

Úvod do teorie měření

opora pro studium

Informace o předmětu

Základní informace

Název předmětu: **Úvod do teorie měření**

Forma výuky: cvičení

Ověření studijních výsledků: zápočet

Anotace

Cílem kurzu je seznámit studenty se základními principy statistického uvažování a s jejich aplikací. Cvičení jsou zaměřena na aplikace statistických metod a na správnou interpretaci výsledků statistických výpočtů. Absolvent kurzu bude schopen zpracovat naměřená data a správně interpretovat získané výsledky. Na cvičeních bude využíván především Microsoft Excel.

Sylabus

- 1.–2. Základní pojmy popisné statistiky, míry polohy, variability, šikmosti, špičatosti
3. Skupinové rozdělení četností, histogram
4. Základní pravděpodobnostní rozdělení
5. Bodové odhady pro parametry a jejich vlastnosti (konzistence, vydatnost)
6. Intervaly spolehlivosti (pro parametry normálního rozdělení a parametry jiných rozdělení)
- 7.–11. Statistické testy (principy, F -test a t -testy, testy pro parametry nenormálních rozdělení, chí-kvadrát test a další neparametrické testy)
12. Korelace, regrese
13. Chyby měření, klasifikace chyb, vyjadřování náhodných a systematických chyb
14. Zápočtová písemná práce

Zdroje pro studium

Základní

- HEJNOVÁ, Eva. *Fyzikální praktikum A* [online]. Dostupné na: http://physics.ujep.cz/~ehejnova/PraktikumA/Fyz_praktikum_A.pdf
- HOLČÍK, Jiří, KOMENDA, Martin (eds.) a kol. *Matematická biologie: e-learningová učebnice* [online]. 1. vyd. Brno: Masarykova univerzita, 2015. ISBN 978-80-210-8095-9. Dostupné na: <http://portal.matematickabiologie.cz/index.php?pg=aplikovana-analyza-klinickyh-a-biologickyh-dat--biostatistika-pro-\matematickou-biologii>

- LITSCHMANNOVÁ, Martina. *Úvod do statistiky*. Ostrava: Vysoká škola báňská – Technická univerzita Ostrava, 2011. Dostupné na: <http://mi21.vsb.cz/modul/uvod-do-statistiky>

Doporučené

- NOVÁK, Rudolf. *Úvod do teorie měření*. Ústí nad Labem: Univerzita J. E. Purkyně v Ústí n. L., 2003. ISBN 80-7044-527-0.

Programové vybavení

V rámci tohoto předmětu budou využívány tyto programy:

- Excel (desktopová varianta, ideálně verze 2013 a vyšší)
- online kalkulačky testů na webu <https://www.statskingdom.com/>

Týden 1 a 2 – Popisná statistika

Základní pojmy

- ✓ statistický soubor (základní soubor, výběrový soubor), statistická jednotka, rozsah souboru
- ✓ metrická/kategoriální veličina
- ✓ ordinální/nominální veličina, spojitá/diskrétní veličina
- ✓ míry polohy: průměr, modus, medián
- ✓ míry variability: rozptyl, směrodatnou odchylku, rozpětí, kvartilové rozpětí
- ✓ další charakteristiky: minimum, maximum, kvartily (kvantily, percentily)
- ✓ krabicový graf
- ✓ vylučování odlehlých hodnot metodou vnitřních hradeb a pomocí Grubbsova testu

Pokyny pro studium

- 1) V knize *Základy statistiky pro biomedicínské obory* se na stranách 21–25 seznámte se základními pojmy používanými ve statistice.
- 2) Následně si na webu *Portál Matematická biologie (typy dat)* projděte základní typy dat, se kterými se ve statistice pracuje.
- 3) Dále si v knize *Základy statistiky pro biomedicínské obory* (strany 70–80) prostudujte základní charakteristiky, které můžeme pro danou veličinu určovat. Rozmyslete si, pro které typy veličin můžeme uvedené charakteristiky určit. Látku z knihy dále doplňte na webech *Portál Matematická biologie* (Popis a vizualizace kvantitativních dat), a *Portál Matematická biologie* (Popis a vizualizace kvalitativních dat), kde se seznámíte s možnostmi grafického znázornění různých druhů dat.
- 4) Naučte se v Excelu počítat jednotlivé charakteristiky (využijte nápovědu k Excelu) a zkonstruovat krabicový graf (návod je zde).
- 5) Na webu *Portál Matematická biologie (Vylučování odlehlých hodnot)* se seznámte s podstatou a důvody pro vylučování odlehlých hodnot v datech. V rámci tohoto kurzu byste se měli naučit používat tzv. metodu vnitřních hradeb (popis metody najdete zde na stranách 36–39) a Grubbsův test (jeho popis je v materiálech pro cvičení na stranách 4 a 5).

Úlohy na procvičení

- 1) Do výzkumu bylo zahrnuto 38 čeledí pavouků žijících v ČR. Specifikujte, co je v tomto případě statistickým souborem, statistickou jednotkou, jaký je rozsah souboru. Určete pět metrických a 5 kategoriálních veličin, které můžete u tohoto statistického souboru zkoumat.

Řešení: Statistický soubor: čeledi pavouků žijící v ČR, statistická jednotka: jedna čeleď, rozsah souboru: 38.

- 2) O následujících veličinách rozhodněte, zda jsou metrické, kategoriální, nominální, nebo ordinální: rasa psa, výška v kohoutku koně, pohlaví narozeného dítěte, barva květu dané rostliny, počet mláďat v jednom vrhu u daného zvířete, kontinent, kde se vyskytuje daný druh rostliny, hodnota systolického tlaku, míra znečištění (v měřítku: malá – střední – velká), krevní skupina, rod dané rostliny.

Řešení: K(N), M(S), K(N), K(N), M(D), K(N), M, K(O), K(N), K(N).

- 3) Určete základní statistické charakteristiky pro veličinu BMI ze souboru DATADETI. Tyto charakteristiky vypočtete pro celý soubor, pouze pro dívky a pouze pro chlapce. Výsledky porovnejte a interpretujte. Pro dívky a pro chlapce sestrojte krabicové grafy, porovnejte je mezi sebou a interpretujte rozdíly.

- 4) Metodou vnitřních hradeb určete odlehlé hodnoty u veličiny SKOKD ze souboru DATADETI. Data znázorněte pomocí krabicového grafu.

Řešení: Odlehlými hodnotami jsou 239, 248 a 250.

- 5) Při rozboru křemičitanu byl nalezen tento obsah SiO_2 : 52,44 %, 53,82 %, 52,91 %, 50,10 %, 54,03 %, 53,89 %. Je některý z výsledků odlehlý na hladině významnosti $\alpha = 0,05$? Použijte Grubbsův test.

Řešení: Odlehlou hodnotou je 50,10 %.

Týden 3 – Skupinové rozdělení četností

Základní pojmy

- ✓ skupinové rozdělení četností
- ✓ Sturgesovo pravidlo
- ✓ absolutní četnost, relativní četnost
- ✓ kumulovaná četnost
- ✓ histogram
- ✓ distribuční funkce

Pokyny pro studium

- V knize *Základy statistiky pro biomedicínské obory* se na stranách 81 a 82 se seznámte s pojmem skupinové rozdělení četností. Pro doplnění použijte první část prezentace *Popisná statistika*, kde si všimněte různých postupů, jak určit počet intervalů skupinového rozdělení četnosti spojité veličiny.
- V Excelu se naučte zkonstruovat histogram (návod najdete tady – pro účely kurzu je vhodné používat konstrukci pomocí doplňku Analytické nástroje).
- Zběžně si projděte definici distribuční funkce na webu *Portál Matematická biologie* (Distribuční funkce). V Excelu se naučte zkonstruovat tzv. empirickou distribuční funkci a odečítat z ní hodnoty kvantilů. Videonávod pro konstrukci je zde.

Úlohy na procvičení

- 1) Určete základní statistické charakteristiky pro veličinu VYSKA ze souboru DATADETI. Hodnoty této veličiny rozdělte do 5 intervalů a určete absolutní a relativní četnosti, kumulované absolutní a kumulované relativní četnosti. Sestrojte histogram.
- 2) Proveďte totéž, co v bodě 1, počet intervalů ale určete podle Sturgesova pravidla.
Řešení: Dle Sturgesova pravidla pracujeme s 9 intervaly.
- 3) Pro veličinu VYSKA ze souboru DATADETI sestrojte distribuční funkci. Kolik procent dětí je menších než 157 cm? Jakou výšku musí dítě mít, aby se pohybovalo na úrovni 85% percentilu?
Řešení: 30 %, 174 cm.

Týden 4 – Základní pravděpodobnostní rozdělení

Základní pojmy

- ✓ disktrétní veličina, spojitá veličina
- ✓ obecné disktrétní rozdělení, alternativní a binomické rozdělení
- ✓ normální rozdělení

Pokyny pro studium

- Zopakujte si pojmy spojité a disktrétní veličiny a následně prostudujte strany 57–66 v knize *Základy statistiky pro biomedicínské obory*. Návazně si projděte další informace k normálnímu rozdělení na webu *Portál Matematická biologie* (Normální rozdělení pravděpodobnosti). Pro porozumění další látce postačuje znát „tvary“ hustoty a distribuční funkce normálního rozdělení a jejich význam (totéž u ostatních rozdělení).
- Zběžně se seznamte se základními rozděleními disktrétní náhodné veličiny – věnujte se pouze rovnoměrnému, alternativnímu s binomickému rozdělení. Ke studiu použijte materiály na webu *Rozdělení disktrétní náhodné veličiny*.
- Naučte se používat Generátor pseudonáhodných čísel v Excelu, návod najdete tady.

Úlohy na procvičení

- 1) O následujících veličinách rozhodněte, zda jsou disktrétní nebo spojité: výška rostliny, počet dětí, délka chemické reakce (v minutách), oxidační číslo, délka ukazováčku na pravé ruce.
Řešení: S, D, D, D, S.
- 2) Modelujte 10, 1000 a 1000 hodů hrací kostkou (použijte Generátor pseudonáhodných čísel, kdy pravděpodobnost je pro všechny počty ok na kostce stejná, $\frac{1}{6}$). Pro každý soubor vytvořte histogram a tabulku relativních četností. Porovnejte se skutečnými pravděpodobnostmi (Pro jaký počet hodů se sobě pravděpodobnosti nejlépe podobají? Dokážete vysvětlit proč?).
- 3) Modelujte výšku 1000 náhodně vybraných osob z populace, kde výška je z normálního rozdělení se střední hodnotou 175 a směrodatnou odchylkou 10. Vytvořte histogram, distribuční funkci a spočítejte základní statistické charakteristiky. Porovnejte je se střední hodnotou a směrodatnou odchylkou výšky základního souboru.

Týden 5 a 6 – Bodové a intervalové odhady

Základní pojmy

- ✓ bodový odhad, intervalový odhad
- ✓ interval spolehlivosti oboustranný, jednostranný
- ✓ interval spolehlivosti pro střední hodnotu μ a rozptyl σ^2 normálního rozdělení
- ✓ interval spolehlivosti pro parametr π binomického rozdělení

Pokyny pro studium

- Prostudujte si význam a konstrukci bodových a intervalových odhadů v knize *Základy statistiky pro biomedicínské obory* na stranách 86–92. Alternativně můžete využít tento a tento text.

Úlohy na procvičení

- 1) Vygenerujte základní soubor (25 veličin z normálního rozdělení s $\mu = 175$, $\sigma^2 = 100$, rozsah souboru 1000). Pro každý řádek (výběrový soubor) určete průměr, rozptyl a směrodatnou odchylku. Pro každý řádek určete interval spolehlivosti pro průměr a rozptyl a zjistěte, v kolika procentech případů pokrývají jednotlivé intervaly spolehlivosti hodnoty $\mu = 175$, $\sigma^2 = 100$. Experimentujte s různými intervaly spolehlivosti (90%, 95%, 99%).
- 2) Byla měřena délka trvání určitého procesu. Z 12 měření byla zjištěna střední doba trvání procesu 44 s a směrodatná odchylka 4 s. Sestrojte 90% a 95% interval spolehlivosti pro očekávanou délku procesu za předpokladu normálního rozdělení.
Řešení: 90% interval spolehlivosti je (41,83; 46,17), 95% interval spolehlivosti je (41,35; 46,65).
- 3) Předpokládejme, že mezi 200 dotazovanými (náhodně vybranými z dané populace) je 26 leváků. Sestrojte 95% interval spolehlivosti pro podíl leváků v celé dané populaci.
Řešení: Zkoumáme alternativní rozdělení s parametrem $p = 0,13$. Interval spolehlivosti je (0,08; 0,18).

Týden 7 – Princip testování hypotéz

Základní pojmy

- ✓ nulová hypotéza, alternativní hypotéza
- ✓ hladina významnosti
- ✓ testové kritérium, kritický obor
- ✓ p -hodnota

Pokyny pro studium

- Projděte si tento a tento text, které se věnují testování statistických hypotéz. Následně si prostudujte úvod do testování hypotéz na webu *Portál Matematická biologie* (Úvod do testování hypotéz), věnujte se především p -hodnotě a její interpretaci.

Úlohy na procvičení

- 1) Ke stanovení obsahu pomocné látky trimethoprimu v léčivu Biseptol byla užitá metoda diferenční pulzní polarografie. K ověření byl analyzován standard se známou koncentrací trimethoprimu $57,66 \mu\text{g} \cdot \text{l}^{-1}$. Pomocí intervalu spolehlivosti proveďte test správnosti devíti opakovaných měření. Test proveďte na hladině významnosti $\alpha = 0,05$.

Koncentrace trimethoprimu [$\mu\text{g} \cdot \text{l}^{-1}$] v biseptolu: 57,64, 56,11, 56,60, 56,23, 57,63, 56,31, 57,64, 54,51, 55,00.

Návod: Pro zadaná data určete interval spolehlivosti, opět půjde o interval spolehlivosti pro střední hodnotu normálního rozdělení. Dále rozhodněte, zda hodnota 57,66 v intervalu spolehlivosti leží, nebo ne. O výsledku rozhodněte analogicky s úlohou o spotřebě automobilu. Nezapomeňte formulovat nulovou a alternativní hypotézu.

Řešení: Interval spolehlivosti je (55,54, 57,28).

Týden 8 – chí-kvadrát test

Základní pojmy

- ✓ chí kvadrát-test na kontingenční tabulce

Pokyny pro studium

- V knize *Základy statistiky pro biomedicínské obory* se na stranách 130–139 seznámte s použitím chí-kvadrát testu pro účely zjišťování souvislosti mezi dvěma kategoriálními veličinami. Alternativně můžete použít i tento text.
- Naučte se provádět chí-kvadrát test v Excelu, pro studium můžete použít toto video.

Úlohy na procvičení

- 1) V populaci byly zjištěny počty praváků a leváků. Na základě tabulky rozhodněte, zda je praváctví/leváctví závislé na pohlaví. Testujte na hladině významnosti $\alpha = 0,05$.

	Praváci	Leváci	Celkem
Muži	43	9	52
Ženy	44	4	48
Celkem	87	13	100

Řešení: Používáme chí-kvadrát test na kontingenční tabulce; pohlaví nemá vliv na levorukost/pravorukost, testové kritérium je rovno 1,78, kritickým oborem je interval $(3,84; \infty)$.

Týden 9 a 10 – F -test a t -testy

Základní pojmy

- ✓ F -test
- ✓ dvouvýběrový t -test s rovností/nerovností rozptylů
- ✓ párový t -test

Pokyny pro studium

- V knize *Základy statistiky pro biomedicínské obory* se na stranách str. 99–120 seznamte s použitím F -testu a t -testů pro účely vyhodnocení souvislostí mezi kategoriální a metrickou veličinou. Pro studium můžete využít i tento text.
- Naučte se provádět F -test a t -test v Excelu (pomocí dopňku Analýza dat), návod najdete v tomto videu.

Úlohy na procvičení

- 1) Na skupině dobrovolníků byl testován prostředek na snížení váhy. Hmotnosti 12 testovaných lidí před a po dietní kúře jsou v tabulce. Rozhodněte na hladině významnosti 0,05 zda je prostředek účinný.

hm. před dietou (kg)	85	75	90	65	150	80	110	56	88	73	67	134
hm. po dietě (kg)	76	75	81	64	155	72	99	45	89	66	56	110

Řešení: Používáme párový t -test; prostředek je účinný, $p = 0,008$.

- 2) Byl měřen obsah vápníku v krevním séru skupiny zdravých lidí a skupiny nemocných. Naměřené hodnoty jsou v tabulce. Porovnejte na hladině významnosti 0,05 obsahy vápníku obou skupin, tj. určete, zda se obě skupiny od sebe statisticky významně liší.

zdraví lidé (mmol/l)	2,15	2,13	2,27	2,52	2,11	2,26	2,34	2,68	2,24
nemocní lidé (mmol/l)	2,09	1,8	1,97	2,35	2,08	1,9	2,06	2,3	2,36

Řešení: Používáme t -test s rovností rozptylů; obsah vápníku je u zdravých lidí vyšší než u nemocných, $p = 0,045$.

Týden 11 – Korelace, regrese

Základní pojmy

- ✓ závislé/nezávislé veličiny, závislost pozitivní/negativní, lineární závislost
- ✓ Pearsonův korelační koeficient
- ✓ regresní model, regresní přímka

Pokyny pro studium

- Na webu *Portál Matematická biologie* (Korelační koeficient) se seznamte s definicí korelačního koeficientu a jeho geometrickou interpretací. Věnujte se pouze Pearsonovu korelačnímu koeficientu, vynechat můžete i interval spolehlivosti pro korelační koeficient. Alternativně můžete ke studiu použít i tento text.
- V Excelu se naučte zkonstruovat bodový graf pro zadaná data, vypočítat Pearsonův korelační koeficient a daty v bodovém grafu proložit regresní přímkou. Pro studium využijte toto video.

Úlohy na procvičení

- 1) Na základě tabulky rozhodněte, zda existuje lineární statistická závislost mezi teplotou vzduchu a počtem prodaných kopečků zmrzliny v daný den. Pomocí regresní přímky odhadněte, kolik kopečků zmrzliny prodáme, pokud bude teplota vzduchu 18 °C.

Den	Teplota ve °C	Prodaných kopečků zmrzliny
Pondělí	20	200
Úterý	19	218
Středa	12	141
Čtvrtek	10	99
Pátek	9	85
Sobota	22	210
Neděle	23	211
Pondělí	19	170
Úterý	22	200
Středa	12	110
Čtvrtek	13	131
Pátek	14	141
Sobota	15	152

Řešení: Korelační koeficient je roven 0,95, jde o pozitivní závislost; prodáme 176 kopečků zmrzliny.

Týden 12 – Další testy

Základní pojmy

- ✓ analýza rozptylu (jednofaktorová ANOVA)

Pokyny pro studium

- V knize *Základy statistiky pro biomedicínské obory* se na stranách 158–166 seznámte s podstatou a provedením analýzy rozptylu. Pro hlubší pochopení můžete použít i tuto prezentaci.
- Po prostudování materiálů byste měli být schopni provést nejen samotnou analýzu rozptylu, ale ověřit i její předpoklady – pro tyto účely použijte např. kalkulačku na webu <https://www.statskingdom.com/180Anova1way.html>.

Úlohy na procvičení

- 1) Byly zkoumány výsledky testu z chemie u třech skupin studentů, přičemž v každé skupině probíhala výuka chemie jinak. Na hladině významnosti 0,05 rozhodněte, zda průměrné výsledky testu jsou u všech skupin stejné. Ověřte předpoklady analýzy rozptylu a v případě potřeby proveďte dílčí testy.

Metoda 1	Metoda 2	Metoda 3
89	104	86
101	120	98
87	98	100
87	110	96

Řešení: Používáme analýzu rozptylu; průměrné výsledky testů nejsou u všech skupin stejné, $p = 0,027$. Dle dílčích testů je nejlepší Metoda 2.

Týden 13 – Chyby měření

Základní pojmy

- ✓ systematická/náhodná chyba měření
- ✓ absolutní/relativní chyba měření
- ✓ vyjadřování náhodných a systematických chyb

Pokyny pro studium

- Nastudujte si prvních sedm stran ve skriptu Fyzicální praktikum A (E. Hejnová). Zaměřte se především na rozdíl mezi absolutní a relativní chybou měření a správný zápis výsledků měření. Pro rozšíření některých témat můžete využít kapitolu 3 Základy teorie měření z knihy *Úvod do teorie měření* (R. Novák).

Úlohy na procvičení

- 1) Projděte si tuto prezentaci a samostatně odpovězte na v ní uvedené otázky.