

Projekt z MSP

Logo VUT! VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ Fakulta informačních technologií Číslo
zadání: 21, 6 Cvičení - skupina: pátek, 9.00

Jakub Zárybnický, xzaryb00

6. 12. 2019

a	b	c
---	---	---

(a) caption

d	e	f
---	---	---

(b) caption

(setq python-shell-interpreter "/nix/store/gpnm7i19lpj8p43mjrdw03d0hjalmskl-python3-3.7.5/bin/python")

1 Zadání projektu z předmětu MSP

Každý student obdrží na cvičení konkrétní data (čísla ze seznamu), pro které vypracuje projekt.

K vypracování můžete použít libovolné statistické programy.

1. Při kontrole výrobků byla sledována odchylka X [mm] jejich rozměru od požadované velikosti. Naměřené hodnoty tvoří statistický soubor v listu Data př. 1.

- (a) Proveďte roztržidění statistického souboru, vytvořte tabulku četností a nakreslete histogramy pro relativní četnosti a relativní kumulativní četnosti.
- (b) Vypočítejte aritmetický průměr, medián, modus, rozptyl a směrodatnou odchylku.
- (c) Vypočítejte bodové odhady střední hodnoty, rozptylu a směrodatné odchylky.
- (d) Testujte předpoklad o výběru z normálního rozdělení Pearsonovým (chí-kvadrát) testem na hladině významnosti 0,05.
- (e) Za předpokladu (bez ohledu na výsledek části d)), že statistický soubor byl získán náhodným výběrem z normálního rozdělení, určete intervalové odhady střední hodnoty, rozptylu a směrodatné odchylky se spolehlivostí 0,95 a 0,99.
- (f) Testujte hypotézu optimálního seřazení stroje, tj. že střední hodnota odchylky je nulová, proti dvoustranné alternativní hypotéze, že střední hodnota odchylky je různá od nuly, a to na hladině významnosti 0,05.
- (g) Ověřte statistickým testem na hladině významnosti 0,05, zda seřazení stroje ovlivnilo kvalitu výroby, víte-li, že výše uvedený statistický soubor 50-ti hodnot vznikl spojením dvou dílčích statistických souborů tak, že po naměření prvních 20-ti hodnot bylo provedeno nové seřazení stroje a pak bylo naměřeno zbývajících 30 hodnot.

Návod: Oba soubory zpracujte neroztržiděné. Testujte nejprve rovnost rozptylů odchylek před a po seřazení stroje. Podle výsledku pak zvolte vhodný postup pro testování rovnosti středních hodnot odchylek před a po seřazení stroje.

2. Měřením dvojice (Výška[cm], Váha[kg]) u vybraných studentů z FIT byl získán dvourozměrný statistický soubor zapsaný po dvojicích v řádcích v listu Data př. 2.

- (a) Vypočítejte bodový odhad koeficientu korelace.
- (b) Na hladině významnosti 0,05 testujte hypotézu, že náhodné veličiny Výška a Váha jsou lineárně nezávislé.
- (c) **Regresní analýza** - data proložte přímkou: $Vaha = \beta_0 + \beta_1 \times Vyska$
 - i. Bodově odhadněte β_0 , β_1 a rozptyl s_2 .

ii. Na hladině významnosti 0,05 otestujte hypotézy:

$$H : \beta_0 = -100, H_A : \beta_0 \neq -100,$$

$$H : \beta_1 = 1, H_A : \beta_1 \neq 1,$$

iii. Vytvořte graf bodů spolu s regresní přímkou a pásem spolehlivosti pro individuální hodnotu výšky.

Termín pro odevzdání práce je 11 týden výuky zimního semestru ve cvičení.

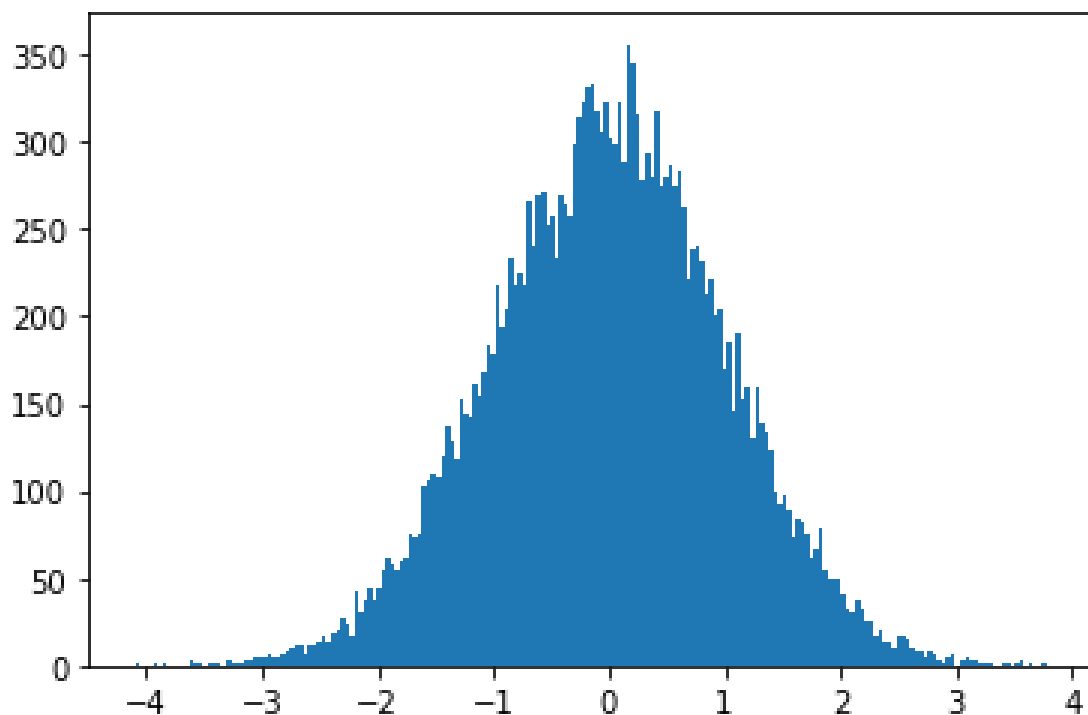
a 1 2
b 2 3
c 3 4

(x, data)

Out[17]:

: (2, [['a', 1, 2], ['b', 2, 3], ['c', 3, 4]])

_ = plt.hist(np.random.randn(20000), bins=200)



2 Příklad 1

Při kontrole výrobků byla sledována odchylka X [mm] jejich rozměru od požadované velikosti. Naměřené hodnoty tvoří statistický soubor v listu Data př. 1.

X [mm] 1,83 0,98 -0,09 -0,23 2,56 0,31 1,06 0,01 0,75 2,26 -0,59 0,90 1,66 0,36 2,19 1,24 -0,58 0,79 0,02 0,31
1,61 0,75 2,46 0,86 0,63 -0,98 -0,75 2,67 1,79 1,84 0,49 1,68 0,39 -0,84 1,49 1,50 1,70 3,40 1,40 0,27 0,48 0,27 1,41
0,55 1,20 -0,68 1,59 0,80 1,21 -1,31

3 Příklad 2

Měřením dvojice (Výška[cm], Váha[kg]) u vybraných studentů z FIT byl získán dvourozměrný statistický soubor zapsaný po dvojicích v řádcích v listu Data př. 2.

Výška [cm] Váha [kg] 150 50 177 73 154 53 152 44 169 69 200 94 196 99 181 74 152 50 172 74 152 58 150 46 178 78 154 57 190 90 195 98 182 80 184 88 156 42 154 66