# Formulas to tests

## z-test

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

## t-test

$$T = \frac{\bar{x} - \mu_0}{s}\sqrt{n} \sim St\left(n - 1\right)$$

- **two samples, paired test**
  The same as one sample test but the difference of samples is tested
- two samples

## test of variance

$$T = \frac{n - 1}{s_0^2}s^2 \sim \chi^2\left(n - 1\right)$$

## test of proportion

$$T = \frac{p - p_0}{\sqrt{p_0\left(1 - p_0\right)/n}} \sim N\left(0, 1\right)$$

## two expectations - independent, equal variances

$$T = \frac{\bar{x}_1 - \bar{x}_2}{S_p\sqrt{1/n_1 + 1/n_2}}$$

$$S_p^2 = \frac{\left(n_1 - 1\right)s_1^2 + \left(n_2 - 1\right)s_2^2}{n_1 + n_2 - 2} \sim St\left(n_1 + n_2 - 2\right)$$

## two expectations - independent, unequal variances

$$T = \frac{\bar{x}_1 - \bar{x}_2}{S_p\sqrt{1/n_1 + 1/n_2}} \sim St\left(\delta\right)$$

$$\delta = \frac{\left[\left(s_1^2/n_1\right) + \left(s_2^2/n_2\right)\right]^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$$

## two expectations - paired samples

$$T = \frac{\bar{D}}{s_D\sqrt{1/n}} \sim St\left(n - 1\right)$$

$$D = x_1 - x_2$$

## two proportions

$$T = \frac{(p_1 - p_2)}{\sqrt{p_P (1 - p_P)} \sqrt{1/n_1 + 1/n_2}} \sim N(0,1)$$

$$p_P = \frac{x_1 + x_2}{n_1 + n_2}$$

## $\chi^2$- goodness-of-fit test

$$T = \sum_i \frac{(O_i - E_i)^2}{E_i} \sim \chi_2 (n-1)$$

## $\chi^2$- independence test

$$T = \sum_{i,j} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \sim \chi_2 (n_1 - 1)(n_2 - 1)$$

## $\chi^2$- homogeneity test

Is the same as independence but it fits to the situation:

We monitor the salaries of men and women in big cities, small towns and in in the villages. We are interested in whether the salary ratios between men and women are the same in all locations.

## Anderson test

$$T = -n - \frac{1}{n} \sum_i (2i-1) [\ln F(x_1) + \ln (1 - F(x_{n-i+1}))]$$

$x$ are sorted

## Shapiro Wilk test

$$T = \frac{\left(\sum_n a_{(i)} x_i\right)^2}{\sum_i (x_i - \bar{x})^2}$$

– $a_{(i)}$ is the ith-smallest number in the sample (the order statistics) and

$a = (a_1, a_2 \cdots a_n) = \sqrt{m' V^{-1} V^{-1} m}$

– $m$ are the expected values of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution

– $V$ is the covariance matrix of $m$

## correlation test

$$T = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \sim St\left(n-2\right)$$

## Spearman test

$$r_x = \text{ranks}\left(x\right), \;\; r_y = \text{ranks}\left(y\right)$$

$$r = \text{cor}(r_x, r_y)$$

$$T = |r\sqrt{n-1}| \sim N\left(o, 1\right)$$

## Keanall test

We have $(x_i, y_i)$ $i = 1, 2, \cdots, n$
$C$ - concordant pairs $i$ and$j$: $\quad x_i < x_j \rightarrow y_i < y_j$ or $x_i > x_j \rightarrow y_i > y_j$
$D$ - discordant pairs are the rest

$$\tau = \frac{|C| - |D|}{n}$$

$|\cdot|$ is number of elements.

$$T = \frac{\tau}{2\left(2n+5\right)/\left(9n\left(n-1\right)\right)} \sim St\left(\text{abs}\left(T\right)\right)$$

neověřeno !!!

## Cramer

[nr,nc]=size(O);
[pval,chi2,df]=chisquare_test_i(O); // chi2 of independence
V=sqrt((chi2/sum(O))/min(nr-1,nc-1)); // coefficient

## Friedman test

Rj = sum(rnk,1); // column sums (1 × k)
rj = Rj-n*(k+1)/2; // deviation from mean rank
Q = (12/(n*k*(k+1))) * sum(rj.^2); // Friedman statistic
where *rnk* are ranks for columns of the matrix $X$ where in rows are subjects (evaluators) and in columns are treatments (evaluated objects)

## Kolmogorov Smirnov test

$$KS = \text{abs}\left(F - F_e\right) \sim \text{spec.distr.}$$

$F$ is distribution function

$F_e$ is empirical distribution function

## Kruskal Wallis test

We have data as matrix with groups in columns. Rank the entire matrix regardless of groups. $r_{ij}$ is the rank in $j$th group and $i$th measurement. The statistics is

$$T = \left(N-1\right) \frac{\sum_{j=1}^{g} n_j \left(\bar{r}_j - \bar{r}\right)^2}{\sum_{j=1}^{g} \sum_{i=1}^{n_j} \left(r_{ij} - \bar{r}\right)^2} \sim \text{special distribution}$$

$N$ is total number of data

$g$ is number of groups

$n_j$ is number of data in group $j$

$\bar{r}_j$ is the average rank of all data in group $j$

$\bar{r}$ is the average of all ranks

## Lambda coefficient

$$\Lambda = \frac{E_y - E_r}{E_y}$$

$E_y$ is number of errors when there is no condition

$E_r$ is number of errors when including the condition

## Mann Whitney test

Rank all data regardless of groups. $r_1$ are ranks belonging to group 1, $r_2$ to group 2.

$$U = \min\left\{n_1 n_2 + \frac{n_1\left(n_1+1\right)}{2}, n_1 n_2 + \frac{n_2\left(n_2+1\right)}{2}\right\} \sim \text{spec.dist.}$$

## MacNemar test

| | Test 2 positive | Test 2 negative |
|---|---|---|
| Test 1 positive | a | b |
| Test 1 negative | c | d |

$$T = \frac{\left(b-c\right)^2}{b+c} \sim \chi^2\left(1\right)$$

## residual tests

### z-test

$$a = x - \text{median}\,(x)$$

$$r\ -\ \text{number of } a_i a_{i+1} \le 0$$

$$T = \frac{2r-(n-2)}{\sqrt{n-1}} \sim N\,(0,1)$$

### f-test

Perform linear regression and prediction. $e_p = x - xp$ is prediction error

$$T = \frac{e_p e_p^{'}/n}{\text{var}\,(x)} \sim F\,(\text{ord}+1, n-\text{ord})$$

where ord is the model order.

## Scheffé test

for all $i$ and $j$ (numbers of groups) do

$$M(j,k) = \text{abs}(mx_j - mx_k);$$

$$F(j,k) = \sqrt{(\text{msw}\, dF\,(nc-1)(1/n_j + 1/n_k))};$$

$$T = F \sim F\,(\text{dfb}, \text{dfw})$$

where msw is mean sum of squares of data (from anova)
$nc$ is number of groups
$dF$ is value of $F$-distribution at the probability $\alpha$ set as input
dfb, dfw are degrees of freedom from the anova test

## sign test

b=min(sum(x>y),sum(x<=y))
T=(2*b-n+1)/sqrt(n);

## Wilcoxon test

$$x = x_1 - x_2$$

$$s = \sum \text{sign}\,(x_i)$$

$$D = \sum \text{abs}(x_i)$$

$n$ length$(D)$

$$R = \text{ranks}\,(D)$$

$jk$ are indexes of $s = 1$

$$W = \sum R(jk)$$

for n>20

$$m = \frac{n(n+1)}{4}$$

$$s = \frac{\sqrt{n(n+1)(2n+1)}}{24}$$

$$T = \frac{W - m}{s} \sim N(0, 1)$$

else

special approach with table

## anova

$$y_{ij} = a_j + k + e_{ij}$$

perform regression and compute predictions $y_p$
MSR = sum((yp - mean(y)).^2)/$(k-1)$;
MSE = sum((y-yp).^2)/$(n-k)$;

$$T = \frac{\text{MSR}}{\text{MSE}} \sim F(k-1, n-k)$$

## anova two factors

$$T_1 = \frac{\text{MS1}}{\text{MSR}}, \quad T_2 = \frac{\text{MS2}}{\text{MSR}}$$

and $F$-test for both.

## anova with repetitions

Similar as the previous - see the program.

## ancova

$$y_{ij} = a_j + x_{ij} + k + e_{ij}$$

regression, prediction and follow anova.

## Bartlett test

```
for i=1:a
ni=length(t(i));
si=variance(t(i));
s1=s1+1/(ni-1);
s2=s2+(ni-1)*log(si);
s3=s3+(ni-1)*si
S(i)=si;
end
C=1+(s1-1/(n-a))/(3*(a-1));
s2c=(n-a)*log(s3/(n-a));
B=(s2c-s2)/C;
```

$a$ number of groups

$t$ data matrix with groups $t_i$ in columns

$n$ vector of lengths of the groups

$s$ variances in groups

$$s_1 = 1/\left(n-1\right)$$

$$s_2 = \left(n-1\right)\ln\left(s\right)$$

$$s_3 = \left(n-1\right)s$$

$$C = 1 + \left[s_1 - \frac{1}{\left(n-a\right)}\right] / \left(3\left(a-1\right)\right)$$

$$s2c = \left(n-a\right)\ln\left(\frac{s_3}{n-a}\right)$$

$$T = \frac{s2c - s_2}{C} \sim \chi^2\left(a-1\right)$$

## factor analysis

### pca

### svd