**Užitečné statistické funkce v Excelu**
Názvy funkcí odpovídají české lokalizaci MS Excel 2007

**POČET(hodnota1; [hodnota2]; ...)** – spočítá počet dat (neprázdných buněk s číselnou hodnotou) v zadané oblasti.
Např. **=POČET(A1:A100)** vrátí počet dat (neprázdných buněk s číselnou hodnotou) v oblasti A1:A100.

**COUNTIF(oblast,"podmínka")** – spočítá počet dat v zadané oblasti splňujích danou podmínku.
Např. **=COUNTIF(A1:A100,">1.5")** vrátí počet dat v oblasti A1:A100, která jsou větší než 1.5.

**MAX(hodnota1; [hodnota2]; ...)** – spočítá maximální hodnotu v zadané oblasti.
Např. **=MAX(A1:A100)** vrátí maximální hodnotu v oblasti A1:A100.

**MIN(hodnota1; [hodnota2]; ...)** – spočítá minimální hodnotu v zadané oblasti.
Např. **=MIN(A1:A100)** vrátí minimální hodnotu v oblasti A1:A100.

**MODE(hodnota1; [hodnota2]; ...)** – spočítá modus, tj. nejčastěji se vyskytující hodnotu v zadané oblasti. Pokud není žádná duplicitní hodnota vrátí MODE chybovou hodnotu #NENÍ_K_DISPOZICI.
Např. **=MODE(A1:A100)** vrátí nejčastěji se vyskytující hodnotu v oblasti A1:A100.

**PRŮMĚR(oblast)** – spočítá aritmetický průměr dat v zadané oblasti, $\frac{1}{N}\sum_{i=1}^{N} x_i$
Např. =**PRŮMĚR(A2:A100)** vrátí aritmetický průměr dat v oblasti A2:A100

**MEDIAN(oblast)** – spočítá medián dat v zadané oblasti
Např. =**MEDIAN(A2:A100)** vrátí medián dat v oblasti A2:A100

**GEOMEAN(oblast)** – spočítá geometrický průměr dat v zadané oblasti $\sqrt[N]{\prod_{i=1}^{N} x_i}$
Např. =**GEOMEAN(A2:A100)** vrátí geometrický průměr dat v oblasti A2:A100

**HARMEAN(oblast)** – spočítá harmonický průměr dat v zadané oblasti $\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{x_i}\right)^{-1}$
Např. =**HARMEAN(A2:A100)** vrátí geometrický průměr dat v oblasti A2:A100

**SUMA.ČTVERCŮ(oblast)** – spočítá součet kvadrátů dat v zadané oblasti $\sum_{i=1}^{N} x_i^2$
Např. = **SUMA.ČTVERCŮ(A2:A100)** vrátí součet kvadrátů dat v oblasti A2:A100

**STDEVA(oblast)** – spočítá standardní odchylku dat (nepředpojatý odhad) v zadané oblasti $\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2}$, kde $\bar{x}$ je aritmetický průměr.
Např. = **STDEVA(A2:A100)** vrátí standardní odchylku dat v oblasti A2:A100

**STDEVPA(oblast)** – spočítá standardní odchylku dat (předpojatý odhad) v zadané oblasti $\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2}$, kde $\bar{x}$ je aritmetický průměr.
Např. = **STDEVPA(A2:A100)** vrátí standardní odchylku dat v oblasti A2:A100
**SMODCH(oblast)** je stejné jako **STDEVPA(oblast)**

**VAR(oblast)** – spočítá rozptyl dat (předpojatý odhad) v zadané oblasti $\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$, kde $\bar{x}$ je aritmetický průměr. Odmocnina z rozptylu je standardní odchylka.
**ODMOCNINA(STDEVPA(oblast)) = STDEVPA(oblast) = SMODCH(oblast)**

**SKEW(oblast)** – spočítá nepředpojatý odhad šikmosti dat v zadané oblasti
$\frac{N}{(N-1)(N-2)}\sum_{i=1}^{N}\left(\frac{x_i-\bar{x}}{s}\right)^2$,
kde $\bar{x}$ je aritmetický průměr $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
a $s$ je $s = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$.

**KURT(oblast)** – spočítá nepředpojatý odhad špičatosti dat v zadané oblasti
$\frac{N(N+1)}{(N-1)(N-2)(N-3)}\sum_{i=1}^{N}\left(\frac{x_i-\bar{x}}{s}\right)^4 - 3\frac{(N-1)^2}{(N-2)(N-3)}$,
kde $\bar{x}$ je aritmetický průměr $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
a $s$ je $s = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$.

## Generování náhodných čísel v Excelu

**NÁHČÍSLO()** – vrátí výběr z rovnoměrného rozdělení U(0,1), tj. číslo, které se nachází se stejnou pravděpodobností kdekoliv v intervalu (0,1)

**RANDBETWEEN(a,b)** – vrátí náhodně vybrané celé číslo ležící mezi *a* a.*b*.
Např. = **RANDBETWEEN(10,20)** může vrátit třeba 12

## Rozdělení diskrétní náhodné proměnné

**POISSON(k,ν,kumulativní)** – pokud je parametr kumulativní=0 vypočítá pravděpodobnost, že *k*-krát nastane úspěch podle Poissonova rozdělení s očekávanou hodnotou ν, tj. $P(k|\nu) = \frac{\nu^k e^{-\nu}}{k!}$
Pokud je kumulativní=1 vypočítá hodnotu distribuční funkce Poissonova rozdělení s očekávanou hodnotou ν, tj. $F(k|\nu) = \sum_{l=0}^{l=k}\frac{\nu^l e^{-\nu}}{l!}$
Např. = **POISSON(2,1.5,0)** vrátí 0.251

**BINOMDIST(k,N,p,kumulativní)** – pokud je parametr kumulativní=0 vypočítá pravděpodobnost, že *k*-krát nastane úspěch při *N* opakování experimentu, když pravděpodobnost úspěchu při jednom opakování je *p*, tj. $P(k|N,p) = \frac{N!}{(N-k)!\,k!}p^k(1-p)^{N-k}$
Pokud je kumulativní=1 vypočítá hodnotu distribuční funkce Binomického rozdělení s očekávanou
$F(k|N,p) = \sum_{l=0}^{k}\frac{N!}{(N-l)!\,l!}p^l(1-p)^{N-l}$.
Např. = **BINOMDIST(5,10,0.5,0)** vrátí 0.246

## Rozdělení spojité náhodné proměnné

**NORMDIST(x,μ,σ,kumulativní)** – pokud je parametr kumulativní=0 vrátí hodnotu hustoty pravděpodobnosti normálního rozdělení s očekávanou hodnotou μ a standardní odchylkou σ
$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$

Pokud je parametr kumulativní=1, tak vrací hodnotu distribuční funkce normálního s očekávanou hodnotou $\mu$ a standardní odchylkou $\sigma$
$F(x|\mu,\sigma) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right] dt$ .
Např. = **NORMDIST(2,0,2,1)** vrátí 0.8413

**NORMINV(P,μ,σ)** – vypočítá inverzní funkci k distribuční funkci normálního rozdělení s očekávanou hodnotou $\mu$ a standardní odchylkou $\sigma$, tj. takové $x$, že $F(x|\mu,\sigma) = P$
Např. = **NORMINV(0.8413,0,1)** vrátí 1.9996
Pomocí této funkce je možné vygenerovat v Excelu náhodnou proměnou z normálního rozdělení použitím metody inverzní funkce = **NORMINV(NÁHČÍSLO(),0,1)** – vrátí náhodný z normálního rozdělení s očekávanou hodnotou 0 a standardní odchylkou 1.

Korelace a kovariance

**COVAR(oblast1,oblast2)** – vypočítá kovarianci náhodných proměnných v oblasti 1 a 2
$\widehat{\text{cov}}(x,y) = \langle xy\rangle - \langle x\rangle\langle y\rangle$,
kde lomená závorka $\langle\ \ \rangle$ označuje aritmetický průměr, např. $\langle x\rangle = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{x_i}$
Např. = **COVAR(A2:A100,B2:B100)**

**CORREL(oblast1,oblast2)** – vypočítá korelaci náhodných proměnných v oblasti 1 a 2
$\hat{\rho}_E(x,y) = \frac{\widehat{\text{cov}}(x,y)}{\hat{\sigma}_x\hat{\sigma}_y}$,
Např. = **CORREL(A2:A100,B2:B100)**
*Poznámka:*
Excel používá v odhadu korelačního koeficientu předpojaté odhady standardní odchylky **STDEVPA**
tj. $\hat{\sigma}_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \langle x\rangle)^2}$
Korelační koeficient tedy Excel počítá jako
**=COVAR(oblast1,oblast2)/(STDEVPA(oblast1)*STDEVPA(oblast2))**
My na přednášce počítáme odhad Pearsonova korelačního koeficientu s použitím nepředpojatých odhadů standardní odchylky, tj. jako $\hat{\rho}(x,y) = \frac{\widehat{\text{cov}}(x,y)}{\hat{s}_x\hat{s}_y}$, kde
$\hat{s}$ je nepředpojatý odhad standardní odchylky, tj. $\hat{s}_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \langle x\rangle)^2}$
V Excelu by se to tedy spočítalo jako
**=COVAR(oblast1,oblast2)/(STDEVA(oblast1)*STDEVA(oblast2))**
Přepočet mezi tím „naším“ a „Excelovským“ korelačním koeficientem je $\hat{\rho} = \frac{N-1}{N}\hat{\rho}_E(x,y)$.

**TDIST(t,ν,chvosty)** – vypočítá pravděpodobnost, že výběr se Studentova rozdělení s počtem volnosti $\nu$ bude větší než $t$, tj. 1 - hodnota distribuční funkce $T(t|\nu)$ Studentova rozdělení s počtem stupňů volnosti $\nu$ v bodě $t$,
$P(x > t) = 1 - T(t|\nu) = 1 - \int_0^t \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{z^2}{\nu}\right)^{-\frac{\nu+1}{2}} dz$,
Takto se to vypočítá, pokud je parametr chvosty =1.
Pokud je parametr chvosty = 2 vypočítá se $P(|x| > t) = 2\,P(x > t)$.
Takže **TDIST(t,ν,2)=2*TDIST(t,ν,1)**
*Poznámka:*
Musí být $t \geq 0$. Protože Studentovo rozdělení je symetrické okolo nuly,. Můžeme pravděpodobnost, že $x < t$ pro záporné $t$ spočítat jednoduše jako $P(x > |t|)$, tj. pokud máme $t$ záporné a zajímá nás pravděpodobnost, že $x < t$, tak se to spočítá jednoduše jako **=TDIST(ABS(t),ν,1)**

**`TINV(P,ν)`** – spočítá takovou hodnotu $t_P$, pro kterou platí, že pravděpodobnost, že náhodná proměnná ze studentova rozdělení s počtem stupňů volnosti $\nu$ se bude od nuly lišit o víc než $t_P$ je P, tj.
1-$P(|t| > t_P|\nu) = 2[1 - T(t_P|\nu)] = P$.
Např. = **`TINV(0.05,20)`** vrátí 2.086