V těch zápiskách hodně referencuju “zkouška” a PAS ultimate. To jsou taháky, které jsem napsal (s pomocí dalších lidí) minulý rok na boďovo zkoušku. Kdyby je někdo potřeboval stačí napsat

* @palacinka

## Okruh I: praktické znalosti popisné statistiky

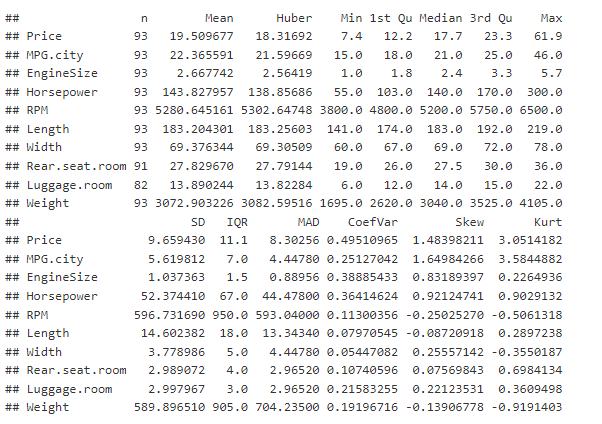
Za správné zodpovězení student získá **maximálně 30 bodů.**

Student si vylosuje jednu proměnnou z určitého datasetu. Pro tuto proměnnou pak vypočítá základní popisné statistiky, nakreslí základní grafy a výsledky bude interpretovat.

* U číselné proměnné se jedná o popisné statistiky polohy a variability, histogram a boxplot.
* U kategorické proměnné se jedná o absolutní a relativni četnosti, sloupcový a koláčový graf

-> huberův odhad

-> nechce hledání modusu

-> Všechny popisné statistiky co jsou na zápočtu na začátku: 

-> U koláčového a bar plotu vyčíst co tam vidíme za informace

-> pro kategorickou proměnnou udělat frekvenční graf

-> říct kdy dává smysl uspořádat hodnoty

## Okruh II: kombinace praktických a teoretických znalosti vybraných statistických metod

Za správné zodpovězení student získá **maximálně 35 bodů.**

Student si vylosuje jedno z následujících témat, jehož praktický výpočet předvede na vybrané proměnné / vybraných proměnných z databáze a prokáže i znalost teoretických vlastností daného tématu (postup výpočtu a jeho odůvodnění, interpretaci možných výsledků atd.)

**Témata okruhu II:**

* bodové a intervalové odhady střední hodnoty a rozdílu středních hodnot
  + spočítat
  + vzorečky a vysvětlit je co tam je proč tam je
  + najít ve vzorci pro odhad střední hodnoty variabilitu - vychází z binomického rozdělení
  + rozdíl mezi bodovým a intervalovým odhadem
  + jejich výhody a nevýhody
  + kdy se používají
* bodové a intervalové odhady podílu a rozdílu podílů
  + spočítat
  + vzorečky a vysvětlit je co tam je proč tam je
  + najít ve vzorci pro odhad střední hodnoty variabilitu - vychází z binomického rozdělení
  + rozdíl mezi bodovým a intervalovým odhadem
  + jejich výhody a nevýhody
  + kdy se používají
* testování statistických hypotéz v základních kontextech
  + v základních kontextech = nevíme je to boďovina
  + Jak obecně sestavit test a příklad = vyberu proměnnou a budu ověřovat tvrzení co si vymyslím (normalita -> správný test -> interpretace)
  + jak fungují hypotézy a proč se dělají
  + ideálně dělat test normality, korelační a jednovýběrový
  + vysvětlit jejich předpoklady
  + vyhodnotit p hodnotu správně
  + co je to hladina významnosti ?
  + co je to p hodnota ? - boďova definice z prezentace
* hodnocení vzájemné souvislosti dvou číselných proměnných (tvar, směr, sila)
  + korelační koeficient - jak vypadá pro sinusoidu nebo pro nelinearni zavislost (ty zvláštní obrazky z bodovo prezentace)
  + korelační matice
  + korelační tabulka a (kontigenční tabulka - kategorický
  + z čeho se to počítá a proč
  + co je to kovariance
  + možné testy ?
  + prakticky využít
* regresní přímka (rovnice regresní přímky)
  + spočítat
  + popsat co jsem spočítal
  + odvodit
  + metoda nejmenších čtverců
  + interpretacee koeficientů
  + významnost koeficientů - p hodnota
* identifikace vhodného podkladového rozdělení dat
  + odhadování rozdělení
  + kontrola pomocí grafů
  + je to ve cvičení z loňska i letos (lognorm, log, atd.)
* hodnocení normality a tvaru rozdělení
  + testy
  + qq plot, histogram
  + špičatost šikmost
  + jak má vypadat normální rozdělení
  + normování - převod na symetrii, aby se rozdělení blížili gaussovému
  + když mám hodně hodnot blížím se k normálnímu rozdělení průměrem
  + základní teorie o normálním rozdělení
* identifikace odlehlých hodnot
  + boxplot, histogram
  + zešikmené rozdělení - v něm najít (musím transformovat jak to vypadá, musí to jít do symetrie abych mohl hledat outliery)
  + rezistence/robustnost
  + co je to odlehlé pozorování

-> vzorce (interval spol.), předpoklady a jak se počítají daný věci (variabilita - je potřeba nějaký počet pozorování atd.)

-> Co to je

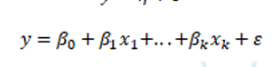
, k čemu to je, proč to je, s čím to souvisí?

## Okruh III: teoretické znalosti vybraných statistických pojmů

Za správné zodpovězení student získá **maximálně 35 bodů.**

Student si vylosuje jedno z následujících témat, u nějž prokáže teoretické znalosti.

**Témata okruhu III:**

* klasifikace proměnných a typů dat
  + Dělení proměnných - kateg./čísel., časové řady, průřez. data,.... + převádění mezi proměnnými
  + v čem se liší a proč to rozlišujeme - je to k něčemu v praxi? - mám jiný popisný statistiky, jiný grafy, atd…, modus je jinak definovany pro spojity a pro diskretni
* rozdělení náhodné veličiny
  + Definice hustoty, distribuční funkce, pravděpodobnostní funkce
  + střední hodnota rozptyl, mám to definovaný vždycky? musí být stejný/různý? co to vypovídá
* spojité náhodné veličiny
  + jakými funkcemi jsou definovaný (hustota, pravděpob.)
  + příklady - rozdělení + proměnnou
  + Nakreslit obrázek grafů rozdělení aspon dva příklady
  + čím je definované, dva parametry, jak se mění tvar hustoty když měním parametry - měnit rozptyl např.
* diskrétní náhodné veličiny
  + to samé jako u spojité
* tradiční versus robustní přístupy k odhadování
  + kdy se co používá, na čem to závisí, příklady - průměr/median
  + huberuv odhad
  + uskenuty prumer
  + nemusíme robustní odhad šikmosti a špičatosti !!!
  + pro střední hodnotu a variabilitu hlavně
  + co je to robustní proč to máme k čemu to máme
* bodový versus intervalový odhad
  + ve spojitosti z okuhu 2
  + jak spočítám jaké má vlastnosti
* tradiční versus bootstrapový přístup k statistické inferenci
  + co je to statistická inference,
  + **Bootstrap** - výhody nevýhody proti klasickýmu, jak se počítá
  + jak funguje ten tradiční
  + co čekat když mám klasický/bootstr odhad intervalu spolehlivosti -> Bootstrap bude vždycky trochu jiný, např.
* zákon velkých čísel a jeho využití, centrální limitní věta a její využití
  + napsat ty věty a k čemu se používají + co to znamená jednoduše
* přístupy k testování statistických hypotéz
  + testová statistika - jak pomocí ní vyhodnotit test, jak pomocí p hodnoty, interval spolehlivosti - o 1 konkrétním parametru
  + statisticky významný, hladina významnosti
  + chyba 1. a 2. druhu
* interpretační problémy a aspekty intervalového odhadu a p-hodnoty, kovariance a korelace
  + statistická vs věcná významnost
  + když mám hodně/málo dat co vidím a neviídm
  + když mám hodně dat p hodnota
  + když málo je to zkreslený
  + je možný aby vyšla korelační koeficient 0.6 a p hodnota 0.3 - jde to proti sobě, je to možné? kdy?
* jádrový odhad hustoty a modus
  + to co je ve “zkouška” - pas ultimate stačí
  + co v tom vyčteš když to vidíš
* populace, náhodný a nenáhodný výběr, populační a výběrové charakteristiky
  + co je to populace, náhodný a nenáhodný výběr, populační a výběrové charakteristiky
  + k čemu se používají výběrové a že jsou odhad populačních
  + reprezentativní výběr
  + velikost výběru
  + jak se to sbírá - celá populace, ne-náhodný výběr, náhodný výběr, experimentální data
* frekvenční rozdělení a frekvenční křivka
  + u diskrétní a spojitý se to zobrazuje jinak, jinak to, co je ve “zkouška”
  + souvisí s jadrovým odhadem
* histogram a jeho citlivost na volbu offsetu a šířky okna
  + vše v “zkouška”
  + tvary histogramů(histogram normálního rozdělení), kdy se to používá a na co, jaké proměnné
  + šikmost špičatost, jde vidět an symm a ne na asym, jak vypadá
* vlastnosti popisných statistik, jejich reakce na posunutí a změnu měřítka
  + střední hodnota, rozptyl jejich reakce
  + a+b\*x
  + šikmost špičatost
  + popisný statistiky a odhad - co odhadují (mimo to že popisují soubor)
* normování proměnné a význam
  + standardizace, ne přiblížení k normálnímu rozdělení
  + z-skóre
  + výpočty
  + jednotkový rozptyl
  + ze všech veličin udělám standardizací něco co se chová stejně
  + k výpočtu šikmosti a špičatosti
  + potřebuju z toho vycházet kvůli intervalu spolehlivosti, proměnná se musí nejdřív standardizovat - proč, musel bych používat jiný kvantil vždycky
  + (pas ultimate má tohle špatně)
* regresni model, jeho účel a odhad
  + bodový graf, prokládáme křivku odhaduje funkční předpis lineární regrese
  + jednoduchý lineární regresní model - hlavně vzorec
  + 
  + ten vzoreček má jen to b0 + b1x1 + e (nechci hledat řecký písmenka :-) )
  + metoda nejmenších čtverců
  + vysvětlit jednotlivé členy ve všech vzorečkách
  + proč to používám a co to počítá,
  + co by se stalo kdybych neměl na druhou ve vzorci metody nejmenších čtverců
  + koeficient determinace
* předpoklady lineární regrese
  + **Lineární vztah** = Vztah mezi X a Y je lineární
  + o **Normalita chyb** = Chybové termíny �� jsou normálně rozdělené
  + o **Homoskedasticita** = Chyby mají konstantní směrodatnou odchylku
  + o **Nezávislost chyb** = Chyby jsou nezávislé
  + žádný špatný outliery - mimo přímku linearni regrese
  + stabilita rozptylu, nesmí být do trychtýře
  + má to biďa v prezentaci nějaký vše
  + nezávislost měření
  + pokud je víc proměnných nesmí být korelovaný u nejednoduchý lineární regrese