

A gyakorló feladatok megoldásai (2. alkalom)

1. A kérdés leegyszerűsítve: a nők átlagosan több egységre emlékeznek, mint a férfiak? A nők és a férfiak nem rendelhetők párokba egyértelműen. A kérdés ezért kétmintás t-próbával vizsgálható. A való életben normalitásvizsgálatot kellene előbb végeznünk, mindkét változóra, de ez nem volt feladat.

Ha a C meghajtóra mentettem az adatbázist, akkor az r-parancs pl. a következő:

```
a=read.table("C:\\memoria.txt", sep=";", header=F)
```

```
a=as.matrix(a)
```

```
t.test(a[,1],a[,2])
```

Itt nem voltak a változók elnevezve!

Eredmény: $t = -2.286$, $p\text{-value} = 0.037$, azaz szignifikáns a különbség, a populáció szintjén is eltérést jelez. Az átlagok 12 és 18 voltak.

2. Egy minta átlagértékét vetjük össze egy feltételezett populációbeli értékkel, ezért egymintás t-próbát kellene alkalmaznunk. Ezt akkor tehetjük meg, ha a változónk normál eloszlású, erre a minta alapján következtetünk. Grafikusan a hisztogramot vizsgálhatjuk, de a ferdeségi és csúcsossági együttható segítségével tesztelhetjük a változó normalitását.

Ha a C meghajtóra mentettem az adatbázist, akkor az r-parancs pl. a következő:

```
a=read.table("C:\\kosar.txt", sep=";", header=T)
```

```
a=as.matrix(a)
```

```
hist(a)
```

A látvány nem túl biztató, de ne feledjük, hogy igen kicsi a mintanagyság.

g1 számítása:

```
b=sum(((a-mean(a))/sd(a))**3)/15
```

```
g1=0.038
```

a hozzá tartozó standard hiba számítása:

```
c=sqrt(6/15)
```

```
eredmény:0.63
```

```
z-érték:0.06
```

```
p-érték:
```

$1\text{-pnorm}(0.060,0,1)=0.476$, azaz az eloszlás szimmetrikus lehet a populáció szintjén vizsgálva.

g2 számítása:

```
b=sum(((a-mean(a))/sd(a))**4)/15-3
```

```
g2=-0.659
```

a hozzá tartozó standard hiba számítása:

```
c=sqrt(24/15)
```

```
eredmény:1.26
```

```
z-érték:-0.52
```

```
p-érték:
```

$\text{pnorm}(-0.52,0,1)=0.302$, azaz az eloszlás csúcsossága a normál eloszlásnak megfelelő a populáció szintjén vizsgálva.

Mindezek alapján elvégezhetjük az egymintás t-próbát:

```
t.test(a, mu=72)
```

$t = -3.3$, $df = 14$, $p\text{-value} = 0.005$

Azaz szignifikáns mértékben különbözik a minta átlaga a populációban feltételezett értéktől. Így vagy a varsói kosarasok különböznek a lengyelektől, vagy a populációban feltételezett érték nem helytálló.

3. Ugyanaz a gyerek játszik a színes és a szürke játékkal is. Így az adatok egyértelműen párokba rendezhetők, a színes és a szürke játékkal töltött idő gyerekenként párokat alkot. Ezért páros t-próbát használhatunk, ha a változók a populáció szintjén normál eloszlásúak.

```
a=read.table("C:\\jatek.txt", sep="," , header=T)
```

```
t.test(a$színes, a$szürke, paired=T)
```

$t = 2.8971$, $df = 16$, $p\text{-value} = 0.011$

Szignifikáns a különbség. És az eltérés iránya is a vártak megfelelő, a színes játékkal játszanak többet.

4. A kettes feladatnak megfelelően végezzük a számításokat, itt csak az eredményeket tüntetem fel.

színes játékhoz tartozó időértékeket tartalmazó változó:

$g1 = -1.967$

$z1 = -3.3$

$s1 = 0.594$

$p1 = 0.00048$

Azaz a változó ferdesége a minta alapján eltér a normál eloszlástól. Ez önmagában azt jelenti, hogy a 3. pontban leírt eredmények torzak lehetnek és t-próba helyett az ennek megfelelő nemparaméteres eljárást kellene használnunk.

$g2 = 2.457$

$z2 = 2.07$

$s2 = 1.188$

$p2 = 0.019$

Azaz a változó csúcsossága sem a normál eloszlásnak megfelelő a minta alapján.

Szürke játékhoz tartozó időértékek változója:

$g1 = 1.9$

$z1 = 3.2$

$p1 = 0.00067$

Azaz a változó ferdesége a minta alapján eltér a normál eloszlásától.

$g2 = 2.37$

$z2 = 1.996$

$p2 = 0.023$

Azaz a változó csúcsossága sem a normál eloszlásnak megfelelő a minta alapján.

Ez azt jelenti, hogy a 3. feladat megoldásának eredményei nem használhatóak, nemparaméteres eljárás használatára lett volna szükség.

5.

A. Különböző gyerekek vannak a délelőtti csoportban és a délutániban, legalábbis semmi nem utal az ellenkezőjére. Ezért kétmintás t-próbát alkalmazhatunk a paraméteres próbák közül.

B. Ugyanazoknak a termékeknek az árát nézi a két boltban, egyértelmű a hozzárendelés, ezért páros t-próbát kell alkalmaznunk, ha az lehetséges.

- C. Egyetlen változó átlagát egy feltételezett átlagértékhez hasonlítjuk, ha a változó normál eloszlást mutat a populáció szintjén, akkor egymintás t-próbát alkalmazhatunk.