Ogledni primjer pitanja u 1. kolokviju iz vjerojatnosti i statistike

1. Za skup podataka $x_1,x_2,...,x_n$:
   (i) napišite formulu za aritmetičku sredinu i navedite značenje.
   (ii) napišite formule za varijancu i standardnu devijaciju i navedite značenje.
   (iii) napišite formule za korigiranu varijancu i standardnu devijaciju i navedite uporabu.
   (iv) izračunajte sve za skup podataka 1,1,2,3,4,4,4,5.

2. Definirajte slučajnu varijablu *X* zadanu oznakom, opišite značenje parametara i okolnosti u kojima se X pojavljuje:
   (i) $X \sim B(n,p)$
   (ii) $X \sim P(a)$
   (iii) $X \sim N(\mu, \sigma^2)$
   (iv) $X \sim E(\lambda)$

Izdvojite kontinuirane od diskretnih i za kontinuirane nacrtajte grafove funkcija gustoće i distribucije.

3. (i) Formulirajte, objasnite i grafički predočite pravilo *tri sigme*.
   (ii) Posebno za $X \sim N(5,2^2)$.

4. (i) Kako procjenjujemo očekivanje, a kako varijancu populacije?
   (ii) Napišite formule i predočite interval pozdanosti za očekivanje uz vjerojatnost 1-2p, posebno ako je vjerojatnost 0.95.
   (iii) Objasnite značenje intervala pouzdanosti.

5. (i) Napišite formulu za t_{exp} i za broj stupnjeva slobode, te predočite kritično područje (područje prihvaćanja hipoteze) za hipotezu $\mu = \mu_0$ (uz razne alternativne hipoteze).
(ii) Objasnite značenje nivoa signifikantnosti i kritične vrijednosti.

6. (i) Napišite formulu za t_{exp} i za broj stupnjeva slobode, te predočite kritično područje (područje prihvaćanja hipoteze) za hipotezu $\mu_1 = \mu_2$ (uz razne alternativne hipoteze).
(ii) Objasnite značenje nivoa signifikantnosti i kritične vrijednosti.

7. Zapišite formulu za χ^2_{exp}, za broj stupnjeva slobode, objasnite značenje kritične vrijednosti i predočite kritično područje pri testiranju *hikvadrat* testom.

8. (i) Navedite načelo na kojemu se zasniva metoda najmanjih kvadrata.
   (ii) Zapišite funkciju cilja za linearnu vezu i podatke:

x_i	1	2	4	5	6
y_i	-2	1	8	10	13

(iii) Predočite grafički ove podatke i procijenite parametre.
(iv) Objasnite značenje koeficijenta korelacije.
(v) Izračunajte parametre i koeficijent korelacije za navedene podatke.

Napomene: 1. Bit će postavljeno 5 pitanja odprilike iz predloženog skupa pitanja, odnosno dijelova tih pitanja.

2. KI i EI imat će jednu težinu pitanja, a PK i KiM drugu.

3. PK i KiM neće imati pitanja poput: 1 (iii), 4 (iii), 5 (ii), 6, 8 (v).

Neka rješenja, odnosno skice i naputci (slike su vrlo važne, ali su ovdje izostavljene; potražite ih u lekcijama). Za potpunost rješenja važne su točne formule, precizni crteži i jasan tekst (odgovori riječima).

1. (i) $\bar{x} = \dfrac{x_1 + x_2 + ... + x_n}{n}$

(ii) **Varijanca je prosječno kvadratno odstupanje od prosjeka**:

$$(s')^2 := \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + ... + (x_n - \bar{x})^2}{n}$$

Standardna devijacija uzorka s' je drugi korijen iz varijance uzorka:

$$s' := \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + ... + (x_n - \bar{x})^2}{n}}$$

I varijanca i standardna devijacija su mjere rasipanja podataka oko aritmetičke sredine (što je standardna devijacija veća, podatci su raspršeniji).

(iii) **Korigirana varijanca**

$$s^2 := \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + ... + (x_n - \bar{x})^2}{n-1}$$

Korigirana standardna devijacija

$$s := \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + ... + (x_n - \bar{x})^2}{n-1}}.$$

Korigirane varijanca i standardna devijacija imaju slične uloge kao i obična varijanca i standardna devijacija. Dodatno pomoću njih procjenjujemo varijancu i standardnu devijaciju populacije.

(iv) $\bar{x} = 3$, $(s')^2 = 2$, $s^2 = \dfrac{16}{7}$.

2. (i) To je binomna razdioba s parametrima n (u pravilu ima značenje broja izvođenja pokusa) i p (u pravilu ima značenje vjerojatnosti uočenog događaja); $p(X=i) = \binom{n}{i} p^i(1-p)^{n-i}$

U ovoj interpretaciji slučajna varijabla X registrira broj pojavljivanja uočenog događaja. Vrijedi $E(X) = np$, $V(X)=np(1-p)$.

(ii) To je Poissonova razdioba s parametrom $a>0$ (koji ima značenje očekivanja od X); $p(X=i) = e^{-a}\dfrac{a^i}{i!}$. Pojavljuje se pri brojanju poruka na nekoj adresi u fiksiranom vremenskom intervalu i sl.

(iii) To je normalna razdioba (i spada u kontinuirane) s parametrima μ (ima značenje očekivanja) i σ^2 (ima značenje varijance). Pojavljuje se (približno) pri mjerenju mase, visine, inteligencije, prema njoj se ponašaju grješke pri mjerenju i sl.

Funkcija gustoće vjerojatnosti je $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, a funkcija distribucije ne može se eksplicitno izraziti.

(iv) To je eksponencijalna razdioba (i spada u kontinuirane) s parametrom λ (vrijedi $E(X) = 1/\lambda$); s funkcijom gustoće $f(x):= \lambda e^{-\lambda x}$, za $x>0$ i funkcijom distribucije $F(x)= 1- e^{-\lambda x}$, za $x>0$.

Pojavljuje se (približno) pri mjerenju vremena između dviju poruka na nekoj adresi, vremena između dvaju uzastopnih kvarova na nekom uređaju i sl.

3. (Skica). (i) Vjerojatnost da slučajna varijabla $X \sim N(\mu, \sigma^2)$ poprimi vrijednost u intervalu $<\mu-3\sigma, \mu+3\sigma>$ približno je jednaka 1 (oko 0.997). Kraće, $p(\mu-3\sigma < X < \mu+3\sigma)>0.997$.
Slično treba napraviti i za *dva sigma* i *jedan sigma*.
(ii) Tu je $<\mu-3\sigma, \mu+3\sigma> = <-1,11>$ itd.

4. (i) Očekivanje populacije procjenjujemo aritmetičkom sredinom uzorka $\bar{x}$ (napisati formulu!!), a varijancu korigiranom varijancom uzorka s^2 (napisati formulu!!).
(ii) (skica) To je interval $<\bar{x} - z_p\cdot\dfrac{\sigma}{\sqrt{n}}, \bar{x} + z_p\cdot\dfrac{\sigma}{\sqrt{n}}>$ (ako je poznata standardna devijacija σ), odnosno $<\bar{x} - t_p(k)\dfrac{s}{\sqrt{n}}, \bar{x} + t_p(k)\dfrac{s}{\sqrt{n}}>$ ako σ nije poznato, već ga procjenjujemo pomoću s iz n mjerenja.
(Za potpuni odgovor, u koji trebaju biti uključena značenja svih oznaka, treba pogledati sažetak i odgovarajuću lekciju).
(iii) Pogledajte u odgovarajućem sažetku (str. 4 i 5) ili u lekciji.

5. Pogledajte sažetak (str. 6 i 7).

6. Pogledajte sažetak (str. 7 i 8).

7. Pogledajte sažetak (str. 8 i 9).

8. (i) Metoda najmanjih kvadrata zasniva se na načelu da

suma kvadrata razlika eksperimentalnih i teoretskih podataka bude minimalna.

Kraće, $\sum D_i^2 \rightarrow$ min, gdje je $D_i := y_i - f(x_i, a,b)$ odstupanje između eksperimentalnih i teoretskih podataka.

(ii) Tu je funkcija cilja $F(a,b) := \sum D_i^2$, dakle

$F(a,b)= [-2-(a\cdot 1+b)]^2+[1-(a\cdot 2+b)]^2 +[8-(a\cdot 4+b)]^2 +[10-(a\cdot 5+b)]^2 +[13-(a\cdot 6+b)]^2$

(iii) pogledajte odgovarajuće mjesto u lekciji.

(iv) Koeficijent korelacije *r* je mjera linearne zavisnosti dviju veličina (dviju serija podataka). Taj je broj između -1 i 1. Ako je *r* blizu 1, to je visoka pozitivna, a ako je blizu -1 to je visoka negativna koreliranost. Ako je, pak, *r* blizu nule koreliranost je vrlo niska.

(v) Pogledajte formule iz sažetka i rješenje primjera u lekciji.