Matematici speciale
Seminar 12

Mai 2017

"Statistica este arta de a minti prin intermediul cifrelor."

Wilhelm Stekel

12

Notiuni de statistica

Datele din dreapta arata temperaturile de racire ale unei cesti de cafea, care tocmai a fost preparata. Temperatura la care ajunge aparatul de cafea este 180 de grade Fahrenheit (aproximativ $82°C$).

Time (mins)	Temp (º F)
0	179.5
5	168.7
8	158.1
11	149.2
15	141.7
18	134.6
22	125.4
25	123.5
30	116.3
34	113.2
38	109.1
42	105.7
45	102.2
50	100.5

In anul 1992 o femeie a dat in judecata McDonald's pentru ca au servit cafeaua la temperatura $180°F$ si aceasta i-a cauzata arsuri serioase in momentul in care a incercat sa o bea (vezi Liebeck vs. McDonald's). Un expert adus din partea acuzarii a sustinut la proces ca lichidele care se afla la aceasta temperatura pot cauza distrugerea totala a pielii umane in doua pana la sapte secunde. S-a stabilit si ca daca ar fi fost servita la $155°F$ ($68°C$) s-ar fi racit la timp si ar fi fost evitat tot incidentul. Femeia a primit in prima instanta o

despagubire de 2.7 milioane de dolari. Ca urmare a acestui caz faimos multe restaurante servesc acum cafeaua la o temperatura de aproximativ $155°F$. Cat de mult ar trebui sa astepte restaurantele din momentul in care cafeaua este turnata in ceasca din aparat si pana cand ea poate fi servita, pentru a se asigura ca nu este mai fierbinte de $155°F$?

- Determinati ecuatia unui model de regresie exponentiala pentru a reprezenta datele
- Reprezentati grafic curba obtinuta
- Decideti daca ecuatia obtinuta este buna pentru a reprezenta datele existente in tabel
- Interpolare: Cand ajunge temperatura cafelei la $106°F$?
- Extrapolare: Care este temperatura prezisa, de modelul gasit, peste o ora?

Notiuni teoretice:

• Statistica descriptiva: populatie statistica, esantion statistic, serie statistica, frecventa abosluta, frecventa relativa, histograma, media $\bar{x}$, mediana m_3, amplitudinea A, dispersia σ^2, deviatia standard σ, moda (modulul) m_o, dispersia de selectie s^2, deviatia standard de selectie s, cuartilele Q_1, Q_2, Q_3, indicatorul de asimetrie sk (skewness), indicatorul de aplatizare k (kurtosis)

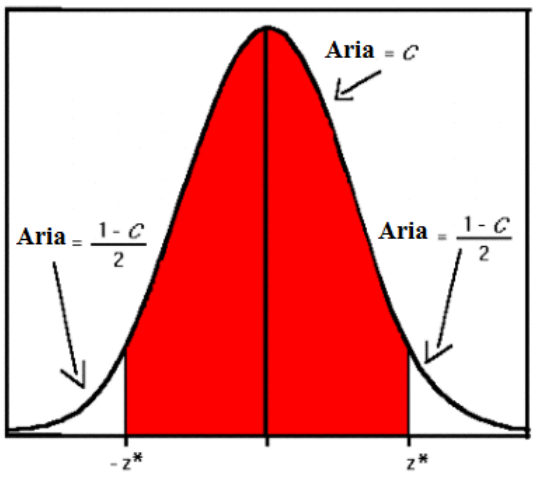
Intervale de incredere

• confidence intervals are used when we want to estimate a population parameter from a sample. The parameter may be estimated by a single value (a point estimate) but it is usually preferable to estimate it by an interval which will give some indication of the amount of uncertainty attached to the estimate.

• the common notation for the parameter in question is θ. Often, this parameter is the population mean μ , which is estimated through the sample mean $\bar{x}$.

• the level C of a confidence interval gives the probability that the interval produced by the method employed includes the true value of the parameter.

The selection of a confidence level for an interval determines the probability that the confidence interval produced will contain the true parameter value. Common choices for the confidence level C are 0.90, 0.95, and 0.99. These levels correspond to percentages of the area of the normal density curve. For example, a 95% confidence interval covers 95% of the normal curve. The probability of observing a value outside of this area is less than 0.05. Because the normal curve is symmetric, half of the area is in the left tail of the curve, and the other half of the area is in the right tail of the curve. As shown in the diagram , for a confidence interval with level C, the area in each tail of the curve is equal to $(1 - C)/2$. For a 95% confidence interval, the area in each tail is equal to $0.05/2 = 0.025$.



The value z^* representing the point on the standard normal density curve such that the probability of observing a value greater than z^* is equal to p

is known as the upper p critical value of the standard normal distribution. For example, if $p = 0.025$, the value z^* such that $P(Z > z^*) = 0.025$, or $P(Z < z*) = 0.975$, is equal to 1.96. For a confidence interval with level C, the value p is equal to $(1-C)/2$. A 95% confidence interval for the standard normal distribution is then the interval $(-1.96, 1.96)$, since 95% of the area under the curve falls within this interval.

Medie necunoscuta si deviatie standard cunoscuta

Teorema:

Pentru o populatie cu media μ necunoscuta si deviatie standard σ cunoscuta, un interval de incredere pentru media populatiei, construit pe baza unui esantion de volum n, este:

$$(\bar{x} - z^* \frac{\sigma}{\sqrt{n}}, \bar{x} + z^* \frac{\sigma}{\sqrt{n}})$$

unde z^* este valoarea critica corespunzatoare lui $\frac{1-C}{2}$ pentru distributia normala standard, adica $z^* = \Phi(\frac{1-C}{2})$.

Medie necunoscuta si deviatie standard necunoscuta

• cand deviatia standard σ este necunoscuta este estimata de obicei prin s numita eroarea standard /deviatia standard de selectie , unde:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

si n este volumul selectiei.

Teorema:

Pentru o populatie cu media necunoscuta μ si deviatia standard σ necunoscuta, un inteval de incredere pentru media populatiei, construit pe baza unui esantion de volum n, este:

$$(\bar{x} - t^* \frac{s}{\sqrt{n}}, \bar{x} + t^* \frac{s}{\sqrt{n}})$$

unde t^* este valoarea critica corespunzatoare lui $\frac{1-C}{2}$ pentru distributia t-Student cu n-1 grade de libertate.

• Pasul final consta in interpretarea rezultatului: pe baza datelor avute suntem $C\%$ siguri ca adevarata medie a populatiei se afla intre valorile date de intervalul gasit

De retinut

• valorile critice z^* si t^* se pot gasi in tabelul urmator z-t-table

• distributia t sau distributia Student este data de catre urmatoarea densitate de probabilitate:

$$f(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

unde n este numarul de grade de libertate si Γ este functia lui Euler.

Exemplu:

Suppose a student measuring the boiling temperature of a certain liquid observes the readings (in degrees Celsius) 102.5, 101.7, 103.1, 100.9, 100.5, and 102.2 on 6 different samples of the liquid. He calculates the sample mean to be 101.82. If he knows that the standard deviation for this procedure is 1.2 degrees, what is the confidence interval for the population mean at a 95% confidence level?

In other words, the student wishes to estimate the true mean boiling temperature of the liquid using the results of his measurements. If the measurements follow a normal distribution, then the sample mean will have the distribution $N(\mu, \frac{\sigma^2}{n})$. Since the sample size is 6, the standard deviation of the sample mean is equal to $\frac{1.2}{\sqrt{6}} = 0.49$.

The critical value for a 95% confidence interval is 1.96, where $(1 - C)/2 = (1 - 0.95)/2 = 0.025$. A 95% confidence interval for the unknown mean is:

$$(101.82 - 1.96 \cdot 0.49, 101.82 + 1.96 \cdot 0.49) = (100.86, 102.78)$$

As the level of confidence decreases, the size of the corresponding interval will decrease. Suppose the student was interested in a 90% confidence interval for the boiling temperature. In this case, $C = 0.90$, and $(1 - C)/2 = 0.05$. The critical value z^* for this level is equal to 1.645, so the 90% confidence interval is:

$$(101.82 - 1.645 \cdot 0.49, 101.82 + 1.645 \cdot 0.49) = (101.01, 102.63)$$

An increase in sample size will decrease the length of the confidence interval without reducing the level of confidence. This is because the standard deviation decreases as n increases. The margin of error e of a confidence interval is defined to be the value added or subtracted from the sample mean which determines the length of the interval: $e = z^* \frac{\sigma}{\sqrt{n}}$.

Suppose in the example above, the student wishes to have a margin of error equal to 0.5 with 95% confidence. Substituting the appropriate values into the expression for m and solving for n gives the calculation $n = (1.96 \cdot 1.2/0.5)^2 = 22.09$. To achieve a 95% confidence interval for the mean boiling point with total length less than 1 degree, the student will have to take 23 measurements. □

Testarea ipotezelor statistice

In a decision-making process managers make hypotheses which afterwards can be tested using the tools of statistics. A hypothesis test examines two opposing hypotheses about a population: the null hypothesis and the alternative hypothesis. How you set up these hypotheses depends on what you are trying to show.

Null hypothesis H_0

• the null hypothesis states that a population parameter is equal to a value. The null hypothesis is often an initial claim that managers specify using previous research or knowledge.

Alternative Hypothesis H_a

• the alternative hypothesis states that the population parameter is different than the value of the population parameter in the null hypothesis. The alternative hypothesis is what you might believe to be true or hope to prove true.

What are some common hypotheses?

E.g.: Hypothesis to determine whether a population mean μ, is equal to some target value μ_0 include the following:

$\Rightarrow$ for a big sample size n or σ known

· we use the z test and compute:

$$z_{calc} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$\Rightarrow$ for a sample size $n < 30$ and σ unknown

· we use the t test and compute:

$$t_{calc} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Two-tailed test:

$$H_0 : \mu = \mu_0$$

$$H_a : \mu \neq \mu_0$$

$\Rightarrow$ the critical region/ region of rejection, when we reject H_0 is given by:

$$z_{calc} < -z^*_{\frac{\alpha}{2}} \textbf{ or } z_{calc} > z^*_{\frac{\alpha}{2}} \qquad t_{calc} < -t^*_{\frac{\alpha}{2},n-1} \textbf{ or } t_{calc} > t^*_{\frac{\alpha}{2},n-1}$$

Upper-tailed test:

$$H_0 : \mu = \mu_0$$

$$H_a : \mu > \mu_0$$

$\Rightarrow$ the critical region/ region of rejection, when we reject H_0 is given by:

$$z_{calc} > z^*_\alpha \qquad\qquad t_{calc} > t^*_{\alpha,n-1}$$

Lower-tailed test:

$$H_0 : \mu = \mu_0$$

$$H_a : \mu < \mu_0$$

$\Rightarrow$ the critical region/ region of rejection, when we reject H_0 is given by:

$$z_{calc} < -z^*_\alpha \qquad\qquad t_{calc} < -t^*_{\alpha,n-1}$$

$\Rightarrow$ in all these examples α is the significance level corresponding to a confidence level $C = 1 - \alpha$

$\Rightarrow$ the critical values z^* and t^* for different confidence intervals are shown in the z-t-table

Estimarea parametrilor prin metoda momentelor

The method of moments is a method of estimation of population parameters. The method is based on the assumption that the sample moments are good estimates of the corresponding population moments.

- for a population X the moments μ_k (or M_k) of order k are defined as:

$$\mu_k = M(X^k) = \begin{cases} \int_{-\infty}^{\infty} x^k f(x)dx, & \text{if } X \text{ is continuous} \\ \sum_{i \in I} x_i^k p_i, & \text{if } X \text{ is discrete} \end{cases}$$

- the sample moment m_k of order k of a sample of size n is defined as:

$$m_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

The method of moments estimation simply equates the moments of the distribution with the sample moments $\mu_k = m_k$ and solves for the unknown parameters. (the distribution must have finite moments)

Method of moments:

1. we want to estimate a parameter θ
2. calculate low-order moments μ_k as functions of θ
3. set up a system of equations setting the population moments μ_k equal to the sample moments m_k, and derive expressions for the parameter as functions of the sample moments m_k.

Exemplu:

Let $X_1, X_2, \dots X_n$ a sample from a binomial distributed population $X \sim Bi(n_0, p)$ with parameters n_0 and p. Estimate these parameters using the method of moments.

Solutie: Since

$$M(X) = n_0 p$$

and:

$$M_2(X) = M(X^2) = D^2(X) + M(X)^2 = n_0 p(1-p) + n_0^2 p^2,$$

we can write $n_0 p(1-p) = M_2(X) - M(X^2)$.

Equating:

$$M(X) = m_1 \left(= \frac{X_1 + X_2 + \ldots + X_n}{n} \right)$$

and

$$M_2(X) = m_2 \left(= \frac{X_1^2 + X_2^2 + \ldots + X_n^2}{n} \right)$$

one can observe:

$$1 - p = \frac{m_2 - m_1^2}{m_1}$$

thus:

$$p = \frac{m_1 + m_1^2 - m_2}{m_1}$$

can be used as an estimator for the parameter p.

In the same context:

$$n_0 = \frac{m_1}{p} = \frac{m_1^2}{m_1 + m_1^2 - m_2}.$$

□

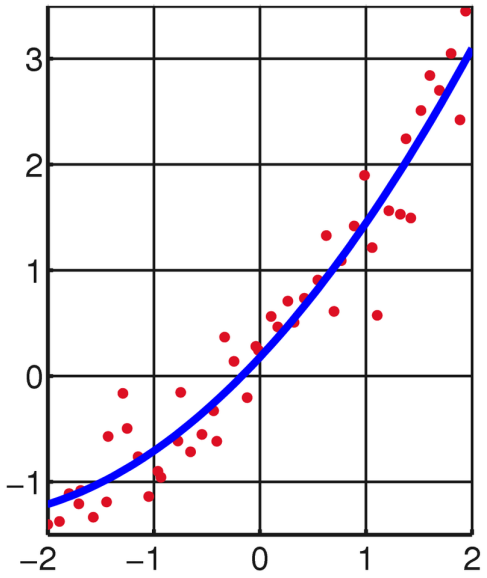
Analiza regresiva prin metoda celor mai mici patrate

• in sectiunile anterioare am considerat experimente pentru care am observat o singura cantitate (variabila) aleatoare, iar esantioanele respective au constat din date reprezentate de numere reale $x_1, x_2, \ldots, x_n$

• in aceasta sectiune vom considera experimente în care suntem interesati de doua cantitati (variabile) aleatoare, deci esantioanele respective vor fi reprezentate de perechi de numere reale $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

• in analiza regresiva una din cele doua variabile (spre exemplu X) este privita ca o variabila ce poate fi masurata (determinata) cu precizie, numita variabila independenta si suntem interesati de modul cum cealalta variabila Y (numita variabila dependenta) depinde de aceasta: spre exemplu suntem interesati de modul de aportul de crestere Y al animalelor în functie de cantitatea zilnica de hrana X.

• in general, intr-un anumit experiment alegem valorile $x_1, x_2, \ldots, x_n$ apoi observam valorile $y_1, y_2, \ldots, y_n$ ale unei variabile aleatoare Y, obtinand astfel un esantion $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$



Se pune problema gasirii unei curbe care sa aproximeze cat mai bine datele obitnute experimental (norul de puncte)

• aceasta aproximare se face de obicei impunand conditia ca suma patratelor distantelor de la puncte la curba sa fie minima (metoda celor mai mici patrate)

Regresia liniara

• estimam norul de puncte printr-o dreapta $y = f(x) = a + bx$

• impunand conditia data de metoda celor mai mici patrate se obtine sistemul:

$$\begin{cases} a + b \cdot \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{n} y_i}{n} \\ a \cdot \frac{\sum_{i=1}^{n} x_i}{n} + b \cdot \frac{\sum_{i=1}^{n} x_i^2}{n} = \frac{\sum_{i=1}^{n} x_i y_i}{n} \end{cases}$$

care are solutia:

$$b = \frac{n \sum xy - \sum x \cdot \sum y}{n \sum x^2 - (\sum x)^2}$$

si:

$$a = \frac{\sum_{i=1}^{n} y_i}{n} - b \frac{\sum_{i=1}^{n} x_i}{n} = \bar{Y} - b\bar{X}.$$

Regresia parabolica

- estimam norul de puncte printr-o parabola $y = f(x) = a + bx + cx^2$
- impunand conditia data de metoda celor mai mici patrate se obtine sistemul:

$$\begin{cases} a \cdot n + b \cdot \sum x + c \cdot \sum x^2 = \sum y \\ a \cdot \sum x + b \cdot \sum x^2 + c \cdot \sum x^3 = \sum xy \\ a \cdot \sum x^2 + b \cdot \sum x^3 + c \cdot \sum x^4 = \sum x^2 y \end{cases}$$

Regresia hiperabolica

- estimam norul de puncte printr-o hiperbola $y = f(x) = a + \frac{b}{x}$
- impunand conditia data de metoda celor mai mici patrate se obtine sistemul:

$$\begin{cases} a \cdot n + b \cdot \sum \frac{1}{x} = \sum y \\ a \cdot \sum \frac{1}{x} + b \cdot \sum \frac{1}{x^2} = \sum \frac{y}{x} \end{cases}$$

Regresia exponentiala

- estimam norul de puncte printr curba $y = f(x) = a \cdot b^x$
- se logaritmeaza relatia si obtinem:

$$\ln y = \ln a + \ln b \cdot x$$

care are forma unui model de regresie liniara pentru datele $(x_i, \ln y_i), \quad i = \overline{1, n}$ deci a si b se determina din:

$$\ln b = \frac{n \sum x \ln y - \sum x \cdot \sum \ln y}{n \sum x^2 - (\sum x)^2}$$

si:

$$\ln a = \frac{\sum_{i=1}^{n} \ln y_i}{n} - \ln b \cdot \frac{\sum_{i=1}^{n} x_i}{n}.$$

prin intermediul formulelor $a = e^{\ln a}$ si $b = e^{\ln b}$

Probleme rezolvate

Problema 1. *Calculaţi cuartilele Q_1, Q_2, Q_3 pentru următoarea serie statistica simplă*

$$X: 1,\ 2,\ 5,\ 7,\ 11,\ 21,\ 22,\ 23,\ 29$$

şi abaterea cuartilică.

Solutie: Facem mai întâi observaţia că mediana m_e coincide cu cuartila Q_2.
Deoarece seria statistică dată are un număr impar de termeni (9 mai exact), vom folosi formula corespunzătoare pentru a determina cuartila Q_2 şi avem

$$x_{\frac{9+1}{2}} = x_5 = 11 \Rightarrow m_e = Q_2 = 11.$$

Mai departe pentru a determina prima cuartilă ţinem cont de seria statistică simplă

$$1,\ 2,\ 5,\ 7,\ 11$$

care are tot un număr impar de termeni şi obţinem

$$x_{\frac{5+1}{2}} = x_3 = 5 \Rightarrow Q_1 = 5.$$

Analog procedăm pentru a treia cuartilă ţinând cont de seria statistică simplă

$$11,\ 21,\ 22,\ 23,\ 29$$

care are tot un număr impar de termeni şi rezultă

$$x_{\frac{5+1}{2}} = x_3 = 22 \Rightarrow Q_3 = 22.$$

Atunci rezultă că abaterea cuartilică este

$$Q = Q_3 - Q_1 = 22 - 5 = 17.$$

Problema 2. *Fie seria statistică*

$$X: 1,\ 5,\ 4,\ 20,\ 3,\ 16.$$

Determinaţi:
a) amplitudinea absolută A.
b) abaterea medie pătratică $\bar{a}(X)$.
c) dispersia $\sigma^2(X)$.
d) deviatia standard $\sigma(X)$.
e) coeficientul de variaţie $cv(X)$.

Solutie: **a)** Amplitudinea absolută A este

$$A = X_{\max} - X_{\min} = 20 - 1 = 19.$$

b) Abaterea medie pătratică $\bar{a}\,(X)$ se obţine astfel

$$\bar{a}\,(X) = \frac{|1-\overline{x}| + |5-\overline{x}| + |4-\overline{x}| + |20-\overline{x}| + |3-\overline{x}| + |16-\overline{x}|}{6},$$

unde media $\overline{x}$ este

$$\overline{x} = \frac{1+5+4+20+3+16}{6} = 8,16.$$

Atunci rezultă

$$\bar{a}\,(X) \simeq 6,55.$$

c) Dispersia este

$$\begin{aligned}\sigma^2\,(X) &= \frac{1}{6}\sum_{i=1}^{6}(x_i-\overline{x})^2 = \\ &= \frac{1}{6}\left(7,16^2+3,16^2+4,16^2+11,84^2+5,16^2+7,84^2\right) \\ &= 51,138 \simeq 51.\end{aligned}$$

d) deviatia standard rezultă imediat de mai sus

$$\sigma\,(X) = \sqrt{\sigma^2(X)} = \sqrt{51} = 7,14 \simeq 7.$$

e) Din cele de mai sus, rezultă coeficientul de variaţie

$$cv(X) = \frac{\sigma\,(X)}{\overline{x}} \cdot 100 = 85,78.$$

Problema 3. *Pe o perioadă de mai mulţi ani, un profesor a înregistrat rezultatele elevilor şi a obţinut ca media μ a acestor rezultate este 72 şi abaterea standard $\sigma = 12$. Clasa de 36 de elevi pe care-i învaţă în prezent are o medie $\overline{x} = 75,2$, iar profesorul afirmă ca ea este superioară celor de până acum. Întrebarea care se pune este dacă media clasei $\overline{x}$ este un argument suficient pentru a susţine afirmaţia profesorului la un nivelul de semnificaţie dat $\alpha = 0,05$ (95% sigur).*

Solutie: Etapa 1: Formularea ipotezei nule H_0

$$H_0 : \overline{x} = \mu = 72 \quad \Leftrightarrow \text{ clasa nu este superioară.}$$

Etapa 2: Formularea ipotezei alternative H_a

$$H_a : \overline{x} = \mu > 72 \quad \Leftrightarrow \text{ clasa este superioară.}$$

Etapa 3: Metodologia de verificare a ipotezelor

a) Când în ipoteza nulă media populaţiei şi deviaţia standard sunt cunoscute, atunci folosim scorul standard z ca şi test statistic.

b) Nivelul de semnificaţie este dat şi este $\alpha = 0,05$.

c) În baza teoremei limită centrală distribuţia mediilor eşantioanelor este aproape normală, deci prin urmare distribuţia normală va fi folosită pentru

determinarea regiunii critice. Regiunea critică este egală cu mulţimea valorilor scorului standard z care determină respingerea ipotezei nule şi este situată la extremitatea dreaptă a distribuţiei normale. Regiunea critică este la dreapta deoarece valori mari ale mediei eşantionului susţin ipoteza alternativă în timp ce valori apropiate valorii 72 susţin ipoteza nulă.

Valoarea critică ce desparte zona valorilor "nu este superior" de zona valorilor "este superior" este determinată de probabilitatea $\alpha = 0,05$ de a comite o eroare de tip I (*eroarea de tip I apare când ipoteza nulă este adevărată şi tot ea este respinsă*).

Etapa 4: Determinarea valorii testului statistic

Valoarea testului statistic este dată de formula

$$z_{calc} = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{75,2 - 72}{\frac{12}{\sqrt{36}}} = 1,6.$$

Etapa 5: Luarea unei decizii şi interpretarea ei

Dacă comparăm valoarea găsită cu valoarea critică observăm că:

$$1,6 < 1,65$$

Conform celor stabilite in sectiunea ipotezelor statistice respingem ipoteza H_0 daca:

$$z_{calc} > z^*_\alpha$$

Decizia: nu putem respinge ipoteza nulă !

În final, tragem concluzia că probele nu sunt suficiente pentru a susţine că actuala clasă este superioară celor anterioare.

Problema 4. *Noua dintre studenţii unei facultati cu profil sportiv au fost selectaţi pentru a da un test de alergare pe distanţă mare. Masurătorile pentru acest grup au condus la un timp mediu de 12,87 minute cu o abatere standard $s = 1,3$. Să se aproximeze, cu o probabilitate de 90%, timpul mediu pe care studentii intregii facultati il vor inregistra pe acea distanta .*

Solutie: Deoarece nu se cunoaşte dispersia populaţiei iar eşantionul are volumul mai mic dacât 30, intervalul de încredere este dat de formula

$$\left(\overline{x} - \frac{s}{\sqrt{n}} t_{n-1,\frac{\alpha}{2}}, \overline{x} + \frac{s}{\sqrt{n}} t_{n-1,\frac{\alpha}{2}}\right),$$

unde $\overline{x} = 12,87$; $s = 1,3$; $n = 9$; $\alpha = 0,10$; iar $t_{n-1,\frac{\alpha}{2}}$ este valoarea critică a repartiţiei Student (statisticianul William Sealy Gosset folosea acest pseudonim in articolele sale) cu $n-1$ grade de libertate corespunzătoare valorii $\frac{\alpha}{2} = \frac{1-C}{2}$ care în cazul nostru este $t_{9-1,\ 0.05} = t_{8,\ 0,05} = 1,860$ conform tabelului z-t-table

Obtinem intervalul

$$(12.064,\ 13.676)$$

In concluzie suntem 90% siguri ca timpul mediu inregistrat de un student pe acea distanta va fi in acest interval !

Probleme propuse

Problema 1. *Fiind date seriile statistice simple*

$$X : 1,\ 5,\ 7,\ 8,\ 10,$$

$$Y : 1,\ 6,\ 100,\ 135$$

determinaţi mediana în ambele cazuri.

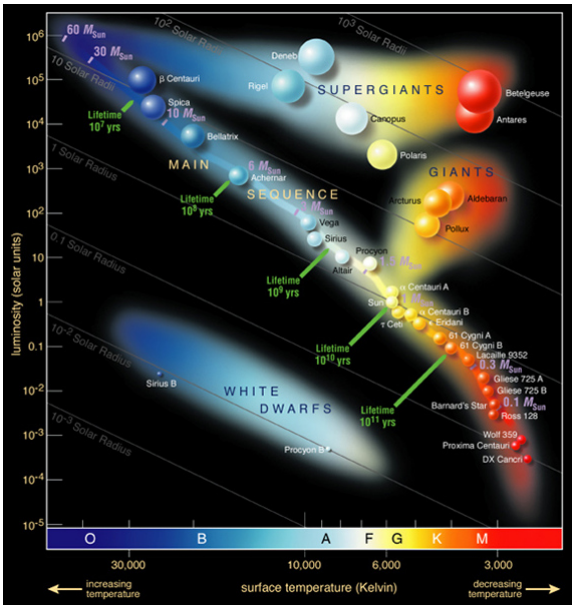
Problema 2. *Într-o colectivitate s-au ales date statistice numerice obţinându-se*

$$X : 4,\ 1,\ 1,\ 5,\ 6,\ 3,\ 2,\ 1,$$

$$Y : 100,\ 90,\ 40,\ 80,\ 70,\ 50,\ 100,\ 70.$$

Aflaţi după care din variabilele de mai sus, colectivitatea este mai omogenă.

Problema 3. *Diagrama Herzsprung-Russell arata dependenta dintre magnitudinile absolute si temperaturile efective de la suprafata stelelor:*



Pentru un grup de stele din sirul principal al diagramei astronomii au inregistrat cu ajutorul telescopului Keck urmatoarele date:

$$(+5, 5000^\circ K),\ (+10, 3000^\circ K),\ (0, 10000^\circ K),\ (-5, 25000^\circ K),\ (+6, 7500^\circ K)$$

Cautati un model de regresie adecvat pentru aceste date.

Problema 4. *The operations manager of a large production plant would like to estimate the mean amount of time a worker takes to assemble a new electronic component. Assume that the standard deviation of this assembly time is* 3.6 *minutes.*

a) After observing 120 *workers assembling similar devices, the manager noticed that their average time was* 16.2 *minutes. Construct a* 95% *confidence interval for the mean assembly time.*

b) How many workers should be involved in this study in order to have the mean assembly time estimated up to ± 15 *seconds with* 95% *confidence?*

Problema 5. *In order to ensure efficient usage of a server, it is necessary to estimate the mean number of concurrent users. According to records, the sample mean and sample standard deviation of number of concurrent users at* 100 *randomly selected times is* 37.7 *and* 9.2*, respectively.*

Construct a 90% *confidence interval for the mean number of concurrent users.*

Problema 6. *Let* $X_1, X_2, ..., X_n$ *be normal random variables with mean* m *and variance* σ^2*. What are the method of moments estimators of the mean* m *and variance* σ^2*?*

Problema 7. *A consumer group, concerned about the mean fat content of a certain steakburger submits to an independent laboratory a random sample of* 12 *steakburgers for analysis. The percentage of fat in each of the steakburgers is as follows:*

21 18 19 16 18 24 22 19 24 14 18 15

The manufacturer claims that the mean fat content of this steakburger is around 20%*. Assuming percentage fat content to be normally distributed with a standard deviation of* 3*, carry out a hypothesis test, with significance level* $\alpha = 0.05$*, in order to advise the comsumer group as to the validity of manufacturer's claim.*

Problema 8. *During a particular week,* 13 *babies were born in a maternity unit. Part of the standard procedure is to measure the length of the baby. Given below is a list of the lengths, in centimetres, of the babies born in this particular week.*

49 50 45 51 47 49 48 54 53 55 45 50 48

Assuming that this sample came from an underlying normal population, test, at the 5% *significance level, the hypothesis that the population mean length is* 50 *cm.*

Problema 9. $X_1, X_2, \dots X_n$ *represents a selection from a population* X *with exponential distribution, i.e. the probability density function is:*

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0, \\ 0, & \text{otherwise} \end{cases}$$

Estimate the parameter λ *using the method of moments.*

Problema 10. $X_1, X_2, \dots X_n$ *represents a selection from a population* X *with Poisson distribution, i.e. the probability mass function is:*

$$P(X = k) = \begin{cases} e^{-\lambda} \frac{\lambda^k}{k!}, & \text{if } k = 0, 1, \dots \\ 0, & \text{otherwise} \end{cases}$$

Estimate the parameter λ using the method of moments.