

## Tematika, fogalmak, mintafeladatok

A zh-t 1-2 elméleti kérdés és olyan feladatok fogják alkotni amelyek az órán illetve házikban előfordult/előforduló feladatokhoz hasonlóak.

### Tematika, fogalmak

Véletlen jelenség, esemény, eseménytér, kísérlet, kísérletsorozat.

Eseményalgebra: események uniója, metszete, különbsége, komplementer esemény; biztos esemény, lehetetlen esemény.

Valószínűség fogalma, valószínűséggel kapcsolatos számolási szabályok.

Diszkrét modell. Klasszikus probléma (minden kimenetel azonos valószínűségű).

Feltételes valószínűség, független események.

Valószínűségi változó. Diszkrét valószínűségi változó súlyfüggvénye. Nevezetes diszkrét valószínűségű változók: egyenletes eloszlás, binomiális eloszlás, hipergeometriai eloszlás. Folytonos eloszlások: eloszlásfüggvény, sűrűségfüggvény. Nevezetes folytonos eloszlások: intervallumon egyenletes eloszlás,  $\lambda$  paraméterű exponenciális eloszlás, standard normális és  $(m, \sigma)$  paraméterű normális eloszlás.

Várható érték, valószínűségi változó függvénynek várható értéke, szórás fogalma. Kiszámítása diszkrét és folytonos esetben. Várható érték és szórás tulajdonságai. Nevezetes eloszlások várható értékei, szórásai.

Kétdimenziós eloszlások. Független valószínűségi változók.

Normális eloszlás fontossága: Centrális határeloszlás tétel.

Közelítő statisztikák: átlag, empirikus szórásnégyzet, ferdeség, csúcsosság, átlag standard hibája (standard error of mean), medián, alsó és felső kvartilis.

Kétoldali konfidenciaintervallum normális eloszlás várható értékére ismert és ismeretlen szórás esetén.

Hipotézisvizsgálat: elsőfajú és másodfajú hibavalószínűség. Egymintás, kétoldali u- és t-próbák.

Összetartozó és független kétmintás t-próbák.  $\chi^2$ -próbák illeszkedés és függetlenség vizsgálatra.

Lineáris regresszió.

Becslélmélet: Maximum-likelihood becslés.

**Néhány papíron kidolgozandó mintafeladat (természetesen az Excel számolási segédeszközként használható)**

1. Egy szabályos kockával dobunk. Tekintsük az alábbi eseményeket: A: 4-nél kevesebbet dobunk, B: négyzetszámot dobunk, C: páros számot dobunk, D: 6-ot dobunk.

(a) Határozza meg azon eseménypárokat, melyek kizáróak.

(b) Határozza meg azon eseménypárokat, melyek függetlenek.

- (c) Határozza meg a  $P((B \cup C) \setminus D)$  valószínűséget.
2. Fogalmazzon meg olyan feladatot, amit a binomiális eloszlással/hipergeometriai eloszlással/exponenciális eloszlással modellezünk.
3. Egy piros és egy kék szabályos kockával dobunk. Mi a valószínűsége, hogy
- a kék kockával nagyobbat dobunk mint a pirossal?
  - dobunk 6-ost?
4. Egy szabályos pénzérmével kétszer dobok, majd ahány fejet dobtam, annyiszor újra dobok.
- Határozza meg a második dobássorozatnál az írások számának eloszlását, várható értékét.
  - Mia valószínűsége, hogy ha a második dobássorozatban 0 írást dobtam, akkor az első dobássorozatban két fejet dobtam?
5. Tekintsük az  $F(x) = \begin{cases} 0 & \text{ha } x < 0 \\ x^2 & \text{ha } 0 \leq x \leq 1 \\ 1 & \text{ha } x > 1 \end{cases}$  eloszlásfüggvénnyel rendelkező  $X$  valószínűségi változót. Határozza meg a
- $P(X < 0,5)$
  - $P(0,3 < x < 0,7)$
  - $P(X > 0,2)$  valószínűségeket
  - az  $X$  valószínűségi változó sűrűségfüggvényét
6. Határozza meg az  $RND^2$
- eloszlásfüggvényét
  - sűrűségfüggvényét
  - várható értékét
  - szórását
7. Legyen  $X$  egy  $(2,3)$  paraméterű normális eloszlású valószínűségi változó. Határozza meg a
- $P(X < 4)$
  - $P(2 < X < 5)$
  - $P(X > 0,5)$  valószínűségeket.
  - Határozza meg az  $x$  számot úgy, hogy  $P(X > x) = 0,1$  teljesüljön.

8. Egy szabályos kockával 720-szor dobunk.
  - (a) Határozza meg pontosan annak a valószínűségét, hogy legfeljebb 110 hatost dobunk.
  - (b) Közelítse a fenti valószínűséget a normális eloszlással a Centrális határeloszlás tételt használva.
9. Az építőmérnöki vállalkozásunk egy alvállalkozó révén jut 25 kg-os cementes zsákokhoz. A legutóbbi beszállításnál próbaképpen lemértük 10 zsák tömegét. A tömegekre a következők adódtak: 24.5, 25.1, 24.7, 24.9, 25.2, 24.8, 24.4, 24.9, 24.8, 24.7. Adjunk meg a zsákok várható értékére 0,95 megbízhatóságú konfidenciaintervallumot. Oldjuk meg a feladatot amikor
  - (a) tudjuk, hogy a cementes zsákok tömegének szórása 0.2 kg;
  - (b) nem ismerjük a cementes zsákok tömegének szórását.
10. Az  $X_1, X_2, \dots, X_n$  mintát a  $\lambda$  paraméterű exponenciális eloszlású háttérváltozó esetén kaptuk. Adjunk maximum-likelihood becslést a  $\lambda$  paraméterre.

### Néhány Excel mintafeladat

1. Feldobunk 3 szabályos dobókockát. Jelölje  $X$  a három kockadobás maximumát. Közelítsd  $X$  súlyfüggvényét 1500 elemű szimulációra támaszkodva! Ábrázold oszlopdiagramon is az adódó relatív gyakoriságokat!
2. Nyisd meg Excel-ben a házi feladat II-ben is szereplő teszteredmények.xlsx fájlt! Válaszolj meg független kétmintás t-próbával a következő kérdést (a próbák legyenek kétoldaliak továbbá az előírt elsőfajú hibavalószínűség legyen 0.03, ne tételezd fel a szórások egyezőségét): A megírt tesztek közül (utolsó 5 változó) melyek azok, amelyeknél szignifikánsan különbözik a lányok és fiúk teljesítménye?  
  
Megjegyzés: A független kétmintás t-próbának feltétele a csoportok normalitása. De a centrális határeloszlás tétel miatt nagy minta esetén akkor is használhatók a várható értékek egyezőségének tesztelésére ha nem teljesül a normalitás.
3. Töltsd be a honlapon található dieata.xlsx file-t! Sikeres betöltés után 15 kísérleti alany diéta előtti és diéta utáni 9 perc alatti futásteljesítményét találod a file-ban (az adatok nem valódiak). Vizsgáld meg t-próbával (tedd fel, hogy a szükséges normalitási feltétel teljesül, a szórás nem ismert), hogy a diéta befolyásolta-e a teljesítményt! Dolgozz 0.04 elsőfajú hibavalószínűséggel!
4. Az előző feladatban szereplő futásteljesítményeket tartalmazó file-ban számold ki a különbségváltozót majd szerkessz a különbségváltozó várható értékére 96%-os konfidencia

intervallumot (vagyis a hibavalószínűség 0.04 legyen)! Itt is tedd fel, hogy a szükséges normalitási feltétel teljesül (a szórás itt sem ismert).

Megjegyzés: A konfidencia intervallum és az egymintás t-próba kapcsolatából és abból a tényből, hogy az összetartozó kétmintás t-próba igazából a különbségváltozó 0 várható értékűségét tesztelő egymintás t-próba következik, hogy a 3-as feladatot úgy is meg lehet oldani, hogy azt nézzük meg, hogy a 0 beleesik-e az itt kapott intervallumba.

5. Legyenek  $X_1, X_2, \dots, X_k$  független standard normális eloszlású valószínűségi változók. A  $Z = X_1^2 + \dots + X_k^2$  négyzetösszeg eloszlását  $k$  szabadságfokú khi-négyzet eloszlásnak nevezzük. A feladat szempontjából nem az előbbi definíció hanem az a lényeges, hogy az Excel tartalmazza beépítve a khi-négyzet eloszlás eloszlás- illetve sűrűségfüggvényét. Az Excel beépített KHINÉGYZET.ELOSZLÁS() függvényét használva készíts egy olyan ábrát amely rögzített szabadságfok mellett kirajzolja az eloszlás sűrűségfüggvényét! Mindezt tedd úgy, hogy a rögzített szabadsági fok változtatható paraméter legyen. Az ábrára második görbeként rajzold rá a  $\frac{1}{2}$  paraméterű exponenciális eloszlás sűrűségfüggvényét is (pl az EXP.ELOSZLÁS() függvény használatával). Mely szabadsági fokra egyezik meg a két görbe?

Megjegyzés: A nyolcadik órán készített Excel file harmadik munkalapjához hasonló ábrát várunk (itt a különböző szabadságfokú Student t eloszlások sűrűségfüggvényei vannak a standard normális eloszlás sűrűségfüggvényével közös ábrán ábrázolva) .

6. Egy az 1, 2 illetve 3 lehetséges értékekkel rendelkező diszkrét háttérváltozóból van 400 elemű mintánk. Az a hipotézisünk, hogy a háttérváltozó az 1-est, 2-est illetve a 3-ast  $\frac{1}{4}, \frac{1}{4}$  illetve  $\frac{1}{2}$  valószínűségekkel veszi fel. A mintában 87 1-es, 94 2-es és 219 3-as van. Teszteld le chi-négyzet próbával a hipotézisünket 0.05 elsőfajú hibavalószínűség mellett!