-->

math.e

Hrvatski matematički elektronički časopis

O primjeni programa Maxima u nastavi predmeta Vjerojatnost i statistika na specijalističkom stručnom studiju graditeljstva

Maxima **nastava matematike** **statistia**

Mandi Orlić Bachler

Tehničko veleučilište u Zagrebu, Graditeljski odjel, 10000 Zagreb, Av. V. Holjevca 15, Hrvatska
mandi.orlic@tvz.hr

Sažetak

U članku se iznose dosadašnja iskustva o primjeni računalnog programa Maxima u nastavi predmeta Vjerojatnost i statistika, koji se izvodi na Politehničkom diplomskom specijalističkom stručnom studiju graditeljstva Tehničkog veleučilišta u Zagrebu. Opisuju se načini postizanja ishoda učenja navedenog predmeta primjenom programa Maxima, te se predlažu mogućnosti dodatne modernizacije spomenutog predmeta korištenjem ovog ili sličnih besplatnih računalnih programa. S ciljem boljeg prikaza teme u članak su uvršteni i zadaci, pri čemu je dana diskusija rješenja, koja su dobivena na dva načina: primjenom programa Maxima i klasičnim ("ručnim") računanjem.

Ključni pojmovi: vjerojatnost, statistika, Maxima, besplatni računalni program

1 Uvod

U prvom semestru Politehničkog diplomskog specijalističkog stručnog studija graditeljstva Tehničkog veleučilišta u Zagrebu izvodi se predmet Vjerojatnost i statistika. Cilj tog predmeta je upoznati studente s osnovnim pojmovima vjerojatnosti i metodama za statističku obradu podataka. Fond sati tog predmeta je 15+13+2, od čega je 15 sati predviđeno za predavanja, 13 sati za auditorne vježbe, a 2 sata za laboratorijske vježbe koje se održavaju u računalnom laboratoriju. Teme predviđene nastavnim planom i programom odnose se na: klasičnu definiciju vjerojatnosti, operacije među događajima, uvjetnu i totalnu vjerojatnost, diskretne i kontinuirane slučajne varijable, statističku populaciju i slučajni uzorak, grafičko prikazivanje statističkih podataka, procjenitelje, intervalnu procjenu očekivanja i varijance te testiranje hipoteza. Do ove akademske godine u sklopu laboratorijskih vježbi zadaci su se rješavali pomoću MS Excela. Međutim, kako se od akademske godine 2015./2016. u sklopu predmeta Računarstvo u graditeljstvu, koji se izvodi u prvom semestru preddiplomskog stručnog studija graditeljstva, obrađuje program Maxima, ove akademske godine umjesto MS Excela koristili smo program Maxima. Osim na predmetu Vjerojatnost i statistika Maximu koristimo i u sklopu ostalih matematičkih predmeta koji se izvode na studijima graditeljstva Tehničkog veleučilišta u Zagrebu, o čemu smo pisali u [1].

Računalni program Maxima, besplatni je program koji se može preuzeti na web adresi http://maxima.sourceforge.net/. Općenito, program je pogodan za različite algebarske operacije sa simboličkim i numeričkim izrazima, kao što su deriviranje, integriranje, razvoj u Taylorov red, Laplaceova transformacija, sustavi linearnih jednadžbi, vektori, matrice, statistika itd.

U ovom članku prikazano je kako se pomoću programa Maxima mogu riješiti neki od zadataka koji se obrađuju u sklopu nastave predmeta Vjerojatnost i statistika.

2 Statističke funkcije programa Maxima

U ovom dijelu rada dan je pregled i opis onih funkcija programa Maxima, koje koristimo u sklopu nastave predmeta Vjerojatnost i statistika. To su funkcije iz paketa `descriptive`, `distrib` i `stats`.

Paket `descriptive` sadrži funkcije deskriptivne statistike i funkcije za grafički prikaz podataka. Sljedeće funkcije samo su neke od dostupnih funkcija ovog paketa. Njihov argument x može biti lista ili matrica.

- Funkcija `mean(x)` računa uzoračku aritmetičku sredinu.
- Funkcija `var(x)` računa uzoračku varijancu.
- Funkcija `var1(x)` računa korigiranu uzoračku varijancu.
- Funkcija `std(x)` računa uzoračku standardnu devijaciju.
- Funkcija `std1(x)` računa korigiranu standardnu devijaciju.
- Funkcija `histogram(x, opcije)` koristi se za izradu histograma.

Paket `distrib` sadrži funkcije za izračun vjerojatnosti diskretnih i kontinuiranih univarijatnih modela. Dalje su opisane one funkcije koje se odnose na slučajne varijable s normalnom odnosno binomnom razdiobom.

- Za slučajnu varijablu s normalnom razdiobom $N(m, s)$ funkcija:
  - `pdf_normal(x,m,s)` za podatak x računa vrijednost funkcije gustoće vjerojatnosti,
  - `cdf_normal(x,m,s)` za podatak x računa vrijednost funkcije razdiobe vjerojatnosti,

  - mean_normal(m,s) računa vrijednost očekivanja,
  - var_normal(m,s) računa vrijednost varijance,
  - std_normal(m,s) računa vrijednost standardne devijacije.
- Za slučajnu varijablu s binomnom razdiobom $B(n,p)$ funkcija:
  - pdf_binomial(x,n,p) za podatak x računa vrijednost funkcije vjerojatnosti,
  - cdf_binomial(x,n,p) za podatak x računa vrijednost funkcije razdiobe,
  - mean_binomial(n,p) računa vrijednost očekivanja,
  - var_binomial(n,p) računa vrijednost varijance,
  - std_binomial(n,p) računa vrijednost standardne devijacije.

Paket stats sadrži funkcije za izvođenje zaključaka o populaciji na temelju svojstava uzorka.

- Funkcija test_mean(x,opcije) koristi se za testiranje hipoteza i određivanje intervala povjerenja za očekivanje s poznatom ili nepoznatom varijancom. Argument x može biti lista ili stupac matrice, a odnosi se na jednodimenzionalni slučajni uzorak normalno distribuirane slučajne varijable, dok se pod opcijama mogu unijeti sljedeći podaci:
  - 'mean poprima vrijednost očekivanja koju se želi provjeriti.
  - 'alternative odnosi se na alternativnu hipotezu H_1 i mogu se pridodati vrijednosti 'twosided, 'greater i 'less.
  - 'dev poprima vrijednost 'unknown ili pozitivan realan broj, ovisno je li varijanca (standardna devijacija) uzorka poznata ili nepoznata.
  - 'conflevel je razina pouzdanosti za interval povjerenja. Poprima vrijednosti iz intervala $[0,1]$. Ako se vrijednost razine pouzdanosti ne unese podrazumijeva se da iznosi 0.95.

  Izlaz funkcije test_mean je objekt koji prikazuje sljedeće rezultate:

  - 'mean_estimate prikazuje vrijednost uzoračke aritmetičke sredine,
  - 'conf_level prikazuje razinu pouzdanosti definiranu od korisnika,
  - 'conf_interval daje interval povjerenja za očekivanje normalne razdiobe,
  - 'method prikazuje osnovne podatke o korištenoj razdiobi,
  - 'hypotheses prikazuje vrijednosti hipoteza H_0 i H_1,
  - 'statistic daje vrijednost test-statistike,
  - 'distribution prikazuje podatke o razdiobi statističkog uzorka,
  - 'p_value daje p-vrijednost testa.
- Funkcija test_variance(x, opcije) koristi se za testiranje hipoteza te određivanje intervala povjerenja za varijancu. Argument i opcije definiraju se isto kao i kod funkcije test_mean.

Prije zadavanja bilo koje funkcije iz navedenih paketa potrebno je

učitati odgovarajući paket i to jednom od naredbi:
`load("descriptive")$`, `load("distrib")$` ili `load("stats")$`.
Unutar jednog radnog lista paket je dovoljno učitati jednom bez obzira koliko se funkcija iz njega poziva. Više o ovim i ostalim funkcijama dostupnim u navedenim paketima može se pronaći u [2].

3 Primjeri zadataka

U ovom poglavlju dani su primjeri zadataka i njihova rješenja koji se rješavaju u sklopu auditornih ("ručnim" putem) i laboratorijskih vježbi (upotrebom računalnog programa) na predmetu Vjerojatnost i statistika. Budući da se zbog ograničena vremena na laboratorijskim vježbama ne stigne u sklopu zadataka ponavljati teorijska podloga, rješavamo iste zadatke kao i na auditornim vježbama. Na taj način ne "trošimo" vrijeme na razumijevanje teksta zadatka i postavljanje problema, već su studenti usmjereni na razumijevanje naredbi programa Maxima i interpretaciju dobivenih rezultata. Najčešći problem s kojim se susrećemo na laboratorijskim vježbama vezan je uz poznavanje sintakse programa odnosno uz ispravan unos početnih podataka i interpretaciju dobivenih rezultata.

Zadatak 1. *Godišnja količina oborina u nekom mjestu, izražena u l/m^2, je normalno distribuirana slučajna varijabla $X \sim N\left(250, 40\right)$. Kolika je vjerojatnost da godišnja količina oborina bude između 235 i 260?*

Rješenje:

Za slučajnu varijablu $X \sim N\left(\mu, \sigma^2\right)$, slučajna varijabla sa standardnom normalnom razdiobom je $X^* = \frac{X-\mu}{\sigma}$, a vjerojatnost da normalno distribuirana slučajna varijabla X poprimi vrijednosti unutar segmenta $\left[a, b\right]$ je

$$P\left(a \le X \le b\right) = P\left(\frac{a-\mu}{\sigma} \le X^* \le \frac{b-\mu}{\sigma}\right) = F^*\left(\frac{b-\mu}{\sigma}\right) - F^*\left(\frac{a-\mu}{\sigma}\right)\,, \tag{1}$$

$$\text{pri čemu je} \quad F^*(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} t\,.$$

Podintegralna funkcija $f^*(x)$ je funkciju gustoće vjerojatnosti standardne normalne razdiobe ([3]). Vrijednosti funkcija F^* i $f^*(x)$ čitaju se iz odgovarajućih tablica (vidjeti tablice u [4]).
Stoga, primjenom formule 1 vjerojatnost da godišnja količina oborina bude između 235 i 260 je:

$$\begin{aligned} P\left(235 \le X \le 260\right) &= P\left(\frac{235-250}{\sqrt{40}} \le X^* \le \frac{260-250}{\sqrt{40}}\right) = P\left(-2.37 \le X^* \le 1.58\right) \\ &= F^*(1.58) - F^*(-2.37) = 0.942947 - 0.008894 = 0.9341\,. \end{aligned}$$

U programu Maxima vrijednosti funkcije $F^*(x)$ mogu se izračunati, i to nešto preciznije, na sljedeći način. Najprije izračunamo vrijednosti

$\frac{a-\mu}{\sigma}$ i $\frac{b-\mu}{\sigma}$ unošenjem naredbi:

$$(235-250)/\mathtt{sqrt(40),numer;}$$
$$(260-250)/\mathtt{sqrt(40),numer;}$$

Nakon što stisnemo zajedno tipke `SHIFT+ENTER` Maxima će ispisati:

$$-2.371708245126$$
$$1.58113883008\,.$$

Za izračun vrijednosti funkcije F^* koristimo funkciju `cdf_normal(x,m,s)`, gdje `x` odgovara vrijednosti $\frac{X-\mu}{\sigma}$, $\mathtt{m}=0$ i `s=1`.
U novi red unesemo naredbu za učitanje paketa `distrib`, a potom zadamo odgovarajuću naredbu.

$$\mathtt{load(distrib)\$}$$
$$\mathtt{cdf\_normal}(1.58113883008,0,1)-\mathtt{cdf\_normal}(-2.371708245126,0,1);$$

Kao izlaz zadane naredbe Maxima će ispisati:

$$0.9342238180929876\,.$$

Zadatak 2. *Broj pogodaka u cilj u 10 gađanja je binomno distribuirana slučajna varijabla $X\sim B\left(10,\frac{57}{100}\right)$. Kolika je vjerojatnost da se pogodi cilj barem 2 puta?*

Rješenje:

Za binomno distribuiranu slučajnu varijablu $X\sim B\left(n,p\right)$ funkcija vjerojatnosti dana je izrazom

$$P\left(X=k\right)=b\left(k;n,p\right)=\binom{n}{k}p^k\left(1-p\right)^{n-k}, \qquad (2)$$

gdje je n proizvoljan prirodan broj, $k\leq n$ prirodan broj i $p(0<p<1)$ realan broj ([3]).
Vjerojatnost da se pogodi cilj barem 2 puta, uključuje događaje da se cilj pogodi dva ili tri ili četiri,..., ili deset puta. U ovom slučaju jednostavnije je izračunati vjerojatnost suprotnog događaja, koji uključuje događaje da se cilj nije pogodio te da se pogodio jednom.
Stoga, primjenom formule 2 tražena vjerojatnost iznosi:

$$P\left(X\geq 2\right)=1-P\left(X=0\right)-P\left(X=1\right)=$$
$$=1-\binom{10}{0}\left(\frac{57}{100}\right)^0\left(1-\frac{57}{100}\right)^{10}-\binom{10}{1}\left(\frac{57}{100}\right)^1\left(1-\frac{57}{100}\right)^9=0.9969\,.$$

U programu Maxima ovo možemo izračunati unošenjem sljedećih naredbi:

$$\mathtt{load(distrib)\$}$$
$$1-\mathtt{pdf\_binomial}(0,10,57/100)-\mathtt{pdf\_binomial}(1,10,57/100),\mathtt{numer};$$

Izlaz ove naredbe je:

$$0.9969191072888272\,.$$

Zadatak 3. *Stroj proizvodi limene ploče debljine 9.5 mm. Slučajno odabrane ploče na tom stroju imaju debljinu redom*

$$8.65,\ 9.90,\ 9.71,\ 9.60,\ 9.61,\ 10.21,\ 11.29,\ 9.70,\ 11.24,\ 9.10,\ 11.35,\ 9.43\,.$$

Pretpostavimo li da je debljina ploče slučajna varijabla normalne razdiobe može li se uz razinu značajnosti od $\alpha = 0.05$ *tvrditi da stroj proizvodi ploče debljine manje od propisane?*

Rješenje:

Test-statistika

$$T = \frac{\bar{X} - \mu}{S}\sqrt{n} \sim t(n-1)$$

ima Studentovu t-razdiobu s $n-1$ stupnjeva slobode. Za određivanje vrijednosti test-statistike potrebno je odrediti aritmetičku sredinu i korigiranu standardnu devijaciju:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} x_i \cdot f_i = 9.9825$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{m} (x_i - \bar{x})^2 \cdot f_i = 0.7744 \quad \Rightarrow \quad s = 0.87999.$$

Tada test-statistika iznosi:

$$t = \frac{\bar{x} - \mu_0}{s}\sqrt{n} = \frac{9.9825 - 9.5}{0.87999}\sqrt{12} = 1.8994\,.$$

Digresija: Vrijednosti aritmetičke sredine i standardne devijacije u programu Maxima mogu se izračunati na sljedeći način:

```
load(descriptive)$
x: [8.65, 9.90, 9.71, 9.60, 9.61, 10.21, 11.29, 9.70, 11.24, 9.10, 11.35, 9.43] ;
mean (x);
std1 (x);
```

Izlazi ovih naredbi su:

$$9.9825$$
$$0.8799909606973759\,.$$

Postavljamo hipoteze

$$\begin{cases} H_0: & \mu = 9.5 \\ H_1: & \mu < 9.5 \end{cases}$$

Za ovako postavljene hipoteze kritično područje ili područje odbacivanja hipoteze H_0 je $[-\infty, t_\alpha\rangle$. Vrijednost t_α odredimo iz tablice Studentove t-razdiobe za $n-1 = 11$ stupnjeva slobode (vidjeti tablicu u [4] na stranici 231.):

$$t_\alpha = t_{0.05} = -t_{1-0.05} = -t_{0.95} = -1.85955\,,$$

odnosno kritično područje je $\langle -\infty, -1.85955\rangle$. Kako vrijednost test-statistike t ne upada u kritično područje prihvaćamo hipotezu H_0, te odbacujemo hipotezu H_1, na razini značajnosti od 5%. Prema tome, možemo reći da stroj ne proizvodi ploče debljine manje od propisanih 9.5 mm.

Zadatak ćemo u programu Maxima riješiti tako da ćemo prvo učitati paket `stats`, definirat ćemo skup podataka te potom zadati naredbu za lijevi jednostrani test. To činimo na sljedeći način:

```
load("stats")$
x: [8.65, 9.90, 9.71, 9.60, 9.61, 10.21, 11.29, 9.70, 11.24, 9.10, 11.35, 9.43];
test_{m}ean(x,'mean=9.5,'alternative='less,'conflevel=0.95);
```

Napomena: `'conflevel=0.95` je razina pouzdanosti i njegova vrijednost jednaka je $1-\alpha$.

Maxima će rezultat zadane naredbe ispisati u obliku:

```
MEAN TEST
mean_estimate=9.9825
conf_level=0.95
conf_interval=[-inf,10.43871131438427]
method="Exact t-test. Unknown variance."
hypotheses="H0: mean = 9.5 , H1: mean \lt 9.5"
statistic = 1.899370679875383
distribution = [student_t,11]
p_value = 0.9579800520853199
```

Na temelju dobivenih rezultata odgovor o odbacivanju ili prihvaćanju hipoteze H_0 dajemo na temelju $p-$vrijednosti testa. Kako na auditornim vježbama, kao što je prethodno prikazano, odgovor o prihvaćanju hipoteze H_0 dajemo na temelju vrijednosti test-statistike i kritičnog područja te se pojam $p-$vrijednosti ne spominje, ovdje je potrebno studentima dodatno pojasniti što ona predstavlja.
U slučaju lijevog jednostranog t-testa p-vrijednost jednaka je $p = \mathbb{P}\left(T \le t \middle|\, H_0\right)$, desnog jednostranog testa jednaka je $p = \mathbb{P}\left(T \ge t \middle|\, H_0\right)$, dok je kod dvostranog testa p-vrijednost manja od brojeva $2\cdot\mathbb{P}\left(T \le t \middle|\, H_0\right)$ i $2\cdot\mathbb{P}\left(T \ge t \middle|\, H_0\right)$. Na temelju p-vrijednosti možemo dati sljedeće zaključke:

- ako je $p \le \alpha$, onda se t nalazi u kritičnom području, pa odbacujemo H_0 na razini značajnosti α,
- ako je $p > \alpha$, onda se t ne nalazi u kritičnom području, pa prihvaćamo H_0 na razini značajnosti α.

U našem slučaju je $p = 0.96 > 0.05$, pa stoga hipotezu H_0 prihvaćamo na razini značajnosti od 0.05.

Zadatak 4. *Za podatke dane u Zadatku 3. uz pouzdanost od 95% odrediti interval povjerenja za očekivanje $\mu = 9.5$.*

Rješenje:

Interval povjerenja za očekivanje s nepoznatom varijancom i pouzdanošću γ je

$$\left\langle \bar{X} - t_{\frac{1+\gamma}{2}} \cdot \frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{1+\gamma}{2}} \cdot \frac{S}{\sqrt{n}} \right\rangle , \qquad (3)$$

gdje je $t_{\frac{1+\gamma}{2}}$ kvantil Studentove t-razdiobe s $n-1$ stupnjeva slobode. Vrijednost kvantila $t_{\frac{1+\gamma}{2}}$ odredimo iz tablice Studentove t-razdiobe za $n-1=11$ stupnjeva slobode (vidjeti tablicu u [4] na stranici 231.):

$$t_{\frac{1+\gamma}{2}} = t_{\frac{1+0.95}{2}} = t_{0.975} = 2.201\,.$$

Tada je interval povjerenja na temelju formule 3 jednak:

$$\left\langle 9.9825 - 2.201 \cdot \frac{0.87999}{\sqrt{12}},\ 9.9825 + 2.201 \cdot \frac{0.87999}{\sqrt{12}} \right\rangle = \langle 9.4234,\ 10.5416\rangle\ .$$

U programu Maxima interval povjerenja određuje se pomoću naredbe za određivanje dvostranog t-testa. U novi red unesemo naredbu:

```
test_ {m}ean(x,'mean=9.5,'alternative='twosided,'conflevel=0.95);
```

Maxima će rezultat zadane naredbe ispisati u obliku:

```
MEAN TEST
mean_estimate=9.9825
conf_level=0.95
conf_interval=[9.423380425654932,10.54161957434507]
method="Exact t-test. Unknown variance."
hypotheses="H0: mean = 9.5 , H1: mean ≠ 9.5 ”
statistic = 1.899370679875383
distribution = [student_t,11]
p_value = 0.08403989582936022
```

Sada je u četvrtom redu ispisan interval povjerenja za očekivanje s nepoznatom varijancom s pouzdanošću od 95%.

4 Zaključak

Od akademske godine 2015./2016. na stručnom studiju graditeljstva Tehničkog veleučilišta u Zagrebu računalni program Maxima obrađuje se u sklopu predmeta Računarstvo u graditeljstvu. U posljednje tri godine pokazalo se da ga studenti s lakoćom savladavaju te da ga uspješno primjenjuju u rješavanju zadataka u sklopu ostalih matematičkih predmeta (bilo na nastavi ili samostalno u sklopu seminarskih radova). Iz tog razloga, ali i zato što imamo mali broj sati predviđen za rad na računalu iz predmeta Vjerojatnost i statistika, odlučili smo se koristiti program Maxima, a ne neki, možda, prikladniji program za statističku obradu podataka (R, SPS i sl.).

U ovom trenutku postoje prijedlozi da se broj laboratorijskih vježbi poveća. Na taj način imali bi mogućnost raditi s oba programa, Excelom (kojeg smo do ove godine koristili) i Maximom. Iako se pokazalo da program Maxima u potpunosti zadovoljava sve nastavne potrebe matematičkih predmeta, koji se izvode na studijima graditeljstva, Excel je program koji će studenti u svom budećem poslu koristiti na svakodnevnoj razini i stoga je dobro u što većoj mjeri i u što različitije svrhe koristiti ga u sklopu nastave.

Bibliografija

[1] L. Marohnić, M. Orlić Bachler: Applications of free computational software in math courses at Zagreb University of Applied Sciences, Matematics Education as a Science and a Profession, Mathematics and Children 2017, Osijek, Hrvatska, 2017.

[2] Maxima Manual, Version 5.41.0
Dostupno na: http://superk.physics.sunysb.edu/ mcgrew/phy310 /documentation/maxima-reference.pdf

[3] M. Orlić, T. Perkov: Repetitorij matematike za studente graditeljstva, Tehničko veleučilište u Zagrebu, Zagreb, 2014.

[4] S. Suljagić: Vjerojatnost i statistika, interna skripta, Tehničko veleučilište u Zagrebu, Zagreb, 2003.

ISSN 1334-6083