Probability Theory:

$P(\text{something}) = \dfrac{\text{positiv outcome}}{\text{all outcome}}$

Ω ... Set of all possible outcomes (sample space) (Unterscheidung zwischen diskrete und continous sample space)
abzählbar ↓ unendlich ↓
$\omega \in \Omega$ ... elementary event

$A \subseteq \Omega$ ... event
- $\emptyset$ ... impossible event
- Ω ... sure event
- $A^c = \Omega \setminus A$ ... opposite event
- $A \cap B = \emptyset$ ... A and B are disjoint (mutually exclusive)

$(\Omega, \mathcal{F}, P)$ ... probability space

Rules of Probability:

1. $P(A^c) = 1 - P(A)$
2. $P(B \cup C) = P(B) + P(C) - P(B \cap C)$
3. $A \subseteq B \Rightarrow P(A) \leq P(B)$
4. $A \cap B = \emptyset \Rightarrow P(A \cap B) = \emptyset$ (dijoint)

Set properties:

$A \cap B = \{x : x \in A \text{ and } x \in B\}$ ... intersection
$A \cup B = \{x : x \in A \text{ or } x \in B\}$ ... union
$A \triangle B = \{x : x \in A \text{ and } x \notin B\}$ ...symmetric difference
$|A|$ ... number of elements in A

$A \cap \emptyset = \emptyset$, $A \cup \emptyset = A$
$A \cup B = B \cup A$ $A \cap B = B \cap A$
$(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$
$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$
Complement Laws: $A \cup A^c = S$, $A \cap A^c = \emptyset$, $(A^c)^c = A$, $\emptyset^c = S$, $S^c = \emptyset$
De Morgan's laws: $(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$

Inclusion-Exclusion Principle: $|A \cup B| = |A| + |B| - |A \cap B|$

Cartesian prodct: $A \times B = \{(a,b) : a \in A \text{ and } b \in B\}$

Counting:

Permutation:
Bei Permutation ist die Reihenfolge wichtig.

Permutation von k vielen Elementen = $k!$

Combinations:
Bei Combinations ist die Reihenfolge nicht wichtig.

Combinations von k vielen Elementen von n vielen Elementen $= \binom{n}{k} = \dfrac{n!}{k! \cdot (n-k)!}$

Production Rule: $n \cdot m$
n ways to perform Action 1
m ways to perfor Action 2
$n \cdot m$ ways to perform combined Actions

Probability:

Probability mass Function (pmF):
Wahrscheinlichkeits gewichtungs Funktion

$P : \mathcal{F} \to [0,1]$

gibt die Wahrscheinlichkeit eines bestimmten Ereignis X = Ereigniss wieder.

Axiome:

- $P(\Omega) = 1$
- $P(A) \geq 0 \ \forall A \in \mathcal{F}$
- $P(A \cup B) = P(A) + P(B)$ wenn independent

Properties of P:

- $P(A^c) = 1 - P(A)$, special $P(\emptyset) = 0$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Probability table:

If $\Omega = \{\omega_1, \omega_2, \ldots\}$ is countable and p_i the probability $p_i = P(\omega_i)$

then you have a probability table:

elementary event ω_i	ω_1	ω_2	.....	ω_n
probability p_i	p_1	p_2	...	p_n

The table has following propertie:
- $0 \leq p_i \leq 1$
- $\sum_i p_i = 1$

Probability of an event A:

$P(A) = \sum_{\omega_i \in A} p_i$

Conditional probability:

The conditional probability is the probability of A given, that an event B is happening.

$P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ with $P(B) > 0$

A given B

If A and B are disjoint, then $P(A|B) = P(B|A) = 0$ holts.

Usefull transformation:

- $P(A \cap B) = P(A|B) \cdot P(B)$ ... $P(B) > 0$
- $P(A \cap B) = P(B|A) \cdot P(A)$ ... $P(A) > 0$
- $P(A|B) = \dfrac{P(B|A) \cdot P(A)}{P(B)}$
- $P(B|A) = \dfrac{P(A|B) \cdot P(B)}{P(A)}$

$P(A) = P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)$

Independence:

If $P(A|B) = P(A)$ then A and B are independent

If A and B are independent. Then following holds:

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

Bayes theorem for two or more events:

$P(C_i|A) = \dfrac{P(A \cap C_i)}{P(A)} = \dfrac{P(A|C_i) \cdot P(C_i)}{\sum_{j=1} P(A|C_j) \cdot P(C_j)}$

Cumulative distribution Function (cdF):
The cdf of a random variable X is the Integral of pdf

$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)$

Properties:

- $0 \leq F(x) \leq 1$ for all $x \in \mathbb{R}$
- F is monotonically increasing ie from $x \leq y$ Follows $F(x) \leq F(y)$
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$
- F is right continuous, i.e. $\lim_{h \searrow 0} F(x+h) = F(x)$ for all $x \in \mathbb{R}$

This means: $P(a \leq X \leq b) = P(a < X \leq b) = F(b) - F(a)$
since $P(X = a) = \int_a^a f(x)\, dx = 0$

Random variables and distributions:

Random variables:
Random variables assign numbers to each outcome (probability) in the sample space.

$X : \Omega \to \mathbb{R}$ with $X(\omega) = x$ $\forall \omega \in \Omega$

$x \in \mathbb{R}$ is a realization of X

Two types of random variables:

- Discrete
- Continuous

Quantile function:

The quantile function is the Invers of the cdF ($F(p)^{-1}$). It is defined as

$F^{-1}(p) := \inf\{x | F(x) \geq p\}$ for $p \in (0,1)$

$F^{-1}(p)$ gives you the x where holds $F(x) = p$

Types of quantiles:

25% quantile ($x_{0.25}$) = lower quantile

50% quantile ($x_{0.5}$) = median

75% quantile ($x_{0.75}$) = upper quantile

Probability mass Funktion (pmF):
The pmf of a discrete random variable X is the function:

$p(a) = P(X = a)$

Expected value:

The Expected value is the most likely random Variable X.

Calculated for discrete random Variable X:

$\mathbb{E}(X) = \sum_{i=1} x_i \cdot p_i(x_i)$

Calculated for expected random Variable X:

$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$

Properties of $\mathbb{E}(X)$:

- $\mathbb{E}(aX + b) = a \cdot \mathbb{E}(X) + b$
- $\mathbb{E}(aX + bY) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y)$
- $\mathbb{E}(h(X)) = \sum_{i=1} h(x_i) \cdot p(x_i)$ for diskrete X
- $\mathbb{E}(h(X)) = \int_{-\infty}^{\infty} h(x) \cdot f(x)\, dx$ for contious X

- $\mathbb{E}(X^k)$ ... moment order k of X
- $\mu_k = \mathbb{E}((X - E(X))^k)$ ... central moment order k of X
- $\mu_2 = \mathbb{E}((X - E(X))^2) = Var(X)$ ... variance of X
- $\sigma = \sqrt{Var(X)}$ ... standard deviation of X

Probability density function (pdf):
The pdf for a continous variable X is defined as:

$P(c \leq X \leq d) = \int_c^d f(x)\, dx$ $f(x)$... pdf

Properties:
- f is nonnegative $f(x) \geq 0$
- The area und f is one: $\int_{-\infty}^{\infty} f(x) dx = 1$

Properties of Var(X):

- $Var(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - (E(X))^2$
- $Var(aX + b) = a^2 \cdot Var(X)$
- $Var(X + b) = Var(X)$
- Let X and Y be independent:

$Var(X + Y) = Var(X) + Var(Y)$

Transformations:

Transformation is used if a random Variable X and a distribution of a transformation ($Y = g(X)$) is know. And you want the cdF of Y

Steps to do:
- Determine F_X
- Determine F_Y for $Y = g(X)$
- Find $f_Y(y) = F'_Y(y)$

Normal (Gaussian) distribution: Is the most important one

$X \sim N(\mu, \sigma^2)$ $\mathbb{E}(X) = \mu$

pdf: $f(x) = \dfrac{1}{\sqrt{2\pi \sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $Var(X) = \sigma^2$

cdF is hard to do. There for standardisation is used

In R:

dnorm(x, mean, sd) $= P(X = x)$ ... is the pdF
pnorm(x, mean, sd) $= P(X \leq x)$ ... is the cdF
qnorm(p, mean, sd) $= F^{-1}(p)$ ... is the quantile Function
rnorm(n, mean, sd) ... creates n numbers with this N distribution

Standard Normal distribution:

$Z \sim N(0,1)$ with $Z = \dfrac{X - \mu}{\sigma}$

cdF: $\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} f(t) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt$

And $\Phi(-z) = 1 - \Phi(z)$

Common Families of Distributions:

- Discrete distributions
  - Bernoulli
  - Binomial
  - Geometric
  - Poisson

- Continuous distributions
  - Uniform
  - Exponential
  - Normal (Gaussian)
  - X^2 - distribution
  - t - distribution

Bernoulli distribution:

$X \sim ber(p)$

Models 1 trial that can either be success ($P(p)$) or Failure ($P(1-p)$)

X takes 1 for success takes 0 for failure

$P(X=1) = p$ $P(X=0) = 1-p$

	x	0	1
pmF	p(x)	1-p	p
cdF	F(x)	1-p	1

Binomial distribution: $X \sim B(n,p)$ or $X \sim b(n,p)$

The binomial distribution models the number of sucesses in n independent ber(p) trials.

in R: dbinom(x, n, p)

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

$\mathbb{E}(X) = n \cdot p$ $Var(X) = np \cdot (1-p)$

Geometric distribution:
The geometric distribution models the number of tries before a success.

X is the number of tries before success.

$p(x) = P(X = x) = (1-p)^x \cdot p$

$\mathbb{E}(X) = \dfrac{1-p}{p}$ $Var(x) = \dfrac{1-p}{p^2}$

in R: dgeom(x, p)

Poisson distribution:

$X \sim P(\lambda)$ or $X \sim Poi(\lambda)$

The poisson distribution models the probablity that an event happens X times

λ... intensity parameter
X... How often something should happen

$p(x) = P(X = x) = \dfrac{\lambda^x}{x!} \cdot e^{-\lambda}$

Exponential Distribution:
X has exponential distribution with $\lambda > 0$ if following holds:

$X \sim exp(\lambda)$

pdF: $f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

cdF: $F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

$\mathbb{E}(X) = \tau = \frac{1}{\lambda}$ $Var(X) = \tau^2 = \frac{1}{\lambda^2}$

Relation between Binomial and Poison:
If n is large ($n \geq 50$) and p is small $p \leq \frac{1}{10}$. Then you can approx. Binomial Distribtion with posion distribution

$X \sim B(n, \frac{\lambda}{n})$

Uniform distribution:

X is a random variable with uniform distribution over the interval (a,b) if following holds:

$X \sim U(a,b)$ $a < b,\ a, b \in \mathbb{R}$ ∧ pdF: $f(x) = \begin{cases} \frac{1}{b-a} & x \in (a,b) \\ 0 & \text{otherwise} \end{cases}$ ∧ cdF: $F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$ ∧ $\mathbb{E}(X) = \frac{a+b}{2}$ $Var(X) = \frac{(b-a)^2}{12}$

pbinom()	dbinom()	qbinom()	rbinom()
pexp()	dexp()	qexp()	rexp()
pnorm()	dnorm()	qnorm()	rnorm()
punif()	dunif()	qunif()	runif()
ppois()	dpois()	qpois()	rpois()
pgeom()	dgeom()	qgeom()	rgeom()

The 68-95-99.7 Rule:

$P(\mu - \sigma \leq X \leq \mu + \sigma) = 68\%$
$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 95\%$
$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 99.7\%$

Sample mean:

Let $X_1, X_2, X_3, \ldots, X_n$ be independent identically distributed random variables.

Then the sample mean is:

$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} (X_i) \sim N(\mu, \frac{\sigma^2}{n})$

and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

Central limit theorem and Law of large number:

Law of Large Numbers:

Let $X_1, \ldots, X_n$ independent and identically distributed random variables with:

$\mathbb{E}(X_1) = \mu$ and Finite $Var(X_1) = \sigma^2 < \infty$

$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

$X_1, \ldots, X_n$ are not necessarily normally distributed

Then the Law of large Numbers say, that: as n grows, the probability of $\bar{X}_n$ being near μ tends to 1.

Central Limit Theorem:

The central limit theorem states that the sum (also mean) of many independent copies of a random variable is approximately a normal random variable

Let $X_1, X_2, \ldots$ be a sequence of i.i.d with expectation μ and standard deviation σ

For each n:
- The sum $S_n = X_1 + X_2 + \ldots + X_n$
- The sample mean $\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$

The CLT says:

$S_n \approx N(n \cdot \mu, n \cdot \sigma^2)$

$\bar{X}_n \approx N(\mu, \frac{\sigma^2}{n})$

And standardized:

$\dfrac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} \approx N(0,1)$ $\dfrac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \approx N(0,1)$

Covariance and Correlation:

Covariance:

Covariance of two random variables are given by:

$Cov(X,Y) = \mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y)))$

Variance is a special case of covariance.

$Cov(X,X) = Var(X)$

Properties:

- $Cov(X,Y) = \mathbb{E}(X\cdot Y) - \mathbb{E}(X)\cdot\mathbb{E}(Y)$
- $Cov(aX+b, cY+d) = a\cdot c\cdot Cov(X,Y)$
- $Cov(X_1+X_2, Y) = Cov(X_1,Y) + Cov(X_2,Y)$
- $Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$
- if X and Y are independent:
  - $Var(X+Y) = Var(X) + Var(Y)$
  - $Cov(X,Y) = 0$ (return is not necessary true)

Correlation (coefficient):

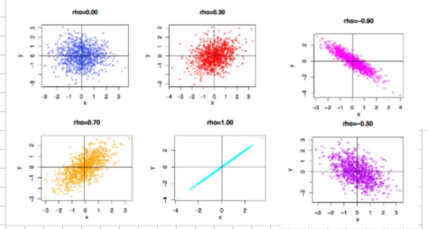
The correlation coefficient between X and Y is defined by:

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\cdot\sqrt{Var(Y)}}$$

Properties: • $\rho(X,Y)$ is the covariance of the standardized versions of X and Y

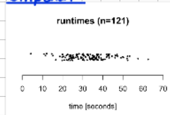
$$\rho(X,Y) = Cov\left(\frac{X-\mathbb{E}(X)}{\sqrt{Var(X)}}, \frac{Y-\mathbb{E}(Y)}{\sqrt{Var(Y)}}\right)$$

• $-1 \le \rho(X,Y) \le 1$ $\begin{cases} \rho(X,Y) = 1 \Leftrightarrow Y = aX+b \text{ with } a>0 \\ \rho(X,Y) = -1 \Leftrightarrow Y = aX+b \text{ with } a<0 \end{cases}$



Descriptiv Statistics:

Stripchart:



Is in R used as:
`x <- c(..., ..., ...)`
`stripchart(x, method = "jitter")`

We can see:
- Many data lies close to 30
- Minimum is about 5
- Maximum is about 65

Histogram:

A Histogramm shows the Frequency of a x_i or an intervall $[a,b]$

In R: `hist(x, las = 1)`

Norming: Is used when you have to different Datas with differen n.
It norms it, so that the total area is 1.

So $\sum H_i \cdot \Delta = 1$

In r with `(..., prob=True)`

Mean and empirical standard deviation:

If Data $x_1, x_2, \dots, x_n$

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

$\mathbb{E}(X) = \sum x\cdot\mathbb{P}(X=x) \quad Var = \mathbb{E}[(X-\mathbb{E}(X))^2]$

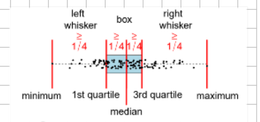
$\sigma_x = \sqrt{Var(x)}$

In R: $\bar{x}$ = mean(x) $\quad s^2$ = var(x) $\quad \sigma$ = sd(x)

About $\frac{2}{3}$ of data lies in the s-neighborhood of $\bar{x}$

Boxplot:

You can read:
- Median
- Minimum
- Maximum
- 1st quatile
- 3st quatile



In R: `boxplot(x, horizontal = TRUE, range = 0)`

One-sample t-test:

The one sample t-test is used, if we don't know the standard deviation σ of the H_0.

Instead of σ we can use $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$. So we calculate $T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t(n-1)$

$t(n-1)$ is the t-distribution with n = number of elements in $\bar{X}$ (degrees of Freedom)

$t(n)$: $\Leftrightarrow$ X has density $f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\cdot\Gamma(n/2)}\cdot\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}$

Standart error of the mean:

The SEM says how much the mean of the sample can deviate from the actual mean.

$$SEM := \frac{S}{\sqrt{n}}$$

Confidence interval:

Is the probability $1-\alpha$. So the negative of the significance level. It says determines an intervall where $\bar{x}$ will be with a probability $1-\alpha$.

$$I := (\bar{X} - q_{1-\alpha/2}\cdot SEM,\ \bar{X} + q_{1-\alpha/2}\cdot SEM)$$

Hypothesis testing:

With hypothesis testing you can test if a hypothesis is true.

You start with a Hypothesis, that μ_0 is a certain value. And its alternativ Hypothesis:

$H_0: \mu_0 = \dots \quad H_1: \mu_0 \neq 0 \Rightarrow$ double sided Test
$H_0: \mu_0 = \dots \quad H_1: \mu_0 > 0 \Rightarrow$ right sided Test
$H_0: \mu_0 = \dots \quad H_1: \mu_0 < 0 \Rightarrow$ left sided Test

And then you have a sample $\bar{x}$ with $\mu = \bar{x} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Then you have to calculate the standardized $\bar{x}$:

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

In order to accept/reject a hypothesis, you have to calculate the p-value or the rejection Range. Both depend on the type of test:

Two-sided Test $\Rightarrow \begin{Bmatrix} R(-\infty, q_{\alpha/2}] \cup [q_{1-\frac{\alpha}{2}}, \infty) \\ p = \mathbb{P}_{H_0}(|Z| \ge |z|) \end{Bmatrix}$

Left sided Test $\Rightarrow \begin{Bmatrix} R(-\infty, q_\alpha] \\ p = \mathbb{P}_{H_0}(Z \le z) \end{Bmatrix}$

Right-sided Test $\Rightarrow \begin{Bmatrix} R[q_{1-\alpha}, \infty) \\ p = \mathbb{P}_{H_0}(Z \ge z) \end{Bmatrix}$

You reject if:

$p \le \alpha \quad \vee \quad z \in R$

Error and Testpower:

Null hypothesis	rejected	not rejected
holds true	α (Type 1 error)	$1-\alpha$
dosnt holds true	testpower $= 1-\beta$	β (Type 2 error)

Type 1 error: Is the probability, that a hypothesis is getting rejected even though its true.

Type 2 error: Is the probability, that a the hypothesis H_0 is false and H_1 is true. But H_0 is wrongfully kept.

Testpower: The probability, that H_0 and H_1 are correctly differentiated.

Surrounding the two-sample t-test:

The two-sample t-test is used to compare two samples. And if they are different or not (then $H_0: \mu_0 = \mu_1$ and $H_1: \mu_0 \neq \mu_1$)

$$t = \frac{\bar{y}-\bar{x}}{\sqrt{sem_y^2 + sem_x^2}} \sim t(v)$$

this is given in the test

... measures the discrepancy of the mean in relation of the SEM

If the Hypothesis H_0 is, that $\bar{y}$ and $\bar{x}$ are distanced with the distance $d = d_0$ then following holds

$$t = \frac{(\bar{y}-\bar{x}) - d_0}{\sqrt{sem_y^2 + sem_x^2}} \quad \text{with } d = |\mu_y - \mu_x|$$

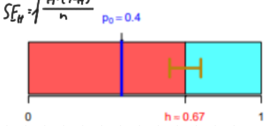
The Rest is the same as in one-sample t-test

Proportions:

With that you can test a hypothesis of a Bernulli distribution. Example: H_0: 33% say yes. And actual Frequency: And the question to answer: is the observed proportion h is for away from the H_0

$h := \frac{1}{n}\sum_{i=1}^{n} y_i$ with $y_i = \begin{cases} 1 & \text{if the i-te observation is positiv} \\ 0 & \text{else} \end{cases}$ and $\mathbb{E}(H) = p$ and $Var(H) = \frac{p(1-p)}{n}$ and $\underbrace{\frac{H-p}{\sqrt{p\cdot(1-p)/n}}}_{z} \sim N(0,1)$ and $SE_H = \sqrt{\frac{H\cdot(1-H)}{n}}$

$z = \frac{h-p_0}{se_h}$



And for 2 samples:

Under: $H_0: p_2 - p_1 = 0$

$$Z := \frac{(H_2-H_1) - 0}{\sqrt{SE_{H_2}^2 + SE_{H_1}^2}} \sim N(0,1) \quad \text{with } se_{h_1} = \sqrt{\frac{h_1(1-h_1)}{n_1}} \text{ and } se_{h_2} = \sqrt{\frac{h_2\cdot(1-h_2)}{n_2}}$$

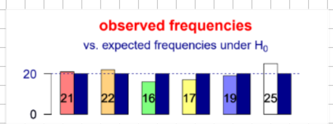
The χ^2-tests:

The χ^2 test is used if you want to test if a group of frequenzies is like an expected group of frequenzies. (Expected group = H_0)
This is called goodness of fit.

And you can use χ^2 test if you want to test if 2 different groups of frequenzies are alike. (not same n but proportion of frequenzies)
This is called for independence.

Goodness of Fit:

If you have 5 frequencies colors with n = 120.
Then the outcome is red, blue, blue, white ...
And the resulting frequencies is:



And for example H_0 = every color has same frequency $\Rightarrow n = 120 \Rightarrow$ every color should have 20
Then you have to make this tabel:

k	1	2	3	4	5	6	Σ
x_k	21	22	16	17	19	25	120
$\mathbb{E}[X_k]$	20	20	20	20	20	20	120

χ^2 is based on the binomial distribution. But transformed to multinomial distribution:

$$\mathbb{P}(\mathfrak{X} = (x_1,\dots,x_d)^t) = \binom{n}{x_1, x_2, \dots, x_d}\prod_{k=1}^{d} p_k^{x_k}$$

with

$$\binom{n}{x_1, x_2, \dots, x_d} := \frac{n!}{x_1!\cdots x_d!} = \binom{n}{x_1}\binom{n-x_1}{x_2}\binom{n-x_1-x_2}{x_3}\cdots\binom{n-x_1-\cdots-x_{d-1}}{x_d}$$

And $\chi^2 := \sum_{k=1}^{d}\frac{(x_k - \mathbb{E}_{H_0}[X_k])^2}{\mathbb{E}_{H_0}[X_k]}$

with $\mathbb{E}_{H_0}[X_k] = n\cdot p_{0,k}$

This χ^2 measures the discrepany between H_0 and real values.

And $\chi^2 \overset{d}{\sim} \chi^2(d-1)$
with d = number of frequenzies
and $\mathbb{E}[X] = d$ and $Var(X) = 2d$

In R: `rchisq`, `pchisq()` etc...

Be carefull!: Only positive values ⇒ R is always $= [q_{1-\alpha}, \infty)$

For independence:

Example: Colors on two differen simulations:

k	1	2	3	4	$x_{j,\cdot}$
simulation1	25	22	19	14	80
simulation2	27	19	44	60	150
$x_{\cdot,k}$	52	41	63	74	n=230

with $d = d_1\cdot d_2 = 4\cdot 2 = 8$
And at of this you get

$\mathbb{P}_{H_0}$	1	2	3	4	$p_{j,\cdot}$
simulation1	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	$p_{1,4}$	p_1
simulation2	$p_{2,1}$	$p_{2,2}$	$p_{2,3}$	$p_{2,4}$	p_2
$p_{\cdot,k}$	$p_{\cdot,1}$	$p_{\cdot,2}$	$p_{\cdot,3}$	$p_{\cdot,4}$	$\Sigma = 1$

$p_{j,k} = \mathbb{P}(\text{row } j \text{ and column } k) = \mathbb{P}(\text{row } j)\cdot\mathbb{P}(\text{column}) = p_{j,\cdot}\cdot p_{\cdot,k}$

And:

$\mathbb{E}_{H_0}(\cdot)$	1	2	3	4
simulation1	$n\cdot p_{1,\cdot}\cdot p_{\cdot,1}$	$n\cdot p_{1,\cdot}\cdot p_{\cdot,2}$	$n\cdot p_{1,\cdot}\cdot p_{\cdot,3}$	$n\cdot p_{1,\cdot}\cdot p_{\cdot,4}$
simulation2	$n\cdot p_{2,\cdot}\cdot p_{\cdot,1}$	$n\cdot p_{2,\cdot}\cdot p_{\cdot,2}$	$n\cdot p_{2,\cdot}\cdot p_{\cdot,3}$	$n\cdot p_{2,\cdot}\cdot p_{\cdot,4}$

$$\chi^2 = \sum_{j,k}\frac{\left(x_{j,k} - \frac{x_{j,\cdot}\cdot x_{\cdot,k}}{n}\right)^2}{\frac{x_{j,\cdot}\cdot x_{\cdot,k}}{n}} \approx 19{,}3$$

And $\chi^2 \overset{d}{\sim} \chi^2((d_1-1)\cdot(d_2-1))$

Linear regression:

ρ is the correlation of y and x

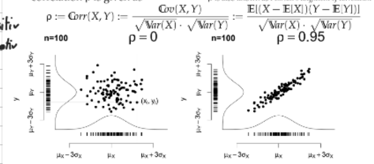
if ρ = positiv $\Rightarrow (X-\mathbb{E}[X])\cdot[Y-\mathbb{E}[Y]]$ positiv
and if ρ = negativ $\Rightarrow (X-\mathbb{E}[X])\cdot[Y-\mathbb{E}[Y]]$ negativ

$\rho = 0 \Leftrightarrow$ no correlation
$\rho > 0 \Leftrightarrow$ positiv linear relation
$\rho < 0 \Leftrightarrow$ negativ linear relation
$|\rho| = 1 \Leftrightarrow$ perfect linear relation

- $X_1, \dots, X_n$ i.i.d. RVs, $X_1 \sim N(\mu_x, \sigma_x^2)$. Here $n = 100$
- $Y_1, \dots, Y_n$ i.i.d. RVs, $Y_1 \sim N(\mu_y, \sigma_y^2)$
- also let the pairs $(X_i, Y_i)_{i=1,\dots,n}$ be independent over $i = 1, 2, \dots$
- So long, nothing said about the *relation* between X_i and Y_i
- This is accomplished (e.g.,) through the notion of *correlation*
- Definition: For the RVs X and Y (with $\mathbb{V}ar(X), \mathbb{V}ar(Y) \in (0,\infty)$) their correlation ρ is given as

ρ is also known as *Pearson's coefficient of correlation*

$$\rho := \mathbb{C}orr(X,Y) := \frac{\mathbb{C}ov(X,Y)}{\sqrt{\mathbb{V}ar(X)}\cdot\sqrt{\mathbb{V}ar(Y)}} := \frac{\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]}{\sqrt{\mathbb{V}ar(X)}\cdot\sqrt{\mathbb{V}ar(Y)}}$$



Empirical correlation:

If you have a realisation of X and Y you can estimate ρ through the empirical correlation:

$$r := \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y})}{s_x\cdot s_y}$$

Create linear relation ($y = b_0 + b_1\cdot x$):

If $|\rho| = 1$ the $y = b_0 + b_1\cdot x$ with $b_1 = \frac{s_y}{s_x}$ and $b_0 = \bar{y} - b_1\cdot\bar{x}$

If $|\rho| \neq 1$, then you can only approximate a linear regression. This linear function is the function with the least amount of error.

$$y_i = \beta_0 + \beta_1\cdot x_i + e_i \quad \text{with } e_i \text{ beeing the i-th error of the i-th point.}$$

So you have to seach β_0 and β_1 such that $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - [\beta_0 + \beta_1 x_i])^2$. And the minimal b_0 and b_1 yield the regression line $y = b_0 + b_1\cdot x$

You can calculate b_1 and b_0 like that: $b_1 = r\cdot\frac{s_y}{s_x}$ and $b_0 = \bar{y} - b_1\cdot\bar{x}$

! Important info ! The linear regression only holds in the intervall from the x_{min} to the x_{max}

In R: `lm(y ~ x)`