# „Co neumím změřit, neumím řídit“

**Úvod:**

Hlavní role statistiky, ukázky využití a použití statistiky v běžném životě – sociální sítě, televize. Statistika jako nástroj marketingu, ekonomiky, politiky. Jak, čím a co měřit – validita, reliabilita, objektivita

**Obsah:**

***Sázíte-li ve Sportce, je to hazard.***

***Sázíte-li se, že vám v kartách přijdou tři postupky po sobě, je to zábava.***

***Vsadíte-li se, že cena plynu stoupne o 10 %, je to podnikání.***

***Vidíte ten rozdíl?***

Předmět předkládá možnosti využití základního matematicko-statistického aparátu pro potřeby běženého studenta na Masarykově univerzitě. Možnost statistického zpracování dat je stále velmi dynamicky se rozvíjející disciplína. S rostoucí dostupností výkonné měřící a výpočetní techniky se statistické metody začaly ve stále větším měřítku prosazovat v běžném životě. Smyslem předloženého materiálu je porozumět mechanismu, na jehož základě jsou základní statistické výpočty prováděny. Přiměřená znalost statistiky pak pomůže studentům lépe chápat zákonitosti naměřených dat. Aplikovat statistické metody a postupy znamená zaznamenávat data o jevech a zpracovávat je, tj. třídit, vyhodnocovat a interpretovat. Statistika se tak nachází v úzkém kontaktu s informačními technologiemi (informatika, výpočetní technika).

1. příprava výzkumného šetření je nejdůležitější část
2. sběr a analýza dat slouží k zamítnutí/nezamítnutí předem stanovených úkolů práce a hypotéz (explorační vs. konfirmační přístup)
3. vždy mít na paměti věcné hledisko výzkumu, zejména v souvislosti s interpretací statistických výsledků

**Role statistiky**

* Porozumění a zkoumání hromadných jevů
* Zjišťování zákonitostí
* V kvantitativním výzkumu (deduktivní princip) – pojítko mezi teorií a výzkumem
* Zpracování, popsání a analyzování dat

Statistický sw na Masarykově univerzitě:

Multilicence dostupné v systému inet.muni.cz

* IBM SPSS, verze 29
	+ https://acrea.cz/software/ibm-spss-statistics/
* TIBCO Statistica, verze 14
	+ https://www.statistica.pro/
* Matlab R2022a
	+ https://www.humusoft.cz/
* Prostředí R
	+ https://www.r-project.cz/

**Co je statistika?**

Statistika je metoda analýzy dat, která nachází široké uplatnění v celé řadě odvětví, oblast sportu, tělesné výchovy a kinantropologie nevyjímaje. Její význam s rozvojem výpočetní techniky a specializovaných software roste, což umožňuje urychlení a zkvalitnění při sběru a přenosu dat a také při zpracování a ukládání informací.

Role statistiky je nezastupitelná, neboť nepřetržité vyhodnocování informací o celku i jeho částech dává důležité informace použitelné pro další rozhodovací procesy použitelné v běžné práce vysokoškolského pracovníka, studenta, managementu fakulty.

Přiměřená znalost základních statistických pojmů pomáhá porozumět odborným textům, kde je statistiky v hojné míře obsažena.

Aplikovat statistické metody a postupy znamená zaznamenávat data o jevech a zpracovávat je, tj. třídit, vyhodnocovat a interpretovat. Statistika se tak nachází v úzkém kontaktu s informačními technologie (informatika, výpočetní technika).

**Co je popisná statistika**

* Popisná statistika je aplikovaná matematická disciplína, která se zabývá popisem a zpracováváním informací obsažených v datech (tabulky, grafy, charakteristiky). Využívá základních matematických operací, které mohou být následné i složitější (statistické postupy). Úkolem a cílem popisné statistiky je najít informace „ukryté´” v datech. Popisná statistika už všichni používáme (každý zná charakteristiky jako průměr, směrodatná odchylka, histogram, sloupcový graf aj.)

Procedury popisné statistiky použijeme k prvotnímu posouzení předložených dat. Nejčastěji používané statistické charakteristiky jsou

1. aritmetický průměr 

Definice následujících charakteristik předpokládají uspořádaný výběr, tj. 

1. minimální hodnota  = 
2. maximální hodnota = 
3. medián  pro n sudé ,

pro n liché 

1. dolní kvartil  =, kde pro pořadový index *k* platí

 n .0,25 < k < n . 0,25 + 1

1. horní kvartil  =, kde pro pořadový index *k* platí

 n .0,75 < k < n . 0,75 + 1

**Charakteristiky variability**

1. variační rozpětí R = xmax - xmin
2. kvartilové rozpětí 
3. výběrový rozptyl 
4. výběrová směrodatná odchylka 
5. variační koeficient  nebo 

**Charakteristiky kategoriální proměnné**

* *Modus* - hodnota nejčetnější kategorie
* *Četnost*- počet pozorování spadajících do příslušné kategorie
* Stanovení četností – absolutní a relativní

**Příklad 1**

|  |  |
| --- | --- |
| **Pozice** | **Měsíční plat (v tis. Kč)** |
| Ředitel | 100 |
| Dělník | 10 |
| Dělník | 10 |
| Dělník | 10 |
| Dělník | 10 |
| Dělník | 10 |
| Dělník | 10 |
| Dělník | 10 |
| Dělník | 10 |
| Dělník | 10 |
| Dělník | 10 |

Kolik je průměrný plat v tomto podniku?

**Příklad 2**

Máme k dispozici údaje o 1411 respondentech a to konkrétně proměnnou BMI (Body Mass Index). Vypočítáme základní statistické charakteristiky (tab. 1).



Tab. 1 Výsledky popisné statistiky

|  | **BMI** | Popis |
| --- | --- | --- |
| **N platných** | 1411 | počet hodnot |
| **Aritmetický****průměr** | 25,04 | * statistická veličina, která v jistém smyslu vyjadřuje typickou hodnotu popisující soubor mnoha hodnot
* nejčastější chybou je aplikace aritmetického průměru tam, kde je na místě využít jinou statistiku. Např. aritmetickým průměrem souboru { 1, 1, 1, 1, 16 } je 4, přestože čtyři z pěti hodnot tohoto souboru je menších. V obdobných případech je mnohem vhodnější použít pro vyjádření typické hodnoty medián (který je u této množiny roven 1, což je mnohem lepší popis střední hodnoty)
 |
| **Minimum** | 16,03 | * nejmenší hodnota
 |
| **Maximum** | 181,61 | * nejvyšší hodnota
 |
| **Medián** | 24,35 | * medián (označován Me nebo) je hodnota, jež dělí řadu podle velikosti seřazených výsledků na dvě stejně početné poloviny.
* není ovlivněný extrémními hodnotami.
* medián lze definovat na každém souboru uspořádaném relací „menší nebo rovno“, i když se nejedná o soubor čísel. Například medián souboru {absolvent ZŠ, vyučen, vyučen s maturitou, vysokoškolák} je roven hodnotě „vyučen“, pokud kategorie vzdělání považujeme za seřazené podle náročnosti školy.
 |
| **Spodní kvartil** | 22,13 | * kvartily oddělují ze statistického souboru čtvrtiny. Rozlišuje se spodní kvartil Q0,25 a horní kvartil Q0,75. Data předpokládají uspořádaný výběr.
 |
| **Horní kvartil** | 27,17 |
| **Rozpětí** | 165,58 | * rozdíl mezi maximem a minimem
 |
| **Kvartilové** **rozpětí** | 5,04 | * pomocí horního a spodního kvartilu lze zavést mezikvartilové rozpětí, které definujeme jako hodnotu Q0,75 − Q0,25.
 |
| **Rozptyl** | 32,16 | * rozptyl - jedná se o charakteristiku variability rozdělení pravděpodobnosti náhodné veličiny, která vyjadřuje variabilitu rozdělení souboru kolem střední hodnoty.
 |
| **Směrodatná** **odchylka** | 5,67 | * jedná se o kvadratický průměr odchylek hodnot znaku od jejich aritmetického průměru. Vypovídá o tom, jak moc se od sebe navzájem liší typické případy v souboru zkoumaných čísel. Je-li malá, jsou si prvky souboru většinou navzájem podobné, a naopak velká směrodatná odchylka signalizuje velké vzájemné odlišnosti.
 |
| **Variační** **koeficient** | 22,65 | * variační koeficient je použitelný i při porovnávání variability proměnných, které jsou v různých jednotkách
 |

**Debata:**

* Je tady něco podezřelého?
* Co?
* Proč?
* Jak?
* Co s tím?

**Shrnutí:**

Hlavní role statistiky, ukázky využití a použití statistiky v běžném životě – sociální sítě, televize. Statistika jako nástroj marketingu, ekonomiky, politiky. Jak, čím a co měřit – validita, reliabilita, objektivita

**Domácí úkol:**

* Jak změřit lásku? Navrhněte jakékoliv způsoby na měření lásky. Nabízíme definici: „Láska označuje **silný pozitivní vztah náklonnosti, oddanosti nebo touhy, který se může lišit jak svojí povahou, tak také předmětem**. Někdy převažuje volní, jindy citová nebo sexuální stránka vztahu.“ Můžete přijít i se svou definici, která umožní měřit ☺
* Vezměte svá data a spočítejte (jakkoliv) základní statistické charakteristiky. Výsledky vložte do odevzdávárny předmětu.

**Otázky k procvičení:**

1. Uveďte rozdíl mezi aritmetickým průměrem a mediánem.
	1. Vyjadřují střední hodnotu, liší se způsobem výpočtu
	2. Hodnoty jsou většinou shodné
	3. Rozdíl mezi hodnotami mediánu a průměru je obvykle významný na 5 % hladině statistické významnosti
2. Uveďte nevýhody aritmetického průměru
	* OTEVŘENÁ OTÁZKA
	* Není-li uvedena směrodatná odchylka, ztrácím informaci o variabilitě dat
	* Nevyjadřuje vždy „přibližnou“ střední hodnotu a to v případě vypočítání z dat obsahující extrémní hodnoty
3. Uveďte výhody mediánu
	* OTEVŘENÁ OTÁZKA
	* Je robustní vůči extrémním hodnotám
4. Uveďte vztah mezi kvartilem, kvantilem a mediánem
	* OTEVŘENÁ OTÁZKA
	* Druhý kvartil je 50. kvantil a to je medián
5. Jaký je rozdíl mezi rozptylem a směrodatnou odchylkou?
	1. Rozptyl je směrodatná odchylka „na druhou“
	2. Rozptyl je dvojnásobek směrodatné odchylky
	3. Rozptyl nijak nesouvisí se směrodatnou odchylkou
6. K čemu slouží variační koeficient?
	* OTEVŘENÁ OTÁZKA
	* Uvádí míru relativního rozptýlení dat