**Statistická analýza dat s podporou SPSS (APS300166)**

Vyučující: Petr Boschek, Marek Vranka

Čas a místo: UT 18:20 – 19:50, ST 15:00 – 16:30, místnost 119/C

Tento volitelný předmět má za cíl seznámit studenty se statistickým programem IBM SPSS (příprava a transformace datových souborů, použití základních statistických procedur). Nebudeme předpokládat znalosti o tomto programu. Podmínka pro atestaci: v průběhu kurzu budou studenti samostatně prezentovat zadané příklady statistické analýzy.

Datové soubory pro výuku budou umístěny na http://dl1.cuni.cz/course/view.php?id=3254, kam má každý student přístup (i bez přihlašování se).

Harmonogram:

 1. (08.04.) Seznámení se s programem SPSS 19, pracovní prostředí a hlavní ovládací prvky, data sheet, variable sheet, definování proměnných, import dat z formátu xls

 2. (09.04.) Syntaxový editor , deskriptivní statistika

 3. (15.04.) Transformace dat, t-testy a pořadové testy

 4. (16.04.) Regresní a korelační analýza

 5. (22.04.) Analýza kontingečních tabulek

 6. (23.04.) Psychometrická analýza testu (dotazníku)

 7. (29.04.) Základní modely analýzy rozptylu

 8. (30.04.) Prezentační grafy, testy shody s modelem

 9. (06.05.) Rezerva - opakování

10. (07.05.) Závěrečný test

**STATISTICKÁ ANALÝZA – SPSS (příklady a soubory)**

**t-test pro dva nezávislé výběry**

datový soubor: PŘIJÍMAČKY

1. Vypočtěte hrubé skóry (HS) subtestů „BIOLOGIE“ a „PKV“ z „nula-jedničkových“ položek (1=správně, 0=nesprávně). pozor: Biologii si zvolili pouze někteří !
2. Prezentujte tabulku distribuce HS a uveďte 90% kvantil distribuce
3. Testujte normalitu distribuce HS subtestu „BIOLOGIE“
4. Testujte hypotézu μ = 19 , tj. je test stejně obtížný jako loni (jako v originále) ?
5. Testujte pomocí t-testu shodu průměrů obou subtestů
	1. (muži vers. ženy) !! Je nutno překódovat z části rodného čísla !!
	2. (prezenční vers. kombinovaní)
	3. (forma A vers. Forma B) tedy analýzu stejné obtížnosti
6. Statistickou analýzu doplňte Box-plot diagramem
7. Výsledky analýzy, jako součást nějakého textu, převeďte do formátu WORD.

# Neparametrické testy pro nezávislé výběry

datový soubor: PŘIJÍMAČKY

1. Zkontrolujte případné chybné údaje v položkách Vypočtěte hrubé skóry (HS) subtestů „PSYCHOLOGIE“ a „NUMERICKÉ VZTAHY“ z „nula-jedničkových“ položek (1=správně, 0=nesprávně), přičemž správná odpověď je u posledních pěti položek hodnocena dvěma body !!!!!!
2. Testujte normalitu distribuce HS subtestu „NUM: VZTAHY“
3. Testujte shodu distribuce HS obou subtestů u mužů a žen pomocí Mann-Whitneyova (=Wilcoxonova) testu a Kolmogorov-Smirnovova testu.
4. Vytvořte nový znak - vážený součet z-skórů obou testů: 0,7\*Z(PSY) + 0,3\*Z(NV). Vypočtěte základní deskriptivní charakteristiky a převeďte tento skór na T-skór
5. Statistickou analýzu doplňte Box-plot diagramem
6. Výsledky analýzy, jako součást nějakého textu, převeďte do formátu WORD

**Regresní a korelační analýza:**

Datový soubor: REGRESE

Analyzujte pomocí lineárního regresního modelu (vícerozměrného) validitu použité testové baterie assesment centra (test\_1, test\_2, test\_3). K dispozici je kriterium „krit“ vyjadřující „konečnou diagnózu“ - objektivní hodnocení uchazečů po určité době.

1. Převeďte data z formátu EXCEL do formátu SPSS
2. Aplikujte lineární regresní model: prediktory - regresory (test\_1, test\_2, test\_3).

záv. proměnná: KRIT

1. Nejdříve zvolte jako regresory všechny tři testy a potom pomocí postupnoé regrese vyberte optimální podmnožinu regresorů (optimalizace modelu).
2. Pro obě metody napište predikční rovnici, testujte signifikantnost regresorů, přesnost celého modelu, uložte predikované hodnoty kriteria, ukažte, že mnohorozměrný koeficient korelce je vlastně „obyčejná resp. párová“ korelace predikovaných a observovaných hodnot.
3. Vyberte nejsilnější prediktor a pomocí regresního grafu prezentujte interval spolehlivosti pro predikovanou hodnotu konkrétní osoby.
4. Výsledky analýzy, jako součást nějakého textu, převeďte do formátu WORD

**Analýza závislosti kategoriálních dat (kontingenční tabulka)**

Datový soubor: SOCIOLOGIE

1. Analyzujte závislost (cca 3) vybraných proměnných (CHOVÁNÍ, HODNOTY, …,které vás zajímají) na
	1. PROFESI
	2. jedné vybrané demografické proměnné
	3. věkových kategoriích, které vytvoříte „rozumně“
2. Použitá metoda: chí-kvadrát test nezávislosti + adjustované residuály pro detailnější analýzu.
3. Pozor:
	1. na očekávané četnosti u χ2 – testu, asi bude nutné sdružit kategorie
	2. v souboru nejsou ošetřeny chybějící hodnoty („neodpověděl“, „neví“)
4. Výsledky analýzy, jako součást nějakého textu, převeďte do formátu WORD

# Psychometrická analýza testu

datový soubor: PŘIJÍMAČKY (vyberte položky libovolného testu)

1. Položková analýza:

obtížnost položek, korelace položek s HS testu (diskriminační síla)

1. Seřaďte položky dle obtížnosti (do tabulky).
2. Analýza reliability testu: Cronbachův α-koeficient a split-half model
3. Komentujte výsledky analýzy
4. Vyberte 4 položky a pomocí Mc-Nemarova testu testujte rozdíl v obtížnosti (vyberte je na základě koeficientů obtížnosti, abyste dostali cit pro „už rozdílnou obtížnost“)
5. Výsledky analýzy, jako součást nějakého textu, převeďte do formátu WORD

# Analýza variance (dvoufaktorový model) vybereme nějaký subtest

Datový soubor: WISC (Wechslerův test inteligence)

1. Analýza se týká pouze dětí 10-letých, 11-letých a 12-letých
2. Vyberte dva subtesty (HS) a testujte pomocí analýzy variance rozdíl průměrného výkonu kluků a holek. Druhým faktorem (jehož efekt chceme eliminovat) je pochopitelně věk (celé roky bez měsíců)
3. Vyberte vhodný graf (z procedury ANOVA) pro ilustraci výsledků
4. Výsledky analýzy, jako součást nějakého textu, převeďte do formátu WORD

**Testy shody s modelem (jednovýběrový t-test, Kolmogorov-Smirnovův test a chí-kvadrát test dobré shody)**

Datový soubor: Interkultural

1. Škála „A“ je T- skór českého převodu škály agresivity dle původních finských norem (převod HS na T-skóry). Ověřte, zda tyto normy (převodní relaci HS → T-skór) lze nadále udržet i pro českou populaci, nebo zda je nutné vytvořit nový „český“ převod.
2. Škály (resp. tzv. komponenty) „volnost“ a „kladný vztah“ v dotazníku stylu výchovy (Čáp) byly standardizované v roce 1991 - definované tertily, které definují intervalové třídění – tři ekvipravděpodobnostní kategorie: nízký, střední a vysoký. Ověřte testem dobré shody, zda původní ekvipravděpodobnostní intervalové třídění odpovídá i dnes