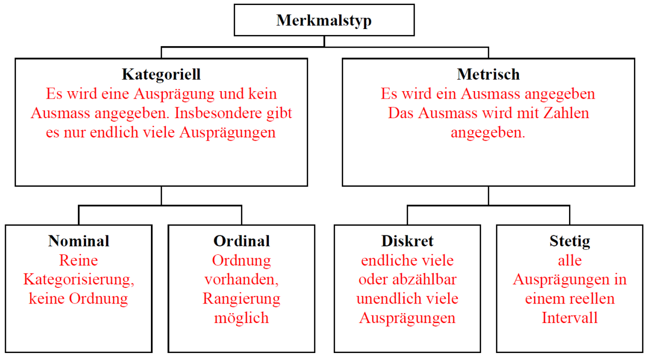
Stochastik und Statistik

Deskriptive Statistik

Begriffe

- **Merkmalsträger**: Objekte, über die Informationen gesammelt werden (z.B. Personen, Tiere, Dinge)
- **Merkmal**: Eigenschaft eines Merkmalsträgers (z.B. Alter, Geschlecht, Größe)
- **Grundgesamtheit**: Gesamte Menge aller Merkmalsträger
- **Stichprobe**: Teilmenge der Grundgesamtheit
- **Vollerhebung**: Untersuchung der gesamten Grundgesamtheit
- **Ausprägung**: Mögliche Werte eines Merkmals



Relative/Absolute Häufigkeit

Nicht klassierte Daten

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

Klassierte Daten

$$f_i = \frac{g(x)}{b_i} \quad \text{mit} \quad g(x) = \frac{h_i}{n}$$

- f_i: relative Häufigkeit der Ausprägung i (PMF)
- h_i: absolute Häufigkeit der Ausprägung i
- n: Gesamtanzahl der Beobachtungen

Summenhäufigkeit

Nicht klassierte Daten

$$H(x) = \sum_{j=1}^{x} h_j \quad \text{bzw.} \quad F(x) = \sum_{j=1}^{x} f_j$$

Klassierte Daten

$$H(x_k) = \sum_{j=1}^{k} h_j \quad \text{bzw.} \quad F(x) = \int_{-\infty}^{x} f(y)\,dy$$

- $H(x)$: Summe aller absoluten Häufigkeiten bis Ausprägung x
- Wenn h absolute Häufigkeit, dann heisst H Summenhäufigkeit
- Wenn f relative Häufigkeit, dann heisst F kumulative Verteilungsfunktion (CDF)

Eigenschaften der PMF und CDF nicht klassierter Daten

- $0 \leq f_i \leq 1$ und $0 \leq F(x) \leq 1$
- $\sum_i f_i = 1$ bzw. $F(x_{\max}) = 1$
- $F(x)$ ist monoton wachsend $\Rightarrow F(x_1) \leq F(x_2)$ für $x_1 < x_2$
- $F(x)$ ist rechtsstetig heisst der Wert an der Stelle x ist gleich dem Grenzwert von rechts

Eigenschaften der PMF und CDF klassierter Daten

- $0 \leq f_i \leq 1$ und $0 \leq F(x) \leq 1$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$ und $F'(x) = f(x)$
- Für $x \in K_i$ gilt: $\frac{F(x)-F(a_i)}{b_i} = \frac{F(a_{i+1})-F(a_i)}{b_i} = f_i$

Klassierte Daten

- Daten werden in Klassen eingeteilt
- Klassenbreite: $b_i = a_{i+1} - a_i$ (mit a_i Klassenuntergrenze)
- Klassenmitte: $M_i = \frac{a_{i+1}-a_i}{2}$ (mit l untere und u obere Klassenbegrenzung)
- Dichte: $d_i = \frac{f_i}{b_i}$

Kenngrössen

- **Arithmetisches Mittel**:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{bzw. wenn klassiert} \quad \bar{x} = \sum_{i=1}^{m} f_i \cdot M_i$$

- **Median**:

$$\tilde{x} = \begin{cases} x_{\frac{n+1}{2}}, & n \text{ ungerade} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}, & n \text{ gerade} \end{cases}$$

- **Modalwert**:

$$\hat{x} = x_i \quad \text{mit } h_i = \max(h_1, h_2, \ldots, h_k)$$

- **Varianz**:

$$\tilde{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (\text{bei klassierten Daten } x_i = M_i)$$

$$\tilde{s}^2 = \overline{x^2} - \bar{x}^2 \quad \text{mit} \quad \overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (\text{bzw.}) \quad s^2 = \frac{n}{n-1}\tilde{s}^2$$

- **Standardabweichung**:

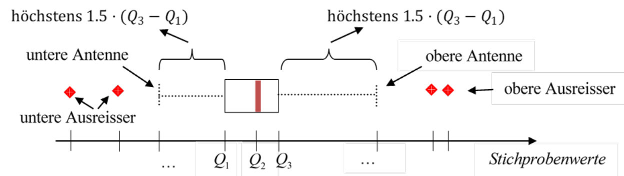
$$s = \sqrt{s^2} \quad \text{bzw.} \quad \tilde{s} = \sqrt{\tilde{s}^2}$$

- **Spannweite**:

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

- **Quantile**:

$$Q_p = \begin{cases} x_k, & \text{für } n \cdot p \text{ ganzzahlig} \\ x_{\lceil n \cdot p \rceil}, & \text{sonst} \end{cases}$$



Kovarianz und Korrelation

- **Kovarianz**:

$$\tilde{s}_{xy} = \operatorname{Cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \overline{xy}-\bar{x}\cdot\bar{y}$$

- **Korrelationskoeffizient nach Pearson**:

$$r = \frac{\operatorname{Cov}(X,Y)}{s_X s_Y} = \frac{\tilde{s}_{xy}}{\sqrt{\overline{x^2}-\bar{x}^2}\cdot\sqrt{\overline{y^2}-\bar{y}^2}}$$

- r liegt im Intervall $[-1,1]$
- $r>0$: positive Korrelation, $r<0$: negative Korrelation, $r=0$: keine Korrelation
- $\tilde{s}_{xy}$ und r beschreiben die lineare Abhängigkeit zwischen zwei Merkmalen d.h. wie stark die Merkmale von der Geraden $y=mx+b$ abweichen
- Ist nicht robust gegenüber Ausreissern

Rangkorrelation nach Spearman

- Rang:

$$rg(x_i) = \begin{cases} k, & \text{wenn } x_i \text{ der} \\ & k\text{-te kleinste Wert ist} \\ \frac{1}{m}\sum_{j=1}^{m} rg(x_{i_j}), & \text{bei } m \text{ gleichen Werten} \end{cases}$$

- Korrelationskoeffizient nach Spearman:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \quad \text{mit } d_i = rg(x_i) - rg(y_i)$$

- r_s beschreibt die monotone Abhängigkeit zwischen zwei Merkmalen
- Ist robust gegenüber Ausreissern

Kombinatorik

Grundlagen

- **Multiplikationsregel**: Wenn ein Vorgang in m Arten und ein zweiter Vorgang in n Arten durchgeführt werden kann, dann können beide Vorgänge in $m\cdot n$ Arten durchgeführt werden.
- **Additionsregel**: Wenn ein Vorgang in m Arten und ein zweiter Vorgang in n Arten durchgeführt werden kann, und beide Vorgänge sich gegenseitig ausschließen, dann können beide Vorgänge in $m+n$ Arten durchgeführt werden.

Binomialkoeffizient

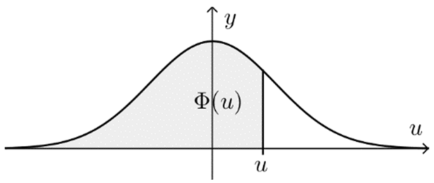
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- Anzahl der Möglichkeiten, aus n Elementen k Elemente ohne Zurücklegen und ohne Beachtung der Reihenfolge auszuwählen
- n: Gesamtanzahl der Elemente
- k: Anzahl der auszuwählenden Elemente
- $\binom{n}{0}=1$, $\binom{n}{n}=1$, $\binom{n}{1}=n$
- Symmetrie: $\binom{n}{k}=\binom{n}{n-k}$
- Rekursion: $\binom{n}{k}=\binom{n-1}{k-1}+\binom{n-1}{k}$
- Binomischer Lehrsatz: $(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^{n-k}y^k$
- Summe der Binomialkoeffizienten: $\sum_{k=0}^{n}\binom{n}{k}=2^n$

Kombinatorische Auswahlmöglichkeiten

	Mit Zurücklegen	**Ohne Zurücklegen**
Mit Reihenfolge	n^k *Zahlenschloss*	$\frac{n!}{(n-k)!}$ *Rennen*
Ohne Reihenfolge	$\binom{n+k-1}{k}$ *Warenwahl*	$\binom{n}{k}$ *Gruppenwahl*

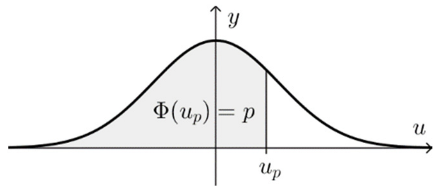
Tabelle 1: CDF $\Phi(u)$ der Standardnormalverteilung



$$P(U \le u) = \Phi(u)$$
$$P(U \ge u) = 1 - \Phi(u)$$
$$P(-u \le U \le u) = 2 \cdot \Phi(u) - 1$$
$$\Phi(-u) = 1 - \Phi(u)$$

u	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Tabelle 2: Quantile der Standardnormalverteilung

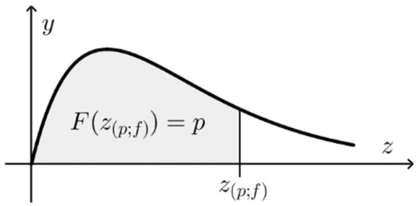


p : vorgegebene Wahrscheinlichkeit

u_p : zur Wahrscheinlichkeit p gehöriges Quantil

p	u_p	p	u_p
0.90	1.282	0.10	–1.282
0.95	1.645	0.05	–1.645
0.975	1.960	0.025	–1.960
0.99	2.326	0.01	–2.326
0.995	2.576	0.005	–2.576
0.999	3.090	0.001	–3.090

Tabelle 3: Quantile der Chi-Quadrat-Verteilung

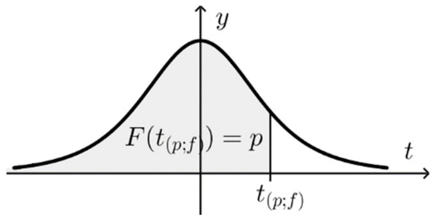


p : vorgegebene Wahrscheinlichkeit

$z_{(p;f)}$: zur Wahrscheinlichkeit p gehöriges Quantil bei f Freiheitsgraden

	p									
f	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995
1	0.000	0.000	0.001	0.004	0.016	2.71	3.84	5.02	6.63	7.88
2	0.010	0.020	0.051	0.103	0.211	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
22	8.6	9.5	11.0	12.3	14.0	30.8	33.9	36.8	40.3	42.8
24	9.9	10.9	12.4	13.8	15.7	33.2	36.4	39.4	43.0	45.6
26	11.2	12.2	13.8	15.4	17.3	35.6	38.9	41.9	45.6	48.3
28	12.5	13.6	15.3	16.9	18.9	37.9	41.3	44.5	48.3	51.0
30	13.8	15.0	16.8	18.5	20.6	40.3	43.8	47.0	50.9	53.7
40	20.7	22.2	24.4	26.5	29.1	51.8	55.8	59.3	63.7	66.8
50	28.0	29.7	32.4	34.8	37.7	63.2	67.5	71.4	76.2	79.5
60	35.5	37.5	40.5	43.2	46.5	74.4	79.1	83.3	88.4	92.0
70	43.3	45.4	48.8	51.7	55.3	85.5	90.5	95.0	100.4	104.2
80	51.2	53.5	57.2	60.4	64.3	96.6	101.9	106.6	112.3	116.3
90	59.2	61.8	65.6	69.1	73.3	107.6	113.1	118.1	124.1	128.3
100	67.3	70.1	74.2	77.9	82.4	118.5	124.3	129.6	135.8	140.2

Tabelle 4: Quantile der t-Verteilung von «Student»



p : vorgegebene Wahrscheinlichkeit

$t_{(p;f)}$: zur Wahrscheinlichkeit p gehöriges Quantil bei f Freiheitsgraden

	p				
f	0.9	0.95	0.975	0.99	0.995
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
22	1.321	1.717	2.074	2.508	2.819
24	1.318	1.711	2.064	2.492	2.797
26	1.315	1.706	2.056	2.479	2.779
28	1.313	1.701	2.048	2.467	2.763
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
60	1.296	1.671	2.000	2.390	2.660
100	1.290	1.660	1.984	2.364	2.626
200	1.286	1.653	1.972	2.345	2.601
500	1.283	1.648	1.965	2.334	2.586
⋮	⋮	⋮	⋮	⋮	⋮
∞	1.282	1.645	1.960	2.326	2.576

$$t_{(1-p;f)} = -t_{(p;f)}$$