- Intenzita závislosti a regresní funkce
- Statistická významnost parametrů regresního modelu
- Posouzení regresního modelu jako celku

2. Kvalita regresní funkce

Cíl kapitoly:

Poté co je vytvořena regresní funkce je nutno zhodnotit její kvalitu. Tato kapitola Vám nabídne několik statistických metod, které jsou k těmto účelům používány. Kritéria se pohybují od jednoduchých popisných veličin až po poměrně sofistikované metody jako jsou například statistické testy. Z těchto důvodů je teoretický popis doplněn podrobným příkladem jejich výpočtu na konkrétním příkladu.

Časová zátěž

6 hodin (1. týden v březnu)

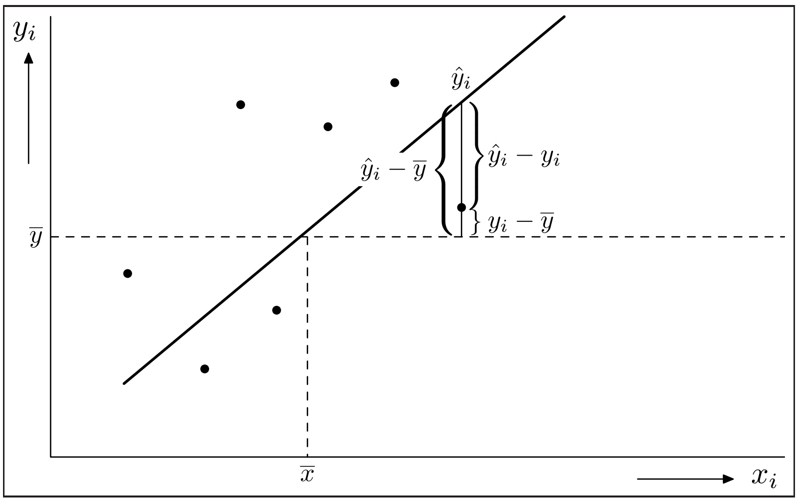
Uvod

Nezbytnou součástí procesu vytváření regresní funkce je posouzení jeho kvality. K vlastnímu posouzení je možno použít několik kritérií. Jedná se především o zhodnocení statistické významnosti vytvořeného modelu i jednotlivých vypočítaných parametrů regresní funkce.

Východiskem k posouzení kvality regresní funkce je zhodnocení těsnosti (resp. volnosti) posuzované závislosti. K tomuto účelu slouží několik popisných statistických veličin. Jedná se především o index determinace, korelační koeficient a korelační poměr.

2.1 Intenzita závislosti a regresní funkce

Posuzovaný vztah popisovaný regresní funkcí je tím kvalitnější (silnější), čím více jsou empirické veličiny více soustředěné kolem vytvořené regresní funkce. Východiskem odvození charakteristik kvality regresní funkce je následující obrázek:



Obrázek 2.1: Rozdíly a rozptyly v regresní funkci

Pro každou z empirických (naměřených, původních) hodnot y_i (v grafu naznačeny plný mi body) nalezneme právě jednu teoretickou (vyrovnanou, vypočítanou) hodnotu $\hat{y}_i$ (nalezneme je jako kolmý průmět empirické hodnoty na regresní přímku). Obě tyto hodnoty lze srovnat s aritmetickým průměrem všech empirických hodnot $\overline{y}$, jak je naznačeno v obrázku 2.1. Tyto rozdíly pak konstruujeme pro všechny dvojice empirických a teoretických hodnot a dostáváme tak dva druhy rozptylů – rozptyl empirických a teoretických hodnot. Tyto rozptyly jsou dále doplňovány tzv. reziduálním rozptylem.

2.1.1 Rozptyl empirických hodnot

Rozptyl empirických hodnot měří průměrnou odchylku naměřených hodnot (y_i), jež jsou východiskem k odhadu regresní funkce, od jejich aritmetického průměru ($\overline{y}$). Jedná se tedy o klasickou popisnou statistickou veličinu, která naznačuje rozvrstvení původních hodnot vysvětlující proměnné vzhledem k jejich aritmetickému průměru.

$$s_y^2 = \frac{\sum\limits_{i=1}^{n}(y_i - \overline{y})^2}{n}$$

2.1.2 Rozptyl teoretických (vyrovnaných) hodnot

Rozptyl teoretických hodnot měří průměrnou odchylku hodnot, jež jsou vypočítány pomocí regresní funkce ($\hat{y}_i$), od aritmetického průměru empirických hodnot ($\overline{y}$). Jedná se tedy opět o popisnou statistickou veličinu.

$$s_{\hat{y}}^2 = \frac{\sum\limits_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{n}$$

2.1.3 Reziduální rozptyl

Tato popisná statistická veličina je měřítkem průměrné odchylky empirických (y_i) a vypočítaných hodnot ($\hat{y}_i$). Reziduální rozptyl je tedy rozptylem empirických hodnot kolem regresní funkce.

$$s_{(y-\hat{y})}^2 = \frac{\sum\limits_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}$$

Lze dokázat, že mezi uvedenými rozptyly platí vztah $s_y^2 = s_{\hat{y}}^2 + s_{(y-\hat{y})}^2$. Rozptyl empirických hodnot tedy lze rozložit na součet zbývajících dvou rozptylů.

Kdyby mezi závisle proměnnou y a nezávisle proměnnou x byla **dokonalá funkční závislost** (y byla závislá na x), **ležely by naměřené hodnoty na regresní přímce**. Vyrovnané a naměřené hodnoty by z těchto důvodu v ideálním případě splynuly, což by značilo, že **reziduální rozptyl by byl nulový**.

Naopak pokud by mezi x a y nebyla funkční závislost (uvedené proměnné by byly naprosto nezávislé), byly by vyrovnané hodnoty všechny stejné. Jejich rozptyl (označujeme jej jako rozptyl teoretických hodnot) se v tomto případě bude rovnat nule. Při naprosté nezávislosti mezi posuzovanými veličinami se tedy sobě rovnají reziduální rozptyl s rozptylem empirických hodnot.

Index determinace

Uvedené případy – naprostá závislost a naprostá nezávislost skutečnosti jsou hraniční a regresní modely se pohybují v těchto mantinelech. K posouzení, ke kterému z obou hraničních stavů má posuzovaná regresní přímka blíže, se používá zejména veličina nazývaná **index determinace**. Je definován jako

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2}.$$

Závislost proměnné x a proměnné y bud tím silnější, čím více se rozptyl teoretických hodnot bude svou hodnotou blížit k rozptylu empirických hodnot. V tomto případě se tedy index determinace bude blížit k jedné. Naopak, čím více se rozptyl vyrovnaných hodnot bude blížit k nule (všechny vyrovnané hodnoty jsou stejné), tím slabší bude tato vzájemná závislost. V případě naprosté nezávislosti bude index determinace roven nule.

Index determinace tedy udává tu část rozptylu závisle proměnné, kterou je možno vysvětlit pomocí regresní funkce.

Při hodnocení intenzity závislosti pomocí indexu determinace je nutno mít na zřeteli, že tato statistická veličina je svázána se zvolenou regresní funkcí. Nízké hodnoty indexu ještě nemusí znamenat, že mezi posuzovanými veličinami není vzájemný vztah. Nízká hodnota indexu determinace může být způsobena špatně zvoleným typem regresní funkce (závislost mezi veličinami není lineární ale např. exponenciální).

Korelační koeficient

V praxi je často místo indexu determinace používána jeho druhá odmocnina nazývaná jako **index korelace**. Pro přímkovou (lineární) regresi je index korelace označován jako **korelační koeficient**.

Pro výpočet korelačního koeficientu lze využít zjednodušeného vztahu, který lze odvodit dosazením regresní funkce do vztahu pro koeficient korelace. Toto odvození, stejně jako interpretaci zjednodušeného (výpočtového) tvaru korelačního koeficientu naleznete v učebnici Seger, Hindls Statistické metody v tržním hospodářství na stranách 275–276.

2.2 Statistická významnost parametrů regresního modelu

Druhou možností jak ověřit kvalitu regresní funkce je použití intervalů spolehlivosti, respektive principu statistického testování pro regresní parametry. Jedná se o složitější statistické postupy než je výpočet výše uvedených popisných veličin. Z těchto důvodů k těmto metodám obvykle přistupujeme až tehdy, když známe hodnoty indexu determinace, resp. korelačního koeficientu. V případě, že hodnoty těchto veličin jsou blízké jedné (hovořící o dobrém vystižení zkoumaného vztahu), je vhodné otestovat regresní model na jeho statistickou významnost. Teprve na základě výsledku obou těchto analýz lze formulovat skutečný závěr o kvalitě regresního modelu.

2.2.1 Interval spolehlivosti pro jednotlivé regresní parametry (b_0, b_1)

Interval spolehlivosti je interval, ve kterém se nachází hodnota závisle proměnné s danou pravděpodobností (hladinou spolehlivosti). Konstrukce intervalu vychází předpokladu, že regresní parametry mají při splnění základních podmínek modelu Studentovo t-rozdělení s $n-p$ stupni volnosti (n=počet pozorování; p= počet parametrů regresní funkce (u lineární=2)).

Pro jednotlivé regresní parametry lze interval spočítat podle vztahu:

$$IS_{b_0} = \left(b_0 - s_{b_0} t_{1-\frac{\alpha}{2}};\, b_0 + s_{b_0} t_{1-\frac{\alpha}{2}}\right),$$
$$IS_{b_1} = \left(b_1 - s_{b_1} t_{1-\frac{\alpha}{2}};\, b_1 + s_{b_1} t_{1-\frac{\alpha}{2}}\right),$$

kde b_0, b_1 jsou vypočítané regresní parametry a s_{b_0}, s_{b_1} jsou příslušné směrodatné odchylky těchto parametrů. Veličiny $t_{1-\alpha/2}$ nalezneme jako příslušný kvantil studentova rozdělení s daným počtem stupňů volnosti.

Mimo intervalu spolehlivosti pro jednotlivé regresní parametry je možno konstruovat podobné intervaly i pro výše uvedené popisné veličiny – např. koeficient korelace. Jejich princip, stejně jako odvození (včetně vztahu pro interval spolehlivosti regresních parametrů) je naznačeno v učebnici Seger, Hindls Statistické metody v tržním hospodářství na stranách 300–313.

Cílem vytváření intervalu spolehlivosti pro regresní parametry je provést test na nulovost (nenulovost) regresního parametru. Pokud není možno vyloučit skutečnost, že parametr je roven nule, nelze takový parametr ponechat v regresním modelu.

(Uvědomte si, že nula násobená libovolným číslem je vždy rovna nule.)

Sestrojíme-li interval spolehlivosti pro daný regresní koeficient, mohou nastat dva případy:

- Interval **obsahuje 0**
  V tomto případě nelze zamítnout hypotézu, že je koeficient roven nule. Neprokázali jsme tedy, že je tento koeficient nutno zahrnout do modelu (**koeficient není statisticky významný**).
- Interval **neobsahuje 0**
  Není-li v IS obsažena nula (tedy interval tvoří buď jenom kladné nebo jenom záporné hodnoty), zamítáme hypotézu, že regresní koeficient je roven 0. V tomto případě je regresní koeficient nutno zahrnout do modelu (**koeficient je statisticky významný**).

2.2.2 Statistický test na nenulovost regresních parametrů b_0 a b_1

Stejný výsledek jako posouzení pomocí intervalu spolehlivosti dává i statistický test. Provedeme-li statistický test s nulovou hypotézou, že regresní koeficient je roven nule ($b_i = 0$), lze do regresního modelu zahrnout pouze ty koeficienty, pro které výsledek testu hovoří ve prospěch zamítnutí hypotézy H_0.

Jelikož je opět možno využít vlastností t-rozdělení, k testování použijeme t-test. Testové kritérium má tvar:

$$T = \frac{|b_i|}{s_{b_i}} > t_{\frac{\alpha}{2}}.$$

$t_{\frac{\alpha}{2}}$ – kvantil t-rozdělení pro hladinu významnosti α a počet stupňů volnosti$=p$,
b_i – testovaný koeficient
s_{b_i} – směrodatná odchylka tohoto koeficientu.

Je-li hodnota testového kritéria T vyšší než příslušný kvantil t-rozdělení, padá výsledek testu do kritického oboru. V tomto případě zamítáme hypotézu H_0 o nulovosti koeficientu a konstatujeme, že daný parametr je statisticky významný na hladině spolehlivosti (obvykle je volena 5% - tzn. s 95% pravděpodobností).

Postup testování statistických hypotéz je součástí prvního dílu tohoto textu (Aplikovaná statistika I). Stejně tak jej lze nalézt ve většině statistických učebnic – viz např. Seger, Hindls: *Statistické metody v tržním hospodářství* na stranách 159–224. V této učebnici také můžete nalézt odvození uvedeného testového kritéria pro jednotlivé parametry regresní funkce (na stranách 300–313).

2.3 Posouzení regresního modelu jako celku

V případě, že je za regresní funkci zvolena přímka (lineární regrese), je vhodné otestovat vhodnost volby tohoto tvaru. Zajímá nás tedy, zda je přímka vhodnou funkcí pro odhad dané závislosti, či zda by bylo vhodnější zvolit jiný tvar regresní funkce (např. exponenciální funkci). V tomto případě provádíme statistický test na celkovou statistickou významnost regresního modelu. Zajímá nás, zda jsou koeficienty zvolené regresní přímky statisticky významné jako sada.

Nejsou-li regresní koeficienty statisticky významné jako celek, nemá význam testovat jejich statistickou významnost odděleně. V případě neúspěšného testu na celkovou statistickou významnost modelu jsou t-testy pro jednotlivé koeficienty bezpředmětné.

Pro testování statistické významnosti celého regresního modelu vycházíme z nulové hypotézy o nulovosti všech regresních koeficientů. Tedy:

H_0: všechna $b_i = 0$
H_1: alespoň jedno b_i je různé od nuly

Testové kritérium má mít F-rozdělení o $p-1$ a $n-p$ stupních volnosti (n=počet pozorování; p=počet parametrů regresní funkce – u lineární=2).

H_0 zamítáme v případě, že $F > F_{1-\alpha}$, kde $F_{1-\alpha}$ je kvantil F-rozdělení. Jeho tvar je:

$$F = \frac{\frac{s_t^2}{p-1}}{\frac{s_r^2}{n-p}} > F_{\frac{\alpha}{2}}$$

$F_{\frac{\alpha}{2}}$ – kvantil F-rozdělení pro hl. význ. α a počet stupňů volnosti $p-1$ a $n-p$
s_t^2 – rozptyl teoretických hodnot
s_r^2 – reziduální rozptyl

V praktických úlohách většinou vystačíme s následujícím pomocným kritériem: Koeficienty regresní funkce jsou jako celek statisticky významné v případě, že dostáváme extrémně vysoké hodnoty statistiky F (testového kritéria).

Je-li výsledek F-testu statisticky významný, je vhodné provádět t-testy, v opačném případě ani úspěšné t-testy nehovoří nic o dobré kvalitě regresního modelu. Pokud je F-test statisticky významný a t-testy nejsou, pak lze říci, že je model špatně určen (je nutno vypočítat jiné regresní koeficienty). V případě, že je F-test neúspěšný, není model vůbec možno použít (např. to znamená, že nelze použít regrese lineární ale např. polynomickou).

Příklad 2.1

(Pokračování z kapitoly 1)

Prozkoumejte vztah mezi výdaji veřejných rozpočtů (VV) a HDP v ČR v letech 1993–2001 pomocí regresního modelu.

V kapitole 1 jsme vztah mezi veřejnými výdaji a hrubým domácím produktem v ČR popsali pomocí regresní funkce ve tvaru: VV = 23,11 + 0,37 HDP. Na základě výše uvedených možností posoudíme kvalitu a použitelnost uvedené funkce.

Jelikož by výpočty jednotlivých charakteristik byly poměrně technické a zdlouhavé, určíme je přímo v prostředí programu EXCEL. V případě, kdy neuvádíme postup, jsou hodnoty uvedeny ve výstupu funkce LINREGRESE, tak jak je ukázáno v příloze první kapitoly.

Posouzení kvality:

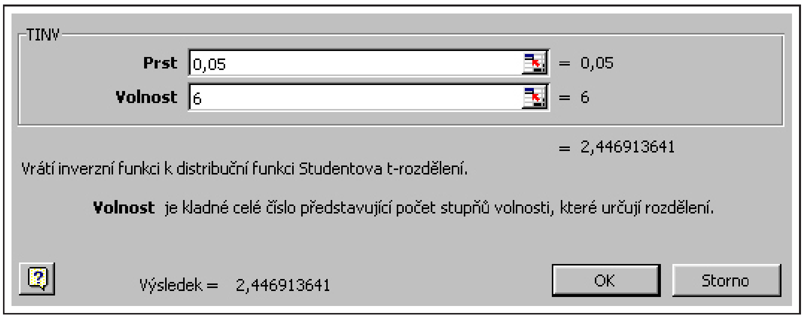
1) **Popisné veličiny – index determinace** $R^2 = 0{,}987$
   Hodnota indexu determinace je blízká jedné, proto lze říci, že mezi oběma proměnnými je velmi těsná závislost. Tato skutečnost může však být ovlivněna některými skutečnostmi:
   - Jednak jsou to vlivy matematicko-statistické (špatný model, který vede k nadějným výsledkům, jež nejsou v souladu se skutečností). Je proto nutné prověřit daný model dalšími statistickými kritérii (viz body b) až d)).
   - Dalším důvodem může být i skutečnost, že obě proměnné jsou ovlivněny jinou proměnnou (v tomto případě lze mít podezření na vliv růstu cenové hladiny – inflaci). Proměnné se vlivem tohoto skrytého faktoru vyvíjejí velmi podobně a model dosahuje poměrně dobrých výsledků. Jeho použitelnost je však oslabena. Abychom pomocí modelu mohli dělat predikce (odhady vývoje do

budoucna), museli bychom vycházet ze znalosti, či predikce vývoje této skryté proměnné (inflace).

2) **Statistická významnost jednotlivých parametrů**
   a) **intervaly spolehlivosti**
   Základní předpis pro interval spolehlivosti (IS) koeficientů je

$$b_0 - t_{1-\frac{\alpha}{2}} s_{b_0} < b_1 < b_0 + t_{1-\frac{\alpha}{2}} s_{b_0}$$
$$b_1 - t_{1-\frac{\alpha}{2}} s_{b_1} < b_1 < b_1 + t_{1-\frac{\alpha}{2}} s_{b_1}$$

Kritickou hodnotu t-rozdělení $t_{1-\frac{\alpha}{2}}$ v EXCELu vypočítáme pomocí funkce TINV.



Obrázek 2.2: Kvantil t-rozdělení pomocí funkce TINV

Jako první se zadá hladina spolehlivosti (Prst), proměnná volnost označuje počet stupňů volnosti ($n - p$, kde n=počet pozorování (zde 8); p=počet parametrů regresní funkce – u lineární=2).

Směrodatná odchylka s_{b_0} je ve výstupu fce LINREGRESE. (v našem případě má hodnotu 27,97). Interval spolehlivosti má tedy tvar:

$$b_0 \in \langle -45{,}3;\ 91{,}5 \rangle$$

Po dosazení do vztahu pro IS pro b_1 dostáváme (směrodatná odchylka s_{b_1} má hodnotu 0,0175):

$$b_1 \in \langle 0{,}331;\ 0{,}416 \rangle$$

Z výsledků můžeme vyvodit následující závěry:

Pro b_0:
Interval obsahuje 0 – **nelze zamítnout hypotézu**, že je koeficient roven nule – neprokázali jsme, že je tento koeficient nutno zahrnout do modelu (**koeficient není statisticky významný**)

Pro b_1: Interval neobsahuje 0 – **zamítáme hypotézu**, že regresní koeficient je roven 0 a tento koeficient je nutno zahrnout do regresního modelu (**koeficient je statisticky významný**).

b) **statistický test pro jednotlivé koeficienty** Testují se hypotézy o nulové hodnotě jednotlivých regresních koeficientů – tedy zda by je bylo možné z regresní funkce vyloučit (jestliže jsou rovny 0).

H_0: $b_0 = 0$ (resp. $b_1 = 0$)
H_1: $b_0 \neq 0$ (resp. $b_1 \neq 0$)

kritérium

$$\frac{|b_i|}{s_{b_i}} > t_{\frac{\alpha}{2}}$$

Po dosazení dostáváme

pro b_0

$$\frac{|23{,}08|}{27{,}97} = 0{,}83 < t_{\frac{\alpha}{2}} = 2{,}45$$

∗ tedy pro b_0 **nelze zamítnout** hypotézu, že $b_0 = 0$ a koeficient proto **není statisticky významný**.

pro b_1

$$\frac{|0{,}37|}{0{,}018} = 21{,}31 > t_{\frac{\alpha}{2}} = 2{,}45$$

∗ tedy pro b_1 **zamítáme** hypotézu, že $b_1 = 0$ a koeficient proto **je statisticky významný**.

Hodnoty koeficientů a jejich směrodatných odchylek lze nalézt přímo ve výstupu funkce LINREGRESE.

Kritickou hodnotu t-rozdělení $t_{1-\frac{\alpha}{2}}$ v EXCELu vypočítáme opět pomocí funkce TINV.

3) **Kvalita modelu jako celku (F-test celého modelu)**

Test:
H_0: všechna $b_i = 0$
H_1: alespoň jedno b_i je různé od nuly

Testové kritérium má mít F-rozdělení o $p-1$ a $n-1$ stupních volnosti (n=počet pozorování; p=počet parametrů regresní funkce – u lineární=2).

H_0 zamítáme v případě, že $F > F_{1-\alpha}$, kde $F_{1-\alpha}$ je kvantil F-rozdělení.

Výpočet v EXCELu:

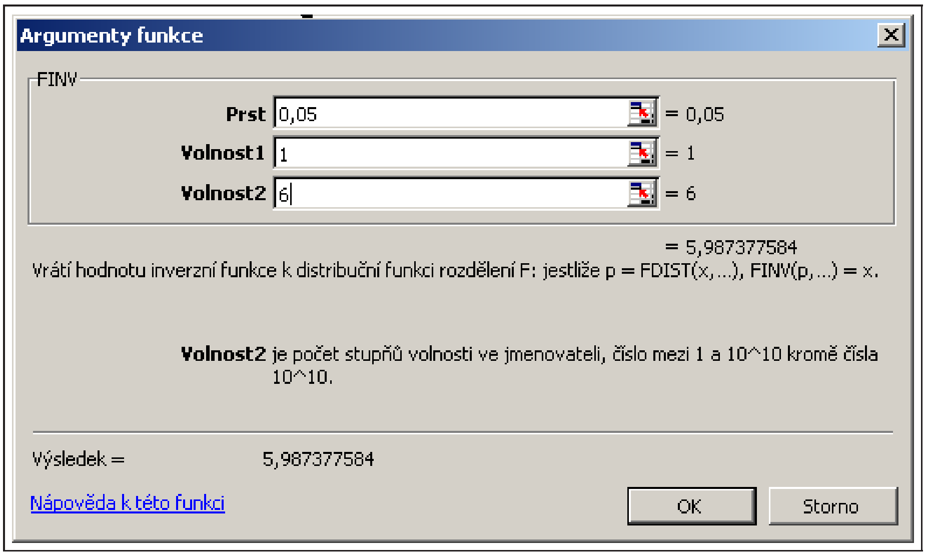
Hodnota testového kritéria F je přímo ve výstupu funkce LINREGRESE (viz příloha předchozí kapitoly). Kritickou hodnotu vypočítáme pomocí funkce FINV.

Prst — hladina významnosti α (0,05)
Volnost 1 — počet stupňů volnosti $p-1$
Volnost 2 — $n-p$ stupních volnosti (n=počet pozorování; p=počet parametrů regresní funkce – u lineární=2).

Výsledek testu:

$$F > F_{1-\alpha} \Longrightarrow 454{,}17 > 5{,}99$$

Zamítáme hypotézu o nulovosti všech (obou) koeficientů $\Longrightarrow$ alespoň jeden z nich je nenulový. Kolik jich je a které to jsou udávají t-testy (V tomto případě je nenulový pouze b_1).

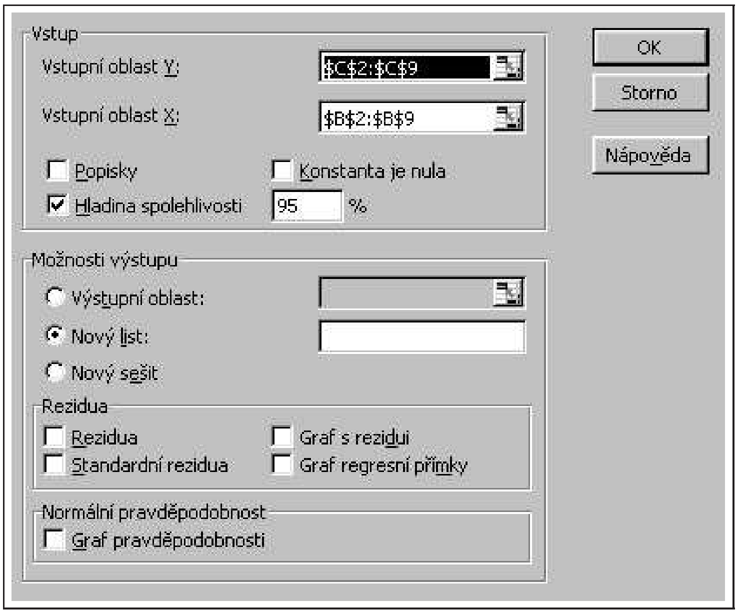


Obrázek 2.3: Kvantil F-rozdělení pomocí funkce FINV

Použití analytického nástroje Regrese v EXCELu

Druhou možností jak k uvedeným výsledkům dojít je využít zabudované procedury EXCELu Regrese. Je ji možno nalézt v nabídce Analytické nástroje/ Analýza dat/Regrese.

Vstupní okno má následující tvar:



Obrázek 2.4: Analytický nástroj Regrese v Excelu

Vstupní oblast Y – označíme buňky, kde jsou hodnoty y (VV), ze kterých chceme vytvořit model
Vstupní oblast X – označíme buňky, kde jsou hodnoty x (HDP), ze kterých chceme vytvořit model

Hladina spolehlivosti – obvykle 95% nebo 99%
Možnosti výstupu – pro přehlednost je vhodné umístit výstup na nový list

Výstup analytického nástroje Regrese pro uvedený příklad je v tabulce 2.1.

VÝSLEDEK

Regresní statistika	
Násobné R	0,993459
Hodnota spolehlivosti R	0,986961
Nastavená hodnota spolehlivosti R	0,984788
Chyba stř. hodnoty	15,86321
Pozorování	8

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	114287,7637	114287,8	454,1692	6,96E-07
Rezidua	6	1509,848228	251,6414		
Celkem	7	115797,6119			

	Koeficienty	*Chyba stř. hodnoty*	*t stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*
Hranice	23,08331	27,9685001	0,825331	0,440757	-45,3533	91,51996
Soubor X 1	0,373505	0,017526199	21,31125	6,96E-07	0,33062	0,41639

Tabulka 2.1: Výstup analytického nástroje Regrese pro příklad 2.1

Významné buňky:

a) **Regresní statistika** – charakteristika intenzity závislosti
Hodnota spolehlivosti R (0,986961) – hodnota indexu determinace R^2
Pozorování – n (využití pro výpočet stupňů volnosti)

b) ANOVA – **test významnosti celého modelu**
Sloupec rozdíl obsahuje v prvních dvou řádcích $p - 1$ a $n - p$
F – je hodnota testového kritéria Významnost F – tzv. p-hodnota statistického testu. Pokud je p-hodnota nižší než hladina významnosti (obvykle 0,05) pak zamítáme hypotézu o nulovosti všech koeficientů a model je statisticky významný.

c) (bez názvu) – test významnosti jednotlivých koeficientů
Koeficienty – hodnota jednotlivých koeficientů – první řádek (Hranice) se týká koef. b_0, druhý řádek (Soubor X1) se týká koef. b_1.
Chyba střední hodnoty – směrodatné odchylky jednotlivých koeficientů
t-stat – hodnoty testových kritérií t-testu významnosti jednotlivých koeficientů
Hodnota p – p-hodnoty pro t-testy (opět platí, že jsou-li nižší než hladina významnosti, koef. je statisticky významný) Dolní 95% a Horní 95% – hranice intervalu spolehlivosti jednotlivých koeficientů (obsahují-li nulu, příslušný koeficient není statisticky významný).

Shrnutí výsledků příkladu:

A. Regresní model:

$$\mathbf{VV = 23{,}08 + 0{,}37\,HDP}$$

tedy $b_0 = 23{,}08,\ s_{b_0} = 27{,}97$
$b_1 = 0{,}37,\ s_{b_1} = 0{,}018$

B. Index determinace

$$R^2 = 0{,}987$$

C. Významnost modelu

a) **jako celku**
Model jako celek je statisticky významný na hladině spolehlivosti 95%.

b) **jednotlivých koeficientů**
Regresní koeficient b_1 je statisticky významný na hladině spolehlivosti 95%. Koeficient b_0 není statisticky významný na hladině spolehlivosti 95% a proto jeho zahrnutí do modelu není vhodné.

Celkově lze konstatovat, že výše veřejných výdajů (VV) velmi těsně závisí na objemu HDP. Tuto závislost je možno charakterizovat číslem 0,37. Tedy veřejné výdaje v ČR tvoří přibližně 0,37 (37%) hrubého domácího produktu.

Pro výše uvedené výsledky však platí velmi výrazné omezení. Oba ukazatele jsme do modelu zahrnuli v běžných cenách. Je proto vysoká pravděpodobnost, že je v modelu zahrnut významný vliv inflace. Pro vyloučení tohoto vlivu by bylo nutno použít ukazatele ve stálých cenách.

Shrnutí kapitoly

Je-li vytvořena regresní funkce, je vhodné zhodnotit její kvalitu a tedy i vypovídající hodnotu. Základní veličinou, která dává rychlou a poměrně významnou informaci je index determinace. Dostáváme-li pro posuzovanou závislost hodnoty indexu determinace blízké jedné, lze (s jistou mírou zjednodušení) hovořit o poměrně silné závislosti mezi uvedenými jevy. Hodnoty indexu blízké nule hovoří spíše o nezávislosti mezi uvedenými veličinami.

Údaj o indexu determinace je obvykle doplňován statistickými testy významnosti jednotlivých regresních parametrů i celého regresního modelu. K těmto účelům se používají standardní statistické testy – t-test pro jednotlivé koeficienty a F-test pro model jako celek.

Teprve souhrn těchto výsledků dává informaci o kvalitě závislosti, kterou jsme popisovali pomocí regresní funkce. Kapitola také prezentuje jednoduché metody, jak uvedená kritéria prakticky spočítat pomocí výpočetní techniky.

Otázky k zamyšlení

1. Odvoďte a zdůvodněte vztah pro index determinace. O čem hovoří jeho hodnota blízká jedné?
2. Vysvětlete jaký je rozdíl mezi výsledky, které o kvalitě regresní funkce podává t-test a F-test.
3. Co je kvantil F-rozdělení a jak jej určíte?

POT1

Součástí studia předmetu je i vypracování a odevzdání dvou krátkých samostatných prací, které jsou označovány jako POT. Obě samostatné práce mají formu příkladu, který by Vám měl dát možnost otestovat vědomosti nabyté v předchozí části studijní opory. Výsledky obou POTů odevzdáte ve stanovených termínech tutorovi v elektronické podobě (soubor v MS EXCEL + případný doprovodný text).

Termíny odevzdání jednotlivých úkolů jsou následující:
POT 1 2. týden v březnu
POT 2 2. týden v dubnu

Odevzdání POTů a jejich správné řešení je podmínkou připuštění ke zkoušce z předmětu.

Zadání POT 1

1. Na základě hodnot ukazatelů, jež jsou uvedeny v následující tabulce, vytvořte regresní funkci (přímku) pro závislost průměrné hrubé mzdy v ČR na hrubém domácím produktu v letech 1993–2000.

	HDP b.c	**průměrná hrubá mzda**
1993	1 020 278	5 817
1994	1 182 784	6 894
1995	1 381 049	8 172
1996	1 566 968	9 676
1997	1 679 921	10 691
1998	1 837 060	11 693
1999	1 887 325	12 666
2000	1 959 585	13 490

2. Znáte-li hodnotu HDP v roce 2001, jež je rovna 2146103 mil. Kč, odhadněte hodnotu průměrné hrubé mzdy v tomto roce.
3. Pro vypočítanou regresní funkci popisující závislost průměrné hrubé mzdy v ČR na hrubém domácím produktu v letech 1993–2000 vypočítejte kritéria její kvality a výsledek zhodnoťte.