SVEUČILIŠTE U SPLITU
FAKULTET ELEKTROTEHNIKE,
STROJARSTVA I BRODOGRADNJE
U SPLITU

Teorija iz predmeta
Vjerojatnost i statistika
Studij Računarstva 750/3 - šk. god. 2004./2005.

Student: A. Vujević

Split, kolovoz 2005.

1. Nabrojite srednje vrijednosti i njihovu interpretaciju. Koje srednje vrijednosti se upotrebljavaju kod redosljednih nizova?

Najčešće su u potrebi slijedeće srednje vrijednosti:

a) Aritmetička sredina

b) Harmonijska sredina

c) Geometrijska sredina

d) Medijan

e) Mod

Izrazi za **jednostavnu** i **vaganu aritmetičku sredinu** glase:

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N} \quad , \quad \overline{x} = \frac{\sum_{i=1}^{k} f_i \cdot x_i}{\sum_{i=1}^{k} f_i}$$

Izrazi za **jednostavnu** i **vaganu harmonijsku sredinu** glase:

$$H = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}} \quad , \quad H = \frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$$

Izrazi za **jednostavnu geometrijsku sredinu** glasi: $G = \sqrt[N]{\prod_{i=1}^{N} x_i}$

Medijan je ona srednja vrijednost koja frekvencije statističkog niza dijeli na dva jednaka dijela.

Mod je ona vrijednost numeričkog obilježja koja je s obzirom na svoje susjedne vrijednosti **najčešća**. Mod dijeli distribuciju frekvencija na rastuću i padajuću stranu. Mod ima smisla računati samo kod tzv. unimodalnih distribucija. Kod bimodalne distribucije postoje glavni mod i lokalni mod.

Aritmetička, harmonijska i geometrijska sredina su tzv. **izračunate** srednje vrijednosti.

Medijan i mod su tzv. **položajne** srednje vrijednosti.

Kod **redosljednih nizova** od navedenih srednjih vrijednosti koriste se **medijan** i **mod**.

2. Objasnite razliku između medijana i aritmetičke sredine. Kada je bolja uporaba medijana kao srednje vrijednosti?

Medijan je tzv. položajna srednje vrijednosti, dok je aritmetička sredina tzv. izračunata srednja vrijednost.

Medijan je ona srednja vrijednost koja frekvencije statističkog niza dijeli na dva jednaka dijela.

Izrazi za **jednostavnu** i **vaganu aritmetičku sredinu** glase:

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N} \quad , \quad \overline{x} = \frac{\sum_{i=1}^{k} f_i \cdot x_i}{\sum_{i=1}^{k} f_i}$$

Uporaba medijana se preporuča:

- Kada su vrijednosti numeričkog obilježja poredane po veličini i poznat je njihov ukupan broj, ali su nepoznate sve vrijednosti numeričkog obilježja
- Kod distribucija frekvencija koje imaju otvorene razrede

Prednosti medijana u odnosu na aritmetičku sredinu su:

- Manje osjetljiv na ekstreme
- Reprezentativniji kod izrazito asimetričnih distribucija frekvencija

3. Koje su mane aritmetičke sredine? Da li je prikladna uporaba aritmetičke sredine kod redosljednoga obilježja? Odgovor obrazložite.

Mane aritmetičke sredine su:

- Osjetljiva na ekstreme
- Ima malu reprezentativnost kod izrazito asimetričnih distribucija frekvencija

Uporaba aritmetičke sredine nije prikladna kod **redoslijednog** obilježja. Aritmetička sredina je tzv. izračunata srednja vrijednost, a ne položajna srednja vrijednost. Kod **redosljednih nizova** koriste se položajne srednje vrijednosti **medijan** i **mod**.

4. Objasnite klasičnu definiciju vjerojatnosti. Kakva je to vjerojatnost "a priori", a kakva "a posteriori"? Kako glasi zakon o zbrajanju, a kako zakon o množenju vjerojatnosti?

Prema **klasičnoj definiciji,** vjerojatnost realizacija slučajnog događaja A jednaka je:

$$P(A) = \frac{m(A)}{n}$$

gdje je: m(A) - ukupan broj povoljnih ishoda,

n - ukupan broj svih mogućih ishoda.

Vrijedi da je: $0 \leq P \leq 1$.

Sigurnome događaju odgovara vjerojatnost 1, a nemogućem vjerojatnost 0.

Vjerojatnost da se slučajni događaj A ne realizira je:

$$P\left(\overline{A}\right) = 1 - P(A) = \frac{n - m(A)}{n}$$

Vjerojatnost "a priori" je ona vjerojatnost kod koje je unaprijed poznata vjerojatnost realizacije slučajnog događaja.

Vjerojatnost "a posteriori" je ona vjerojatnost kod koje nije unaprijed poznata vjerojatnost realizacije slučajnog događaja. Ona je jednaka graničnoj vrijednosti relativne frekvencije kada ukupan broj pokusa teži u beskonačnost:

$$P(A) = p \cdot \lim_{n \to \infty} \frac{m(A)}{n}$$

Izraz $p \cdot \lim$ se čita: "granična vrijednost po vjerojatnosti.

Zakon o zbrajanju (aditivnosti) vjerojatnosti glasi:

a) Događaji se međusobno isključuju (ne mogu nastupiti istodobno):

P(ili A ili B) = P $(A \cup B)$ = P (A) + P (B)

b) Događaji se međusobno ne isključuju:

P(ili A ili B ili oba) = P $(A \cup B)$ = P(A) + P(B) – P(A i B)

Zakon o množenju (multiplikativnosti) vjerojatnosti glasi:

Ako se događaji A i B međusobno ne isključuju (mogu se dogoditi istodobno) te ako vjerojatnost realizacije jednog događaja ne utječe na vjerojatnost realizacije drugog događaja, tada je vjerojatnost istovremene realizacije obaju događaja:

$$P(A \text{ i } B) = P(A \cap B) = P(A) \cdot P(B)$$

5. Objasnite definiciju uvjetne (kondicionalne) vjerojatnosti. Kako glasi Bayesov teorem?

Ako je realizacija događaja A uvjetovana realizacijom događaja B, tada govorimo o uvjetnoj ili kondicionalna vjerojatnosti događaja A. Vrijedi izraz:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) > 0$$

Ako su događaji A i B nezavisni, onda je: $P(A/B) = P(A)$.

Bayesov teorem glasi:

Ako se događaj B realizira samo tada kad nastupi jedan od n disjunktnih događaja A_1, A_2, ..., A_n za koje je $\sum_{i=1}^{n} A_i = 1$, onda vrijede izrazi:

$$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B/A_i)$$

$$P(A_k/B) = \frac{P(A_k) \cdot P(B/A_k)}{P(B)} = \frac{P(A_k) \cdot P(B/A_k)}{\sum_{i=1}^{n} P(A_i) \cdot P(B/A_i)}$$

6. Napišite i objasnite Bernoullijev zakon velikih brojeva.

Bernoullijev zakon velikih brojeva glasi:

Za događaj A kojem u svakom eksperimentu pripada ista vjerojatnost $P(A) = p$ te za proivoljni $\varepsilon > 0$ vrijedi:

$$\lim_{n \to \infty} P\left\{ \left| f_{r\,n} - p \right| \geq \varepsilon \right\} = 0$$

gdje je $f_{r\,n}$ relativna frekvencija toga događaja u n pokusa. Ako događaj A nastupi x puta u n pokusa, onda je relativna frekvencija toga događaja:

$$f_{r\,n} = \frac{x}{n}$$

Dakle, ako $n \to \infty$, onda relativna frekvencija događaja A **teži po vjerojatnosti** stvarnoj vjerojatnosti p događaja A.

Drugim riječima, što više pokusa izvodimo tim možemo biti sigurniji da će relativna frekvencija $f_{r\,n}$ biti bliža stvarnoj vjerojatnosti p događaja A.

7. Koja varijabla se može smatrati diskontinuiranom slučajnom varijablom? Koje uvjete mora zadovoljavati diskontinuirana slučajna varijabla?

Diskontinuirana (diskretna) slučajna varijabla je ona varijabla koja može poprimiti najviše prebrojivo beskonačno mnogo vrijednosti s određenom vjerojatnošću:

$$P(x_1),\ P(x_2),\ldots$$

pri čemu mora biti:

a) $\sum_{i=1}^{\infty} P(x_i) = 1$

b) $P(x_i) \geq 0 \quad \forall i$

8. **Napišite sve što znate o diskontinuiranoj slučajnoj varijabli (definiciju, svojstva, očekivanje, varijancu, funckiju distribucije).**

Diskontinuirana (diskretna) slučajna varijabla je ona varijabla koja može poprimiti najviše prebrojivo beskonačno mnogo vrijednosti s određenom vjerojatnošću:

$$P(x_1),\ P(x_2),\ \dots$$

pri čemu mora biti:

a) $$\sum_{i=1}^{\infty} P(x_i) = 1$$

b) $$P(x_i) \geq 0 \quad \forall i$$

Zakon po kojem se svakoj vrijednosti slučajne varijable X pridjeljuje vjerojatnost $P(x_i)$ zove se **zakon vjerojatnosti** slučajne varijable X.

Očekivanje diskontinuirane slučajne varijable X je njena aritmetička sredina, tj. vrijedi da je:

$$E(X) = \mu = \sum_{i=1}^{\infty} x_i \cdot P(x_i)$$

Varijanca diskontinuirane slučajne varijable je opisana izrazom:

$$V(X) = \sigma^2 = E\left[(X-\mu)^2 \right] = \sum_{i=1}^{\infty} \left(x_i - \mu \right)^2 \cdot P(x_i)$$

gdje je σ standardna devijacija diskontinuirane slučajne varijable X.

Vrijedi da je:

$$V(X) = E(X^2) - \left[E(X) \right]^2 = \sum_{i=1}^{\infty} x_i^2 \cdot P(x_i) - \mu^2$$

Funkcija distribucije diskontinuirane slučajne varijable predstavlja vjerojatnost da slučajna varijabla X ne premaši neku zadanu vrijednost $x_k \in \mathbb{R}$:

$$F(x_k) = \sum_{i=1}^{k} P(x_i)$$

Funkcija distribucije je neopadajuća funkcija.

9. Diskontinuirana slučajna varijabla. Svojstva i teorijske distribucije.

Diskontinuirana (diskretna) slučajna varijabla je ona varijabla koja može poprimiti najviše prebrojivo beskonačno mnogo vrijednosti s određenom vjerojatnošću:

$$P(x_1),\ P(x_2),\ \ldots$$

pri čemu mora biti:

a) $$\sum_{i=1}^{\infty} P(x_i) = 1$$

b) $$P(x_i) \geq 0 \quad \forall i$$

Teorijske distribucije diskontinuirane slučajne varijable su:

- Binomna distribucija
- Poissonova distribucija
- Jednolika distribucija
- Hipergeometrijska distribucija

Binomna distribucija ima slijedeći zakon vjerojatnosti:

$$P(X = x) = \binom{n}{x} p^x \cdot q^{n-x} \quad , \qquad x = 0, 1, 2, ..., n$$

Poissonova distribucija ima slijedeći zakon vjerojatnosti:

$$P(X = x) = \frac{\mu^x \cdot e^{-\mu}}{x!} \quad , \qquad x = 0, 1, 2, ..., \infty$$

Jednolika distribucija ima slijedeći zakon vjerojatnosti:

$$P(X = x_i) = \frac{1}{n} \quad , \qquad \forall x_i \in \{x_1, x_2, \ldots, x_n\}$$

Hipergeometrijska distribucija ima slijedeći zakon vjerojatnosti:

$$P(X = x) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}} \quad , \qquad x = 0, 1, 2, ..., n$$

10. **Napišite sve što znate o binomnoj distribuciji.**

Binomna distribucija je teorijska distribucija diskontinuirane slučajne varijable. Ona se koristi kada je vjerojatnost **p** nastanka slučajnog događaja **konstantna** tijekom izvođenja pokusa, a traži se vjerojatnost da u seriji od **n** pokusa događaj X nastupi baš x puta.

Binomna distribucija ima slijedeći zakon vjerojatnosti:

$$P(X = x) = \binom{n}{x} p^x \cdot q^{n-x} \quad , \qquad x = 0, 1, 2, ..., n$$

gdje je: n - ukupan broj pokusa,

x - ukupan broj povoljnih ishoda u n pokusa,

p - vjerojatnost realizacije slučajnog događaja,

q = 1 - p.

Očekivanje slučajne varijable je:

$$E(X) = \mu = n \cdot p$$

Varijanca slučajne varijable je:

$$V(X) = \sigma^2 = n \cdot p \cdot q$$

Koeficijent asimetrije je:

$$\alpha_3 = \frac{q - p}{\sigma}$$

Koeficijent zaobljenosti je:

$$\alpha_4 = 3 + \frac{1 - 6 \cdot p \cdot q}{\sigma^2}$$

11. Napišite sve što znate o Poissonovoj distribuciji.

Poissonova distribucija je teorijska distribucija diskontinuirane slučajne varijable. Poissonova distribucija se koristi za opis **rijetkih događaja,** tj. vjerojatnost nastanka slučajnog događaja je **veoma malena i konstantna** tijekom izvođenja pokusa. Poissonova distribucija je granični slučaj binomne distribucije kada ukupan broj pokusa $n \to \infty$, uz uvjet da je očekivanje $\mu = n \cdot p$ konstantno.

Binomna distribucija se može dobro aproksimirati Poissonovom distribucijom ako je $n \geq 50$ i $p \leq 0{,}1$. Time se olakšava izračunavanje pripadnih vjerojatnosti.

Poissonova distribucija ima slijedeći zakon vjerojatnosti:

$$P(X = x) = \frac{\mu^x \cdot e^{-\mu}}{x!} \quad , \qquad x = 0, 1, 2, ..., \infty$$

Očekivanje slučajne varijable je:

$$E(X) = \mu$$

Varijanca slučajne varijable je:

$$V(X) = \sigma^2 = \mu$$

Koeficijent asimetrije je:

$$\alpha_3 = \frac{1}{\sqrt{\mu}}$$

Koeficijent zaobljenosti je:

$$\alpha_4 = 3 + \frac{1}{\mu}$$

Poissonova distribucija je uvijek pozitivno asimetrična.

12. Napišite sve što znate o jednolikoj distribuciji.

Jednolika distribucija je teorijska distribucija diskontinuirane slučajne varijable.

Ona se koristi kada varijabla X poprima s istom vjerojatnošću bilo koju od n vrijednosti.

Jednolika distribucija ima slijedeći zakon vjerojatnosti:

$$P(X = x_i) = \frac{1}{n} \quad , \qquad \forall x_i \in \{x_1, x_2, \ldots, x_n\}$$

Najbolji primjer je bacanje kocke. Vjerojatnost da kocka padne na bilo koji od 6 brojeva je 1/6.

13. Napišite sve što znate o hipergeometrijskoj distribuciji.

Hipergeometrijska distribucija je teorijska distribucija diskontinuirane slučajne varijable.

Koristi se kad se osnovni skup sastoji iz dva dijela:

- skupa elemenata koji imaju neko obilježje A,
- skupa elemenata koji nemaju to obilježje.

Iz osnovnog skupa koji ima N elemenata uzima se uzorak od n elemenata, a traži se vjerojatnost da u uzorku ima x elemenata s obilježjem A.
Pritom se element koji je uzet u uzorak više **ne vraća** u osnovni skup.

Hipergeometrijska distribucija ima slijedeći zakon vjerojatnosti:

$$P(X = x) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}} \quad , \qquad x = 0, 1, 2, ..., n$$

gdje je: N - ukupan broj elemenata u osnovnom skupu,

M - ukupan broj elemenata u osnovnom skupu s obilježjem A,

n - ukupan broj elemenata u uzorku,

x - ukupan broj elemenata u uzorku s obilježjem A.

Očekivanje slučajne varijable je: $E(X) = \mu = n \cdot p = n \cdot \frac{M}{N}$

Varijanca slučajne varijable je: $V(X) = \sigma^2 = n \cdot p \cdot q \cdot \frac{N-n}{N-1} = n \cdot \frac{M}{N} \cdot \frac{N-M}{N} \cdot \frac{N-n}{N-1}$

14. Kada se upotrebljava binomna distribucija? Koliko iznosi njeno očekivanje i varijanca?

Binomna distribucija se koristi kada je vjerojatnost **p** nastanka slučajnog događaja **konstantna** tijekom izvođenja pokusa, a traži se vjerojatnost da u seriji od **n** pokusa događaj X nastupi baš x puta.

Binomna distribucija ima slijedeći zakon vjerojatnosti:

$$P(X = x) = \binom{n}{x} p^x \cdot q^{n-x} \quad , \qquad x = 0, 1, 2, ..., n$$

gdje je: n - ukupan broj pokusa,

x - ukupan broj povoljnih ishoda u n pokusa,

p - vjerojatnost realizacije slučajnog događaja,

q = 1 - p.

Očekivanje slučajne varijable je: $E(X) = \mu = n \cdot p$

Varijanca slučajne varijable je: $V(X) = \sigma^2 = n \cdot p \cdot q$

15. Kada se upotrebljava Poissonova distribucija? Koliko iznosi njeno očekivanje i varijanca?

Poissonova distribucija se upotrebljava kada je vjerojatnost nastanka slučajnog događaja **veoma malena i konstantna** tijekom izvođenja pokusa.

Poissonova distribucija je granični slučaj binomne distribucije kada ukupan broj pokusa $n \to \infty$, uz uvjet da je očekivanje $\mu = n \cdot p$ konstantno.

Binomna distribucija se može dobro aproksimirati Poissonovom distribucijom ako je $n \geq 50$ i $p \leq 0{,}1$. Time se olakšava izračunavanje pripadnih vjerojatnosti.

Poissonova distribucija ima slijedeći zakon vjerojatnosti:

$$P(X = x) = \frac{\mu^x \cdot e^{-\mu}}{x!} \quad , \qquad x = 0, 1, 2, ..., \infty$$

Očekivanje slučajne varijable je: $E(X) = \mu$

Varijanca slučajne varijable je: $V(X) = \sigma^2 = \mu$

16. Koja je razlika između Poissonove i binomne distribucije?

Binomna distribucija se upotrebljava kada je vjerojatnost p nastanka slučajnog događaja **konstantna** tijekom izvođenja pokusa. Ako je vjerojatost p **konstantna** i **veoma malena**, onda se umjesto binomne distribucije može koristiti Poissonova distribucija što olakšava izračunavanje pripadnih vjerojatnosti.

Teorijski gledano, Poissonova distribucija je granični slučaj binomne distribucije kada ukupan broj pokusa $n \to \infty$, uz uvjet da je očekivanje $\mu = n \cdot p$ konstantno.

Binomna distribucija se može dobro aproksimirati Poissonovom distribucijom ako je $n \geq 50$ i $p \leq 0{,}1$. Time se olakšava izračunavanje pripadnih vjerojatnosti.

17. Kada se upotrebljava hipergeometrijska distribucija? Kojom distribucijom se ona može aproksimirati i pod kojim uvjetima?

Hipergeometrijska distribucija se koristi kad se osnovni skup sastoji iz dva dijela:

- skupa elemenata koji imaju neko obilježje A,
- skupa elemenata koji nemaju to obilježje.

Iz osnovnog skupa koji ima N elemenata uzima se uzorak od n elemenata, a traži se vjerojatnost da u uzorku ima x elemenata s obilježjem A.

Pritom se element koji je uzet u uzorak više ne vraća u osnovni skup.

Hipergeometrijska distribucija se može aproksimirati **binomnom** distribucijom ako:

- ukupan broj elemenata osnovnog skupa $N \to \infty$,
- ukupan broj elemanata u uzorku n = konst.

Dakle, binomna distribucija je granični slučaj hipergeometrijske distribucije kada ukupan broj elemenata osnovnog skupa $N \to \infty$, uz uvjet n = konst.

Aproksimacijom hipergeometrijske distribucije pomoću binomne distribucije olakšava se izračunavanje pripadnih vjerojatnosti.

18. Koja je razlika između hipergeometrijske i binomne distribucije? Kada se one približavaju (aproksimiraju jedna drugom)?

Hipergeometrijska distribucija se koristi kad se osnovni skup sastoji iz dva dijela:

- skupa elemenata koji imaju neko obilježje A,
- skupa elemenata koji nemaju to obilježje.

Iz osnovnog skupa koji ima N elemenata uzima se uzorak, a traži se vjerojatnost **da u uzorku od n elemenata ima x elemenata s obilježjem A**. Pritom se element koji je uzet u uzorak više ne vraća u osnovni skup.

Binomna distribucija se koristi kada je vjerojatnost **p** nastanka slučajnog događaja **konstantna** tijekom izvođenja pokusa, a traži se vjerojatnost **da u seriji od n pokusa događaj X nastupi baš x puta**.

Hipergeometrijska distribucija se može aproksimirati **binomnom** distribucijom ako:

- ukupan broj elemenata osnovnog skupa $N \to \infty$,
- ukupan broj elemanata u uzorku n = konst.

Dakle, binomna distribucija je granični slučaj hipergeometrijske distribucije kada ukupan broj elemenata osnovnog skupa $N \to \infty$, uz uvjet n = konst.

19. Pojam vjerojatnosti. Dvodimenzionalna distribucija vjerojatnosti. Marginalna distribucija vjerojatnosti.

Prema klasičnoj definiciji**, vjerojatnost** realizacija slučajnog događaja A jednaka je:

$$P(A) = \frac{m(A)}{n}$$

gdje je: m(A) - ukupan broj povoljnih ishoda,

n - ukupan broj svih mogućih ishoda.

Vrijedi da je: $0 \leq P \leq 1$.

Kod **dvodimenzionalne distribucije vjerojatnosti** diskontinuirana slučajna varijabla X može poprimiti vrijednosti:

$$x_1, x_2, \ldots, x_n$$

a diskontinuirana slučajna varijabla Y vrijednosti:

$$y_1, y_2, \ldots, y_m$$

Vjerojatnost da slučajna varijabla X poprimi vrijednost x_i, a istovremeno varijabla Y vrijednost y_j označava se:

$$P(X = x_i,\ Y = y_j) = P(x_i,\ y_j)$$

Pritom mora biti zadovoljen uvjet:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,\ y_j) = 1$$

Skup svih uređenih parova $\left\{(x_i,\ y_j),\ P(x_i,\ y_j)\right\}$ tvori **dvodimenzionalnu distribuciju vjerojatnosti** slučajne varijable (X, Y).

Marginalna distribucija slučajne varijable X predstavlja vjerojatnost da slučajna varijabla X poprimi vrijednost x_i bez obzira na to koju će vrijednost poprimiti slučajna varijabla Y. Vrijedi izraz:

$$P(x_i) = \sum_{j=1}^{m} P(x_i,\ y_j)$$

Analogni izraz za **marginalnu distribuciju** slučajne varijable Y glasi:

$$P(y_j) = \sum_{i=1}^{n} P(x_i,\ y_j)$$

20. Napišite što znate o dvodimenzionalnoj diskretnoj (diskontinuiranoj) distribuciji vjerojatnosti. Što je to uvjetna distribucija, a što marginalna distribucija?

Kod **dvodimenzionalne distribucije vjerojatnosti** diskontinuirana slučajna varijabla X može poprimiti vrijednosti: $x_1, x_2, \ldots, x_n$
a diskontinuirana slučajna varijabla Y vrijednosti: $y_1, y_2, \ldots, y_m$.

Vjerojatnost da slučajna varijabla X poprimi vrijednost x_i, a istovremeno varijabla Y vrijednost y_j označava se:

$$P(X = x_i,\ Y = y_j) = P(x_i,\ y_j)$$

Pritom mora biti zadovoljen uvjet: $\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,\ y_j) = 1$.

Skup svih uređenih parova $\left\{(x_i,\ y_j),\ P(x_i,\ y_j)\right\}$ tvori **dvodimenzionalnu distribuciju vjerojatnosti** slučajne varijable (X, Y).

Marginalna distribucija slučajne varijable X predstavlja vjerojatnost da slučajna varijabla X poprimi vrijednost x_i bez obzira na to koju će vrijednost poprimiti slučajna varijabla Y. Vrijedi izraz:

$$P(x_i) = \sum_{j=1}^{m} P(x_i,\ y_j)$$

Analogni izraz za **marginalnu distribuciju** slučajne varijable Y glasi:

$$P(y_j) = \sum_{i=1}^{n} P(x_i,\ y_j)$$

Uvjetna ili kondicionalna distribucija vjerojatnosti je vjerojatnost da slučajna varijabla X poprimi vrijednost x_i pod uvjetom da je slučajna varijabla Y poprimila neku vrijednost y_j. Vrijedi izraz:

$$P(x_i / y_j) = \frac{P(x_i,\ y_j)}{P(y_j)} = \frac{P(x_i,\ y_j)}{\sum_{i=1}^{n} P(x_i,\ y_j)} \quad , \quad P(y_j) > 0$$

Analogni izraz za **uvjetnu distribuciju** slučajne varijable Y glasi:

$$P(y_j / x_i) = \frac{P(x_i,\ y_j)}{P(x_i)} = \frac{P(x_i,\ y_j)}{\sum_{j=1}^{m} P(x_i,\ y_j)} \quad , \quad P(x_i) > 0$$

Ako su slučajne varijable X i Y **nezavisne**, onda vrijedi:

$$P(x_i, y_j) = P(x_i) \cdot P(y_j) \quad ; \quad P(x_i / y_j) = P(x_i) \quad ; \quad P(y_j / x_i) = P(y_j)$$

21. Napišite sve što znate o kontinuiranoj slučajnoj varijabli (svojstva, očekivanje, varijancu, funckiju distribucije).

Kontinuirana slučajna varijabla X može poprimiti neprebrojivo beskonačno mnogo vrijednosti. Stoga se ne računa vjerojatnost u određenoj točki, već vjerojatnost nad određenim intervalom vrijednosti varijable X.

Funkcija (gustoće) vjerojatnosti kontinuirane slučajne varijable ima slijedeća svojstva:

a) $f(x) \geq 0 \quad \forall x$

b) $\int_{-\infty}^{\infty} f(x) \cdot dx = 1$ (površina ispod krivulje vjerojatnosti jednaka je 1)

c) $\int_{x_1}^{x_2} f(x) \cdot dx = P\{x_1 \leq X \leq x_2\} \quad , \quad x_1 < x_2$

Očekivanje kontinuirane slučajne varijable X je:

$$E(X) = \mu = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx$$

Varijanca kontinuirane slučajne varijable X je:

$$V(X) = \sigma^2 = E\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x) \cdot dx = \int_{-\infty}^{\infty} x^2 \cdot f(x) \cdot dx \ - \mu^2$$

gdje je σ standardna devijacija kontinuirane slučajne varijable X.

Funkcija distribucije kontinuirane slučajne varijable predstavlja vjerojatnost da slučajna varijabla X ne premaši neku zadanu vrijednost x:

$$F(x) = \int_{-\infty}^{x} f(x) \cdot dx$$

Funkcija distribucije je neopadajuća funkcija.

Lako se vidi da se funkcija gustoće vjerojatnosti može dobiti deriviranjem funkcije distribucije, odnosno:

$$f(x) = \frac{d\,F(x)}{dx}$$

22. Kontinuirana slučajna varijabla. Svojstva i teorijske distribucije.

Kontinuirana slučajna varijabla X može poprimiti neprebrojivo beskonačno mnogo vrijednosti. Stoga se ne računa vjerojatnost u određenoj točki, već vjerojatnost nad određenim intervalom vrijednosti varijable X.

Funkcija (gustoće) vjerojatnosti kontinuirane slučajne varijable ima slijedeća svojstva:

a) $f(x) \geq 0 \quad \forall x$

b) $\int_{-\infty}^{\infty} f(x) \cdot dx = 1$ (površina ispod krivulje vjerojatnosti jednaka je 1)

c) $\int_{x_1}^{x_2} f(x) \cdot dx = P\{x_1 \leq X \leq x_2\} \quad , \quad x_1 < x_2$

Najpoznatije **teorijske distribucije** kontinirane slučajne varijable su:

- Normalna distribucija
- Studentova distribucija
- Hi-kvadrat distribucija
- F-distribucija

Funkcija gustoće vjerojatnosti **normalne** distribucije je:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad , \ x \in (-\infty, \infty)$$

Studentova distribucije se koristi kod procjene parametara i kod testiranja hipoteza na osnovi uzorka. Njena nezavisna varijabla $t \in (-\infty, \infty)$. Kod uzoraka koji imaju ukupan broja stupnjeva slobode $\nu > 30$, Studentova se distribucija približava oblikom i svojstvima normalnoj distribuciji.

Hi-kvadrat distribucija se koristi kod prilagodbe neke od teorijskih razdioba empirijskim podacima. Kod $\nu > 30$ Hi-kvadrat distribucija se aproksimira normalnom distribucijom.

F-distribucija se uglavnom koristi pri testiranju značajnosti omjera varijanci.

23. Napišite sve što znate o normalnoj distribuciji.

Funkcija gustoće vjerojatnosti **normalne** distribucije je:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad , \ x\in(-\infty,\ \infty)$$

Parametri normalne distribucije su njeno očekivanje μ i varijanca σ^2.

Normalna distribucija je potpuno simetrična distribucija pa je koeficijent simetrije $\alpha_3=0$, dok je koeficijent zaobljenosti $\alpha_4=3$.

Za pronalaženje vjerojatnosti $P\{x_1\le X\le x_2\}$ za bilo koju normalnu distribuciju koristi se tablica **standardizirane ili jedinične normalne distribucije** $N\left(\mu,\ \sigma^2\right)=N(0\ ;\ 1)$ čija je gustoća vjerojatnosti:

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\cdot z^2}$$

gdje je:

$$z=\frac{x-\mu}{\sigma}$$

24. Napišite sve što znate o Studentovoj distribuciji.

Studentova distribucije se koristi kod procjene parametara i kod testiranja hipoteza na osnovi uzorka. Njena nezavisna varijabla $t\in(-\infty,\ \infty)$. Kod uzoraka koji imaju ukupan broja stupnjeva slobode $\nu>30$, Studentova se distribucija približava oblikom i svojstvima normalnoj distribuciji. U tom slučaju se Studentova distribucija može dovoljno dobro aproksimirati normalnom distribucijom.

25. Napišite sve što znate o Hi-kvadrat distribuciji.

Hi-kvadrat distribucija se koristi kod prilagodbe neke od teorijskih distribucija empirijskim podacima.

Veličina $\chi^2 = \sum_{i=1}^{k} \frac{\left(f_i - f_{ti}\right)^2}{f_{ti}}$ pripada Hi-kvadrat distribuciji s ukupnim brojem stupnjeva slobode:

$$\nu = k - \text{ukupan broj parametara teorijske distribucije} - 1.$$

Kod ukupnog broja stupnjeva slobode $\nu > 30$, Hi-kvadrat distribucija se aproksimira normalnom distribucijom:

$$\chi^2 = \frac{1}{2} \cdot \left(z + \sqrt{2 \cdot \nu - 1}\right)^2$$

gdje je z varijabla jedinične normalne distribucije.

26. Napišite sve što znate o F-distribuciji.

F-distribucija se uglavnom koristi pri testiranju značajnosti omjera varijanci.

F-distribucija je definirana izrazom:

$$F = \frac{\chi_1^2 / \nu_1}{\chi_2^2 / \nu_2}$$

gdje su χ_1^2 i χ_2^2 nezavisne Hi-kvadrat distribucije sa stupnjevina slobode ν_1 i ν_2.

27. Koja je razlika između normalne i Studentove distribucije?

Normalne distribucija ima dvije nezavisne varijable (očekivanje μ i varijancu σ^2), dok Studentova distribucija ima samo jednu nezavisnu varijablu (t).

Studentova distribucije se uglavnom koristi samo kod procjene parametara i kod testiranja hipoteza na osnovi uzorka.

Normalna distribucija je najvažnija teorijska distribucija u statističkoj teoriji. Zbog svojih izvanrednih svojstava ona ima vrlo široko područje primjene, a ostale distribucije nastojimo aproksimirati normalnom distribucijom kad god je to moguće. Kod ukupnog broja stupnjeva slobode $\nu > 30$, Studentovu distribuciju možemo dovoljno dobro aproksimirati pomoću normalne distribucije.

28. Objasnite razliku između uzoračne vrijednosti, populacijske vrijednosti i istinite (prave) vrijednosti. Koja svojstva mora posjedovati uzorak da bismo primijenili sve testove koji su u programu ovog kolegija?

Uzorak je podskup osnovnog statističkog skupa koji se uzima u svrhu procjene obilježja elemenata osnovnog skupa.

Populacija je česti naziv za osnovni skup.

Uzoračna vrijednost je vrijednost izabranog parametra za promatrani uzorak.

Populacijska vrijednost je vrijednost izabranog parametra osnovnog skupa. Popupacijska vrijednost izabranog parametra može biti procijenjena na osnovi uzorka ili pak poznata. Ako je populacijska vrijednost izabranog parametra poznata, onda govorimo o **istinitoj (pravoj) vrijednosti** osnovnog skupa.

Uzorak mora biti:

- Reprezentativan,
- Izbor jedinica (elemenata) u uzorak mora biti izvršen na slučajan način, npr. pomoću tablice slučajnih brojeva.

Postoji mogućnost odabira jedinica u uzorak s ponavljanjem i bez ponavljanja. U okviru ovog kolegija ograničili smo se na odabir **bez ponavljanja**.

29. Što je frakcija obabiranja? Koliki je ukupan broj mogućih uzoraka?

Frakcija obabiranja je opisana izrazom:

$$f = \frac{n}{N}$$

gdje je: n - ukupan broj jedinica u uzorku,

N - ukupan broj jedinica u osnovnome skupu.

Ako se odabir jedinica u uzorak vrši bez ponavljanja jedinica, ukupan broj svih mogućih uzoraka veličine n iz osnovnog skupa veličine N jednak je:

$$K = \binom{N}{n}$$

30. Što je to sampling distribucija? Kakva je to pristrana, a kakva nepristrana procjena? Koje su od vama poznatih procjena pristrane, a koje nepristrane?

Sampling distribucija je distribucija vrijednosti izabranog parametra svih mogućih uzoraka veličine n iz osnovnog skupa veličine N. Za svaki parametar osnovnog skupa može se kreirati zasebna sampling distribucija.

Pristranost procjene na bazi slučajnog uzorka predstavlja razliku između očekivane vrijednosti parametra u sampling distribuciji i vrijednosti populacijskog parametra. Općenito vrijedi:

- Ako je $E(\hat{\Theta}) = \Theta$ kažemo da je procjena $\hat{\Theta}$ **nepristrana**.
- Ako je $E(\hat{\Theta}) \neq \Theta$ kažemo da je procjena $\hat{\Theta}$ **pristrana**.

Pristrana je točkasta procjena varijance osnovnog skupa na osnovi slučajnog uzorka.

Nepristrane su točkaste procjene:

- aritmetičke sredine
- proporcije i
- totala

osnovnog skupa na osnovi slučajnog uzorka.

31. Kakva je to nepristrana procjena? Kod kojih procjena je sampling distribucija normalna, kod kojih Studentova a kod kojih Hi-kvadrat?

Pristranost procjene na bazi slučajnog uzorka predstavlja razliku između očekivane vrijednosti parametra u sampling distribuciji i vrijednosti populacijskog parametra. Općenito vrijedi:

- Ako je $E(\hat{\Theta}) = \Theta$ kažemo da je procjena $\hat{\Theta}$ **nepristrana**.
- Ako je $E(\hat{\Theta}) \neq \Theta$ kažemo da je procjena $\hat{\Theta}$ **pristrana**.

Uzima se da je sampling distribucija **normalna** kod svih procjena ako je veličina uzorka $n > 30$. Razlog tome je to što sve sampling distribucije teže normalnom obliku kada veličina uzorka $n \to \infty$.

Ako je veličina uzorka $n \leq 30$, sampling distribucije imaju oblik **Studentove distribucije** kod svih procjena, osim sampling distribucije varijanci koja ima oblik **Hi-kvadrat distribucije**.

32. Napišite sve što znate o procjeni aritmetičke sredine osnovnog skupa na osnovi uzorka.

Sampling distribucija aritmetičkih sredina ima oblik Studentove distribucije, a može se dovoljno dobro aproksimirati normalnom distribucijom ako je veličina uzorka n > 30.

Točkasta procjena aritmetičke sredine osnovnog skupa na osnovi slučajnog uzorka je nepristrana, tj. vrijedi:

$$E\left(\hat{\bar{X}}\right) = \bar{X}$$

Ako je skup uzoraka grupiran u razrede, onda se aritmetička sredina osnovnog skupa procjenjuje pomoću izraza:

$$\hat{\bar{X}} = \frac{\sum_i \bar{X}_i \cdot k_i}{\sum_i k_i}$$

gdje je: $\bar{X}_i$ - aritmetička sredina i-tog razreda,

k_i - ukupan broj uzoraka i-tog razreda.

Standardna greška (standardna devijacija sampling distribucije) računa se iz izraza:

$$Se(\bar{x}) = \begin{cases} \dfrac{\sigma}{\sqrt{n}} & \text{za } f < 0{,}05 \\ \dfrac{\sigma}{\sqrt{n}} \cdot \sqrt{\dfrac{N-n}{N-1}} & \text{za } f \geq 0{,}05 \end{cases}$$

gdje je σ standardna devijacija osnovnog skupa koja može biti poznata ili je pak treba procijeniti na osnovi uzorka.

Nepristrana procjena varijance osnovnog skupa na osnovi uzorka vrši se pomoću izraza:

$$s^2 = \hat{\sigma}^2 \cdot \left(\frac{n}{n-1}\right)$$

gdje je:

$$\hat{\sigma}^2 = \frac{\sum_i \left(\bar{X}_i - \hat{\bar{X}}\right)^2 \cdot k_i}{\sum_i k_i}$$

pristrana procjena varijance osnovnog skupa.

Ako je n > 30, u praksi se može koristiti aproksimacija $s^2 = \hat{\sigma}^2$.

Potrebna veličina uzorka uz zadani nivo pouzdanosti (povjerenja) i zadanu maksimalnu grešku vrši se pomoću izraza:

$$n = \begin{cases} n' & \text{ako je} \quad \dfrac{n'}{N} < 0,05 \\ \dfrac{n'}{1 + \dfrac{n'}{N}} & \text{ako je} \quad \dfrac{n'}{N} \geq 0,05 \end{cases}$$

gdje je: $$n' = \left(\frac{z \cdot \sigma}{\text{greška}} \right)^2 .$$

Izraz za **intervalnu procjenu aritmetičke** sredine osnovnog skupa glasi:

$$\Pr \left\{ \hat{\bar{X}} - z \cdot Se(\bar{x}) \; < \; \hat{\bar{X}} \; < \; \hat{\bar{X}} + z \cdot Se(\bar{x}) \right\} \; = \; 1 - \alpha$$

gdje je $1 - \alpha$ nivo pouzdanosti procjene.

Ako je n > 30, z se dobiva iz tablice površina ispod krivulje normalne distribucije.

Ako je n ≤ 30, umjesto koeficijenta z koristi se koeficijent t iz Studentove distribucije za ν = (n-1) stupnjeva slobode.

33. Napišite sve što znate o procjeni varijance osnovnog skupa na osnovi uzorka.

Sampling distribucija varijanci ima oblik Hi-kvadrat distribucije, a može se dovoljno dobro aproksimirati normalnom distribucijom ako je veličina uzorka n > 30.

Točkasta procjena varijance osnovnog skupa na osnovi uzorka vrši se pomoću izraza:

$$s^2 = \hat{\sigma}^2 \cdot \left(\frac{n}{n-1} \right) \quad ; \qquad \hat{\sigma}^2 = \frac{\sum_i \left(\overline{X}_i - \hat{\overline{X}} \right)^2 \cdot k_i}{\sum_i k_i}$$

gdje je: s^2 - nepristrana procjena varijance osnovnog skupa,

$\hat{\sigma}^2$ - pristrana procjena varijance osnovnog skupa,

$\overline{X}_i$ - aritmetička sredina i-tog razreda,

k_i - ukupan broj uzoraka i-tog razreda,

$\hat{\overline{X}}$ - procjena aritmetičke sredine osnovnog skupa.

Ako je n > 30, u praksi se može koristiti aproksimacija $s^2 = \hat{\sigma}^2$.

Izraz za **intervalnu procjenu varijance** osnovnog skupa glasi:

$$\Pr \left\{ \frac{n \cdot s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{n \cdot s^2}{\chi^2_{1-\alpha/2}} \right\} = 1 - \alpha$$

gdje je $1-\alpha$ nivo pouzdanosti procjene.

Ako je $n \leq 30$, vrijednosti Hi-kvadrata očitavamo iz odgovarajućih tablica Hi-kvadrat distribucije na pripadnom nivou α i uz ukupan broj stupnjeva slobobe:

$$df = \nu = n - 1.$$

Ako je n > 30, Hi-kvadrat distribucija se aproksimira pomoću normalne distribucije pomoću izraza:

$$\chi^2_{\alpha/2} = \frac{1}{2} \cdot \left(z_{\alpha/2} + \sqrt{2 \cdot \nu - 1} \right)^2$$

$$\chi^2_{1-\alpha/2} = \frac{1}{2} \cdot \left(- z_{\alpha/2} + \sqrt{2 \cdot \nu - 1} \right)^2$$

34. Napišite sve što znate o procjeni totala osnovnog skupa na osnovi uzorka.

Točkasta procjena totala osnovnog skupa na osnovi uzorka glasi:

$$\sum_{i=1}^{N} x_i = N \cdot \hat{\bar{X}}$$

gdje je N veličina osnovnog skupa, a $\hat{\bar{X}}$ procjena aritmetičke sredine osnovnog skupa.

Izraz za **intervalnu procjenu totala** osnovnog skupa glasi:

$$\Pr\left\{N \cdot \hat{\bar{X}} - z \cdot N \cdot Se(\bar{x}) < \sum_{i=1}^{N} x_i < N \cdot \hat{\bar{X}} + z \cdot N \cdot Se(\bar{x})\right\} = 1 - \alpha$$

gdje je $1 - \alpha$ nivo pouzdanosti procjene.

Ako je n > 30, z se dobiva iz tablice površina ispod krivulje normalne distribucije.

Ako je n ≤ 30, umjesto koeficijenta z koristi se koeficijent t iz Studentove distribucije za ν = (n-1) stupnjeva slobode.

Potrebna veličina uzorka uz zadani nivo pouzdanosti (povjerenja) i zadanu maksimalnu grešku vrši se pomoću izraza:

$$n = \begin{cases} n' & \text{ako je } \dfrac{n'}{N} < 0{,}05 \\[2ex] \dfrac{n'}{1 + \dfrac{n'}{N}} & \text{ako je } \dfrac{n'}{N} \geq 0{,}05 \end{cases}$$

gdje je: $$n' = \left(N \cdot \frac{z \cdot \sigma}{\text{greška}}\right)^2.$$

35. Napišite sve što znate o procjeni proporcije osnovnog skupa na osnovi uzorka.

Potrebna veličina uzorka uz zadani nivo pouzdanosti (povjerenja) i zadanu maksimalnu grešku vrši se pomoću izraza:

$$n = \begin{cases} n' & \text{ako je} \quad \dfrac{n'}{N} < 0{,}05 \\[2ex] \dfrac{n'}{1+\dfrac{n'}{N}} & \text{ako je} \quad \dfrac{n'}{N} \geq 0{,}05 \end{cases}$$

gdje je: $n' = \left(\dfrac{z \cdot \sqrt{P \cdot Q}}{\text{greška}} \right)^2$, gdje je $P \cdot Q$ varijanca osnovnog skupa.

Ako je zadana varijacija osnovnog skupa $\sqrt{\dfrac{Q}{P}}$, onda je: $n' = \left(\dfrac{z \cdot \sqrt{\dfrac{Q}{P}}}{\text{relativna greška}} \right)^2$.

Izraz za **intervalnu procjenu proporcije** (relativne frekvencije) osnovnog skupa glasi:

$$\Pr \left\{ \hat{p} - z \cdot Se(p) < P < \hat{p} + z \cdot Se(p) \right\} = 1 - \alpha$$

gdje je $1-\alpha$ nivo pouzdanosti procjene, $\hat{p}$ točkasta procjena proporcije na osnovi uzorka, a $Se(p)$ standardna greška procjene proporcije.

Ako je n > 30, z se dobiva iz tablice površina ispod krivulje normalne distribucije.

Ako je n ≤ 30, umjesto koeficijenta z koristi se koeficijent t iz Studentove distribucije za $\nu = (n-1)$ stupnjeva slobode.

Standardna greška se računa iz izraza:

$$Se(p) = \begin{cases} \sqrt{\dfrac{\hat{p} \cdot \hat{q}}{n}} & \text{ako je } n > 30 \\[2ex] \sqrt{\dfrac{\hat{p} \cdot \hat{q}}{n-1}} & \text{ako je } n \leq 30 \end{cases} \qquad \text{ako je } f < 0{,}05$$

$$Se(p) = \begin{cases} \sqrt{\dfrac{\hat{p} \cdot \hat{q}}{n}} \cdot \sqrt{\dfrac{N-n}{N-1}} & \text{ako je } n > 30 \\[2ex] \sqrt{\dfrac{\hat{p} \cdot \hat{q}}{n-1}} \cdot \sqrt{\dfrac{N-n}{N-1}} & \text{ako je } n \leq 30 \end{cases} \qquad \text{ako je } f > 0{,}05$$

36. Koje su vrste grešaka pri testiranju hipoteza? Objasnite što znači svaka od njih i njihov međusobni odnos.

Postoje dvije vrste grešaka pri testiranju hipoteza:

- Greška tipa I ili α
- Greška tipa II ili β

Greška tipa I predatavlja vjerojatnost odbacivanja istinite hipoteze H_o. Ta vjerojatnost jednaka je nivou signifikantnosti testa α.

Greška tipa II predatavlja vjerojatnost prihvaćanja lažne hipoteze H_o. Vjerojatnost greške tipa II označava se s β.

Snaga testa = $1-\beta$ je vjerojatnost da odbacimo lažnu hipotezu H_o.

Hipoteza H_o	Istinita	Lažna
Odbacuje se	**Greška tipa I**	Ispravan zaključak
Prihvaća se	Ispravan zaključak	**Greška tipa II**

Iznos greške tipa II ovisi o razlici istinite (prave) i pretpostavljene vrijednost izabranog parametra osnovnog skupa. Što je ta razlika veća, to je iznos greške tipa II manji. U krajnjem slučaju, ako su prava i pretpostavljena vrijednost izabranog parametra jednake, onda je: $\beta = 1 - \alpha$.

Ako se **smanji greška tipa I**, **poveća se greška tipa II** i obratno.

37. Koje su vrste grešaka pri testiranju hipoteza? Kako utječe veličina uzorka, varijanca i signifikantnost testa na interval prihvaćanja nulte hipoteze o nepoznatoj aritmetičkoj sredini populacije? Koja vrsta greške se ne pripisuje uzorku?

Postoje dvije vrste grešaka pri testiranju hipoteza:

- Greška tipa I ili α
- Greška tipa II ili β

Greška tipa I predatavlja vjerojatnost odbacivanja istinite hipoteze H_o. Ta vjerojatnost jednaka je nivou signifikantnosti testa α.

Greška tipa II predatavlja vjerojatnost prihvaćanja lažne hipoteze H_o. Vjerojatnost greške tipa II označava se s β.

Snaga testa = $1-\beta$ je vjerojatnost da odbacimo lažnu hipotezu H_o.

Interval prihvaćanja nulte hipoteze o nepoznatoj aritmetičkoj sredini populacije je širi ako je:

- veličina uzorka n manja,
- signifikantnost testa α manja,
- varijanca, odnosno standardna greška veća.

Uzorku se ne pripisuje greška tipa I. Nivo signifikantnosti testa α obično se zadaje unaprijed.

38. Opišite postupak testiranja hipoteze o nepoznatoj aritmetičkoj sredini osnovnog skupa. Navedite i praktični primjer.

Hipoteze glase: H_o: .... $\overline{X} = \overline{X}_o$

H_1: .... $\overline{X} \neq \overline{X}_o$

Interval prihvaćanja hipoteze H_o je: $\overline{X}_o \pm z \cdot Se(\overline{x})$ ako je $n > 30$

$\overline{X}_o \pm t \cdot Se(\overline{x})$ ako je $n \leq 30$

Testiranje se može provesti i z-testom ili t-testom: $z^* = \dfrac{\hat{\overline{X}} - \overline{X}_o}{Se(\overline{x})}$ ako je $n > 30$

$t^* = \dfrac{\hat{\overline{X}} - \overline{X}_o}{Se(\overline{x})}$ ako je $n \leq 30$

Kritična vrijednost za **z** se vadi iz tablice površine ispod normalne krivulje, dok se kod malog uzorka kritična vrijednost za **t** vadi iz Studentove distribucije sa (n-1) stupnjeva slobode i uz signifikantnost testa α.

Hipoteza H_o se prihvaća ako je $|z^*| < z_{\alpha/2}$, odnosno ako je $|t^*| < t_{\alpha/2}$.

Testiranje hipoteze o nepoznatoj aritmetičkoj sredini može se provesti i **jednosmjerno** kao:

- Testiranje na donju granicu: H_o: .... $\overline{X} \geq \overline{X}_o$

  H_1: .... $\overline{X} < \overline{X}_o$

  Donja granica prihvaćanja H_o je: $DG = \overline{X}_o - z \cdot Se(\overline{x})$ ako je $n > 30$

  $DG = \overline{X}_o - t \cdot Se(\overline{x})$ ako je $n \leq 30$

- Testiranje na gornju granicu: H_o: .... $\overline{X} \leq \overline{X}_o$

  H_1: .... $\overline{X} > \overline{X}_o$

  Donja granica prihvaćanja H_o je: $GG = \overline{X}_o + z \cdot Se(\overline{x})$ ako je $n > 30$

  $GG = \overline{X}_o + t \cdot Se(\overline{x})$ ako je $n \leq 30$

I jednosmjerna testiranja se mogu provesti pomoću z-testa ili t-testa.

Praktični primjer: Na osnovi uzorka treba ispitati da li se može prihvatiti hipoteza da je radni staž svih radnika zaposlenih u nekom poduzeću 10 godina, uz zadani nivo signifikantnosti testa.

39. Opišite postupak testiranja razlike aritmetičkih sredina dvaju osnovnih skupova. Navedite i praktični primjer.

Hipoteze glase: H_o: .... $\overline{X}_1 = \overline{X}_2$; $\overline{X}_1 - \overline{X}_2 = 0$

H_1: .... $\overline{X}_1 \neq \overline{X}_2$; $\overline{X}_1 - \overline{X}_2 \neq 0$

Interval prihvaćanja hipoteze H_o je: $0 \pm z \cdot Se(\overline{x}_1 - \overline{x}_2)$ ako je $df = n_1 + n_2 - 2 > 30$

$0 \pm t \cdot Se(\overline{x}_1 - \overline{x}_2)$ ako je $df = n_1 + n_2 - 2 \leq 30$

gdje su n_1 i n_2 veličine uzoraka iz osnovnih skupova.

Standardna greška se računa pomoću izraza:

$$Se(\overline{x}_1 - \overline{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad \text{ako je} \quad df = n_1 + n_2 - 2 > 30$$

$$Se(\overline{x}_1 - \overline{x}_2) = \sqrt{\frac{n_1 \cdot \hat{\sigma}_1^2 + n_2 \cdot \hat{\sigma}_2^2}{n_1 + n_2 - 2}\left(\frac{n_1 + n_2}{n_1 \cdot n_2}\right)} \quad \text{ako je} \quad df = n_1 + n_2 - 2 \leq 30$$

gdje su s_1^2 i s_2^2 nepristrane procjene varijanci osnovnih skupova na osnovi uzorka, a σ_1^2 i σ_2^2 nepristrane procjene varijanci.

Testiranje se može provesti i pomoću z-testa ili t-testa:

$$z^* = \frac{\left|\hat{\overline{X}}_1 - \hat{\overline{X}}_2\right|}{Se(\overline{x}_1 - \overline{x}_2)} \quad \text{ako je} \quad df = n_1 + n_2 - 2 > 30$$

$$t^* = \frac{\left|\hat{\overline{X}}_1 - \hat{\overline{X}}_2\right|}{Se(\overline{x}_1 - \overline{x}_2)} \quad \text{ako je} \quad df = n_1 + n_2 - 2 \leq 30$$

Hipoteza H_o se prihvaća ako je $z^* < z_{\alpha/2}$, odnosno ako je $t^* < t_{\alpha/2}$.

U ovom testiranju pošlo se od pretpostavki:

- Osnovni skupovi su nezavisni,
- Uzorci potječu iz normalnih populacija,
- Uzorci imaju istu varijancu.

Praktični primjer: Na osnovi uzoraka treba ispitati da li se može prihvatiti hipoteza da nema značajne razlike u prosječnoj težini proizvoda proivedenih na stroju A i stroju B uz odabrani nivo signifikantnosti testa.

Sve prethodno izneseno vrijedi kad su osnovni skupovi nezavisni.

Kod testiranja razlike aritmetičkih sredina dvaju **zavisnih** osnovnih skupova, hipoteze i postupak donošenja zaključaka su isti, a razlika je u izrazu za standardnu grešku koji za veliki uzorak glasi:

$$Se(\overline{x}_1 - \overline{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} - 2 \cdot r_{1,2} \cdot Se(\overline{x}_1) \cdot Se(\overline{x}_2)}$$

gdje je $r_{1,2}$ Pearsonov koeficijent linearne korelacije između dvaju mjerenja iste slučajne varijable na istome uzorku.

40. Opišite postupak testiranja hipoteze o nepoznatoj proporciji osnovnog skupa. Navedite i praktični primjer.

Hipoteze glase: H_o: .... $P = P_o$

H_1: .... $P \neq P_o$

Interval prihvaćanja hipoteze H_o je:

$$P_o \pm z \cdot Se(p) \quad \text{ako je} \quad n > 30 \qquad ; \qquad P_o \pm t \cdot Se(p) \quad \text{ako je} \quad n \leq 30$$

Standardna greška računa se iz izraza:

$$Se(p) = \sqrt{\frac{P_o \cdot Q_o}{n}} \quad \text{ako je} \quad n > 30 \qquad ; \qquad Se(p) = \sqrt{\frac{P_o \cdot Q_o}{n-1}} \quad \text{ako je} \quad n \leq 30$$

Testiranje se može provesti i z-testom ili t-testom:

$$z^* = \frac{\hat{p} - P_o}{Se(p)} \quad \text{ako je} \quad n > 30 \qquad ; \qquad t^* = \frac{\hat{p} - P_o}{Se(p)} \quad \text{ako je} \quad n \leq 30$$

Kritična vrijednost za **z** se vadi iz tablice površine ispod normalne krivulje, dok se kod malog uzorka kritična vrijednost za **t** vadi iz Studentove distribucije sa (n-1) stupnjeva slobode i uz signifikantnost testa α.

Hipoteza H_o se prihvaća ako je $|z^*| < z_{\alpha/2}$, odnosno ako je $|t^*| < t_{\alpha/2}$.

Testiranje hipoteze o nepoznatoj proporciji može se provesti i **jednosmjerno** kao:

- Testiranje na donju granicu: H_o: .... $P \geq P_o$

  H_1: .... $P < P_o$

  Donja granica prihvaćanja H_o je: $DG = P_o - z \cdot Se(p)$ ako je $n > 30$

  $DG = P_o - t \cdot Se(p)$ ako je $n \leq 30$

- Testiranje na gornju granicu: H_o: .... $P \leq P_o$

  H_1: .... $P > P_o$

  Donja granica prihvaćanja H_o je: $GG = P_o + z \cdot Se(p)$ ako je $n > 30$

  $GG = P_o + t \cdot Se(p)$ ako je $n \leq 30$

I jednosmjerna testiranja se mogu provesti pomoću z-testa ili t-testa.

Praktični primjer: Na osnovi uzorka treba ispitati da li se može prihvatiti hipoteza da je prolaznost studenata na ispitu iz predmeta Vjerojatnost i statitika 20 %, uz zadani nivo signifikantnosti testa.

41. Opišite postupak testiranja razlike proporcija dvaju osnovnih skupova.

Hipoteze glase: H_o: .... $P_1 = P_2$; $P_1 - P_2 = 0$

H_1: .... $P_1 \neq P_2$; $P_1 - P_2 \neq 0$

Interval prihvaćanja hipoteze H_o je: $0 \pm z \cdot Se(p_1 - p_2)$ ako je $df = n_1 + n_2 - 2 > 30$

$0 \pm t \cdot Se(p_1 - p_2)$ ako je $df = n_1 + n_2 - 2 \leq 30$

gdje su n_1 i n_2 veličine uzoraka iz osnovnih skupova.

Standardna greška se računa pomoću izraza:

$$Se(p_1 - p_2) = \sqrt{\hat{p} \cdot (1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

gdje je prosječna proporcija za oba uzorka: $\hat{p} = \frac{m_1 + m_2}{n_1 + n_2}$.

Testiranje se može provesti i pomoću z-testa ili t-testa:

$$z^* = \frac{|\hat{p}_1 - \hat{p}_2|}{Se(p_1 - p_2)} \quad \text{ako je} \quad df = n_1 + n_2 - 2 > 30$$

$$t^* = \frac{|\hat{p}_1 - \hat{p}_2|}{Se(p_1 - p_2)} \quad \text{ako je} \quad df = n_1 + n_2 - 2 \leq 30$$

Kritična vrijednost za **z** se vadi iz tablice površine ispod normalne krivulje, dok se kod malog uzorka kritična vrijednost za **t** vadi iz Studentove distribucije sa df stupnjeva slobode i uz signifikantnost testa α.

Hipoteza H_o se prihvaća ako je $z^* < z_{\alpha/2}$, odnosno ako je $t^* < t_{\alpha/2}$.

I ovdje se može provesti jednosmjerno testiranje, tj. testiranje na jednu granicu.

Npr. H_o: .... $P_1 \geq P_2$

H_1: .... $P_1 < P_2$

Praktični primjer: Na osnovi uzoraka treba ispitati da li se može prihvatiti hipoteza da nema značajne razlike u prolaznosti studenata na predmeta Vjerojatnost i statistika i Programiranje uz odabrani nivo signifikantnosti testa.

42. Hi-kvadrat test. Primjena i opis procedure. Navedite i konkretne primjere.

Hi-kvadrat test spada u neparametrijske testove i zasniva se na rasporedu frekvencija.

Najčešća primjena Hi-kvadrat testa:

a) Testiranje hipoteze da distribucija ima određeni oblik

b) Testiranje hipoteze o nezavisnosti obilježja elemenata osnovnog skupa.

U oba slučaja test veličina je empirijska vrijednost Hi-kvadrata:

a) $$\chi^{*2} = \sum_{i=1}^{k} \frac{\left(f_i - f_{ti}\right)^2}{f_{ti}}$$

b) $$\chi^{*2} = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(m_{ij} - e_{ij}\right)^2}{e_{ij}}$$

Empirijska vrijednost Hi-kvadrata (χ^{*2}) se uspoređuje s kritičnom vrijednošću Hi-kvadrata (χ^2) iz tablica za određeni nivo signifikantnosti testa α i uz pripadni ukupan broj stupnjeva slobode.

Ako je $\chi^{*2} < \chi^2$ prihvaćamo hipotezu:

a) da empirijska distribucija ima pretpostavljeni oblik,

b) o nezavisnosti obilježja elemenata osnovnog skupa.

Praktični primjer 1: Na osnovi uzorka treba ispitati da li se može prihvatiti hipoteza da distribucija domaćinstava prema broju zaposlenih ima binomni oblik, uz zadani nivo signifikantnosti testa.

Praktični primjer 2: Na osnovi uzorka treba ispitati da li se može prihvatiti hipoteza da aktivnost stanovništva na nekom području ne zavisi od spola, uz zadani nivo signifikantnosti testa.

43. Testiranje hipoteze da distribucija ima oblik neke teorijske distribucije. Postupak testiranja i primjeri.

Testira se hipoteza da distribucija slučajne varijable X ima oblik neke teorijske distribucije. Obično se koriste: jedniloka, binomna, Poissonova i normalna distribucija.

Test veličina je vrijednost empirijskog hi-kvadrata: $\chi^{*2} = \sum_{i=1}^{k} \frac{\left(f_i - f_{ti}\right)^2}{f_{ti}}$

gdje su: f_i - empirijske (originalne) frekvencije,

f_{ti} - teorijske (očekivane) frekvencije.

Teorijske frekvencije za **diskontinuirane** slučajne varijable računaju se pomoću izraza:

$$f_{ti} = p(x_i) \cdot \sum_{i=1}^{k} f_i$$

gdje $p(x_i)$ predstavlja vjerojatnost da slučajna varijabla X poprimi vrijednost x_i prema zakonu vjerojatnosti teorijske distribucije.

Kod testiranja hipoteze da distribucija ima normalni oblik ne računa se vjerojatnost u točki nego vjerojatnost nad određenim intervalom i također množi sa sumom empirijskih frekvencija.

Empirijska vrijednost Hi-kvadrata (χ^{*2}) se uspoređuje s kritičnom vrijednošću Hi-kvadrata (χ^2) iz tablica za određeni nivo signifikantnosti testa α i uz pripadni ukupan broj stupnjeva slobode.

Ako je $\chi^{*2} < \chi^2$ prihvaćamo hipotezu da empirijska distribucija ima pretpostavljeni oblik.

Ukupan broj stupnjeva slobode je:

$$df = \nu = k - \text{ukupan broj parametara teorijske distribucije} - 1$$

Vrijedi da je:

- Za jednoliku distribuciju $df = \nu = k - 1$
- Za binomnu distribuciju $df = \nu = k - 2$
- Za Poissonovu distribuciju $df = \nu = k - 2$
- Za normalnu distribuciju $df = \nu = k - 3$

Praktični primjer 1: Na osnovi uzorka treba ispitati da li se može prihvatiti hipoteza da distribucija domaćinstava prema broju zaposlenih ima binomni oblik, uz zadani nivo signifikantnosti testa.

Praktični primjer 2: Na osnovi uzorka treba ispitati da li se može prihvatiti hipoteza da distribucija telefonskih impulsa po sekundi na nekom području ima Poissonov oblik, uz zadani nivo signifikantnosti testa.

44. Testiranje hipoteze o nezavisnosti obilježja elemenata osnovnog skupa. Postupak testiranja i primjeri.

Hipoteze glase: H_o: .... $P_{ij} = P_{i\cdot} \cdot P_{\cdot j} \quad \forall i \forall j$

H_1: .... $\exists\, P_{ij} \neq P_{i\cdot} \cdot P_{\cdot j}$

Nultom hipotezom se pretpostavlja da ne postoji zavisnost dvaju obilježja.

Test veličina je vrijednost empirijskog Hi-kvadrata:

$$\chi^{*2} = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(m_{ij} - e_{ij}\right)^2}{e_{ij}}$$

gdje su: m_{ij} - empirijske frekvencije,

e_{ij} - teorijske frekvencije.

Teorijske frekvencije se računaju pomoću izraza:

$$e_{ij} = \frac{m_{i\bullet} \cdot m_{\bullet j}}{n}$$

gdje je: $m_{i\bullet}$ - marginalna frekvencija i-tog retka,

$m_{\bullet j}$ - marginalna frekvencija j-tog stupca,

n - veličina uzorka.

Empirijska vrijednost Hi-kvadrata (χ^{*2}) se uspoređuje s kritičnom vrijednošću Hi-kvadrata (χ^2) iz tablica za određeni nivo signifikantnosti testa α i uz pripadni ukupan broj stupnjeva slobode:

$$df = (r-1) \cdot (c-1)$$

gdje je: r - ukupan broj redaka u tablici kontigence,

c - ukupan broj stupaca u tablici kontigence.

Ako je $\chi^{*2} < \chi^2$ prihvaćamo nultu hipotezu da nema ovisnosti obilježja elemenata osnovnog skupa.

Praktični primjer 1: Na osnovi uzorka treba ispitati da li se može prihvatiti hipoteza da aktivnost stanovništva na nekom području ne zavisi od spola, uz zadani nivo signifikantnosti testa.

Praktični primjer 2: Na osnovi uzorka treba ispitati da li se može prihvatiti hipoteza da nema značajne razlike u interesu za sport između stanovništva sela i grada, uz zadani nivo signifikantnosti testa.

45. Testiranje hipoteze da je koeficijent linearne korelacije jednak nuli.

Hipoteze glase: H_o: .... $r = 0$

H_1: .... $r \neq 0$

Nultom hipotezom se pretpostavlja da ne postoji korelacija između slučajnih varijabli.

Interval prihvaćanja hipoteze H_o je: $0 \pm z \cdot Se(r)$ ako je $df = n-2 > 30$

$0 \pm t \cdot Se(r)$ ako je $df = n-2 \leq 30$

gdje je n veličina uzorka.

Standardna greška se računa pomoću izraza:

$$Se(r) = \sqrt{\frac{1}{n-1}} \qquad \text{ako je} \quad df = n-2 > 30$$

$$Se(r) = \sqrt{\frac{1-\hat{r}^2}{n-2}} \qquad \text{ako je} \quad df = n-2 \leq 30$$

gdje je $\hat{r}$ koeficijent korelacije izračunat na osnovi uzorka.

Testiranje se može provesti i pomoću z-testa ili t-testa:

$$z^* = \frac{\hat{r}}{Se(r)} \qquad \text{ako je} \quad df = n-2 > 30$$

$$t^* = \frac{\hat{r}}{Se(r)} \qquad \text{ako je} \quad df = n-2 \leq 30$$

Kritična vrijednost za **z** se vadi iz tablice površine ispod normalne krivulje, dok se kod malog uzorka kritična vrijednost za **t** vadi iz Studentove distribucije sa **(n-2)** stupnjeva slobode i uz signifikantnost testa α.

Hipoteza H_o se prihvaća ako je $|z^*| < z_{\alpha/2}$, odnosno ako je $|t^*| < t_{\alpha/2}$.

Značajnost Pearsonovog koeficijenta linearne korelacije r kao i **značajnost Spearmanovog koeficijenta** korelacije ranga r_S testiraju se na ovdje opisani način.

46. Objasnite kada se primjenjuje pojedina vrsta analize varijance, za koju vrstu varijabli. Što se dobiva analizom varijance? Koji su to zavisni, a koji nezavisni uzorci?

Analiza varijance s jednim promjenjivim faktorom koristi se u svrhu ispitivanja djelovanja faktora A na vrijednost slučajne varijable X.

Analiza varijance s dva promjenjiva faktora koristi se u svrhu ispitivanja djelovanja faktora A i faktora B na vrijednost slučajne varijable X.

Ova dva testiranja se svode na testiranje značajnosti razlike aritmetičkih sredina više osnovnih skupova (populacija), odnosno testira se hipotezu da svi uzorci potječu iz iste populacije. Testiranje se obično vrši standardnim F-testom, a uzorci su nezavisni.

Analiza varijance kod ponovljenih mjerenja koristi se u slučaju kada se na **istome uzorku** više puta mjeri vrijednost neke varijable. Najčešća primjena je mjerenje neke numeričke varijable na istim ispitanicima (ili proizvodima) u različito vrijeme ili pod različitim uvjetima.

Analiza varijance s jednim promjenjivim faktorom pomoću rangova (Kruskal-Wallisov test**)** koristi se kada nije moguće provesti analizu varijance na osnovi testiranja značajnosti razlike aritmetičkih sredina više osnovnih skupova jer nisu ispunjeni neki od potrebnih uvjeta. Ovaj test je neparametrijski, a uzorci su nezavisni.

Analiza varijance kod ponovljenih mjerenja ako je skala mjerenja ordinalna koristi se u slučaju kada se na **istome uzorku** više puta mjere **rangovi**. Ovaj test je analogan Kruskal-Wallisovome testu, s tom razlikom što su ovdje uzorci zavisni.

Dakle, **analizom varijance** se istražuje djelovanje različitih promjenjivih faktora na vrijednost slučajne varijable X.

Nezavisni uzorci su uzorci odabrani iz međusobno nezavisnih populacija.

Zavisni uzorci su uzorci odabrani iz međusobno zavisnih populacija.

47. Što se sve testira analizom varijance? Navedite sve testove i kada se koji od njih provodi (bez formula).

Analizom varijance se testira djelovanje različitih promjenjivih faktora na vrijednost slučajne varijable X.

Analiza varijance s jednim promjenjivim faktorom koristi se u svrhu ispitivanja djelovanja faktora A na vrijednost slučajne varijable X. Testira se hipotezu da svi uzorci potječu iz iste populacije.

Analiza varijance s dva promjenjiva faktora koristi se u svrhu ispitivanja djelovanja faktora A i faktora B na vrijednost slučajne varijable X.

Analiza varijance kod ponovljenih mjerenja koristi se u slučaju kada se na **istome uzorku** više puta mjeri vrijednost neke varijable. Najčešća primjena je mjerenje neke numeričke varijable na istim ispitanicima (ili proizvodima) u različito vrijeme ili pod različitim uvjetima.

Analiza varijance s jednim promjenjivim faktorom pomoću rangova (Kruskal-Wallisov test**)** koristi se u svrhu neparametrijske analize varijance pomoću rangova. Uzorci su nezavisni.

Analiza varijance kod ponovljenih mjerenja ako je skala mjerenja ordinalna koristi se u slučaju kada se na **istome uzorku** više puta mjere **rangovi**. Ovaj test je analogan Kruskal-Wallisovome testu, s tom razlikom što su ovdje uzorci zavisni.

48. Kako se provodi analiza varijance s jednim i s dva promjenjiva faktora. Postupak testiranja i primjeri. Koje još analize varijance poznajete?

Analiza varijance s jednim i s dva promjenjiva faktora svodi se na testiranje značajnosti razlike aritmetičkih sredina više osnovnih skupova (populacija), odnosno testira se hipotezu da svi uzorci potječu iz iste populacije. Testiranje se obično vrši standardnim F-testom, a uzorci su nezavisni.

Analiza varijance podrazumijeva raščlanjivanje varijance slučajne varijable u komponente, prema određenim izvorima varijacije.

Analiza varijance s jednim promjenjivim faktorom koristi se u svrhu ispitivanja djelovanja faktora A na vrijednost slučajne varijable X.

Hipoteze glase: H_o: ..... $\sigma_A^2 = 0$; H_1: .... $\sigma_A^2 \neq 0$

Tablica analize varijance (ANOVA)

Izvor varijacije	Zbroj kvadrata odstupanja	Stupnjevi slobode	Ocjena varijance
Između uzoraka	$\sum_{j=1}^{k} n_j \left(\bar{X}_{\bullet j} - \bar{X}_{\bullet\bullet}\right)^2$	k - 1	S_A^2
Unutar uzoraka	$\sum_{j=1}^{k}\sum_{i=1}^{n_j} \left(X_{ij} - \bar{X}_{\bullet j}\right)^2$	n - k	S_u^2
Ukupno	$\sum_{j=1}^{k}\sum_{i=1}^{n_j} \left(X_{ij} - \bar{X}_{\bullet\bullet}\right)^2$	n - 1	

Empirijski F-omjer glasi: $$F^* = \frac{\sum_{j=1}^{k} n_j \left(\bar{X}_{\bullet j} - \bar{X}_{\bullet\bullet}\right)^2 \Big/ (k-1)}{\sum_{j=1}^{k}\sum_{i=1}^{n_j} \left(X_{ij} - \bar{X}_{\bullet j}\right)^2 \Big/ (n-k)} = \frac{S_A^2}{S_u^2}$$

Empirijski F-omjer se uspoređuje s tabličnom vrijednošću F-varijable za pripadne stupnjeve slobode i nivo signifikantnosti testa α.

Ako je:

$$F^* > F^{\alpha}_{(k-1;\, n\text{-}k)}$$

prihvaćamo alternativnu hipotezu H_1 da je djelovanje faktora A **značajno**. Ako je signifikantnost testa manja od 5 %, kažemo da je djelovanje faktora A **statistički značajno**.

Analiza varijance s dva promjenjiva faktora koristi se u svrhu ispitivanja djelovanja faktora A i faktora B na vrijednost slučajne varijable X.

Hipoteze glase: H_o: ... $\sigma_A^2 = 0$; H_1: ... $\sigma_A^2 \neq 0$ i H_o: ... $\sigma_B^2 = 0$; H_1: ... $\sigma_B^2 \neq 0$

Tablica analize varijance (ANOVA)

Izvor varijacije	Zbroj kvadrata odstupanja	Stupnjevi slobode	Ocjena varijance
Između redaka	$\sum_{i=1}^{c} n_i \left(\bar{X}_{i\bullet} - \bar{X}_{\bullet\bullet}\right)^2$	c-1	S_A^2
Između stupaca	$\sum_{j=1}^{k} n_j \left(\bar{X}_{\bullet j} - \bar{X}_{\bullet\bullet}\right)^2$	k - 1	S_B^2
Ostatak	$\sum_{j=1}^{k}\sum_{i=1}^{c} \left(X_{ij} - \bar{X}_{i\bullet} - \bar{X}_{\bullet j} + \bar{X}_{\bullet\bullet}\right)^2$	n - k - c + 1	S_R^2
Ukupno	$\sum_{j=1}^{k}\sum_{i=1}^{c} \left(X_{ij} - \bar{X}_{\bullet\bullet}\right)^2$	n - 1	

Za oba parametra empirijski F-omjer se uspoređuje s tabličnom vrijednošću F-varijable za pripadne stupnjeve slobode i nivo signifikantnosti testa α.

Testiranje značajnosti djelovanja faktora A:

$$F^* = \frac{S_A^2}{S_R^2} \quad , \quad F^* > F^{\alpha}_{(c-1;\ n\text{-}k\text{-}c+1)} \Rightarrow$$ djelovanje faktora A je **značajno.**

Testiranje značajnosti djelovanja faktora B:

$$F^* = \frac{S_B^2}{S_R^2} \quad , \quad F^* > F^{\alpha}_{(k-1;\ n\text{-}k\text{-}c+1)} \Rightarrow$$ djelovanje faktora B je **značajno.**

Ostale analize varijance su:

- **Analiza varijance kod ponovljenih mjerenja** koja se koristi se u slučaju kada se na **istome uzorku** više puta mjeri vrijednost neke varijable.
- **Analiza varijance s jednim promjenjivim faktorom pomoću rangova (**Kruskal-Wallisov test**)** koristi se u svrhu neparametrijske analize varijance pomoću rangova. Uzorci su nezavisni.
- **Analiza varijance kod ponovljenih mjerenja ako je skala mjerenja ordinalna** koristi se u slučaju kada se na **istome uzorku** više puta mjere **rangovi**. Ovaj test je analogan Kruskal-Wallisovome testu, s tom razlikom što su ovdje uzorci zavisni.

49. Napišite sve što znate o korelaciji.

Korelacija je međuzavisnost, odnosno pvezanost slučajnih varijabli**.**

Korelacija može biti pozitivna i negativna.

Pozitivna korelacija $\Rightarrow$ rast jedne varijable prati rast druge varijable.

Negativna korelacija $\Rightarrow$ rast jedne varijable prati pad druge varijable.

Ako je korelacija potpuna, govorimo o **funkcionalnoj povezanosti** varijabli.

Ako je korelacija djelomična, govorimo o stohastičkoj ili statističkoj vezi varijabli.

Najvažnija mjera linearne korelacije je **Pearsonov koeficijent linearne korelacije**:

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\cdot\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \cdot \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}} = \frac{\sum_{i=1}^{n} X_i \cdot Y_i - n \cdot \bar{X} \cdot \bar{Y}}{\sqrt{\left(\sum_{i=1}^{n} X_i^2 - n \cdot \bar{X}^2\right)\cdot\left(\sum_{i=1}^{n} Y_i^2 - n \cdot \bar{Y}^2\right)}}$$

Vrijednost koeficijenta $-1 \leq r \leq 1$.

Ako je r = 0, onda korelacija uopće ne postoji.

Ako je r =1, onda se radi o pozitivnoj funkcionalnoj povezanosti.

Ako je r = - 1, onda se radi o negativnoj funkcionalnoj povezanosti.

Spearmanov koeficijent korelacije ranga je opisan izrazom:

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n^3 - n} \quad ; \quad -1 \leq r_s \leq 1$$

gdje je n ukupan broj parova vrijednosti varijabli X i Y, dok je d_i razlika rangova pripadnih vrijednosti varijabli X i Y.

Spearmanov koeficijent korelacije ranga se najčešće koristi za mjerenje korelacije između redosljednih obilježja ili obilježja ranga.

Osnovna razlika između Pearsonovog i Spearmanovog koeficijenta korelacije je u tome što se Pearsonov koeficijent linearne korelacije računa na osnovi numeričkih vrijednosti varijabli X i Y, dok se Spearmanovog koeficijenta korelacije ranga računa na osnovi rangiranja vrijednosti varijabli X i Y.

50. Koja je razlika između Spearmanovoga i Pearsonovoga koeficijenta korelacije?

Osnovna razlika između Pearsonovog i Spearmanovog koeficijenta korelacije je u tome što se Pearsonov koeficijent linearne korelacije računa na osnovi numeričkih vrijednosti varijabli X i Y, dok se Spearmanovog koeficijenta korelacije ranga računa na osnovi rangiranja vrijednosti varijabli X i Y.

51. Opišite (Pearsonov) koeficijent linearne korelacije?

Korelacija je međuzavisnost, odnosno pvezanost slučajnih varijabli**.**

Pearsonov koeficijent linearne korelacije je najvažnija mjera linearne korelacije. Opisan je izrazom:

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\cdot\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \cdot \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}} = \frac{\sum_{i=1}^{n} X_i \cdot Y_i - n\cdot\overline{X}\cdot\overline{Y}}{\sqrt{\left(\sum_{i=1}^{n} X_i^2 - n\cdot\overline{X}^2\right)\cdot\left(\sum_{i=1}^{n} Y_i^2 - n\cdot\overline{Y}^2\right)}}$$

Vrijednost koeficijenta $-1 \le r \le 1$.

Ako je r = 0, onda korelacija uopće ne postoji.

Ako je r =1, onda se radi o pozitivnoj funkcionalnoj povezanosti.

Ako je r = - 1, onda se radi o negativnoj funkcionalnoj povezanosti.

Ako je korelacija djelomična, govorimo o stohastičkoj ili statističkoj vezi varijabli.

52. Opišite (Spearmanov) koeficijent korelacije ranga?

Korelacija je međuzavisnost, odnosno pvezanost slučajnih varijabli**.**

Spearmanov koeficijent korelacije ranga najčešće se koristi za mjerenje korelacije između redosljednih obilježja ili obilježja ranga. Opisan je izrazom:

$$r_s = 1 - \frac{6\cdot\sum_{i=1}^{n} d_i^2}{n^3 - n} \quad ; \quad -1 \le r_s \le 1$$

gdje je n ukupan broj parova vrijednosti varijabli X i Y, dok je d_i razlika rangova pripadnih vrijednosti varijabli X i Y.

Ako je r = 0, onda korelacija uopće ne postoji.

Ako je r =1, onda se radi o pozitivnoj funkcionalnoj povezanosti.

Ako je r = - 1, onda se radi o negativnoj funkcionalnoj povezanosti.

Ako je korelacija djelomična, govorimo o stohastičkoj ili statističkoj vezi varijabli.

Spearmanovog koeficijenta korelacije ranga računa na osnovi rangiranja vrijednosti varijabli X i Y, a ne na osnovi njihovih numeričkih vrijednosti.

53. Napišite sve što znate o linearnoj regresiji.

Modelom jednostruke linearne regresije analitički se izražava odnos između dviju varijabli:

- Y – zavisne varijable (regresanda),
- X – nezavisne varijable (regresora).

Linearna regresijska jednadžba glasi: $\hat{Y} = \hat{\beta}_o + \hat{\beta}_1 \cdot X$

gdje je: $\hat{Y}$ - regresijska funkcija,

X – nezavisna varijabla,

$\hat{\beta}_o$ i $\hat{\beta}_1$ - parametri.

Na osnovi zadanog skupa od n parova vrijednosti varijabli X i Y, parametri regresije se računaju iz izraza:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i \cdot Y_i - n \cdot \overline{X} \cdot \overline{Y}}{\sum_{i=1}^{n} X_i^2 - n \cdot \overline{X}^2} \quad ; \quad \hat{\beta}_o = \overline{Y} - \hat{\beta}_1 \cdot \overline{X}$$

Izrazi za procjenu parametara su dobiveni pomoću metode najmanjih kvadrata, tj. iz kriterija procjene: $\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 = \min$

Linearna regresijska jednadžba je ustvari jednadžba regresijskog **pravca** $\hat{Y}$. Stoga je značenje parametara u regresijskoj jednadžbi:

- $\hat{\beta}_1$ je koeficijent smjera regresijskog pravca $\hat{Y}$,
- $\hat{\beta}_o$ je odsječak regresijskog pravca $\hat{Y}$ na osi Y, odnosno to je vrijednost regresijske funkcije $\hat{Y}$ za X = 0.

U praktičnim primjerima parametar $\hat{\beta}_o$ može ali ne mora imati logičnu interpretaciju.

Analiza varijance linearne regresije vrši se na osnovi izraza: ST = SP + SR

gdje je: ST - suma kvadrata ukupnih odstupanja vrijednosti varijable Y od $\overline{Y}$,

SP - suma kvadrata protumačenog dijela odstupanja,

SR - suma kvadrata neprotumačenog (rezidualnog) dijela odstupanja.

Reprezentativnost regresije može se mjeriti pomoću **koeficijenta determinacije**:

$$r^2 = \frac{SP}{ST} = \frac{\sum_{i=1}^{n} \left(\hat{Y}_i - \overline{Y} \right)^2}{\sum_{i=1}^{n} \left(Y_i - \overline{Y} \right)^2} \quad ; \quad 0 \le r^2 \le 1$$

54. Opišite postupak računanja parametara u linearnoj regresiji. Što znače parametri u regresijskoj jednadžbi?

Linearna regresijska jednadžba glasi:

$$\hat{Y} = \hat{\beta}_o + \hat{\beta}_1 \cdot X$$

gdje je: $\hat{Y}$ - regresijska funkcija,

X – nezavisna varijabla,

$\hat{\beta}_o$ i $\hat{\beta}_1$ - parametri.

Na osnovi zadanog skupa od n parova vrijednosti varijabli X i Y, parametri regresije se računaju iz izraza:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i \cdot Y_i - n \cdot \overline{X} \cdot \overline{Y}}{\sum_{i=1}^{n} X_i^2 - n \cdot \overline{X}^2} \quad ; \quad \hat{\beta}_o = \overline{Y} - \hat{\beta}_1 \cdot \overline{X}$$

Izrazi za procjenu parametara su dobiveni pomoću metode najmanjih kvadrata, tj. iz kriterija procjene:

$$\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 = \min$$

Linearna regresijska jednadžba je ustvari jednadžba regresijskog **pravca** $\hat{Y}$. Stoga je značenje parametara u regresijskoj jednadžbi:

- $\hat{\beta}_1$ je koeficijent smjera regresijskog pravca $\hat{Y}$,
- $\hat{\beta}_o$ je odsječak regresijskog pravca $\hat{Y}$ na osi Y, odnosno to je vrijednost regresijske funkcije $\hat{Y}$ za X = 0.

U praktičnim primjerima parametar $\hat{\beta}_o$ može ali ne mora imati logičnu interpretaciju.