Concepts and Formulas

The pages has been set by support of FRVŠ (project number 1645/2007).

Combinatorics:

• **faktorials** $0! = 1,\ n! = n \cdot (n-1) \cdot \dots \cdot 3 \cdot 2 \cdot 1$ • **binomial coefficients** $\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$

• **combination** - un-ordsered k-tuple (sets) of n elements - without repetition $C_k(n) = \binom{n}{k}$

- with repetition $C'_k(n) = \binom{n+k-1}{k}$

• **permutation** - ordered k-tuple of n elements - without repetition $V_k(n) = \frac{n!}{(n-k)!}$

- with repetition $V'_k(n) = n^k$

• **permutation** - ordered k-tuple of k elements without repetition $P(k) = k!$

- ordered k-tuple of k elements, where $k_1, k_2, \dots, k_r$ elements are not distinct, $k_1 + k_2 + \dots + k_r = k$

$$P(k_1, k_2, \dots, k_r) = \frac{(k_1 + k_2 + \dots + k_r)!}{k_1! \cdot k_2! \cdot \dots \cdot k_r!}$$

$$V_k(n) = C_k(n) \cdot P(k)$$

Probability:

• **random trial** .... trial (process, activity...), in which an occurrance is (= we are not able anticipate uniquely determined result)

possible results are collected in set Ω (úplný a bezesporný systém)

• **random event** .... subset of Ω (corresponding to statement V about result of random trial), $A \subset \Omega$ $A = [V]$

• **probability** .... assign to a random event A a number $P(A)$ besides 0 and 1, expressing how much or how less we expect the occurrance of an event A

$\Omega \neq \emptyset$

$\mathcal{S}$ system of subsets Ω - $\sigma-$algebra

$P : \mathcal{S} \longrightarrow \mathbb{R}$ probability measure

properties:

$P(\emptyset) = 0, \quad P(\Omega) = 1, \quad P(\Omega \setminus A) = 1 - P(A)$

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

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

etc.

• **independence of random events**:

A, B **independent** .... $P(A \cap B) = P(A) \cdot P(B)$

A, B, C **independent** .... $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$

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

etc.

A, B distinguish ....$A \cap B = \emptyset$

• **conditional probability** (for $P(B) \neq 0$) .... $P(A/B) = \frac{P(A \cap B)}{P(B)}$

• **total probability theorem**:

$B_1, B_2, \dots B_n$ factorization $\Omega, P(B_1) \neq 0, \dots P(B_n) \neq 0 \quad \Longrightarrow$

$$\Longrightarrow \quad P(A) = P(A/B_1) \cdot P(B_1) + P(A/B_2) \cdot P(B_2) + \dots + P(A/B_n) \cdot P(B_n)$$

• **Bayes's theorem**:

$B_1, B_2, \dots B_n$ factorization Ω, $P(B_1) \neq 0, \dots P(B_n) \neq 0$, $P(A) \neq 0 \quad \Longrightarrow \quad P(B_m/A) = \dfrac{P(A/B_m) \cdot P(B_m)}{\sum_k P(A/B_k) \cdot P(B_k)}$

(factorization Ω .... evens are pairwise distinguish and $B_1 \cup B_2 \cup \dots \cup B_n = \Omega$)

Random variable:

• **random variable** .... multivalent expression of properties concerning an object of random trial
$X : (\Omega, \mathcal{S}) \longrightarrow \mathbb{R}$ Borel-measurable mapping
• **probability distribution** of random variable X .... assign $P[X \in I]$ to any Borel set $I \in \mathbb{R}$
• **(cummulative) distribution function** .... $F(a) = F_X(a) = P[X < a]$

properties: F monotnicaly increasing, continuous from left, $F(-\infty) = 0$, $F(\infty) = 1$

• **discrete distribution** (of a variable X) .... given by means of **values** x_k and **corresponding probabilities** $p_k = P[X = x_k]$, where $\sum_k p_k = 1$

$$P[X \in I] = \sum_{k;\ x_k \in I} p_k$$

expected value (mean) .... $EX = \sum_k x_k \cdot p_k$

modus $\hat{x}$ .... a number for which $P[X = \hat{x}] \geq P[X = x]$ for all x

variance (disperse) .... $DX = E((X - EX)^2) = \sum_k (x_k - EX)^2 \cdot p_k$

• **continuous distribution** (of a variable X) .... given by a **density function** f (non-negative measurable function), for which $\int_{-\infty}^{\infty} f(x)\, dx = 1$

$$P[X \in I] = \int_I f(x)\, dx$$

expected value (mean) .... $EX = \int_{-\infty}^{\infty} x f(x)\, dx$

modus $\hat{x}$ .... a number for which $f(\hat{x}) \geq f(x)$ for all x

variance (disperse) ....$DX = E((X - EX)^2) = \int_{-\infty}^{\infty} (x - EX)^2 f(x)\, dx$

(for both types of distribution)
p-**kvantil** .... $\tilde{x}_p = \inf\{x; F(x) \geq p\}$
median $\tilde{x}$....... a number for which $F(\tilde{x}^-) \leq \frac{1}{2} \leq F(\tilde{x})$

• **random vector** .... $(X, Y) : (\Omega, \mathcal{S}) \longrightarrow \mathbb{R} \times \mathbb{R}$ Borel-measurable mapping
• **probability distribution** of random vector (X, Y) .... assign $P[(X, Y) \in B]$ to aby Borel set $B \in \mathbb{R} \times \mathbb{R}$

• **(joint) distribution function** .... $F(a, b) = F_{(X,Y)}(a, b) = P[X < a, Y < b]$
• **marginal distribution function** .... $F_X(a) = P[X < a] = F(a, \infty) \quad F_Y(a) = P[Y < a] = F(\infty, a)$

• **discrete distribution** (of a vector (X, Y)) .... given by means of **values** (x_k, y_l) and **corresponding (joint) probabilities** $p_{k,l} = P[X = x_k, y = y_l]$, where $\sum_{k,l} p_{k,l} = 1$

$$P[(X, Y) \in B] = \sum_{k,l;\ (x_k, y_l) \in B} p_{k,l}$$

marginal probabilities .... $p_k = P[X = x_k] = \sum_i p_{ki} \quad q_l = P[X = y_l] = \sum_i p_{il}$
covariance .... $CX = E((X - EX)(Y - EY)) = \sum_{k,l} (x_k - EX)(y_l - EY) \cdot p_{kl}$

• **continuous distribution** (of a vector (X, Y)) .... given by a **(joint) probability density** $f = f_{(X,Y)}$ (non-measurable function), for which $\iint_{\mathbb{R} \times \mathbb{R}} f(x, y)\, dxdy = 1$

$$P[(X, Y) \in B] = \iint_B f(x, y)\, dxdy$$

marginal density function .... $f_X(a) = \int_{-\infty}^{\infty} f_{(X,Y)}(a,t)\,dt \qquad f_Y(a) = \int_{-\infty}^{\infty} f_{(X,Y)}(t,a)\,dt$

covariance .... $C(X,Y) = E((X-EX)(Y-EY)) = \int_{-\infty}^{\infty} (x-EX)(y-EY)f(x,y)\,dxdy$

vector of expected values (means) and covariance matrix .... $(EX\ EY) \qquad \begin{pmatrix} DX & C(X,Y) \\ C(X,Y) & DY \end{pmatrix}$

correlation .... $\rho = \dfrac{C(X,Y)}{\sqrt{DX \cdot DY}}$

- **independence of random variables**

  $X,\ Y$ **independet** ...... $[X \in B_1], [Y \in B_2]$ independent for all Borel sets B_1 and B_2

  $F(a,b) = F_X(a) \cdot F_Y(b)$ for all a and b

  $f(a,b) = f_X(a) \cdot f_Y(b)$ for all a and b (in case of continuous distribution)

  $p_{k,l} = p_k \cdot q_l$ for all k and l (in case of discrete distribution)

  $X,\ Y$ independent $\Longrightarrow$ $C(X,Y) = 0$

similarly for random vectors $(X_1, X_2, \ldots, X_n)$

Types of distributions:

- **alternative** $A(p)$ $\quad 0 < p < 1$

  $x_0 = 0,\ x_1 = 1 \qquad p_0 = 1 - p,\ p_1 = p$

  $EX = p \qquad DX = p(1-p)$

(wheather we draw out black ball of N balls, Z black balls and N-Z white ones, $p = \frac{Z}{N}$)

- **hypergeometrical** $HG(N,Z,n)$ $\quad n < N,\ Z < N$

  $x_k = k = 0, \ldots, n\ , \quad k \le Z,\ n-k \le N-Z \qquad p_k = \dfrac{\binom{Z}{k} \cdot \binom{N-Z}{n-k}}{\binom{N}{n}}$

  $EX = n \cdot \dfrac{Z}{N} \qquad DX = n \cdot \dfrac{Z}{N} \cdot \left(1 - \dfrac{Z}{N}\right) \cdot \dfrac{N-n}{N-1}$

(a number of black balls when we draw out n balls without replacement from N balls, Z black balls and N-Z white ones)

- **binomial** $Bi(n,p)$ $\quad 0 < p < 1$

  $x_k = k = 0, \ldots, n \qquad p_k = \dbinom{n}{k} \cdot p^k (1-p)^{n-k}$

  $EX = np \qquad DX = np(1-p)$

(a number of black balls when we draw out n balls with replacement from N balls, Z black balls and N-Z white ones, $p = \frac{Z}{N}$)
$X_1, \ldots, X_n \sim A(p)$ independent $\Longrightarrow$ $X_1 + \cdots + X_n \sim Bi(n,p)$
$X \sim Bi(m,p), Y \sim Bi(n,p)$ independent $\Longrightarrow$ $X + Y \sim Bi(m+n,p)$
(similarly multinomial distribution)

- **geometrical** $G(p)$ $\quad 0 < p < 1$

  $x_k = k = 1, 2, \ldots \qquad p_k = p \cdot (1-p)^{k-1}$

  $EX = \dfrac{1}{p} \qquad DX = \dfrac{1}{p}$

(order of the first ball when we step-by-step draw out balls wit replacement from N balls, Z black balls and N-Z white ones, $p = \frac{Z}{N}$, discrete failurerate, an order of the first impulse)

- **Poisson** $Po(\lambda)$ $\quad \lambda > 0$

  $x_k = k = 0, 1, 2, \ldots \qquad p_k = \mathrm{e}^{-\lambda} \cdot \dfrac{\lambda^k}{k!}$

  $EX = \lambda \qquad DX = \lambda$

(a number of random impulses in a given time)

$X_1, X_2, \dots \sim Bi(n, p_n)$, $n \cdot p_n \to \lambda \implies \quad X_n \to X \sim Po(\lambda)$

• **exponential** $Exp(\lambda) \quad \lambda > 0$

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

$$EX = \frac{1}{\lambda} \qquad DX = \frac{1}{\lambda^2}$$

(a time to the first random impulse)

it holds $P[X > a + b / X > a] = P[X > b]$ for $X \sim Exp(\lambda)$ (no memory)

• **gama** $\Gamma(n, \lambda) \quad \lambda > 0$

$$f(x) = \begin{cases} \mathrm{e}^{-\lambda x} \cdot \frac{\lambda^n x^{n-1}}{(n-1)!} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$EX = \frac{n}{\lambda} \qquad DX = \frac{n}{\lambda^2}$$

(a time to the n-th random impulse)

• **normal** $N(\mu, \sigma^2) \quad \sigma > 0$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$EX = \mu \qquad DX = \sigma^2$$

(limit distribution, see central limit theorem)

quantile $u(p)$; $P[X < u(p)] = p$ where $X \sim N(0; 1)$

$X \sim N(\mu, \sigma^2) \implies a \cdot X + b \sim N(a \cdot \mu + b, a^2 \cdot \sigma^2)$

$X \sim N(\mu, \sigma^2) \implies \dfrac{X - \mu}{\sigma} \sim N(0, 1)$

$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ nezávislé $\implies \quad X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$X_1, \dots, X_n$ independent $X_k \sim N(\mu_k, \sigma_k^2) \implies \sum_k a_k \cdot X_k + b \sim N(\sum_k a_k \cdot \mu_k + b, \sum_k a_k^2 \cdot \sigma_k^2)$

for distribution function Φ of distribution $N(0, 1)$ it holds $\Phi(-x) + \Phi(x) = 1$, for quantile $u(p) = -u(1 - p)$

(similarly multi-dimensional normal distribution)

• **Pearson chi-square** χ_n^2

$$f_n(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \mathrm{e}^{-\frac{n}{2}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$EX = n \qquad DX = 2n$$

quantile $\chi_n^2(p)$; $P[X < \chi_n^2(p)] = p$, kde $X \sim \chi_n^2$

$X_1, \dots, X_m \sim N(0, 1)$ independent $\implies \quad X_1^2 + \dots + X_m^2 \sim \chi_m^2$

$X \sim \chi_m^2, Y \sim \chi_n^2$ independent $\implies \quad X + Y \sim \chi_{m+n}^2$

(for $n = 2$ Raileigh, pro $n = 3$ Maxwell distribution)

• **Student** t_n

$$f_n(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(\frac{x^2}{n} + 1 \right)^{-\frac{n+1}{2}}$$

$$EX = 0 \qquad DX = \frac{n}{n-2}$$

quantile $t_n(p)$; $P[X < t_n(p)] = p$ where $X \sim t_n$

$X \sim N(0, 1), Y \sim \chi_m^2$ independent $\implies \quad \dfrac{X}{\sqrt{\frac{Y}{m}}} \sim t_m$

for distribution function F of distribution t_n it holds $F(-x) = 1 - F(x)$ for quantile $t_n(p) = -t_n(1 - p)$

• **Fisher - Snedecor** $F_{m,n}$

$$f_{m,n}(x) = \begin{cases} \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \cdot \left(\frac{m}{n} x - +1\right)^{-\frac{m+n}{2}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$EX = \frac{n}{n-2} \qquad DX = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$$

quantile value $F_{m,n}(p)$; $P[X < F_{m,n}(p)] = p$ where $X \sim F_{m,n}$

$X \sim \chi^2_m, Y \sim \chi^2_n$ independent $\Longrightarrow \quad \frac{\frac{X}{m}}{\frac{Y}{n}} \sim F_{m,n}$
for quantile $F_{m,n}(p) = \frac{1}{F_{n,m}(1-p)}$

Limit theorems:

• **convergence** $X_n \to X$ almost surely .... $P[|X_n - X| \to 0] = 1$
$X_n \to X$ in probability .... for all $\epsilon > 0 \quad P[|X_n - X| \geq \epsilon] \to 0$
$X_n \to X$ in distribution .... $F_{X_n}(x) \to F_X(x)$ in all continuity points of F_X

• **Tchebyshew inequality** $\quad P[|X - EX| \geq \epsilon] \leq \dfrac{DX}{\epsilon^2}$ for $\epsilon > 0$

• **Weak Law of Large Numbers**
$X_1, X_2, \ldots$ independent random variables, $\forall k \; EX_k = a$, $DX_k = b (< \infty) \quad \Longrightarrow$
$$\Longrightarrow \quad \frac{1}{n} \sum_{k=1}^{n} X_k \longrightarrow a \text{ in probability}$$
(Strong Law of Large Numbers concerns the convergence almost surely)

• **Bernoulli theorem**
$X_1, X_2, \ldots$ independent random variables, $\forall k \; X_k \sim A(p)$, $0 < p < 1 \quad \Longrightarrow$
$$\Longrightarrow \quad \frac{1}{n} \sum_{k=1}^{n} X_k \longrightarrow p \text{ in probability}$$

• **Central Limit Theorem, Lindeberg - Lévy**
$X_1, X_2, \ldots$ independent random variables with the same distribution, $\forall k \; EX_k = a$, $DX_k = b (< \infty) \quad \Longrightarrow$
$$\Longrightarrow \quad \frac{\frac{1}{n} \sum\limits_{k=1}^{n} X_k - a}{\sqrt{\frac{b}{n}}} \longrightarrow Y \sim N(0,1) \text{ v distribuci}$$

• **Moivre - Laplace theorem**
$X_1, X_2, \ldots$ independent random variables, $\forall k \; X_k \sim A(p)$, $0 < p < 1 \quad \Longrightarrow$
$$\Longrightarrow \quad \frac{\frac{1}{n} \sum\limits_{k=1}^{n} X_k - p}{\sqrt{\frac{p(1-p)}{n}}} \longrightarrow Y \sim N(0,1) \text{ in distribution}$$

Basic statistical tests:

• **one-sample test** $(X_k = a + \epsilon_k)$
$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ independent

• **estimator** pro μ $\quad \bar{X} = \dfrac{1}{n}(X_1 + \cdots + X_n) \qquad \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

• **estimator** pro σ^2 $\quad S_X^2 = \dfrac{1}{n-1}\left((X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\right)$

$$U = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1) \qquad T = \frac{\bar{X} - \mu}{\sqrt{\frac{S_X^2}{n}}} \sim t_{n-1} \qquad \approx N(0,1) \qquad V = \frac{S_X^2}{\frac{\sigma^2}{n-1}} \sim \chi^2_{n-1}$$

$\sum_k (X_k - \bar{X})^2 = \sum_k (X_k - \mu)^2 - n \cdot (\bar{X} - \mu)^2$

- two-tailed test for μ on level α

H_0: $\mu = a$ rejected $\iff$ $|U| \geq u(1 - \frac{\alpha}{2}) \quad \alpha \geq 2(1 - \Phi(|U|))$ for known σ
$|T| \geq t_{n-1}(1 - \frac{\alpha}{2}) \quad \alpha \geq 2(1 - F(|T|))$ for unknown σ and n small
$|T| \geq u(1 - \frac{\alpha}{2}) \quad \alpha \geq 2(1 - \Phi(|T|))$ for unknown σ and n large

- two-tailed confidence interval on level $1 - \alpha$

$\bar{X} \pm \frac{\sigma}{\sqrt{n}} u(1 - \frac{\alpha}{2}) \qquad \bar{X} \pm \frac{S_X}{\sqrt{n}} t_{n-1}(1 - \frac{\alpha}{2}) \qquad \bar{X} \pm \frac{S_X}{\sqrt{n}} u(1 - \frac{\alpha}{2})$

- one-tailed test for μ on level α

H_0: $\mu \geq a$ rejected $\iff$ $U \leq u(\alpha) \quad -U \geq u(1 - \alpha) \quad \alpha \geq \Phi(U) = 1 - \Phi(-U)$ for known σ
$T \leq t_{n-1}(\alpha) \quad -T \geq t_{n-1}(1 - \alpha) \quad \alpha \geq F(T) = 1 - F(-T)$

for unknown σ and n small

$T \le u(\alpha) \quad -T \ge u(1-\alpha) \qquad \alpha \ge \Phi(T) = 1 - \Phi(-T)$

for unknown σ and n large

- one-tailed confidence interval on level $1-\alpha$

$(\bar{X} - \frac{\sigma}{\sqrt{n}}u(1-\alpha); \infty) \qquad (\bar{X} - \frac{S_X}{\sqrt{n}}t_{n-1}(1-\alpha); \infty) \qquad (\bar{X} - \frac{S_X}{\sqrt{n}}u(1-\alpha); \infty)$

H_0: $\mu \le a$ rejected $\iff$ $U \ge u(1-\alpha)$ $\quad \alpha \ge= 1 - \Phi(U)$ for known σ

$T \ge t_{n-1}(1-\alpha)$ $\quad \alpha \ge 1 - F(T)$ for unknown σ and n small

$T \ge u(1-\alpha)$ $\quad \alpha \ge 1 - \Phi(T)$ for unknown σ and n large

- one-tailed confidence interval on level $1-\alpha$

$(-\infty; \bar{X} + \frac{\sigma}{\sqrt{n}}u(1-\alpha)) \qquad (-\infty; \bar{X} + \frac{S_X}{\sqrt{n}}t_{n-1}(1-\alpha)) \qquad (-\infty; \bar{X} + \frac{S_X}{\sqrt{n}}u(1-\alpha))$

- two-tailed test for σ on level α

H_0: $\sigma^2 = b$ rejected $\iff$ $V \le \chi^2_{n-1}(\frac{\alpha}{2})$ or $V \ge \chi^2_{n-1}(1-\frac{\alpha}{2})$ $\qquad \alpha \ge 2F(V)$ or $\alpha \ge 2(1-F(V))$

- one-tailed test for σ on level α

H_0: $\sigma^2 \ge b$ rejected $\iff$ $V \le \chi^2_{n-1}(\alpha)$ $\qquad \alpha \ge F(V)$

H_0: $\sigma^2 \le b$ rejected $\iff$ $V \ge \chi^2_{n-1}(1-\alpha)$ $\qquad \alpha \ge 1 - F(V)$

$X_1, \dots, X_n \sim A(p)$ independent

- **estimator** for p $\qquad \bar{X} = \frac{1}{n}(X_1 + \cdots + X_n) \qquad n\bar{X} \sim Bi(n,p)$

- two-tailed test for p on level α

H_0: $p = a$ rejected $\iff \bar{X} \le \dfrac{k_1}{n}$ or $\bar{X} \ge \dfrac{k_2}{n}$

where $\sum_{k=0}^{k_1} \binom{n}{k} a^k (1-a)^{n-k} \le \frac{\alpha}{2} \qquad \sum_{k=k_2}^{n} \binom{n}{k} a^k (1-a)^{n-k} \le \frac{\alpha}{2}$

$|U| \ge u(1-\frac{\alpha}{2}) \qquad U = \dfrac{\bar{X} - a}{\sqrt{\frac{a(1-a)}{n}}}$

(paired test transformed into na one-sample test)

- **two-sample test**

$X_1, \dots, X_m \sim N(\mu_1, \sigma_1^2)$ independent, $Y_1, \dots, Y_n \sim N(\mu_2, \sigma_2^2)$ independent, both samples independent

$\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \left(\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right))$

for $\sigma_1 = \sigma_2 = \sigma$

$$U = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim N(0,1) \qquad T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t_{m+n-2} \qquad \approx N(0,1)$$

$$S^2 = \frac{1}{m+n-2}\left((m-1)S_X^2 + (n-1)S_Y^2\right) \qquad Z = \frac{S_X^2}{S_Y^2} \sim F_{m-1,n-1}$$

- two-tailed tests on level α

H_0: $\mu_1 - \mu_2 = a$ rejected $\iff$ $|T| \ge t_{m+n-1}(1-\frac{\alpha}{2})$ $\qquad \alpha \ge 2(1 - F(|T|))$ for $\sigma_1 = \sigma_2$

H_0: $\sigma_1 = \sigma_2$ rejected $\iff$ $Z \le F_{m-1,n-1}(\frac{\alpha}{2})$ or $Z \ge F_{m-1,n-1}(1-\frac{\alpha}{2})$ $\qquad \alpha \ge 2F(Z)$ or $\alpha \ge 2(1-F(Z))$

$X_1, \dots, X_m \sim A(p_1)$, $Y_1, \dots, Y_m \sim A(p_2)$ independent

- two-tailed test for p on level α

H_0: $p_1 - p_2 = a$ rejected $\iff |U| \ge u(1-\frac{\alpha}{2}) \qquad U = \dfrac{\bar{X} - \bar{Y} - a}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{m} + \frac{1}{n}\right)}} \qquad \hat{p} = \dfrac{m\bar{X} + n\bar{Y}}{m+n}$

Analysis of variance (ANOVA):

- **one factor ANOVA**

$X_{1,1}, \dots, X_{1,n_1} \sim N(\mu_1, \sigma^2)$ independent $\qquad \mu_1 = a + \alpha_1$

$X_{2,1}, \dots, X_{2,n_2} \sim N(\mu_2, \sigma^2)$ independent $\qquad \mu_2 = a + \alpha_2$

....

$X_{I,1}, \dots, X_{I,n_I} \sim N(\mu_I, \sigma^2)$ independent $\qquad \mu_I = a + \alpha_I$

samples independent, $n_1 + \cdots + n_I = n$, $\alpha_1 + \cdots + \alpha_I = 0$

H_0: $\mu_1 = \mu_2 = \cdots = \mu_I$ or $\alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$
(means)

$$\bar{X}_1 = \frac{1}{n_1}\sum_k X_{1,k}, \ldots, \bar{X}_I = \frac{1}{n_I}\sum_k X_{I,k}, \qquad \bar{X} = \frac{1}{n}\sum_{k,l} X_{k,l}$$

(sums of squares)

$$\begin{array}{lll} S_T = \sum_{k,l}(X_{k,l} - \bar{X})^2 & f_T = n-1 & \\ S_A = \sum_k(\bar{X}_k - \bar{X})^2 \cdot n_k & f_A = I-1 & F_A = \dfrac{\frac{S_A}{f_A}}{\frac{S_e}{f_e}} \quad s^2 = \dfrac{S_e}{f_e} \\ S_e = S_T - S_A = \sum_{k,l}(X_{k,l} - \bar{X}_k)^2 & f_e = n - I & \end{array}$$

H_0: $\alpha_1 = \cdots = \alpha_I = 0$ rejected $\iff$ $F_A \geq F_{f_A,f_e}(1-\alpha)$

(Sheffé method) classes k, l are significantly different $\iff$ $|\bar{X}_k - \bar{X}_l| > \sqrt{\left(\frac{1}{n_k} + \frac{1}{n_l}\right)(I-1)s^2 F_{f_A,f_e}(1-\alpha)}$

(Tukey method for $n_k = p$) classes k, l are significantly different $\iff$ $|\bar{X}_k - \bar{X}_l| > \frac{1}{\sqrt{p}} \cdot s \cdot q_{I,n-I}$

$q_{k,l}$ studentized range (in tables)

- **two factor ANOVA**

$$\begin{array}{llll} X_{1,1,1},\ldots,X_{1,1,p} \sim N(\mu_{1,1},\sigma^2) \text{ ind.} & X_{1,2,1},\ldots,X_{1,2,p} \sim N(\mu_{1,2},\sigma^2) \text{ ind.} & \ldots & X_{1,J,1},\ldots,X_{1,J,p} \sim N(\mu_{1,J},\sigma^2) \text{ ind.} \\ X_{2,1,1},\ldots,X_{2,1,p} \sim N(\mu_{2,1},\sigma^2) \text{ ind.} & X_{2,2,1},\ldots,X_{2,2,p} \sim N(\mu_{2,2},\sigma^2) \text{ ind.} & \ldots & X_{2,J,1},\ldots,X_{2,J,p} \sim N(\mu_{2,J},\sigma^2) \text{ ind.} \\ \cdots & \cdots & \cdots & \cdots \\ X_{I,1,1},\ldots,X_{I,1,p} \sim N(\mu_{I,1},\sigma^2) \text{ ind.} & X_{I,2,1},\ldots,X_{I,2,p} \sim N(\mu_{I,2},\sigma^2) \text{ ind.} & \ldots & X_{I,J,1},\ldots,X_{I,J,p} \sim N(\mu_{I,J},\sigma^2) \text{ ind.} \end{array}$$

samples independent, $I \cdot J \cdot p = n$

(means)

$$\begin{array}{lllll} \bar{X}_{1,1} = \frac{1}{p}\sum_k X_{1,1,k} & \bar{X}_{1,2} = \frac{1}{p}\sum_k X_{1,2,k} & \ldots & \bar{X}_{1,J} = \frac{1}{p}\sum_k X_{1,J,k} & \bar{X}_1^A = \frac{1}{p\cdot J}\sum_{kl} X_{1,k,l} \\ \bar{X}_{2,1} = \frac{1}{p}\sum_k X_{2,1,k} & \bar{X}_{2,2} = \frac{1}{p}\sum_k X_{2,2,k} & \ldots & \bar{X}_{2,J} = \frac{1}{p}\sum_k X_{2,J,k} & \bar{X}_2^A = \frac{1}{p\cdot J}\sum_{kl} X_{2,k,l} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \bar{X}_{I,1} = \frac{1}{p}\sum_k X_{I,1,k} & \bar{X}_{I,2} = \frac{1}{p}\sum_k X_{I,2,k} & \ldots & \bar{X}_{I,J} = \frac{1}{p}\sum_k X_{I,J,k} & \bar{X}_I^A = \frac{1}{p\cdot J}\sum_{kl} X_{I,k,l} \\ \bar{X}_1^B = \frac{1}{p\cdot I}\sum_{k,l} X_{k,1,l} & \bar{X}_2^B = \frac{1}{p\cdot I}\sum_{k,l} X_{k,2,l} & \ldots & \bar{X}_J^B = \frac{1}{p\cdot I}\sum_{k,l} X_{k,J,l} & \bar{X} = \frac{1}{n}\sum_{klm} X_{k,l,m} \end{array}$$

- and interactions

$\mu_{k,l} = a + \alpha_k + \beta_l + \gamma_{k,l}$ $\quad \sum_k \alpha_k = 0$ $\quad \sum_l \beta_l = 0$ $\quad \sum_{k,l} \gamma_{k,l} = 0$
H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$
H_0: $\beta_1 = \beta_2 = \cdots = \beta_I = 0$
H_0: $\gamma_{1,1} = \gamma_{1,2} = \cdots = \gamma_{I,J} = 0$

(sums of squares)

$$\begin{array}{ll} S_T = \sum_{klm}(X_{k,l,m} - \bar{X})^2 & f_T = n-1 \\ S_A = \sum_k(\bar{X}_k^A - \bar{X})^2 \cdot J \cdot p & f_A = I-1 \\ S_B = \sum_k(\bar{X}_k^B - \bar{X})^2 \cdot I \cdot p & f_A = J-1 \\ S_e = \sum_{klm}(X_{k,l,m} - \bar{X}_{k,l})^2 & f_e = n - I \cdot J \\ S_{AB} = S_T - S_A - S_B - S_e & f_{AB} = (I-1)(J-1) \end{array} \qquad F_A = \frac{\frac{S_A}{f_A}}{\frac{S_e}{f_e}} \quad F_B = \frac{\frac{S_B}{f_B}}{\frac{S_e}{f_e}} \quad F_{AB} = \frac{\frac{S_{AB}}{f_{AB}}}{\frac{S_e}{f_e}} \quad s^2 = \frac{S_e}{f_e}$$

H_0: $\alpha_1 = \cdots = \alpha_I = 0$ rejected $\iff$ $F_A \geq F_{f_A,f_e}(1-\alpha)$
H_0: $\beta_1 = \cdots = \beta_I = 0$ rejected $\iff$ $F_B \geq F_{f_B,f_e}(1-\alpha)$
H_0: $\gamma_{1,1} = \cdots = \gamma_{I,J} = 0$ rejected $\iff$ $F_{AB} \geq F_{f_{AB},f_e}(1-\alpha)$

- without interactions

$\mu_{k,l} = a + \alpha_k + \beta_l$ $\quad \sum_k \alpha_k = 0$ $\quad \sum_l \beta_l = 0$
H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$
H_0: $\beta_1 = \beta_2 = \cdots = \beta_I = 0$

$$\begin{array}{ll} S_T = \sum_{klm}(X_{k,l,m} - \bar{X})^2 & f_T = n-1 \\ S_A = \sum_k(\bar{X}_k^A - \bar{X})^2 \cdot J \cdot p & f_A = I-1 \\ S_B = \sum_k(\bar{X}_k^B - \bar{X})^2 \cdot I \cdot p & f_A = J-1 \\ S_e = S_T - S_A - S_B & f_e = n - I - J + 1 \end{array} \qquad F_A = \frac{\frac{S_A}{f_A}}{\frac{S_e}{f_e}} \quad F_B = \frac{\frac{S_B}{f_B}}{\frac{S_e}{f_e}} \quad s^2 = \frac{S_e}{f_e}$$

H_0: $\alpha_1 = \cdots = \alpha_I = 0$ rejected $\iff$ $F_A \geq F_{f_A,f_e}(1-\alpha)$
H_0: $\beta_1 = \cdots = \beta_I = 0$ rejected $\iff$ $F_B \geq F_{f_B,f_e}(1-\alpha)$

$$\text{(Sheffé method) classes } k,\ l \text{ are significantly different} \iff |\bar{X}_k^A - \bar{X}_l^A| > \sqrt{\frac{2(I-1)}{JP} s^2 F_{f_A,f_e}(1-\alpha)}$$

$$\text{classes } k,\ l \text{ are significantly different} \iff |\bar{X}_k^B - \bar{X}_l^B| > \sqrt{\frac{2(J-1)}{IP} s^2 F_{f_B,f_e}(1-\alpha)}$$

$$\text{(Tukey method) classes } k,\ l \text{ are significantly different} \iff |\bar{X}_k^A - \bar{X}_l^A| > \frac{1}{\sqrt{Jp}} \cdot s \cdot q_{I,n-I-J+1}$$

$$\text{classes } k,\ l \text{ are significantly different} \iff |\bar{X}_k^B - \bar{X}_l^B| > \frac{1}{\sqrt{Ip}} \cdot s \cdot q_{J,n-I-J+1}$$

$q_{k,l}$ studentized range

Correlation and regression:

- **correlation coefficient**

$(X_1, Y_1), \ldots, (X_n, Y_n) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ independent, $\boldsymbol{\mu} = (\mu_1, \mu_2)$, $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$, $\rho = \dfrac{\sigma_{12}}{\sqrt{\sigma_1^2 \sigma_2^2}}$

- **estimator** pro σ_{12} $\quad S_{XY} = \frac{1}{n-1}\left((X_1 - \bar{X})(Y_1 - \bar{Y}) + \cdots + (X_n - \bar{X})(Y_n - \bar{Y})\right)$
- **estimator** pro ρ $\quad r = \dfrac{S_{XY}}{S_X \cdot S_Y}$, $T = \dfrac{r}{\sqrt{\frac{1-r^2}{n-2}}} \sim t_{n-2}$

H_0: $\rho = 0$ rejected $\iff$ $|T| \geq t_{n-2}(1 - \frac{\alpha}{2})$ $\qquad \alpha \geq 2(1 - F(|T|))$

$\boldsymbol{R_{XY}}$ correlation matrix of random vectors $\boldsymbol{X} = (X_1, \ldots, X_I)$ a $\boldsymbol{Y} = (Y_1, \ldots, Y_J)$

- **multiple correlation** X a $\boldsymbol{Y} = (Y_1, \ldots, Y_k)$

(maximal correlation X and $c_1 \cdot X_1 + \cdots + c_k \cdot X_k$, equality holds if $c_1, \ldots, c_k$ are regression coefficients)

$$r^2_{X,\boldsymbol{Y}} = \boldsymbol{R}_{X\boldsymbol{Y}} \cdot \boldsymbol{R}^{-1}_{\boldsymbol{YY}} \cdot \boldsymbol{R}_{\boldsymbol{Y},X}, \quad r^2_{X,(Y_1,Y_2)} = \frac{r^2_{XY_1} + r^2_{XY_2} - 2r_{XY_1}r_{XY_2}r_{Y_1Y_2}}{1 - r^2_{Y_1Y_2}}$$

$$\frac{n-k-1}{k} \cdot \frac{r^2_{X,\boldsymbol{Y}}}{1 - r^2_{X,\boldsymbol{Y}}} \sim F_{k,n-k-1}$$

- **partial correlation** X and Y with the effect of $\boldsymbol{Z} = (Z_1, \ldots, Z_k)$ removed

$$r_{X,Y.\boldsymbol{Z}} = \frac{r_{X,Y} - \boldsymbol{R}_{X\boldsymbol{Z}} \cdot \boldsymbol{R}^{-1}_{\boldsymbol{ZZ}} \cdot \boldsymbol{R}_{\boldsymbol{Z},Y}}{\sqrt{(1 - \boldsymbol{R}_{X\boldsymbol{Z}} \cdot \boldsymbol{R}^{-1}_{\boldsymbol{ZZ}} \cdot \boldsymbol{R}_{\boldsymbol{Z},X})(1 - \boldsymbol{R}_{Y\boldsymbol{Z}} \cdot \boldsymbol{R}^{-1}_{\boldsymbol{ZZ}} \cdot \boldsymbol{R}_{\boldsymbol{Z},Y})}}, \qquad r_{X,Y.Z} = \frac{r_{X,Y} - r_{XZ} \cdot r_{YZ}}{\sqrt{(1 - r^2_{XZ})(1 - r^2_{YZ})}}$$

$$\frac{r_{X,Y.\boldsymbol{Z}}}{\sqrt{1 - r^2_{X,Y.\boldsymbol{Z}}}} \sqrt{n-k-2} \sim t_{n-k-2}$$

- **linear regression** $(Y_k = a \cdot x_k + b + \epsilon_k)$

$Y_1 \sim N(a \cdot x_1 + b, \sigma^2), \ldots, Y_n \sim N(a \cdot x_n + b, \sigma^2)$ independent

$\bar{x} = \dfrac{1}{n}(x_1 + \cdots + x_n)$ $\qquad (n-1)s_x^2 = (x_1 - \bar{x})^2 + \cdots + (x_2 - \bar{x})^2$ $\qquad (n-1)S^2_{xY} = (x_1 - \bar{x})(Y_1 - \bar{Y}) + \cdots + (x_n - \bar{x})(Y_1 - \bar{Y})$ $\qquad r_{xY} = \dfrac{S_{xY}}{\sqrt{s_x^2 S_Y^2}}$

estimator pro a $\qquad \hat{a} = \dfrac{n \sum x_k Y_k - \sum x_k \sum Y_k}{n \sum x_k^2 - (\sum x_k)^2} \qquad \hat{a} = \dfrac{S_{xY}}{s_x^2} = r_{xY} \cdot \dfrac{S_Y}{s_x}$

$$E(\hat{a}) = a \quad D(\hat{a}) = \frac{\sigma^2}{(n-1)s_x^2}$$

estimator pro b $\qquad \hat{b} = \dfrac{\sum x_k^2 \sum Y_k - \sum x_k \sum x_k Y_k}{n \sum x_k^2 - (\sum x_k)^2} \qquad \hat{b} = \bar{Y} - \hat{a}\bar{x}$

$$E(\hat{b}) = b \quad D(\hat{b}) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{(n-1)s_x^2} \right)$$

$$\frac{\hat{a} - a}{\sqrt{\frac{s^2}{(n-1)s_x^2}}} \sim t_{n-2} \qquad \frac{\hat{b} - b}{\sqrt{s^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{(n-1)s_x^2} \right)}} \sim t_{n-2} \qquad s^2 = \frac{S_e}{n-2} \qquad s^2 = (1 - R^2) \cdot S_Y^2 \cdot \frac{n-1}{n-2}$$

$$S_T = \sum_k (Y_k - \bar{Y})^2 = (n-1)S_Y^2 \qquad S_{REG} = \sum_k (\hat{a}x_k - \hat{b} - \bar{Y})^2 \qquad S_e = \sum_k (\hat{a}x_k - \hat{b} - Y_k)^2$$

• **coefficient of determination** ...... $R^2 = \frac{S_{REG}}{S_T} = 1 - \frac{S_e}{S_T}$ $\quad R^2 = r_{xY}^2 \quad R^2 = \hat{a}^2 \cdot \frac{S_X^2}{S_Y^2}$

for x_0

$\hat{a}x_0 + \hat{b} \qquad E(\hat{a}x_0 + \hat{b}) = ax_0 + b \qquad D(\hat{a}x_0 + \hat{b}) = \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{(n-1)s_x^2} \right)$

$$\frac{\hat{a}x_0 + \hat{b} - ax_0 - b}{\sqrt{\sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{(n-1)s_x^2} \right)}} \sim t_{n-2}$$

confidence interval $EY(x_0)$ for x_0

$$ax_0 + b = \hat{a}x_0 + \hat{b} \pm s\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{(n-1)s_x^2}} \cdot t_{n-2}(1 - \frac{\alpha}{2})$$

confidence band (Working-Hotelling) $EY(x)$ simultaneous (Sheffé method)

$$ax + b = \hat{a}x + \hat{b} \pm s\sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{(n-1)s_x^2}} \cdot \sqrt{2F_{2,n-1}(1-\alpha)} \qquad \forall x$$

prediction interval $Y(x_0)$ pro x_0

$$ax_0 + b = \hat{a}x_0 + \hat{b} \pm s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{(n-1)s_x^2}} \cdot t_{n-2}(1 - \frac{\alpha}{2})$$

prediction band $EY(x)$ simultaneous

$$ax + b = \hat{a}x + \hat{b} \pm s\sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{(n-1)s_x^2}} \cdot \sqrt{2F_{2,n-1}(1-\alpha)} \qquad \forall x$$

• **case b=0** ($Y_k = a \cdot x_k + \epsilon_k$)

estimator pro a $\qquad \hat{a} = \frac{\sum x_k Y_k}{\sum x_k^2}$

$$\frac{\hat{a} - a}{\sqrt{\frac{s^2}{\sum x_k^2}}} \sim t_{n-1} \qquad s^2 = \frac{1}{n-1}(\sum Y_k^2 - \hat{a}\sum x_k^2)$$

$$\sum x_k^2 = n\bar{x}^2 + (n-1)S_x^2 \qquad \sum Y_k^2 = n\bar{Y}^2 + (n-1)S_Y^2 \qquad \sum x_k Y_k = n\bar{x}\bar{Y} + (n-1)S_{xY}$$

Other tests:

• **contingency tables**

H_0: $X,\ Y$ independent

$X \backslash Y$	1	...	s	
1	n_{11}	...	n_{1s}	$n_{1\cdot}$
$\vdots$		$\cdots$		$\vdots$
r	n_{r1}	...	n_{rs}	$n_{r\cdot}$
	$n_{\cdot 1}$	...	$n_{\cdot s}$	n

$$\chi^2 = \sum_{kl} \frac{\left(n_{kl} - \frac{n_{k\cdot} \cdot n_{\cdot l}}{n}\right)^2}{\frac{n_{k\cdot} \cdot n_{\cdot l}}{n}} = n \sum_{kl} \frac{n_{kl}^2}{n_{k\cdot} \cdot n_{\cdot l}} - n \approx \chi^2_{(r-1)(s-1)}$$

H_0 rejected $\iff \chi^2 \geq \chi^2_{(r-1)(s-1)}(1-\alpha)$

• **Fisher exact test**

H_0: $X,\ Y$ independent

$X \backslash Y$	0	1
0	n_{11}	n_{12}
1	n_{21}	n_{22}

logarithmic interaction $d = \ln \frac{n_{11} n_{22}}{n_{12} n_{21}}$

Fisher (likely similarity) probability $p = \frac{n_{1\cdot}! n_{2\cdot}! n_{\cdot 1}! n_{\cdot 2}!}{n! n_{11}! n_{22}! n_{12}! n_{21}!}$

procedure:
- we list all tables with the same marginal frequencies
- we calculate logarithmic interaction and Fisher probability to all of them
- Q is sum of all p for which $|d| \geq d_0$ (logarithmic interaction of original table)

H_0 rejected $\iff Q < \alpha$h

• **McNemar test**

H_0: $X,\ Y$ have the same distribution

persons	1	2	3	4	...	m	
1. question	0	1	1	0	...		
2. question	1	0	1	0	...		

$X \backslash Y$	0	1
0	n_{11}	n_{12}
1	n_{21}	n_{22}

n_{12} = number of answers 01 n_{21} = number of answers 10 (based on binomial distribution)

H_0 rejected $\iff \left(\frac{1}{2}\right)^n \sum_{k=0}^{s} \binom{n}{k} < \frac{\alpha}{2}$ where $s = \min(n_{12}, n_{21})$, $n = n_{12} + n_{21}$

$\chi^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}} \approx \chi_1^2$ for m large

H_0: rejected $\iff \chi^2 \geq \chi_1^2(1-\alpha)$

• **Cochran Q test** (generalization of McNemar test)

H_0: $X_1, \dots, X_I$ have the same distribution

persons	1	2	...	m	
1th question	0	1	...		n_1
2nd question	1	0	...		n_2
...			...		
r-th question	0	1			n_r
	s_1	s_2	...	s_m	n

$$Q = \frac{r(r-1)\sum_{k=1}^{r} n_k^2 - (r-1)n^2}{rn - \sum_{k=1}^{m} s_k^2} \approx \chi_{r-1}^2$$

H_0 rejected $\iff Q \geq \chi_{r-1}^2(1-\alpha)$

• **Stuart test** (generalization of McNemar test)

H_0: X, Y have the same distribution

$X \backslash Y$	1	...	r	
1	n_{11}	...	n_{1r}	$n_{1\cdot}$
$\vdots$		...		$\vdots$
r	n_{r1}	...	n_{rr}	$n_{r\cdot}$
	$n_{\cdot 1}$	...	$n_{\cdot r}$	n

$a_{kk} = n_{k\cdot} + n_{\cdot k} - 2n_{kk} \quad a_{kl} = -(n_{kl} + n_{lk})$ pro $k \neq l$
$b_k = n_{k\cdot} - n_{\cdot k}$
$k, l = 1, \dots, r-1$
$Q = \boldsymbol{b}^T \boldsymbol{A}^{-1} \boldsymbol{b} \approx \chi_{r-1}^2$

H_0 rejected $\iff Q \geq \chi_{r-1}^2(1-\alpha)$

• **Kolmogorov - Smirnov test** (based on comparison sample and hypothetic distribution function)
- for known parameter values

$X_1, \dots, X_n$ independent, the same distribution, ordered $X_{(1)}, \dots, X_{(n)}$

H_0: X_k has distribution F

$$F_n(x) = \begin{cases} 0 & \text{pro } x \leq x_{(1)} \\ \frac{k}{n} & \text{pro } x_{(k)} < x \leq x_{(k+1)} \\ 1 & \text{pro } x_{(n)} < x \end{cases}$$

$D^+ = \max\left\{\frac{k}{n} - F(x_{(k)})\right\} \qquad D^- = \max\left\{F(x_{(k)}) - \frac{k-1}{n}\right\} \qquad D = \max(D^+, D^-) = \sup\{F_n(x) - F(x)\}$

H_0: $F_n \leq F$ rejected $\iff D^+ \geq D_n^+(1-\alpha)$ (in tables)

H_0: X_k has distribution F rejected $\iff D \geq D_n^+(1-\frac{\alpha}{2}) \approx \sqrt{-\frac{1}{2n}\ln\frac{\alpha}{2}}$

• **Lilliefors test** (based on similar principle)
- for normal distribution, unknown parameters are estimated by $\bar{X}$, S_X^2

• **tests dobré shody**

$n_1, \dots, n_r$ frequencies, $p_1, \dots, p_r$ probabilities

$$\chi^2 = \sum_{k=1}^{r} \frac{(n_k - np_k)^2}{np_k} \approx \chi_{r-1}^2$$

H_0: frequencies correspond to probabilities rejected $\iff \chi^2 \geq \chi_{r-1}^2(1-\alpha)$

- verification of distribution, we does not know m parameters

$X_1, \dots, X_n$ independent, the same distribution
H_0: X_k has given distribution
procedure:

- data divided into r classes
- frequences in classes are investigated
- (respectively) unknown parameters are estimated, e.g. $\bar{X}$, S_X^2,...
- expected frequences np_k are calculated for all classes
- the classes for which $np_k < 5$ are associated (total number of classes is denoted again r)

$$\chi^2 = \sum_{k=1}^{r} \frac{(n_k - n_{p_k})^2}{np_k} \approx \chi^2_{r-1-m}$$

H_0 rejected $\iff \chi^2 \geq \chi^2_{r-1-m}(1-\alpha)$

• **equality of variances tests**

$X_{1,1}, \dots, X_{1,n_1} \sim N(\mu_1, \sigma^2)$ independent
$X_{2,1}, \dots, X_{2,n_2} \sim N(\mu_2, \sigma^2)$ independent
....
$X_{I,1}, \dots, X_{I,n_I} \sim N(\mu_I, \sigma^2)$ independent

independent samples, $n_1 + \cdots + n_I = n$

H_0: $\sigma_1^2 = \cdots = \sigma_I^2$

• **Bartlett test**

H_0: rejected $\iff Q \geq \chi^2_{r-1-m}(1-\alpha)$

$$Q = \frac{(n-I)\ln s^2 - \sum\limits_{k=1}^{I}(n_k-1)\ln s_k^2}{1 + \frac{1}{s(I-1)}\left(\sum\limits_{k=1}^{I}\frac{1}{n_k-1} - \frac{1}{n-I}\right)} \approx \chi^2_{I-1} \quad \text{where } s_k^2 = \frac{1}{n_k-1}\sum_{l=1}^{n_k} x_{kl}^2 - n_k\bar{x}_k^2 \quad s^2 = \frac{1}{n-I}\sum_{k=1}^{I}(n_k-1)s_k^2$$

Non-parametric methods:

(for continuous distribution)

• **sign test**

$X_1, \dots, X_n$ independent, the same distribution
$H_0: \ \tilde{x} = a$
procedure:
- put down $X_1 - a, \dots, X_n - a$
- Y number of positive differences, $U = \frac{2Y-n}{\sqrt{n}} \approx N(0,1)$

H_0: rejected $\iff Y \leq k_1$ or $Y \geq k_2$
$\iff |U| \geq u(1 - \frac{\alpha}{2})$

• **Wilcoxon one-sample test**

$X_1, \dots, X_n$ independent, the same distribution
H_0 : distribution is symmetric about a $(F(-x) + F(x) = 1)$
procedure:
- order is assigned to absolute values of $X_1 - a, \dots, X_n - a$
- $S^+ =$ sum of order of positive $X_k - a$, $S^- =$ sum pof order of negative $X_k - a$, $U = \dfrac{S^+ - \frac{1}{4}n(n+1)}{\sqrt{\frac{1}{24}n(n+1)(2n+1)}} \approx N(0,1)$

H_0: rejected $\iff \min(S^+, S^-) \geq k$
$\iff |U| \geq u(1 - \frac{\alpha}{2})$

• **Wilcoxon two-sample test**

$X_1, \dots, X_m$ independent, the same distribution, $Y_1, \dots, Y_n$ independent, the same distribution, samples independent
H_0 : X_k and Y_k have the same distribution
procedure:

- order is assigned to values $X_1, \dots, X_m, Y_1, \dots, Y_n$
- T_1 = sum of orders X_k, T_2 = sum of orders Y_k,
- $U_1 = mn + \frac{1}{2}m(m+1) - T_1 \quad U_2 = mn + \frac{1}{2}n(n+1) - T_2 \; U = \dfrac{U_1 - \frac{1}{2}mn}{\sqrt{\frac{1}{12}mn(m+n+1)}} \approx N(0,1)$

H_0 rejected $\iff \min(U_1, U_2) \geq k$
$\iff |U| \geq u(1 - \frac{\alpha}{2})$

• Spearmann correlation coefficient
$(X_1, Y_1), \dots, (X_n, Y_n)$ independent, the same distribution
$H_0 : X_k \; Y_k$ independent
procedure:
- values $X_1, \dots, X_n$ are assigned with orders $Q_1, \dots, Q_n$
- values $Y_1, \dots, Y_n$ are assigned with orders $R_1, \dots, R_n$
- $r_S = 1 - \dfrac{6}{n(n^2-1)} \sum_k (R_k - Q_k)^2$

H_0: rejected $\iff |r_S| \geq k$
$\iff |r_S| \geq \dfrac{u(1-\frac{\alpha}{2})}{\sqrt{n-1}}$

• Kendall correlation coefficient
$(X_1, Y_1), \dots, (X_n, Y_n)$ independent, the same distribution
$H_0 : X_k \; Y_k$ independent
procedure:
- couples (X_k, Y_k) ascendently ordered acording to X_k, so in order to $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$
- denoted by s_1 a number of $Y_{(2)}, \dots, Y_{(n)}$ which are greater than $Y_{(1)}$
  s_2 a number of $Y_{(3)}, \dots, Y_{(n)}$ which are greater than $Y_{(2)}$
  ...
  s_{n-1} a number of $Y_{(n)}$ which are greater than $Y_{(n-1)}$
- $\tau = \dfrac{4}{n(n-1)} \sum_k s_k - 1 = \dfrac{1}{n(n-1)} \sum_{k \neq l} \operatorname{sgn}(Q_k - Q_l) \cdot \operatorname{sgn}(R_k - R_l)$ (denotation of preceeding test)

H_0: rejected $\iff |\tau| \geq k$
$\iff |\tau| \geq \dfrac{3\sqrt{n}}{2} u(1 - \dfrac{\alpha}{2})$

• Kruskal - Walis test (one factor ANOVA)

$X_{1,1}, \dots, X_{1,n_1}$ independent, the same distribution
$X_{2,1}, \dots, X_{2,n_2}$ independent, the same distribution
....
$X_{I,1}, \dots, X_{I,n_I}$ independent, the same distribution

samples independent, $n_1 + \dots + n_I = n$

H_0: all distributions are the same
postup:
- values $X_{11}, \dots, X_{1n_1}, \dots, X_{I1}, \dots, X_{In_I}$ are assigned by orders
- T_1 = sum of orders X_{1k}, T_2 = sum of orders X_{2k},...,T_I = sum of orders X_{Ik}
- $Q = \dfrac{12}{n(n+1)} \sum_k \dfrac{T_k^2}{n_k} - 3(n+1) \approx \chi^2_{I-1}$

H_0 rejected $\iff |Q| \geq k$
$|Q| \geq \chi^2_{I-1}(1-\alpha)$
- Neményi method for $n_k = p$ for difference among classes

$P_s = \dfrac{T_s}{n_s}$ difference between classes $\iff |P_s - P_t| > \sqrt{\dfrac{1}{12}\left(\dfrac{1}{n_s} + \dfrac{1}{n_t}\right) n(n+1) h_{I-1}(\alpha)}$

$$|P_s - P_t| > q_{I,\infty}(\alpha)\sqrt{\frac{1}{12} I(Ip+1)}$$

- **Friedmann test** (two factor ANOVA for $n_k = 1$)

$$\begin{array}{l} X_{1,1}, \ldots, X_{1,J} \\ X_{2,1}, \ldots, X_{2,J} \\ \ldots\ldots\ldots\ldots\ldots \\ X_{I,1}, \ldots, X_{I,J} \end{array}$$

independent

H_0: $X_{1l}, X_{2l}, \ldots, X_{Il}$ have the same distribution
postup:
- in every block (column) values are assigned by orders R_{kl}
- $Q = \frac{12}{JI(I+1)} \sum_{l=1}^{I} \left(\sum_{k=1}^{J} R_{kl} \right)^2 - 3J(I+1) \approx \chi^2_{I-1}$

H_0 rejected $\iff$ $|Q| \geq k$
$|Q| \geq \chi^2_{I-1}(1-\alpha)$
- Neményi method for difference among classes

$R_s = \sum_{k=1}^{I} R_{ks}$ significant difference between classes $\iff$ $|R_s - R_t| > q_{m,\infty}(\alpha)\sqrt{\frac{1}{12} IJ(J+1)}$