fernandina	**93**	35	634,49	1494	4,3	95,3	4669,32
Gardner1	**58**	17	0,57	49	1,1	93,1	58,27
Gardner2	**5**	4	0,78	227	4,6	62,2	0,21
Genovesa	**40**	19	17,35	76	47,4	92,2	129,49
Isabela	**347**	89	4669,32	1707	0,7	28,1	634,49
Marchena	**51**	23	129,49	343	29,1	85,9	59,56
Onslow	**2**	2	0,01	25	3,3	45,9	0,1
…	**…**	…	…	…	…	…	…

Beschreibung

- **Verwendung:**
  - Anzahl/Häufigkeit von Ereignissen zu gewisser Zeit und/oder Region
  - Anzahl depressiver Episoden in einem Jahr
- **Schreibweise:**
  - $X \sim Pois(\lambda)$
  - X folgt einer Poissonverteilung mit rate-Parameter (Häufigkeit) λ.
- **Sonstige Eigenschaften:**
  - Erwartungswert: λ
  - Varianz: λ

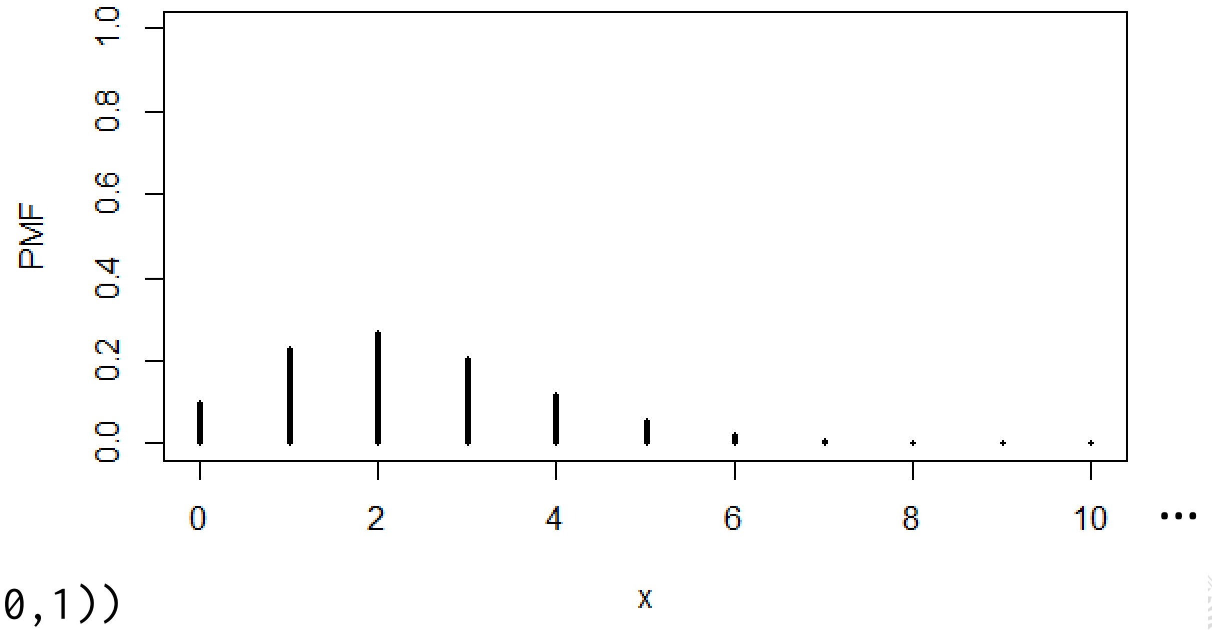
Wahrscheinlichkeitsfunktion (PMF)

- **Formel:**
  - $f_X(x|\lambda) = \frac{\lambda^x \exp(-\lambda)}{x!}$, wobei x in $\{1, 2, 3, \ldots\}$ und $\lambda > 0$
  - Bsp.: x ist die Anzahl depressiver Episoden in einem Jahr
- **R Funktion:**

```
> dpois()
```

- **R Code:**

```
> x <- seq(0, 10, by=1)
> PMF <- dpois(x, lambda=2.3)
> plot(x, PMF, type="h", lwd = 3, ylim=c(0,1))
```



Kurze Aufgabe

- Zeichnet eine Wahrscheinlichkeitsfunktion der Poissonverteilung mit
  - rate-Parameter $\lambda = 4.5$
  - Passt x so an, dass möglichst die ganze W'keitsfunktion zu sehen ist (siehe vorherige Folie)

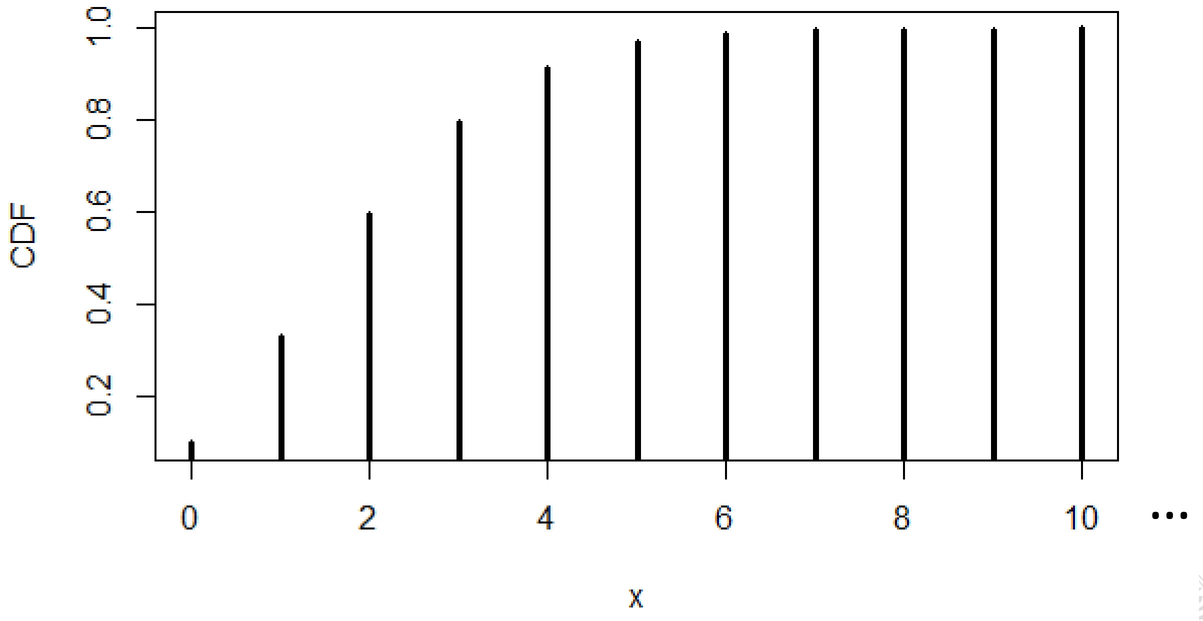
Verteilungsfunktion (CDF)

- **Formel:**
  - $F_X(x|\lambda) = \sum_{j=1}^{x} \frac{\lambda^j \exp(-\lambda)}{j!}$
- **R Funktion:**

```
> ppois()
```

- **R Code:**

```
> x <- seq(0, 10, by=1)
> CDF <- ppois(x, lambda=2.3)
> plot(x, CDF, type="h", lwd = 3)
```



Kurze Aufgabe

- Zeichnet eine Verteilungsfunktion der Poissonverteilung mit
  - rate-Parameter $\lambda = 4.5$
  - Passt x so an, dass möglichst die ganze Verteilungsfunktion zu sehen ist (siehe vorherige Folie)

Exponentialverteilung

kontinuierlich

Datensatz in R: machine (machine.RData)

id	h_failure
1	315
2	249
3	354
4	225
5	1978
6	109
7	867
8	1851
9	17
10	1146
11	653
12	545
13	829
14	2097
15	204
16	297
17	499
18	655
…	…

Beschreibung

- **Verwendung:**
  - Zeit bis neues Ereignis eintritt
  - Bsp.: Zeit bis zur nächsten depressiven Episode
  - Mit dieser Verteilung kann man auch räumliche Abstände abbilden, ist aber unüblich
- **Schreibweise:**
  - $X \sim Exp(\lambda)$
  - X folgt einer Exponentialverteilung mit rate-Parameter λ.
- **Sonstige Eigenschaften:**
  - Erwartungswert: $\frac{1}{\lambda}$
  - Varianz: $\frac{1}{\lambda^2}$

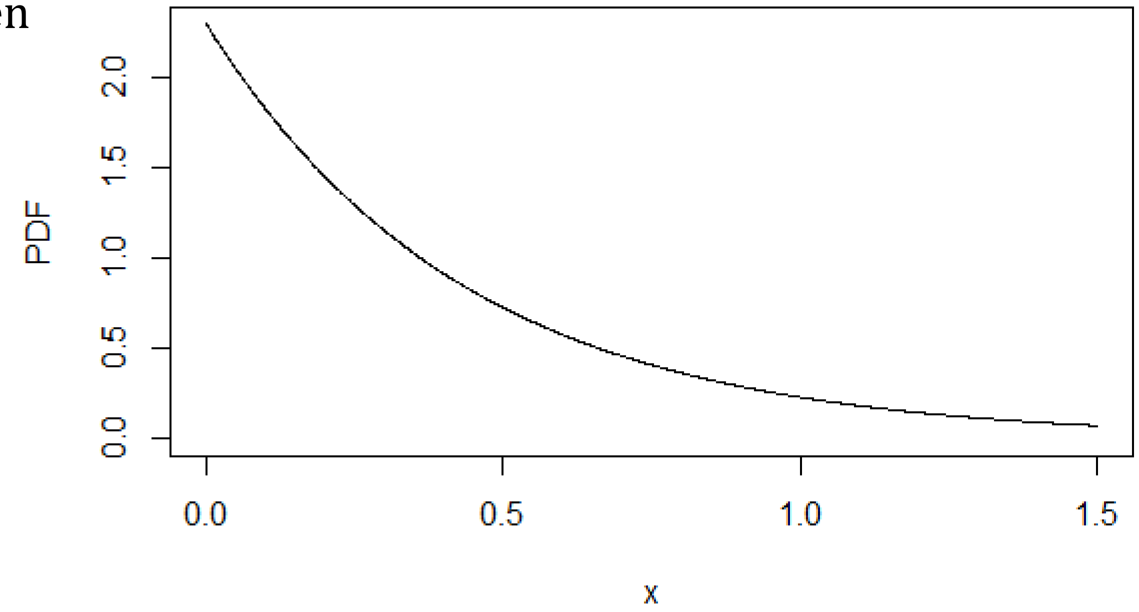
Dichtefunktion (PDF)

- **Formel:**
  - $f_X(x|\lambda) = \lambda \exp(-\lambda x)$, wobei $x > 0$ und $\lambda > 0$
  - Bsp.: x ist die Zeit bis zur nächsten depressiven Episode, wenn λ die Anzahl depressiver Episoden bspw. pro Jahr ist
- **R Funktion:**

```
> dexp()
```

- **R Code:**

```
> x <- seq(0, 1.5, by=.001)
> PDF <- dexp(x, rate=2.3)
> plot(x, PDF, type="l", ylim = c(0,2.3))
```



…

Kurze Aufgabe

- Zeichnet eine Dichtefunktion der Exponentialverteilung mit
    - rate-Parameter $\lambda = 4.5$
    - Passt x so an, dass möglichst die ganze Dichtefunktion zu sehen ist (siehe vorherige Folie)

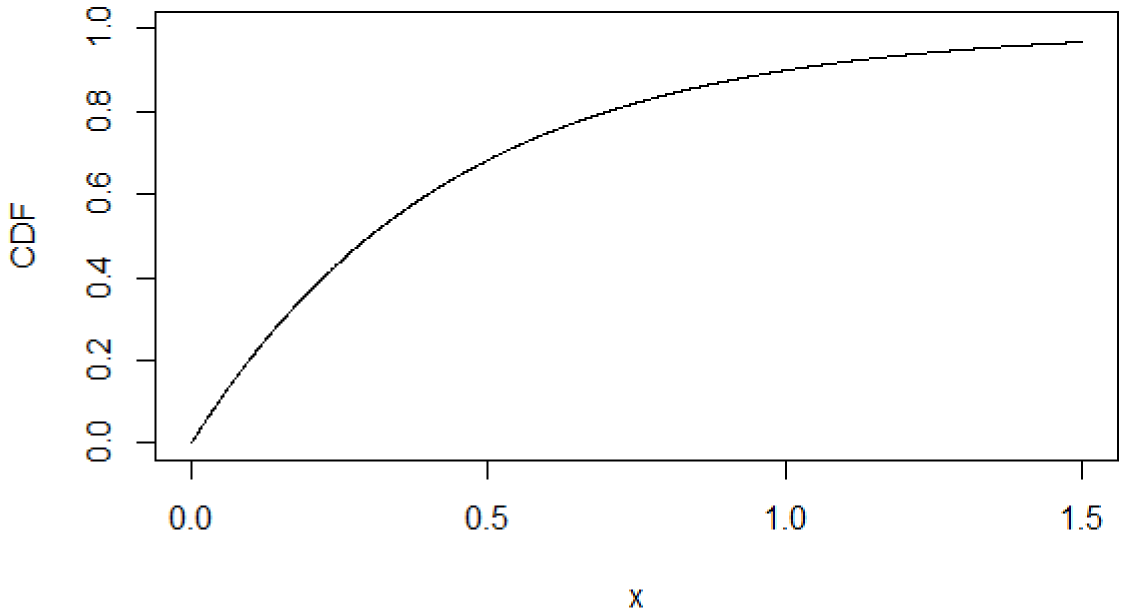
Verteilungsfunktion (CDF)

- **Formel:**
  - $F_X(x|\lambda) = 1 - \exp(-\lambda x)$
- **R Funktion:**

```
> pexp()
```

- **R Code:**

```
> x <- seq(0, 1.5, by=.001)
> CDF <- pexp(x, rate=2.3)
> plot(x, CDF, type="l", ylim=c(0,1))
```



…

Kurze Aufgabe

- Zeichnet eine Verteilungsfunktion der Exponentialverteilung mit
  - rate-Parameter $\lambda = 4.5$
  - Passt x so an, dass möglichst die ganze Verteilungsfunktion zu sehen ist (siehe vorherige Folie)

Beschreibende Modelle

Generalisierte lineare Modelle (GLMs)

Allgemeines lineare Modell

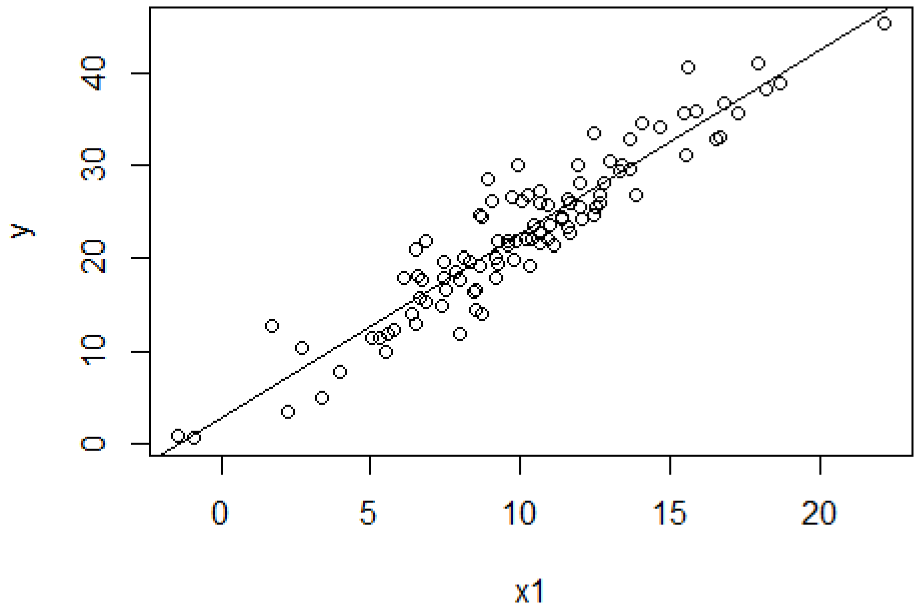
- **Modellgelichung:**
  - $y_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon_i$
  - Die AV Y wird durch eine Gerade vorhergesagt (durch Intercept und Steigungen von Prädiktoren), von der jede Person i zufällig abweicht (Residuen ϵ_i)
- **Matrixschreibweise:**
  - $\boldsymbol{X\beta}$: linearer Prädiktor
- **Prädiktoren:**
  - Kontinuierlich und/oder diskret
- **Annahmen:**
  - Residuen normalverteilt, Homoskedastizität und linearer Zusammenhang

R Code für lineare Regression

```
> # Daten simulieren
> set.seed(1234)
> beta0 <- 3
> beta1 <- 2
> epsilon <- rnorm(n = 100, mean = 0, sd = 3)
> x1 <- rnorm(n = 100, mean = 10, sd = 4)
> y <- beta0 + beta1*x1 + epsilon
> df <- data.frame(y=y, x1=x1)

> # Modell rechnen
> linmod <- lm(formula = y ~ 1 + x1, data = df)
> summary(linmod)

> # Plot zeichnen
> plot(x1, y)
> abline(a = linmod$coefficients[1], b = linmod$coefficients[2])
```



Logistische Regression

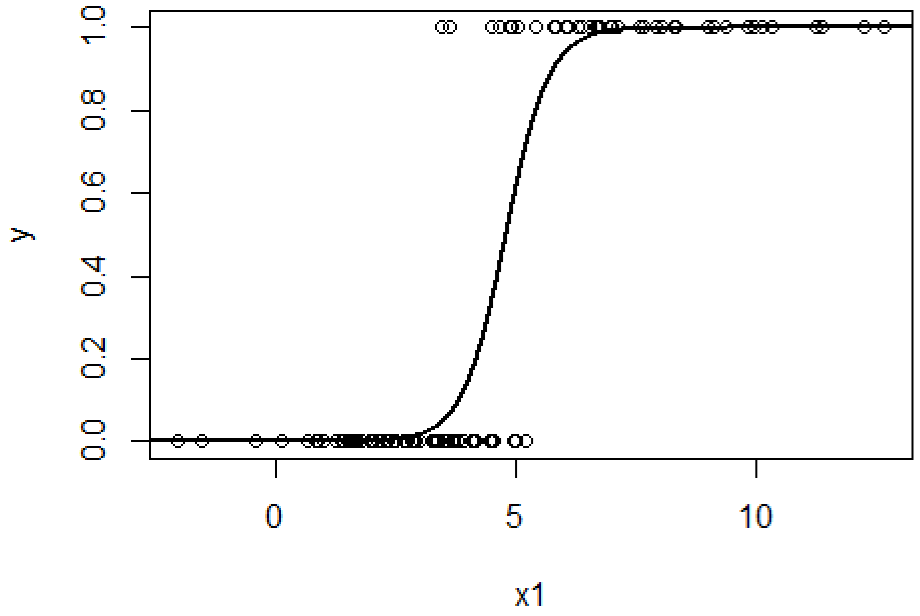
- **Daten:**
  - Annahme: Entstanden aus einer **Bernoulliverteilung**.
  - Modellierung:
    - Erwartungswert von Y ($\mathrm{E}(Y) = \theta$) wird erklärt durch den transformierten linearen Prädiktor. θ ist unser Wahrscheinlichkeitsparameter aus der Bernoulli-/Binomialverteilung oder
    - Transformierter Erwartungswert wird erklärt durch den linearen Prädiktor $\boldsymbol{X\beta}$.
- **Modellgleichung:**
  - $\mathrm{E}(Y) = \theta = \frac{1}{1+\exp(-\boldsymbol{X\beta})}$ (logistische Funktion = inverse Linkfunktion) oder
  - $\mathrm{logit}(\theta) = \log\left(\frac{\theta}{1-\theta}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k = \boldsymbol{X\beta}$ (logit Funktion aka. logodds = Linkfunktion)

R Code für logistisches Modell

```
> # Daten simulieren
> set.seed(1234)
> beta0 <- -10; beta1 <- 2
> x1 <- rnorm(n = 100, mean = 5, sd = 3)
> Xbeta <- beta0 + beta1*x1; theta <- 1/(1+exp(-Xbeta))
> y <- rbinom(n = 100, size = 1, prob = theta)
> df <- data.frame(y=y, x1=x1)

> # Modell rechnen
> GLM <- glm(formula = y ~ 1 + x1, data = df,
>            family = binomial(link = "logit"))
> summary(GLM)

> # Plot zeichnen
> xseq <- seq(-3, 14, by=.001)
> Xbeta_est <- GLM$coefficients[1]+GLM$coefficients[2]*xseq
> plot(x1, y); lines(x = xseq, y = 1/(1+exp(-Xbeta_est)), lwd = 2)
```



Poisson Regression

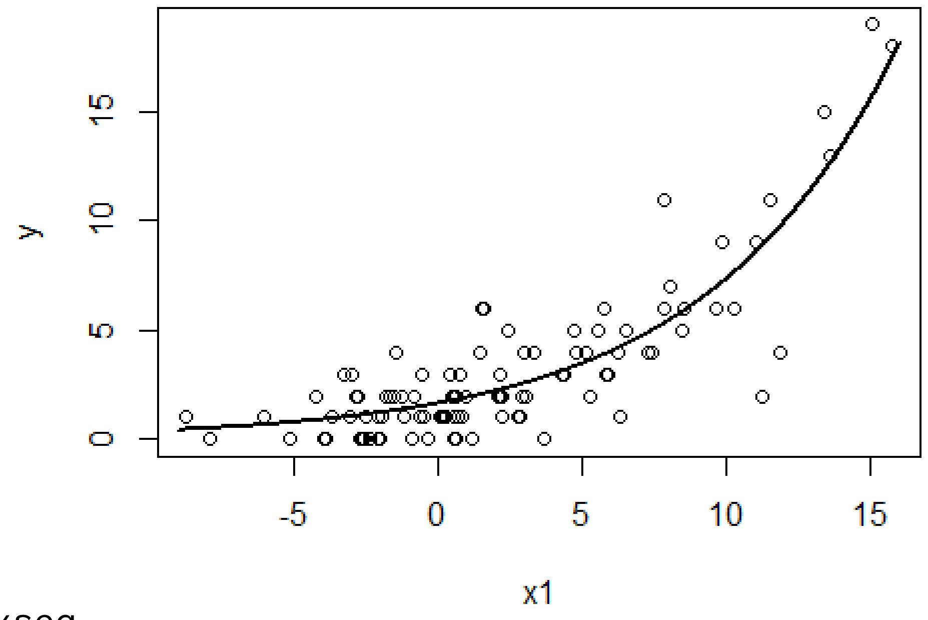
- **Daten:**
  - Annahme: Entstanden aus einer **Poissonverteilung**.
  - Modellierung:
    - Erwartungswert von Y ($\mathrm{E}(Y) = \lambda$) wird erklärt durch den transformierten linearen Prädiktor. λ ist unser rate-Parameter aus der Poissonverteilung oder
    - Transformierter Erwartungswert wird erklärt durch den linearen Prädiktor $\boldsymbol{X\beta}$.
- **Modellgleichung:**
  - $\mathrm{E}(Y) = \lambda = \exp(\boldsymbol{X\beta})$ (exp Funktion = inverse Linkfunktion) oder
  - $\log(\lambda) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k = \boldsymbol{X\beta}$ (log Funktion = Linkfunktion)

R Code für Poisson Regression

```
> # Daten simulieren
> set.seed(1234)
> beta0 <- .5; beta1 <- .15
> x1 <- rnorm(n = 100, mean = 3, sd = 5)
> Xbeta <- beta0 + beta1*x1; lambda <- exp(Xbeta)
> y <- rpois(n = 100, lambda = lambda)
> df <- data.frame(y=y, x1=x1)

> # Modell rechnen
> GLM <- glm(formula = y ~ 1 + x1, data = df,
>            family = poisson(link = "log"))
> summary(GLM)

> # Plot zeichnen
> xseq <- seq(-9, 16, by=.001)
> Xbeta_est <- GLM$coefficients[1]+GLM$coefficients[2]*xseq
> plot(x1, y); lines(x = xseq, y = exp(Xbeta_est), lwd = 2)
```



Modell-Eigenschaften von (G)LMs

- **Varianz aufzuklären:**
  - Welche Prädiktoren sagen die AV gut vorher bzw. korrelieren hoch mit der AV?
- **Daten beschreiben:**
  - Wenn X_1 und eine Einheit steigt, so verändert sich Y im Schnitt um …
- **Problem für kognitive Aufgaben/Daten:**
  - Kognitive Prozesse können wir nicht direkt messen und daher nicht als Variable mitaufnehmen.
  - Solche Prozesse muss man spezifisch in einem Modell über Annahmen einbauen.
  - Generalisierte lineare Modelle (GLMs) können das nur bedingt.
  - Lösung: mathematische/kognitive Modelle, deren Parameter kognitive Mechanismen wiederspiegeln

Erklärende Modelle

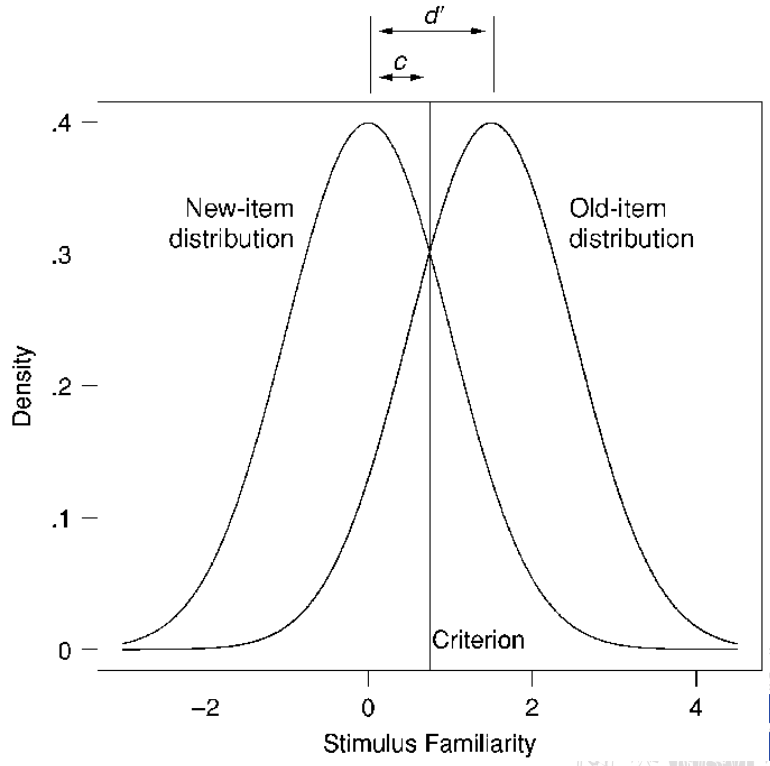
Mathematische/kognitive Modelle

Wozu kognitive Modelle (KMs)?

- Ohne kognitive Modelle können wir nichts über die kognitiven Wirkmechanismen verstehen
- Wir brauchen verschiedene kognitive Modelle, da sie teilweise unterschiedliche Aspekte von kognitiven Wirkmechanismen in Betracht ziehen
- Kognitive Mechanismen werden direkt in das Modell übersetzt/eingebaut
  - Jedes Modell hat unterschiedliche Annahmen über kognitive Wirkmechanismen (siehe nächste zwei Folien)
- Durch das Testen von verschiedenen Modellen können wir Aussagen über deren Plausibilität bzw. deren Nützlichkeit treffen
  - Kein Modell ist wahr, aber einige sind nützlich!

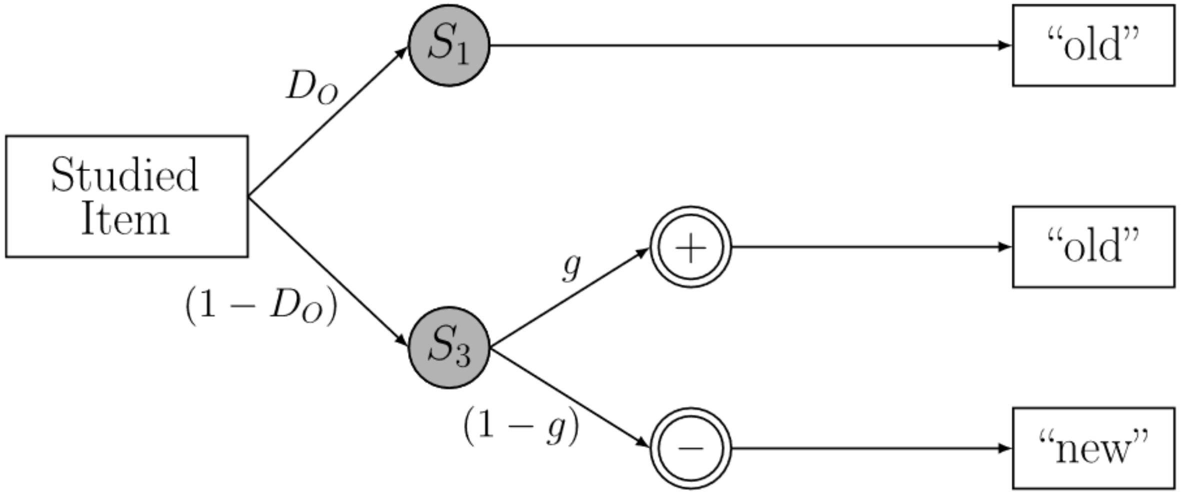
Kognitive Mechanismen – Beispiel 1: SDT

- **Signalentdeckungstheorien:**
  - haben die Annahme, dass Entscheidungen aufgrund von Familiarität von Stimuli (*Stimulus Familiarity*) getroffen werden → Für die Familiarität wird pro Stimulus-Typ eine Normalverteilung mit spezifischem Mittelwert (und Standardabweichung) angenommen. Der Mittelwert ist somit ein Parameter, der im Modell geschätzt wird.



Kognitive Mechanismen – Beispiel 2: MPTs

- **Multinomiale Verarbeitungsbaum-Modelle (MPT):**
  - haben die Annahme, dass Entscheidungen über Entscheidungsbäume (wenn-dann-Verknüpfungen) entschieden werden. Jede mögliche Verzweigung entspricht einem kognitiven Prozess und jeder Knoten einem mentalen Zustand. Die Parameter sind hier die W‘keiten für die verschiedenen Verzweigungen (bzw. die Ausgänge der Prozesse).



Verzweigungen (kogn. Prozesse):
- D_O: Entdecken (detection) von alten (old) Wörtern
- g: Raten (guessing)

Knoten (mentale Zustände):
- S_1: Wort als alt erkannt
- S_3: Unsicherheit über das Wort

Likelihood funktion

Erster Einblick

Definition

- Sei X eine beliebige Zufallsvariable mit der Verteilungsfunktion $F_X(x|\eta)$ und W'keits- bzw. Dichtefunktion $f_X(x|\eta)$, wobei η ein Vektor von Parametern ist, dann ist die Likelihood Funktion definiert als
$$L(\eta|x) = f_X(x|\eta),$$
wobei x die Realisation von X ist, also die beobachteten Daten (heißt x ist fix).

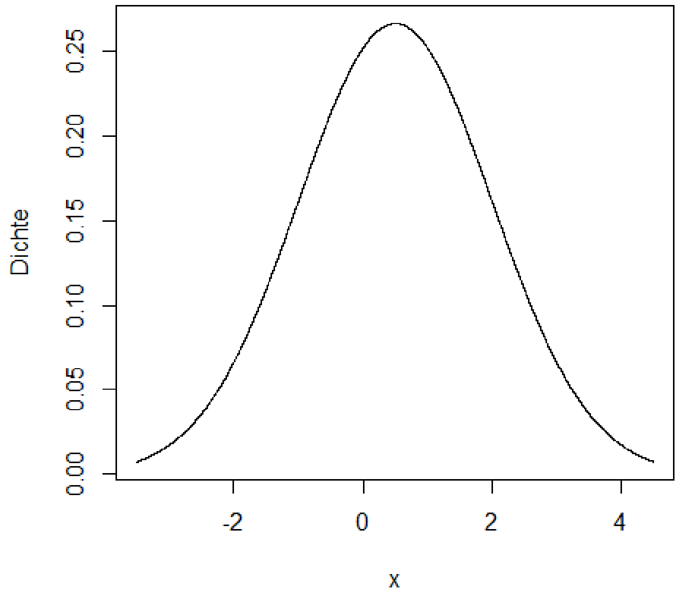
Unterschied zur W'keits- bzw. Dichtefunktion

- **Der Unterschied zwischen der PMF/PDF $\boldsymbol{f_X(x|\eta)}$ und der Likelihood Funktion $\boldsymbol{L(\eta|x)}$:**
  - $f_X(x|\eta)$ ist eine Funktion von x, wobei η konstant/fix gehalten wird und
  - $L(\eta|x)$ ist eine Funktion von η, wobei x konstant/fix gehalten wird.
- Die Likelihood Funktion teil zwar dieselbe Formel wie die PMF/PDF, ist aber eine Funktion von η und nicht von x. Die Daten x sind konstant gehalten auf unsere Beobachtung(en).

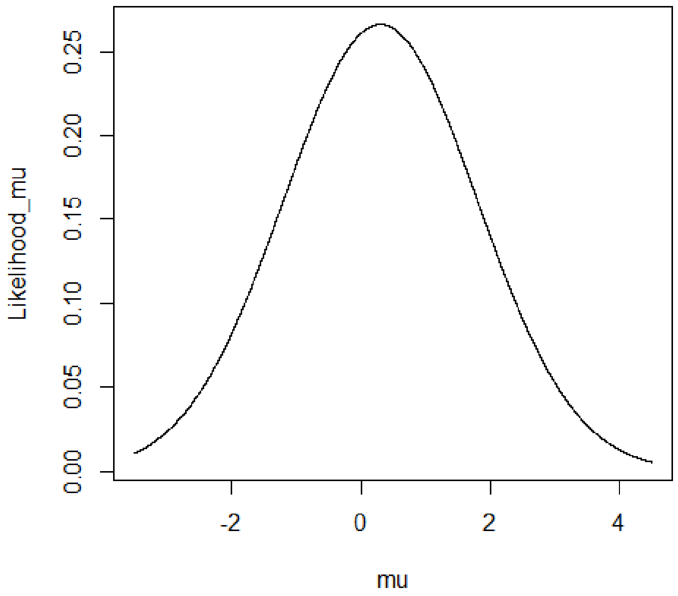
Beispiel – Normalverteilung

- **Formel**: $\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$
- **Dichtefunktion PDF**: $f_X(x|\eta) = \frac{1}{\sqrt{2\pi 1.5^2}} \exp\left(\frac{(x-0.5)^2}{2 \cdot 1.5^2}\right)$, wobei hier $\eta = (\mu, \sigma) = (0.5, 1.5)$
- **Likelihood Funktion**: $L(\eta|x) = L(\mu, \sigma|x = .3) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(0.3-\mu)^2}{2\sigma^2}\right)$
  - Man kann auch einen der Parameter (z. B. σ) konstant halten. Dann ist die Likelihood Funktion eine eindimensionale Funktion: $L(\mu|\sigma = 1.5, x = .3) = \frac{1}{\sqrt{2\pi \cdot 1.5^2}} \exp\left(\frac{(0.3-\mu)^2}{2 \cdot 1.5^2}\right)$

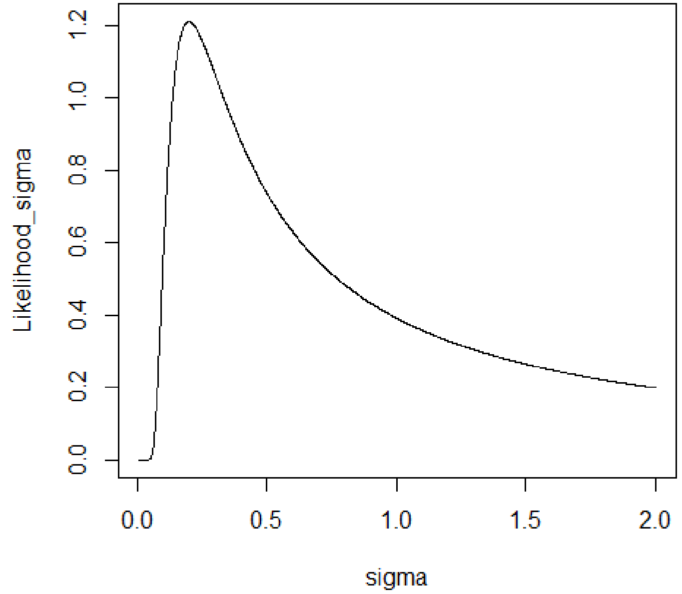
Beispiel – Normalverteilung




```
> x <- seq(-3.5, 4.5, by = .001)
> Dichte <- 1/sqrt(2*pi*1.5^2) *
  exp(-(x-0.5)^2 / (2*1.5^2))
> plot(x, Dichte, type = "l")
```




```
> mu <- seq(-3.5, 4.5, by = .001)
> Likelihood_mu <- 1/sqrt(2*pi*1.5^2) *
  exp(-(0.3-mu)^2 / (2*1.5^2))
> plot(mu, Likelihood_mu, type = "l")
```



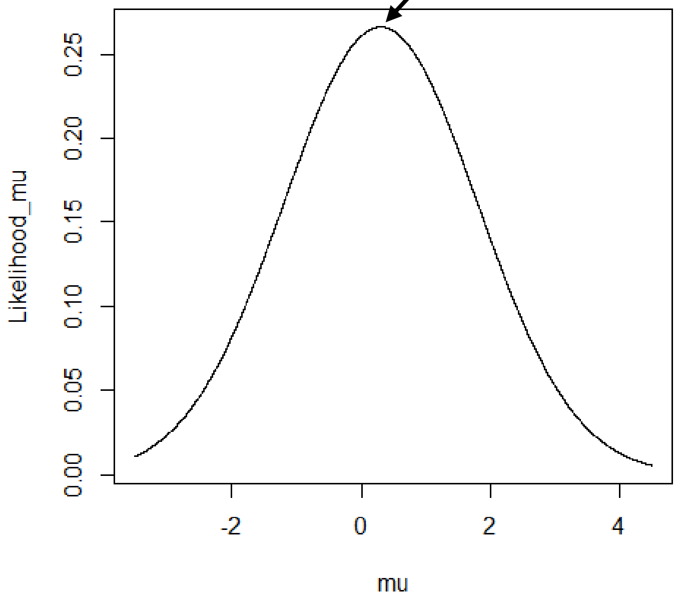

```
> sigma <- seq(0.01, 2, by = .001)
> Likelihood_sigma <- 1/sqrt(2*pi*sigma^2) *
  exp(-(0.3-0.5)^2 / (2*sigma^2))
> plot(sigma, Likelihood_sigma, type = "l")
```

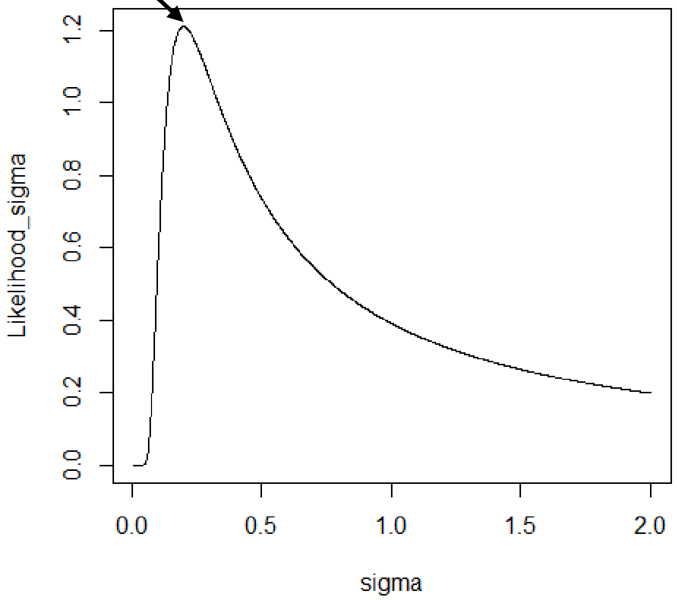
Wozu brauchen wir die Likelihood Funktion?

- **Parameterschätzung:**
  - Maximum Likelihood Schätzmethode: für welches η ist die Likelihood am höchsten?
  - Intuition auf der nächsten Folie
- **Modellgüte:**
  - Die Likelihood Funktion ist ein wesentlicher Bestandteil der Anpassungsgüte eines Modells (siehe übernächste Sitzung)

Beispiel – Normalverteilung

Likelihood Funktion maximal






```
> mu <- seq(-3.5, 4.5, by = .001)
> Likelihood_mu <- 1/sqrt(2*pi*1.5^2) *
  exp(-(0.3-mu)^2 / (2*1.5^2))
> plot(mu, Likelihood_mu, type = "l")
```

```
> sigma <- seq(0.01, 2, by = .001)
> Likelihood_sigma <- 1/sqrt(2*pi*sigma^2) *
  exp(-(0.3-0.5)^2 / (2*sigma^2))
> plot(sigma, Likelihood_sigma, type = "l")
```

Zusammenfassung

R Funktionen für Verteilungen

	Normal	Bernoulli	Binomial	Poisson	Exponential
PDF	dnorm()	dbinom(size=1)	dbinom()	dpois()	dexp()
CDF	pnorm()	pbinom (size=1)	pbinom ()	ppois()	pexp()
Quantil	qnorm()	qbinom (size=1)	qbinom ()	qpois()	qexp()
Zufall	rnorm()	rbinom (size=1)	rbinom ()	rpois()	rexp()

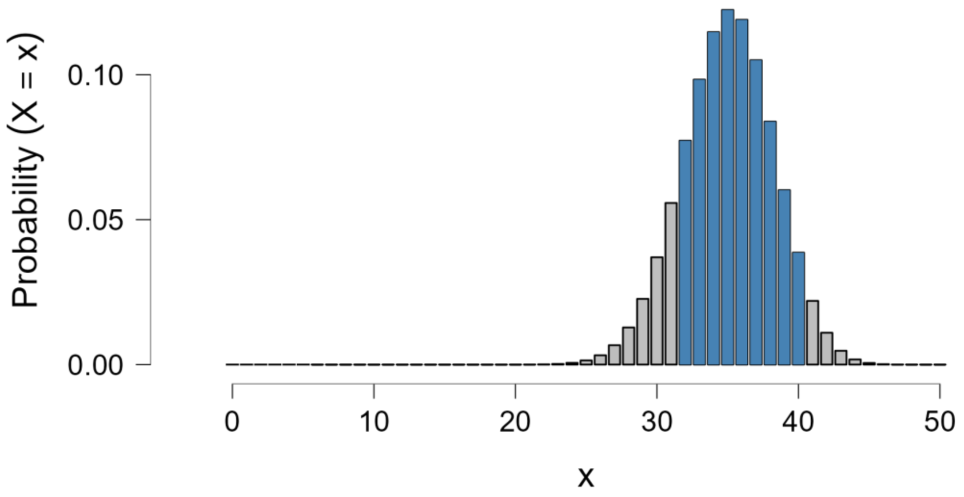
Modelltypen (erklärend, beschreibend)

	GLM	KM
Ziel	Beschreibung von Daten. Im besten Fall beschreiben sie einen kausalen Zusammenhang, wenn das Design und die Modelle richtig gewählt sind.	Erklärung der kognitiven Wirkmechanismen hinter einer/m Entscheidung/Urteil. Damit können kognitive Wirkmechanismen getestet werden.
Parameter	Parameter beschreiben Zusammenhänge von Variablen.	Parameter spiegeln kognitive Prozesse/Mechanismen wieder.

Übungen

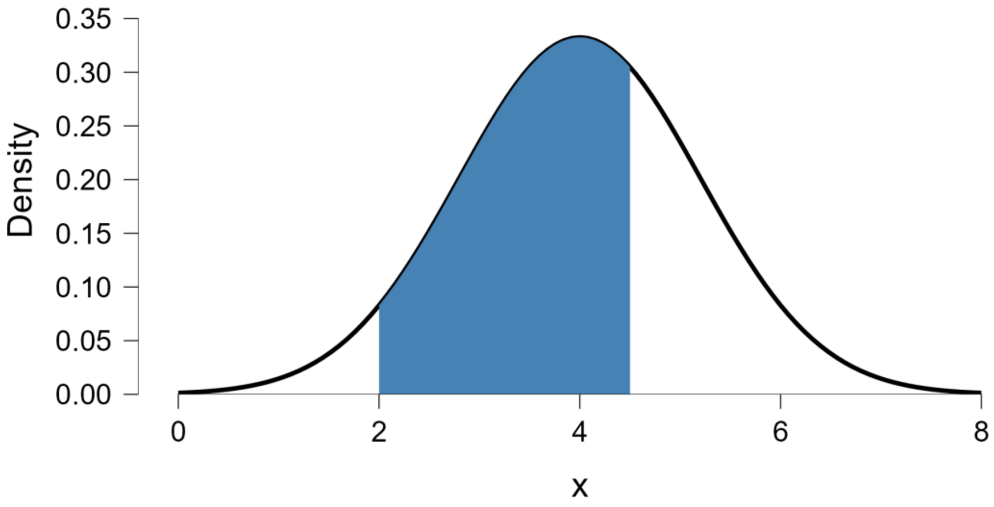
Aufgabe 1

- Wenn wir eine Binomialverteilung mit Wahrscheinlichkeitsparameter 0.7 haben und 50 Beobachtungen, wie hoch ist dann die Wahrscheinlichkeit, dass ein Wert x in dem Intervall $[32, 40]$ liegt?



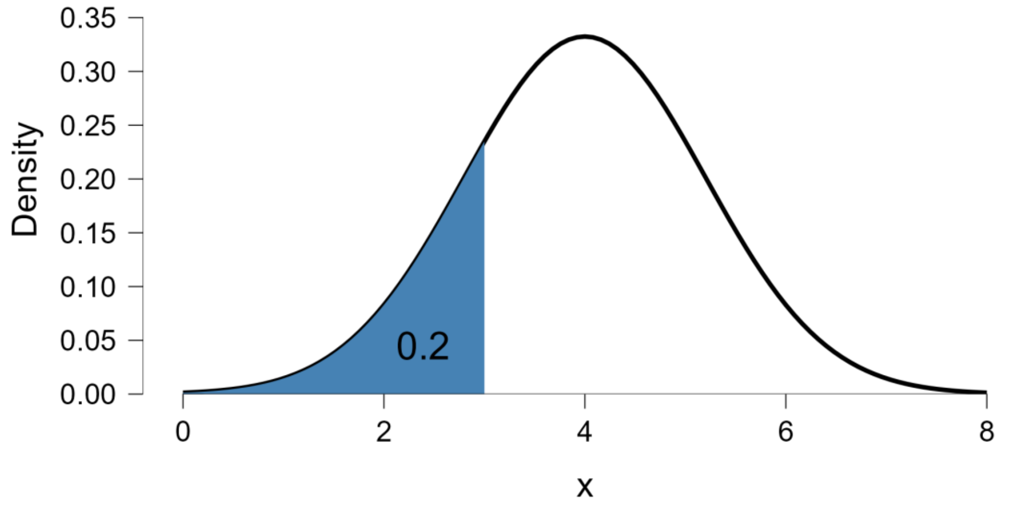
Aufgabe 2

- Wenn wir eine Normalverteilung mit Mittelwert 4 und Standardabweichung 1.2 haben, wie hoch ist dann die Wahrscheinlichkeit, dass ein Wert x in dem Intervall $[2, 4.5]$ liegt?
    - Hinweis: mit `pnorm(x = 2, mean = 4, sd = 1.2)` kann man die Wahrscheinlichkeit für x in $(-\infty, 2]$ berechnen



Aufgabe 3

- Wenn wir eine Normalverteilung mit Mittelwert 4 und Standardabweichung 1.2 haben, was sind dann die Werte für die Quartile?
  - Hinweis: mit `qnorm(p = .2, mean = 4, sd = 1.2)` kann man den Wert $x_{p=.2}$ finden, für den 20% der x-Werte kleiner ist



Aufgabe 4

- In einer Studie wird untersucht, welchen Einfluss Stress und Unterstützung auf den Alkohol-Rückfall von trockenen Alkoholikern hat.
  - Welches beschreibende Modell würden Sie hier wählen (logistische Regression oder Poisson Regression)? Begründen Sie warum.
  - Rechnen Sie eine entsprechende Regression in R mit dem Datensatz `alcohol` von `alcohol.Rdata`
    - Nutzen Sie hierfür beide Prädiktoren (Stress und Unterstützung) und die R Funktion `glm()` mit dem entsprechenden `family` Argument (`binoimal(link = "logit")` oder `poisson(link = "log")`)).
  - Was sagt Ihnen der Output… Braucht es beide Prädiktoren?