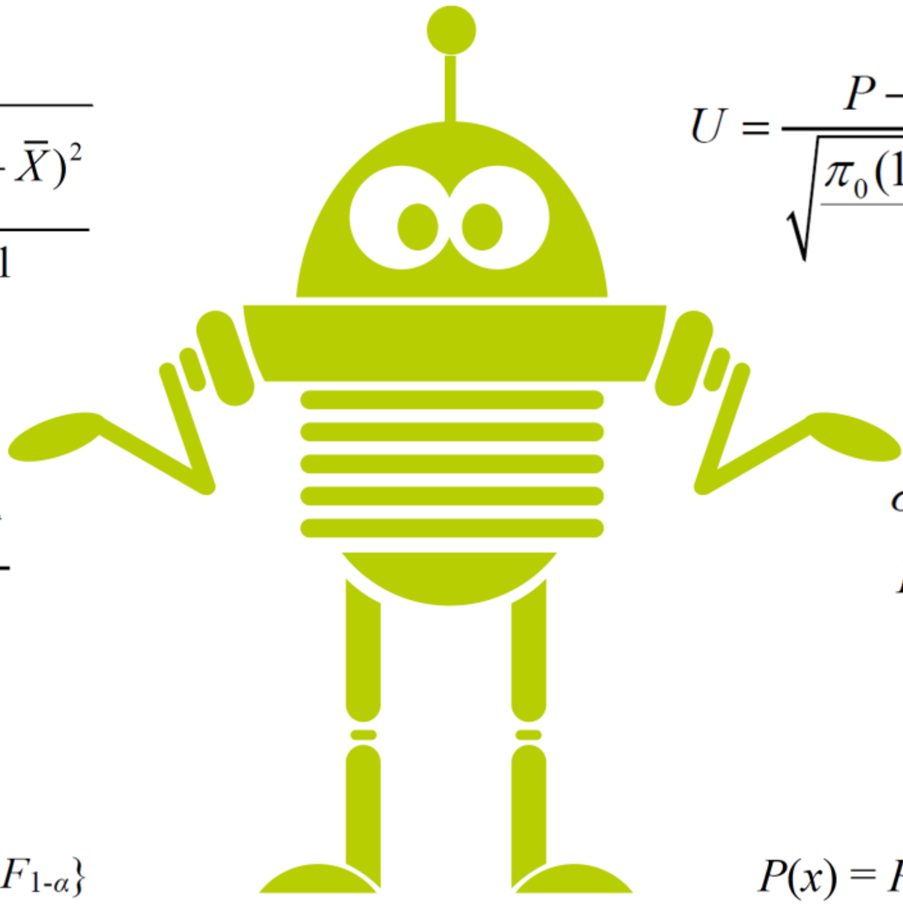
STATISTIKA II

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA

ZBYLÉ UČIVO K ZÁVĚREČNÉ ZKOUŠCE

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

$$U = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$



$$\bar{\Delta} y = \frac{\sum_{t=2}^{T} \Delta y_t}{T-1}$$

$$\sigma = \sqrt{D(X)}$$

$$E(X) = \pi$$

$$W_\alpha = \{F;\ F \geq F_{1-\alpha}\}$$

$$P(x) = P(X = x)$$

EDU FOR LIFE

VZDĚLÁNÍ, KTERÉ SE TI BUDE HODIT

Adriana Řeháčková
www.statistickyneklasicky.cz

STATISTIKA
ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA

Ukážeme si, že statistika není žádná nuda a že má opravdu smysl! Cílem kurzu je, abys dokázal základní až lehce pokročilé statistické metody správně použít při řešení příkladů. Nemine Tě ani osvojení SPSS. Po kurzu bys měl vědět, co, kdy a jak použít a dokázat tyto znalosti využít na reálných datech. Tento kurz slouží jako příprava k závěrečné zkoušce.

OBSAH KURZU:

1) Časová řada - trendová funkce
2) Časové řady - Jednoduché exponenciální vyrovnání
3) Časové řady - Sezónnost
4) Indexní analýza

O TOMTO MATERIÁLU:

Autor materiálu: Adriana Řeháčková
Sazba a grafické úpravy: Adriana Řeháčková

STATISTICKY NEKLASICKY

ČASOVÉ ŘADY

Časová řada - Posloupnost hodnot sledovaného ukazatele, která je jednoznačně uspořádána z hlediska času. Časové řady slouží k popsání a porozumění zákonitostem ekonomických, finančních či jiných skutečností, jenž časovými řadami zachycujeme, zároveň toto porozumění můžeme využít i k budoucím předpovědím.

Dělení: a) Podle rozhodného časového hlediska: **intervalové vs okamžikové**.

Intervalová časová řada obsahuje řadu hodnot sledovaného ukazatele za určitý interval, tj. obsahuje tokové veličiny (př. HDP/rok, průměrný příjem/měsíc, zisk, tržby/rok,měsíc)

Okamžikové časové řada obsahuje řadu hodnot sledovaného ukazatele k určitému okamžiku, tj. stavové veličiny (př. stav bankovnímu účtu k určitému dni, počet zaměstnanců k určitému dni, cena akcie).

b) Podle frekvence s jakou získáváme data: **Dlouhodobé a krátkodobé**
Dlouhodobé - Roční (perioda mezi dvěma po sobě jdoucími záznamy je roční a delší).
Krátkodobé - Čtvrtletní, měsíční, týdenní, denní, atd. (perioda mezi dvěma po sobě jdoucími záznamy je kratší než 1 rok)

ČASOVÉ ŘADY
- u časových řad se pro nezávisle proměnnou místo x používá „t“
- funkce jsou jinak stejné jako pro regresní modely
- když počítáme časové řady, VŽDY se musí nadefinovat, že se o časové řady jedná.

1.Definice datové proměnné
- nejdříve se musí v SPSS nadefinovat, že se jedná o časovou řadu
DATA -> DEFINE DATES
-> zde se zvolí, zda máme časové řady roční, měsíční, apod.

2. Grafické zobrazení časové řady
ANALYZE -> FORECASTING -> SEQUENCE CHARTS

3. Výpočet a interpretace 1. diference
TRANSFORM -> CREATE TIME SERIES

4. Vyrovnání klouzavým průměrem
TRANSFORM -> CREATE TIME SERIES a následně Centered moving average

5. Vyrovnání lineární trendovou funkcí
ANALYZE -> REGRESSION -> CURVE ESTIMATION

INTERPOLACE A EXTRAPOLACE ČŘ

-interpolace

Přibližné určení chybějící hodnoty sledovaného ukazatele uvnitř ČŘ za předpokladu, že známe sousední hodnoty 1. prostřednictvím dvou sousedních hodnot (pomocí aritmetického průměru těchto hodnot * průměrný koeficient růstu)
2. prostřednictvím všech hodnot v ČŘ (pomocí trendové funkce)

-extrapolace

Určení hodnot ČŘ za interval známých hodnot, z pravidla do budoucnosti
1. statické prognózování – pomocí trendových funkcí a sezónních indexů odhadujeme budoucí úroveň ukazatele. Uvedený postup však naráží na předpoklad neměnnosti dosavadního vývoje Uvedený nedostatek (neměnnost) odstraňují metody adaptivního prognózování a ARIMA.

Autokorelace - Hodnoty ukazatele v řadě se vzájemně ovlivňují, jedna má vliv na druhou.
Durbin-Watsonův koeficient autokorelace - Keficientu se pohybuje v rozmezí < 0 , 4 >.
Pokud je tato statistika rovna číslu 2, rezidua **nevykazují žádnou autokorelaci (to chceme),** hodnoty D menší než 2 značí pozitivní autokorelaci a hodnoty větší než 2 značí autokorelaci negativní.

Volba vhodného modelu trendu (míra shody)
1. základem pro rozhodování o vhodném typu trendové funkce jsou věcně ekonomická kritéria
2. druhou možností volby je analýza grafu časové řady (nebezpečí: vizuální výběr funkce může být subjektivní a závisí na měřítku!)
3. elementární charakteristiky časové řady
4. Pomocí indexu determinace
5. Pomocí chyb odhadu (MSE, ME, MAE,...)
6. Pomocí reziduální směrodatné odchylky

Máte k dispozici časovou řadu počtu přínosných společenských inovací v podniku Elona Muska v letech 2009-2018.

rok	Yt
2009	12
2010	13
2011	17
2012	18
2013	22
2014	21
2015	28
2016	26
2017	33
2018	32

a) Vypočtěte lineární trendovou funkci.
b) Určete index korelace a determinace a dále hodnotu MAPE.
c) Určete odhad počtu inovací v roce 2020.

1) Známe počty pracovníků určitého podniku vždy k 1. dni měsíce, a to: k 1.1. - 178 zaměstnanců, k 1.2. - 170, k 1.3. - 180, k 1.4. - 150 zaměstnanců a k 1.5. - 162 zaměstnanců.
Spočtěte průměrný stav zaměstnanců za období od 1.1. do 1.5.

Datum	Počet zaměstnanců	Průměr	Délka intervalu	součin
1.1.	178	-	-	-
1.2.	170	174	31	5394
1.3.	180	175	28	4900
1.4.	150	165	31	5115
1.5.	162	156	30	4680
	-	**-**	**120**	**20089**

Metody exponenciálního vyrovnávání

Jednoduché exponenciální vyrovnávání: trend v krátkých úsecích konstantní, jeden parametr α.

Brownovo exponenciální vyrovnávání: úroveň a trend řady, předpoklad lineárního trendu, jeden parametr α .

Holtovo exponenciální vyrovnávání: úroveň a trend řady, dva parametry α, γ.

Exponenciální vyrovnání: s tlumeným trendem tři parametry α, γ, φ.

1. Vyrovnejte pomocí modelu jednoduchého exponenciálního vyrovnání časovou řadu, která sleduje počty koloběžek v Praze v letech 2009 až 2019.
2. Interpretujte a posuďte hodnotu vyrovnávací konstanty.
3. Odhadněte počet koloběžek v roce 2020.

Rok	Počet koloběžek
2009	5600
2010	14500
2011	19800
2012	34500
2013	33400
2014	56000
2015	63000
2016	61000
2017	91300
2018	93230
2019	89300

Počet koloběžek
14600
14500
15200
15340
14870
15867
14569
16345
13780
17450
16509

ČASOVÉ ŘADY – EXP. VYROVNÁVÁNÍ

Na základě modelu jednoduchého exponenciálního vyrovnání dopočítejte chybějící údaje v tabulce. Vycházejte z předpokladu, že hodnota vyrovnávací konstanty je $\alpha = 0{,}6$. Výchozí hodnotu časové řady (hodnotu pro rok 2010) vypočítejte z prvních 4 pozorování (aritmetický průměr počtu vykouřených cigaret za roky 2010 až 2013). Následně predikujte, jaký byl počet vykouřených cigaret v roce 2020.

rok	Počet vykouřených cigaret	Vyr. hodnoty
2010	8009	
2011	7690	
2012	6890	
2013	7120	
2014	5870	
2015	7206	
2016	6300	
2017	5900	
2018	5578	
2019	7035	

1. Nejdříve se nadefinuje časová řada
-> DATA -> DEFINE DATES
-> zde se zvolí, zda máme časové řady roční, měsíční, hodinové, apod.
2. Grafické zobrazení
- v grafu je dobře vidět sezónní kolísání
-> ANALYZE -> FORECASTING -> SEQUENCE CHARTS
-> do Variables vložíme proměnnou
-> do Time Axis Labels vložíme nově vytvořenou nominální proměnnou
- v Outputu vidíme graf
3. Dekompozice časové řady
- charakteristika časové řady pomocí sezónních odchylek a výpočet klouzavých průměrů
-> tímto očistíme časovou řadu od sezonního prvku a budeme moc vyjádřit trendovou tendenci modelu (trendovou funkci)
-> ANALYZE -> FORECASTING -> SEASONAL DECOMPOSITION
-> do Variables vložíme zkoumanou proměnnou
-> zvolíme v Model Type aditivní (additive) nebo multiplikativní (multiplicative) model, podle toho jestli chceme odchylky nebo indexy
-> dále navolíme klouzavé průměry v Moving Average Weight, podle toho jaký máme počet období, za které se vypočítává klouzavý průměr (3 roky, 4 čtvrtletí, apod.)
-> v Outputu vyjede tabulka sezónních indexů (odchylek) a vidíme shrnuté např. čtvrtletí, kde lze vidět v jakém čtvrtletí to, jak klesá / stoupá
-> v Data view se zobrazí čtyři nové proměnné; nás zajímá pouze sloupek SAS a SAF
SAS – vyrovnané hodnoty na základě klouzavých průměrů -> očištěná časová řada
SAF – sezónní odchylka
4. Odhadnout parametry lineární trendové funkce
- dělá se na základě očištěné časové řady (ta se očistí pomocí klouzavých průměrů)
-> ANALYZE -> REGRESSION -> CURVE ESTIMATION
-> do Dependent vložíme očištěné hodnoty (hodnoty klouzavých průměrů)
-> v Independent zvolíme Time a dáme, že chceme lineární funkci
- v Outputu nám poté vyjedou informace k trendové funkci (podobně jako u regrese)
5. Proveďte predikce (odhad) na 1. a 2. čtvrtletí roku xxxx
-> ANALYZE -> REGRESSION -> CURVE ESTIMATION
-> zde dáme Save a v něm zvolíme Predicted values
-> zadáme, do jakého roku chceme predikci udělat a také do jakého čtvrtletí
- výsledek se zobrazí v Data view pod původními hodnotami

ČASOVÉ ŘADY – SEZÓNNOST

V tabulce jsou uvedeny čtvrtletní údaje o počtu studentů na doučování statistiky za období od prvního čtvrtletí 2017 do čtvrtého čtvrtletí 2019.

	1.čtvrtletí	2.čtvrtletí	3. čtvrtletí	4.čtvrtletí
rok 2017	43	124	167	198
rok 2018	201	245	340	420
rok 2019	501	560	608	674
rok 2020	610	690	700	730

1. Definujte datové proměnné.
2. Graficky zobrazte časovou řadu.
3. Dekompozice časové řady - charakteristika sezónního kolísání pomocí sezónních odchylek
4. Na základě sezónně očištěné časové řady odhadněte parametry lineární trendové funkce.
5. Proveďte predikci na 1. a 2. čtvrtletí roku 2021 zohledněte trendovou i sezónní složku)

ČASOVÉ ŘADY – SEZÓNNOST

Je dána čtvrtletní časová řada počtu motorových člunů, které vyplují z nejmenovaného přístavu. Jsou uvedeny čtvrtletní údaje od 1.čtvrtletí 2008. Soubor **vodni_doprava.sav.** Proveďte následující operace.

1. Definice datové proměnné.
2. Grafické zobrazení časové řady.
3. Dekompozice časové řady - charakteristika sezónního kolísání pomocí sezónních indexů
4. Na základě sezónně očištěné časové řady (proměnná SAS_1) odhadněte parametry nejvhodnější trendové funkce.
5. Vypočítejte predikci na všechny čtvrtletí rok 2020 (zohledněte trendovou i sezónní složku)

Individuální indexy (indexy stejnorodých ukazatelů):
a) jednoduché - Zkoumáme jeden výrobek na jedné pobočce
b) složené - Zkoumáme jeden výrobek na více pobočkách (za více poboček dohromady)

Souhrnné indexy (indexy nestejnorodých ukazatelů):
a) Paascheho – fixuje druhou veličinu v běžném období
b) Laspayresův – fixuje druhou veličinu v základním období
c) Fishrerův – geometrický průměr Paascheho a Laspayresova

Za produkci podniku, který vyrábí dva druhy automobilů, máte tyto údaje za rok 2017 a 2018:

Automobil	Hodnota výroby v mil. Kč		Produkce v tis. ks	
	2017	2018	2017	2018
Peugeot 407	100	145	15	23
Peugeot 308	120	105	32	38

Řešení:

1) Jak se změnil fyzický objem výroby podniku ve sledovaném období měřený Laspeyresovým indexem?
a) Zvýšil se o 0,1345 %
b) Zvýšil se o více než 50 %.
c) Zvýšil se o 34,5 %.
d) Zvýšil se o 3,45 %

2) Jak se změnil fyzický objem výroby podniku ve sledovaném období měřený Paascheho indexem?
a) Zvýšil se o 3,66 %.
b) Zvýšil se o 0,1366 %.
c) Zvýšil se o více než 50 %.
d) Zvýšil se o více % než v případě měření Laspayresovým indexem

3) Jak se změnil fyzický objem výroby podniku ve sledovaném období měřený Fisherovým indexem?
a) Zvýšil se méně než v případě měření Paascheho indexem.
b) Zvýšil se méně než v případě měření Laspayresovým indexem.
c) Zvýšil se o více než 50 %.
d) Zvýšil se o 3,24 %.

V následující tabulce jsou uvedeny údaje o počtu online systémů ve firmě Statistickyneklasicky, která působí na třech školách.

Školy	Počet systémů		Celková produkce – počet prodaných řešení	
	leden	březen	leden	březen
ČZU	12	15	120	134
VSE	8	20	127	130
UK	4	12	57	86

a) Charakterizujte, jak se v relativním a absolutním vyjádření změnila celková produktivita práce ve firmě v březnu oproti lednu.

b) Charakterizujte, jak byla změna celkové produktivity práce v podniku, ovlivněna změnou počtu systémů v jednotlivých školách.

c) Charakterizujte, jak byla změna celkové produktivity práce v podniku, ovlivněna změnou produktivity práce, v jednotlivých školách.