Statistické minimum – příklady zkouškových otázek.

1. Definujte pojem *střední hodnota* náhodné veličiny a *aritmetický průměr*. Vysvětlete rozdíl mezi nimi.

EX=

- (

x*. f*(**x)** *dx*

**hustota pst** náhodné veličiny

-*(X1*, X2,*..., Xn)*

IX

↓

s Xi

E **je** brána z celé dist.

n

je jen z podskuping

**vetsinou** odhadujeme

مسلم

**E**

2. Uvažujte náhodný vektor **X.** Definujte *kovarianční* a *korelační matici.* Jaké mají tyto matice vlastnosti?

K čemu se dají použít?

Σ=COT (X,Y) = E((X*-EX) (*Y-EY)]

Lo Atvercova, symetrická

posit. **semidef.**"

corr(x,t) =

ها

CON *(*X,Y)

**Jvar X √rary'**

Cov(X*,X;)*

E-(") + he

**N**

(-1,1)

Is my normovaná kovariance

**los** corr

=

(^^)

**na**

corr(Xi, X;)

diagonale vozpryly **jednotlivých** složch ↳ cor (X, Y) ~**"**mira" lineární závislosti

corr(x,t**)**

**normolary** ~**rzsah**

mez veličinami

cor (*X,X)* = ror(x) cor (x,t) = cr(X,X)

**TA** CON (A**, B)**

&=

X=(A) **(cor (1,**A) To

Is hovelační hoe fivierry **pro** urdení **závislosti**

3. Co je to *distribuční funkce*? Co je to *kvantil* a *kvantilová funkce*? Co je to *funkce pravděpodobnostní*

*hustoty?* Definujte)pojmy formálně a uveďte jaký je mezi nimi vztah.

Fx(x) = P(x=x)

**P(Xsx**) ↑

s **jakol pst**...

**nah**. vel.

osa **X**

krantilová furlice

X

Ginverzní k Fx (x)

x= F(y) **y**= F(x)

hodnota luahtilové funkce

P

kvantil =

Qx = F" (p)

**s** 20% por **jaká**

bude **hodnota**

**یا**

hustota **pst f(x) =** f(x)-hustota

**pst z** náhodná **veli** haby de **hodholy t**

dane ho invarly

**rozdělen pst**.

4. Vysvětlete vlastními slovy význam *p-hodnoty* statistického testu. Jaký je výklad *hladiny významnosti* a? Předpokládejte, že p-hodnota nějakého testu je 0.045? Jaká je pravděpodobnost, že nulová hypotéza tohoto testu neplatí? Jak to souvisí s a?

(p-value = **pst**., že hodnota statistické ho testu bude taková nebo extrémnější než **mi vyšlo 2.p.** že Ho platí

<< = pot, **že** zamítnes *Ho***, ikdyž** platí → chci Ho zamítnout jeu 5%

LP (X/HO)

P(Ho IX)-> ha toto se

↳ weddražeć určit

x**=**

·0,05 0,045-0,05

·ptají

Ho zamítám

Te správně

to co se Snažím vyrátis

X

5. Vysvětlete pojem *interval spolehlivosti* (confidence interval). Co nejpřesněji popište, jak byste spočítal/a

99% interval spolehlivosti odhadu střední hodnoty normálního rozdělení z malého vzorku velikosti m.

**m-2**

1-2/22. krandil

SE(μ); û+

**99% =** 100 (1-α)

6. Vysvětlete význam pojmů *chyba prvního druhu* a *chyba druhého druhu* používaných k popsání konkrétních

chyb v procesu testování statistických hypotéz. Vysvětlete jak spolu tyto chyby souvisí. Popište, jak se dá výskyt těchto chyb ovlivnit.

7. Formulujte *centrální limitní větu*. Kde se dá využít? pro hodně vzorků

průměry lib. **náhodných** veličin mají Norm

8. Motivujte zavedení *Studentova* t-*rozdělení.* Ke kterému rozdělení se t-rozdělení asymptoticky blíží s rostoucím počtem stupňů volnosti? Vysvětlete za jakých okolností a jak se od tohoto rozdělení odlišuje. K čemu se používá?

**vhodné** pro **malý** počet vzorků

**blíží** se k **normálnímu** rozdělení, když je **vzorků hodně**

**křivka t-rozdělení je definována** počtem **vzorků,** které mám **k dispozici**

odlišuje, když **je** málo **vzorků, má těžší konce**

**využití pro t-**test **- testování** hypotézy, **CI, snažím se zjistit střední hodnotu** malého počtu **vzorků**

9. Definujte *věrohodnostní funkci* (likelihood). K čemu se používá metoda maximální věrohodnosti? Po- jmenujte alespoň dvě metody, které se k maximalizaci věrohodnosti používají. Vysvětlete, proč se často používá logaritmus věrohodnosti. jak se dá věrohodnosti využíét při testování statistických hypotéz?

**P(param x)** =

10. Formálně definujte *multivariátní normální rozdělení.* Kolik parametrů budeme obecně potřebovat pro popis tohoto rozdělení v *d* dimenzích? Lze počet těchto parametrů nějak omezit? Jaké důsledky toto omezení může mít?

**multidimenzionální Norm**

**navíc kovarianční matice - místo rozptylů**

vektory průměrů

**pro popis v d**-dimenzích **-** potřeba parametrů **--> průměr** má d-hodnot + kovarianční **matice dxd,** ale **je symetrická (dx** (d**+1))**

**omezení --**> když vezmu nekorelovanou **kovarianční matici (mimo** a **diagonálu** budou **nuly**) **důsledky** - **přicházíme o informaci z korelace**

11. Definujte alespoň dvě často používaná rozdělení diskrétní náhodné veličiny. Pojmenujte jejich parame-

try. Na příkladech naznačte, kdy se tato rozdělení dají využít.

Binomické

-

Bromide (m,p) = P(*x*=x) = (~~). p\*. (1-p)^

**pdus**

28 hodů

hosthon

3x padhe 6

shed' hod***.***

**X**

h

Poissonovo (2)

2

holikrát zhasne zaborba

P(Y=k) = e

k!

し

počet událostí

12. Vysvětlete rozdíl mezi parametrickým a neparametrickým statistickým testem. Pojmenujte základní

výhody a nevýhody obou přístupů. Jmenujte alespoň jeden paramerický a jeden neparametrický test.

टे

předpoklad, &c

"Blivemane slicing

pochází z

T-best

veryžadují specifice **rozeliten**

Sign test Wilcoxon

**Parametrické metody** (parametric survival **analysis) vyžadují specifikaci konkrétního rozdělení náhodné veličiny, zatímco neparametrické** metody **(nonparametric survival analysis) žádné zvláštní předpoklady ohledně rozdělení pravděpodobnosti náhodné veličiny nevyžadují**

13. K následující statistické úloze přiřaďte vhodný test. Bylo testováno 11 automobilů určité značky. Ověřte, zda se jejich pravé a levé přední pneumatiky ojíždějí srovnatelně. (Předpokládejte, že ojetí pne- umatik [mm] má normální rozdělení. Z následujících testů vyberte ten nejlepší, svou volbu vysvětlete 2. dvouvýběrový t-test. Friedmanův test, jednovýběrový t-test, jednovýběrový Wilcoxonův test, test o parametru alternativního rozdělení, test o rozptylu normálního rozdělení, párový t-test). Co by se stalo pokud byste použili druhý nejlepší test z daného seznamu?

1.

jednovýběrový oboustraný - **liší se krevní tlak (na** obě **strany**)

jednovýběrový **jednostranný - KT** větší **než**