Procjenjivanje i testiranje

1. Uvod

Ponovimo neke pojmove iz deskriptivne statistike.
Populacija je skup svih entiteta koje razmatramo, na primjer svi studenti nekog sveučilišta čine populaciju.
Razmatramo neko statistističko obilježje populacije, na primjer visinu. Visina je slučajna veličina.
Uzorak je neki podskup populacije slučajno odabran, na primjer slučajno odabranih 300 studenata.
Neka je **n** duljina uzorka, na primjer n=300.
Mjerenjem slučajne veličine X na uzorku dobijemo *n* podataka:
$x_1, x_2, \ldots, x_n$.
Interpretiramo ih kao n slučajnih vrijednosti slučajne varijable X.

Primjer 1. Da bismo procijenili količinu kemikalije u posudama koje se automatski pune, izaberemo slučajno 10 posuda i provjeravamo količinu kemikalije u njima. Dobivamo podatke koji (nakon sređivanja, od manjeg prema većem) možemo zapisati ovako:
0.98, 0.98, 0.98, 0.99, 0.99, 1.00, 1.01, 1.01, 1.01, 1.02.
Tu slučajna veličina X mjeri količinu kemikalije u posudi,
uzorak čine odabrane posude,
n=10,
x_1, do x_{10} jesu podatci 0.98,…, 1.02; to su vrijednosti slučajne veličine X na uzorku.

Neka slučajna veličina X (u primjeru ili općenito) ima očekivanje μ i varijancu σ^2:
$E(X) = \mu$
$V(X) = \sigma^2$
(takve ćemo oznake imati i onda ako X nema normalnu razdiobu – približno normalnu razdiobu, već neku drugu, iako u pravilu razmatramo samo slučajne veličine normalno distribuirane).
Ta su nam dva parametra od X nepoznata pa ih **procjenjujemo** na osnovi mjerenja.

Očekivanje E(X) procjenjujemo aritmetičkom sredinom podataka

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

V(X) procjenjujemo izrazom

$$s^2 = \frac{(x_1 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n-1},$$ (u nazivniku je *n*-1, a ne *n*)

U gornjem je primjeru:

$$\overline{x} = \frac{3 \cdot 0.98 + 2 \cdot 0.99 + 1.00 + 3 \cdot 1.01 + 1.02}{10}$$
$$= 0.997$$

$$s^2 = \frac{3(0.98-0.997)^2 + 2(0.99-0.997)^2 + (1.00-0.997)^2 + 3(1.01-0.997)^2 + (1.02-0.997)^2}{10-1}$$

$= 0.000233$

$s = 0.014944$

Dodatak. **Načela na kojima se zasniva procjenjivanje**.

Slučajne vrijednosti slučajne veličine X.

Slučajna veličina X može postići bilo koju svoju vrijednost.

Neka je $x_1, x_2, ..., x_n$ nekih n slučajnih vrijednosti slučajne veličine X. Postavlja se pitanje približne rekonstrukcije slučajne varijable X iz ovih *n* slučajnih vrijednosti. Općenito, pitanje slučajnih vrijednosti slučajne varijable je vrlo važno i teško praktično i teoretsko pitanje. Postoje algoritmi za približno određivanje slučajnih vrijednosti, koje se obično nazivaju **pseudoslučajne vrijednosti**. Na primjer, u programskom paketu Mathematica, pseudoslučajne vrijednosti dobivaju se naredbom `RandomArray.`

Primijenom te naredbe na normalnu distribuciju s očekivanjem $\mu = 175$ i standardnom devijacijom $\sigma = 5$, za 100 slučajnih vrijednosti dobilo se:

```
{172.245,171.528,175.126,181.414,170.207,178.076,172.062,172.4
66,163.106,172.987,178.936,170.424,188.639,174.808,172.607,170
.222,176.149,171.733,179.166,172.677,169.084,179.869,179.148,1
63.325,174.914,170.227,170.328,173.236,169.499,183.918,177.506
,174.083,179.498,163.901,181.032,178.373,180.085,162.944,172.3
93,176.77,183.359,175.51,165.857,175.806,173.678,173.769,170.8
66,165.969,180.366,169.439,178.993,178.954,166.12,173.062,176.
924,179.091,173.304,165.135,181.489,179.646,183.993,169.244,17
2.846,169.152,177.249,173.359,177.106,182.76,174.611,177.011,1
65.135,173.365,170.879,177.681,170.9,177.904,179.597,170.347,1
75.311,176.744,179.578,181.396,178.267,178.185,175.475,184.13,
166.898,178.865,170.939,181.221,175.353,176.94,181.164,177.516
,173.84,171.767,173.072,172.221,172.539,183.831}
```

Da bismo bolje uočavali ove podatke primijenimo naredbu **Sort**, kojom dobijemo uređenu listu:

```
{160.234,163.649,164.078,165.826,165.905,166.895,166.908,166.9
87,167.01,167.226,167.259,167.604,168.073,168.17,168.684,168.6
88,169.749,170.298,170.31,170.398,170.43,170.622,170.778,170.8
34,171.,171.227,171.446,171.549,171.694,171.72,171.758,171.832
,172.038,172.323,172.38,172.81,172.889,173.16,173.194,173.255,
173.357,173.514,173.662,173.858,173.93,173.95,174.034,174.073,
174.115,174.171,174.336,174.545,174.621,174.627,174.713,174.78
1,175.104,175.376,175.571,175.579,175.631,175.714,175.771,176.
047,176.069,176.296,176.332,176.477,176.485,176.56,176.572,176
.632,176.995,177.663,178.582,178.625,178.67,178.718,179.48,179
```

.74,179.778,179.884,179.887,180.1,180.169,180.585,180.62,180.749,180.927,180.994,181.07,181.334,181.863,181.93,182.203,182.534,183.427,184.706,185.716,187.524}

Sad se možemo uvjeriti u funkcioniranje pravila *tri sigme*. Naime, prebrojavanjem dobijemo:
u intervalu <170,180> je 66 podataka (idealno bi trebalo biti 68)
u intervalu <165,185> je 95 podataka (idealno bi trebalo biti također 95)
u intervalu <160,190> je svih 100 podataka (kako bi trebalo biti i idealno).

U ovom primjeru pošli smo od poznate slučajne varijable (tj. poznate distribucije). U praksi se pojavljuje situacija da znademo samo podatke, a da ne znamo izvornu distribuciju.

Tada očekivanje μ procjenjujemo aritmetičkom sredinom podataka $\bar{x} = \dfrac{x_1 + x_2 + ... + x_n}{n}$

To se zasniva na sljedećem razmišljanju. Uz slučajnu varijablu X razmatramo ***n*** **nezavisnih** slučajnih varijabla $X_1, X_2, ..., X_n$ **jednako distribuiranih kao i X**.
Uz takvo razmišljanje na *n* slučajnih vrijednosti $x_1, x_2, ..., x_n$ slučajne varijable X možemo gledati ovako:
x_1 je slučajna vrijednost od X_1
x_2 je slučajna vrijednost od X_2
.
.
.
x_n je slučajna vrijednost od X_n.

Sad se $\bar{x}$ može interpretirati kao slučajna vrijednost od $\bar{X} := \dfrac{X_1 + X_2 + ... + X_n}{n}$.

Kako je $E(\bar{X}) = \mu$, tj. očekivana vrijednost od $\bar{X}$ je μ, kažemo da je $\bar{x}$, kao slučajna vrijednost od $\bar{X}$, **nepristrana procjena** od μ.
U ovom slučaju, primjenom naredbe **Mean**, dobijemo (na jednu decimalu)
$\bar{x}$**=176.0**
Zaključimo: μ = **175**, a procjenom iz 100 slučajno odabranih vrijednosti dobili smo $\bar{x}$**=176.0,** pa procjenjujemo $\mu \approx 176.0$.
Dakle, iako je broj podataka bio relativno velik, pri procjenjivanju je došlo do grješke.

Slično postupamo pri procjeni varijance, odnosno standardne devijacije. Kako smo vidjeli u Dodatku, ako je:

$$S^2 := \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + ... + (X_n - \bar{X})^2}{n-1}.$$

Onda je $E(S^2) = \sigma^2$.

Zato je $s^2 = \dfrac{(x_1 - \bar{x})^2 + ... + (x_n - \bar{x})^2}{n-1}$

nepristrana procjena varijance σ^2. Time se objašnjava činjenica što je u nazivniku *n-1* (korigirana varijanca uzorka), a ne *n* (varijanca uzorka).
Nastavljajući s početnim primjerom, koristeći se naredbom **Variance[data],**

dobivamo, približno na dvije decimale,
$s^2 = 23,80$
odnosno,
$s = 4.88$,
što je vrlo dobra procjena stvarne standardne devijacije $\sigma=5$.
Ovim načelom, koje smo ilustrirali na primjeru procjene očekivanja i varijance iz uzorka, koristimo se općenito pri procjenjivanju. Ubuduće nećemo izvoditi formule koje dobijemo ovim načelom, već ćemo ih samo napisati.

2. Interval pouzdanosti za očekivanje – prava vrijednost mjerene veličine.

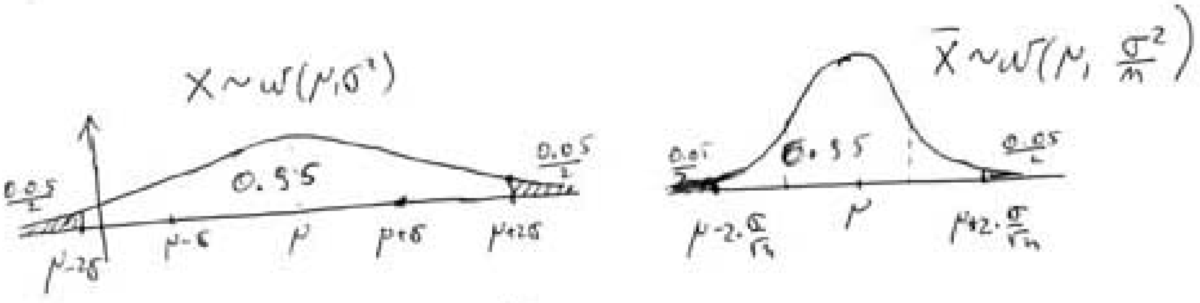
Očekivanje procjenjujemo aritmetičkom sredinom podataka, ali aritmetička sredina ne mora biti (i u pravilu nije) jednaka (nepoznatom) očekivanju. Zato nas zanima **interval** oko $\bar{x}$ unutar kojega će, uz određenu sigurnost, biti očekivanje μ. To je **interval pouzdanosti**.
Jasno je da širina intervala pouzdanosti ovisi o razini sigurnosti da se očekivanje nađe u njemu (što je ta razina veća, interval pouzdanosti je širi).
Interval pouzdanosti se određuje na osnovi sljedećih važnih činjenica (koje se mogu strogo matematički formulirati i dokazati).

Ako je X normalno distribuirana onda je i $\bar{X}$ normalno distribuirana s parametrima μ i $\frac{\sigma^2}{n}$, gdje je $\bar{X} := \frac{X_1 + ... + X_n}{n}$ i gdje su $X_1,...,X_n$ nezavisne slučajne varijable (što aludira na *n* nezavisnih mjerenja) jednako distribuiranih kao i X (što aludira na to da smo svaki put mjerili vrijednost slučajne veličine X).
dakle:

$$\bar{X} \sim N(\mu, (\frac{\sigma}{\sqrt{n}})^2).$$

Zato se $\bar{x}$ može shvatiti kao slučajna vrijednost normalne slučajne varijable s distribucijom $N(\mu, (\frac{\sigma}{\sqrt{n}})^2)$.
Vrijedi i više, naime
$\bar{X}$ ima očekivanje μ i varijancu $\frac{\sigma^2}{n}$, bez obzira kako je X bila distribuirana . Razlika je u tome što za opću X, aritmetička sredina ne mora biti normalno distribuirana.
Veličina $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ zove se **standardna grješka**. Ona je to manja što je *n* veći (što je prirodno, jer što je broj mjerenja veći sigurnost prosjeka treba biti veća).
Kako je $\bar{X}$ normalno distribuirana (sl.1.) ona, prema pravilu "dvije sigme", postiže, s vjerojatnošću većom od 0.95 sve vrijednosti u intervalu $\pm$ *dvije sigme oko sredine*, specijalno, i očekivanje μ bi se tu trebalo naći s tom vjerojatnošću. Zaključak:

$$P(\bar{X} - 2\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 2\frac{\sigma}{\sqrt{n}}) > 0.95$$



Sl. 1.

Zato je, uz 95% vjerojatnost, interval pouzdanosti

$$< \bar{x} - 2\frac{\sigma}{\sqrt{n}}, \bar{x} + 2\frac{\sigma}{\sqrt{n}} >$$

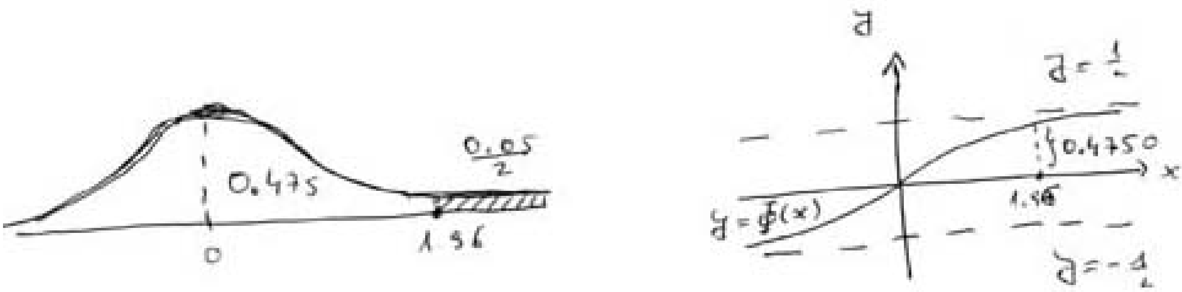
Smisao intervala pouzdanosti nije da se očekivanje μ u njemu nalazi s vjerojatnošću 0.95 (naime μ nije slučajna veličina i nalazi se ili ne nalazi u tom intervalu). Taj se smisao može interpretirati na primjer tako da bi se odprilike u 95 od 100 ponavljanja ovih *n* mjerenja, aritmetička sredina $\bar{x}$ našla u intervalu

$< \mu - 2\frac{\sigma}{\sqrt{n}}, \mu + 2\frac{\sigma}{\sqrt{n}} >$ (što bismo mogli provjeriti da znamo μ i σ),

a to je isto kao da kažemo da bi se odprilike u 95 od 100 ponavljanja, očekivanje μ našlo u intervalu $< \bar{x} - 2\frac{\sigma}{\sqrt{n}}, \bar{x} + 2\frac{\sigma}{\sqrt{n}} >$ (što bismo opet mogli provjeriti da znamo μ i σ).

Umjesto broja 2, za vjerojatnost 0.95, mogli bismo u tablici normalne razdiobe naći precizniji podatak: 1.96. Naime,

P(0<T<1.96) = 0.4750 (broj 0.4750 dobije se kao 0.95/2), gdje je T jedinična normalna razdioba. Dakle $\Phi(1.96) = 0.4750$, odnosno $\Phi^{-1}(0.4750) = 1.96$ (sl.2.).



Sl. 2.

Treba napomenuti da bismo slično mogli odrediti simetrične intervale oko aritmetičke sredine za druge vjerojatnosti, a ne samo za 0.95 .

1. Ako je *n* velik (obično se uzima ako je n>30), onda je veličina $\overline{X}$ približno normalno distribuirana s parametrima μ i $\frac{\sigma^2}{n}$, bez obzira je li X bila normalno distribuirana,dakle:

   $\overline{X} \sim N(\mu, (\frac{\sigma}{\sqrt{n}})^2)$ (približno, ako X nije normalno distribuirana)

Zato u ovom slučaju možemo postupiti kao u 1.

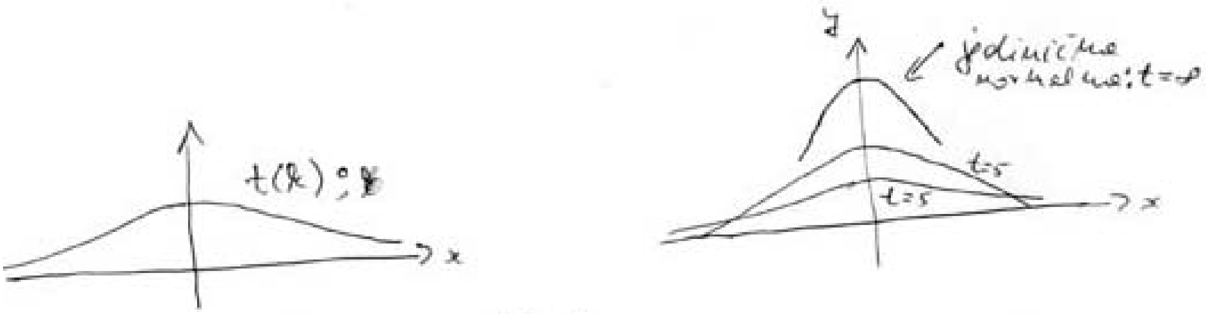
2. Treba napomenuti da je predpostavka da znamo σ (a da μ procijenjujemo iz *n* mjerenja) nerealna, iako nije nemoguća. U praksi smo gotovo uvijek prisiljeni procijeniti σ pomoću *s*. Tada se situacija usložnjava, medjutim za parametre normalne razdiobe, tj. ako predpostavimo da je X normalno distribuirana, problem se može riješiti. Formula iz 1. može se napisati kao

$$\frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim N(0,1^2),$$

medjutim, ako σ zamijenimo sa *S* (slučajna varijabla kojoj je očekivana vrijednost poribližno jednaka *s*), tada jediničnu normalnu razdiobu na desnoj strani treba zamijeniti sa Studentovom t-razdiobom, preciznije:

$$\frac{\overline{X} - \mu}{S}\sqrt{n} \sim t(n-1),$$

gdje je t(n-1) Studentova razdioba s k=n-1 stupnjeva slobode (sl.3.)

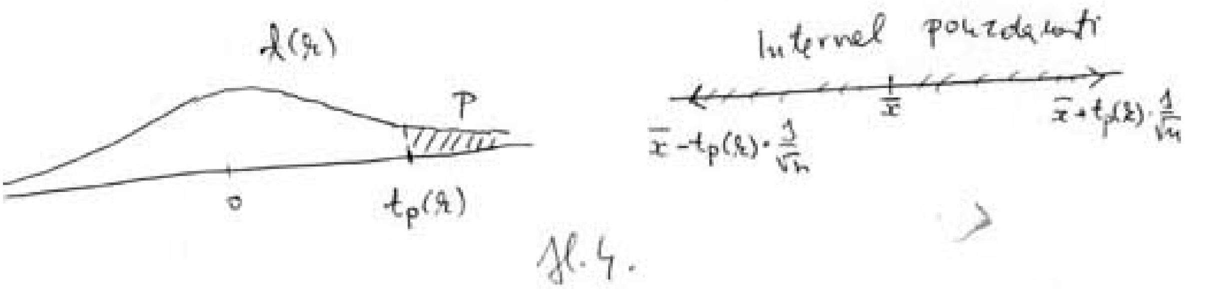


Sl. 3.

Zato je (sl.4.):

$$p(\overline{X} - t_p(k)\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_p(k)\frac{S}{\sqrt{n}}) = 1 - 2p$$

gdje je značenje broja $t_p(k)$ objašnjeno na sl. 4.



Sl. 4.

Tu treba biti pažljiv jer se u literaturi katkad pojavljuju i tzv. dvostrane tablice, uz uobičajene jednostruke.

Interval pouzdanosti, uz vjerojatnost 1-2p, sad je $<\bar{x}-t_p(k)\frac{s}{\sqrt{n}}, \bar{x}+t_p(k)\frac{s}{\sqrt{n}}>$.

Ako je *n* dovoljno velik, recimo oko 30, onda je t(n-1) praktično jednaka jediničnoj normalnoj razdiobi, pa postupamo kao u primjeru.

Primjer 2. U 40 mjerenja neke normalno distribuirane veličine, dobiveno je $\bar{x}$=32.45 i s=2.44. Nađite interval pouzdanosti za očekivanje te slučajne veličine, uz vjerojatnost:

a) 0.95
b) 0.90

Tu je n=40 što je dovoljno veliko da koristimo normalnu razdiobu.

a) Interval pouzdanosti je $<32.45-1.96\frac{2.44}{\sqrt{40}}, 32.45+1.96\frac{2.44}{\sqrt{40}}>$

$$=<31.69; 33.21>$$

b) Faktor kojim ćemo sad množiti (umjesto s faktorom 1.96) naći ćemo u tablici normalne razdiobe kao broj $\Phi^{-1}(0.45)=1.645$ (sl.5.). Dakle, sad je interval pouzdanosti uži:

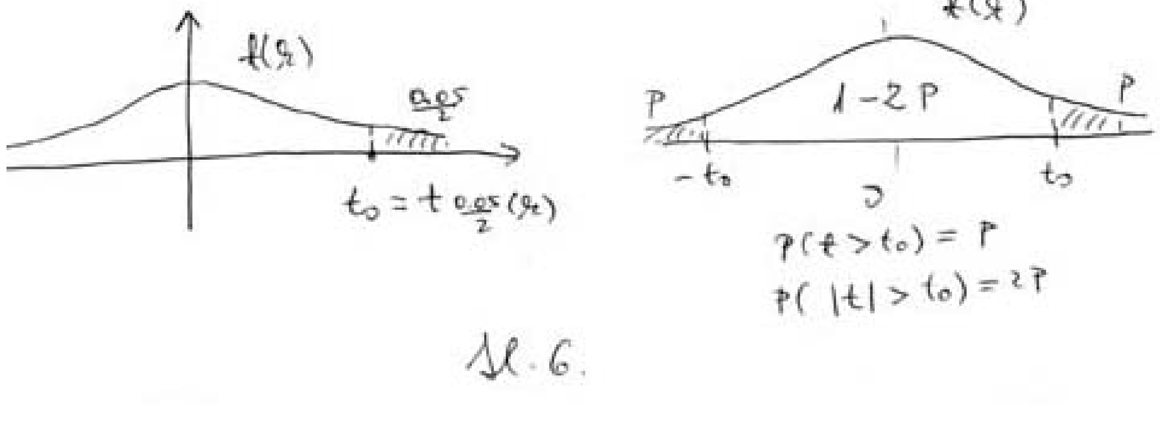
$$<32.45-1.645\frac{2.44}{\sqrt{40}}, 32.45+1.645\frac{2.44}{\sqrt{40}}>$$

$$=<31.82; 33.08>$$

Rezultati su interpretirani geometrijski na slici.

Sl. 5.

Ako je *n* mali (do 30). Tada, uz pretpostavku da slučajna veličina X ima normalnu razdiobu, interval pouzdanosti određujemo ovako.
Faktor kojim ćemo množiti standardnu grješku (točnije, njenu procjenu) određujemo, za vjerojatnost 0.95, iz tablica Studentove (t-razdiobe) s k=n-1 stupnjeva slobode kao broj $t_{0.05/2}(k)$ za kojega vrijedi $P(|t|> t_{0.05/2}(k))<0.05$, a broj 0.05 dobije se kao 1-0.95 (sl.6.).



Sl. 6.

Opet ponavljamo da kod uporabe tablica t-razdiobe treba paziti jer su u nekima tabelirane vrijednosti t_0 za koje je $P(|t|>t_0) = p$, gdje je p=0.05 ili 0.025 ili 0.01 itd.,
a u nekima su tabelirane vrijednosti t_0 za koje je $P(t>t_0)=p$ (tu nema apsolutne vrijednosti) pa su vrijednosti u prvim tablicama za, recimo p=0.05, iste kao i u drugim tablicama za p=0.025 (naravno uz isti broj stupnjeva slobode *k*).
Naravno, ako se služimo određenim statističkim paketom, onda interval pouzdanosti dobijemop izravno.

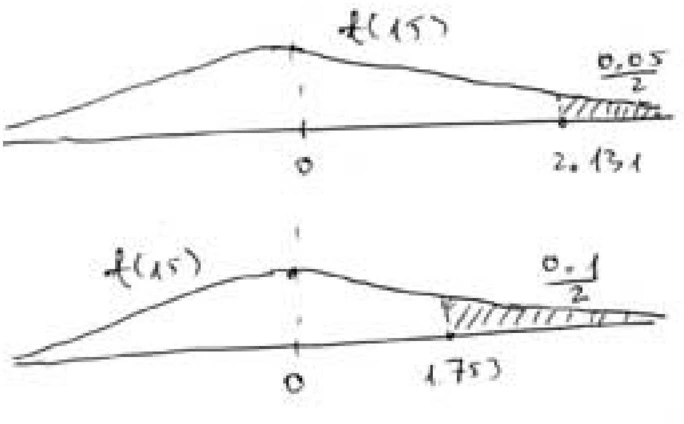
Primjer 3. Iz n=16 mjerenja dobiveno je $\bar{x} = 12.44, \quad s = 1.54$.
Odredimo interval pouzdanosti za vjerojatnost:
a) 0.95
b) 0.90

k=16-1 = 15

a) Tu je, prema prihvaćenim oznakama, 2p=0.05 , $t_{0.05/2}(15)=2.131$, $\frac{s}{\sqrt{n}} = \frac{1.54}{4}$

Interval pouzdanosti je:
$$< 12.44 - 2.131\frac{1.54}{4}, 12.44 + 2.131\frac{1.54}{4} >$$
= <11.62; 13.26>.

b) 2p=0.1, $t_{0.1/2}(15) = 1.753$ (sl.7.).



sl.7.

Interval pouzdanosti je
$$< 12.44 - 1.753\frac{1.54}{4}, 12.44 + 1.753\frac{1.54}{4} >$$
= <11.77; 13.11>.

Taj je interval uži nego prethodni (što je jasno jer je sad vjerojatnost manja)

Da je bilo n=4, a ostali podatci isti kao i prije, intervali pouzdanosti, uz istu vjerojatnost bili bi dva puta širi (jer bismo u standardnoj grješki dijelili s 2 umjesto s 4). To je prirodno (jer interval pouzdanosti treba biti to uži što je broj mjerenja veći).

3. Testiranje varijance i očekivanja

Skicirat ćemo postupak testiranja očekivanja i varijance normalno distribuiranih slučajnih varijabla. U mnogim slučajevima u praksi važno je da varijanca ne bude prevelika (jer to znači preveliko rasipanje). Zato bi, pri ozbiljnom poslu, testiranje varijance u pravilu trebalo prethoditi testiranju očekivanja.

3.1. Testiranje varijance.

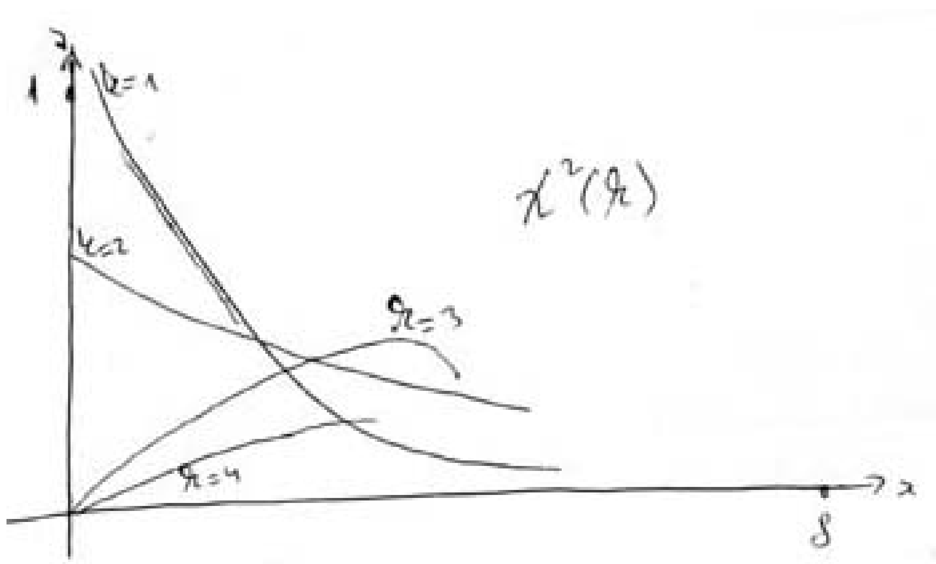
A. Predpostavimo da je X normalno distribuirana slučajna veličina s nepoznatom varijancom σ^2.

Nepoznatu varijancu procijenili smo sa s^2 na osnovi n mjerenja.

Testiramo hipotezu:
H_0: $\sigma^2 = \sigma_0^2$,
za neku deklariranu vrijednost σ_0^2.
Testiranje se zasniva na činjenici iz teorije vjerojatnosti da je:

$$k\frac{S^2}{\sigma^2} \sim \chi^2(k),$$

gdje je $\chi^2(k)$ hi-kvadrat razdioba s k:=n-1 stupnjeva slobode (sl.8.) i ona se zove **test-statistika**. Zato $k\frac{s^2}{\sigma^2}$ možemo interpretirati kao slučajnu vrijednost slučajne veličine $\chi^2(k)$.



Sl.8.

To znači, ako je H_0 istinita hipoteza (slutnja), onda je $k\dfrac{s^2}{\sigma_0^2}$ slučajna vrijednost slučajne veličine $\chi^2(k)$ (dodali smo indeks 0), pa se lijeva strana, kao pozitivan broj ponaša prema njoj.

Postoje dvije mogućnosti.

(I) $s^2 > \sigma_0^2$ (koja je u praksi češća). Tada je, u pravilu, kontrahipoteza $\sigma^2 > \sigma_0^2$, dakle imamo:

H_0: $\sigma^2 = \sigma_0^2$

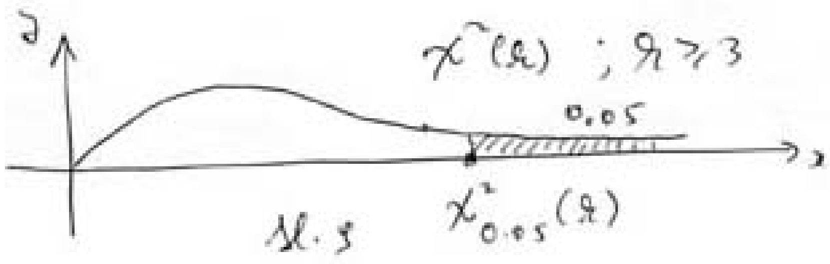
H_a: $\sigma^2 > \sigma_0^2$,

Tada računamo:

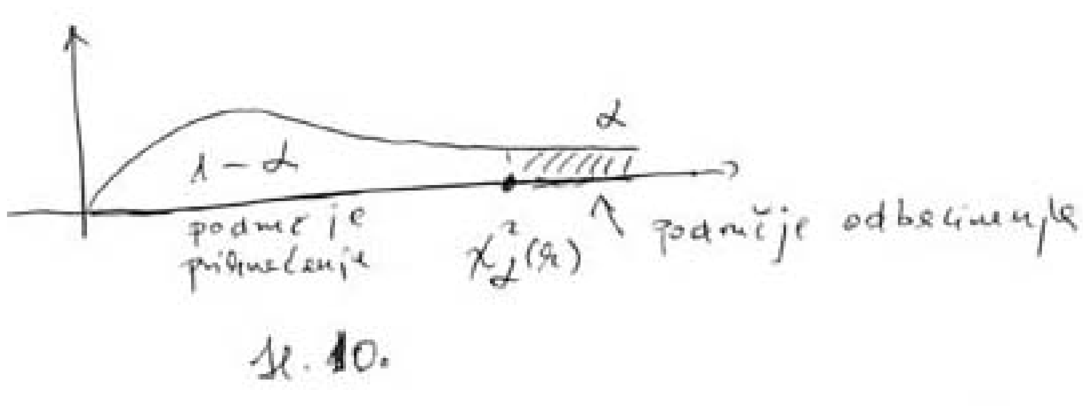
$$W_{\exp} = k\frac{s^2}{\sigma_0^2}$$, gdje je k=n-1.

Ako je $W_{\exp} < (Hi)^2_{0.05}(k)$ hipoteza se prihvaća, inače se odbacuje (sl.9.).

Broj na desnoj strani dobije se iz tablica "hikvadrat" razdiobe za *k* stupnjeva slobode i smisao je da je vjerojatnost da ta razdioba poprimi rezultat veći od tog broje jednaka 0.05 (tako bi bilo i za neki drugi nivo signifikantnosti).



Nivo signifikantnosti (razina značajnosti). Broj α =0.05 zove se nivo signifikantnosti. To je općeprihvaćena vrijednost, medjutim, ona može biti, ovisno o problematici, 0.1, 0.01, 0.025 itd. Područje ispod grafa funkcije gustoće test-statistike (u ovom slučaju $(Hi)^2$ razdiobe), dijeli se na dva dijela (sl.10.), jedan manji površine α (to je područje odbacivanja), jedan veći površine $1-\alpha$ (to je područje prihvaćanja).



Smisao je, za α =0.05 sljedeći:
Ako je nula hipoteza istinita onda će se, odprilike, u 95 od 100 ponavljanja po n mjerenja, eksperimentalni podatak W_{exp} naći u području prihvaćanja, a oko 5 puta u području odbacivanja.
Općenito, α **je pogrješka prve vrste**, tj.
α := vjerojatnost da hipotezu H_0 odbacimo pod uvjetom da je istinita.
Analogno:
1-α := vjerojatnost da hipotezu H_0 prihvatimo pod uvjetom da je istinita.
Dakle, **pogrješno je shvaćanje**, inače široko rasprostranjeno, da je to vjerojatnost da je nulta hipoteza istinita. Naprotiv, ako je α manje, tj. 1-α veće, onda ćemo biti tolerantniji prema razlici. Konkretno, na razini značajnost α =0.01, možda nećemo odbaciti nula hipotezu, koju smo odbacili za α =0.05.

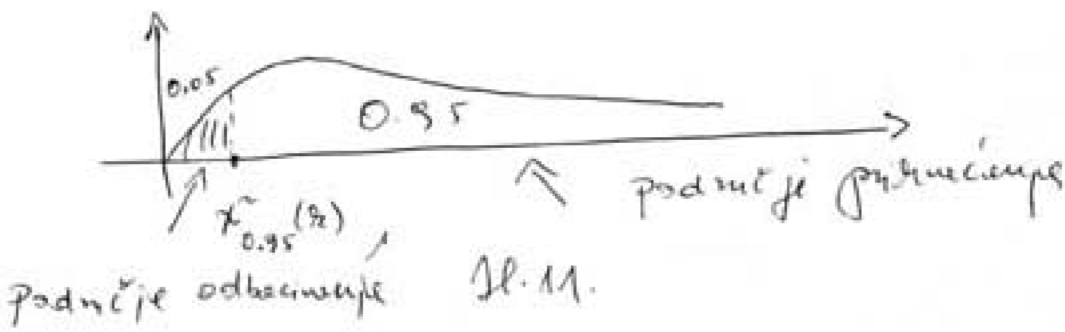
(II) $s^2<\sigma_0^2$

Tada je, u pravilu, kontrahipoteza $\sigma^2 < \sigma_0^2$, dakle imamo:

H_0: $\sigma^2 = \sigma_0^2$

H_a: $\sigma^2 < \sigma_0^2$

Tada hipotezu prihvaćamo ako je $W_{exp} > (Hi)^2_{0.95}(k)$
(znak nejednakosti se mijenja i umjesto 0.05 stavljamo 0.95).
Geometrijsko je tumačenje dano slikom 11.



(B) Testiranje hipoteze $\sigma_1^2 = \sigma_2^2$ (F-test)

Predpostavimo da imamo dvije normalno distribuirane slučajne veličine:

X s očekivanjem μ_1 i varijancom σ_1^2
Y s očekivanjem μ_2 i varijancom σ_2^2.
Očekivanja i varijance tih slučajnih varijabla su nam nepoznate i procjenjujemo ih redom:
Za X iz n_1 mjerenja s $\bar{x}_1$, odnosno s s_1^2,
Za Y iz n_2 mjerenja s $\bar{x}_2$, odnosno s s_2^2.

Testiramo hipotezu o jednakosti tih varijanca. Pri tom predpostavimo da su indeksi odabrani tako da bude $s_1^2 > s_2^2$ i da smo za kontrahipotezu odabrali $\sigma_1^2 > \sigma_2^2$. Dakle imamo:

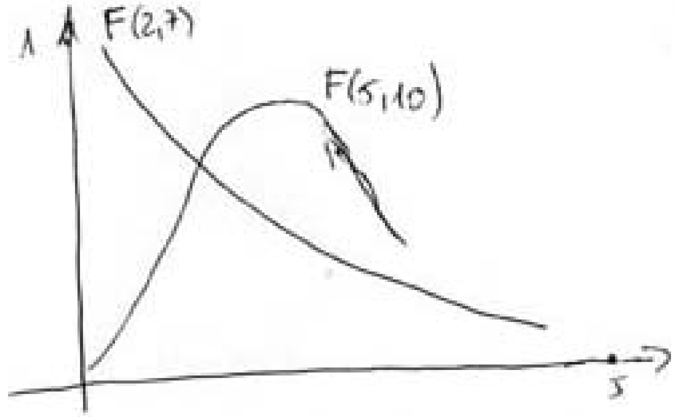
H_0: $\sigma_1^2 = \sigma_2^2$

H_a: $\sigma_1^2 > \sigma_2^2$.

Testiranje se zasniva na činjenici, da je, uz pretpostavku da je nulta hipoteza istinita:

$$\frac{S_1^2}{S_2^2} \sim F(k_1,k_2),$$

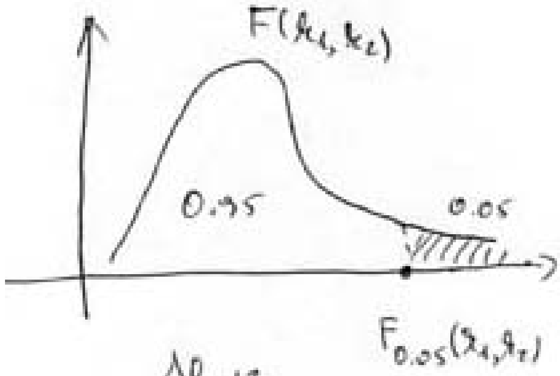
Fisherova razdioba s $(k_1,k_2)= (n_1-1,n_2-1)$ stupnjeva slobode (sl.12.). Tu velika slova primijenjujemo, kao i prije, za odgovarajuće slučajne varijable.



Sl. 12.

Hipotezu, primjenom F-testa (u pojednostavljenom obliku), provjeravamo ovako:

1. Računamo $F_{\exp} = \dfrac{s_1^2}{s_2^2}$
2. U tablici *F* razdiobe očitavamo broj $F_{0.05}(k_1,k_2)$, gdje je $k_1=n_1-1$, $k_2=n_2-1$ (sl.13.).



Sl. 13

3. Ako je $F_{\exp} < F_{0.05}(k_1, k_2)$ hipotezu o jednakosti prihvaćamo, a u suprotnome odbacujemo (tj. smatramo da je razlika među njima bitna).

Napominjemo opet da je postupak testiranja varijance sofisticiraniji od ovog pojednostavljenog pristupa.

3.2. Testiranje očekivanja

(A) Testiranje hipoteze $\mu = \mu_0$ (t-test)

Predpostavimo da je X normalno distribuirana slučajna veličina s očekivanjem μ i varijancom σ^2.

Neka smo na osnovi n mjerenja dobili procjene:
$\bar{x}$ za njeno očekivanje μ,
s^2 za njenu varijancu σ^2.

Testiramo hipotezu:

H_0: $\mu = \mu_0$,
gdje je μ_0 neka deklarirana vrijednost.
Napominjemo da bismo prije toga trebali provjeriti hipotezu o bliskosti varijanca (koju treba formulirati), a nakon što testiranje varijanaca pozitivno prođe, možemo pristupiti testiranju očekivanja.

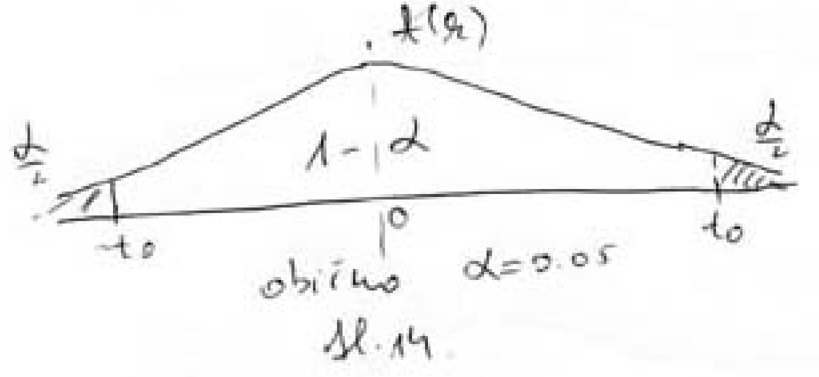
Testiranje nulte hipoteze zasniva na činjenici iz teorije vjerojatnosti, da je
$\dfrac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$ ~t(n-1), Studentova razdioba s k:=n-1 stupnjeva slobode. Zato je, uz predpostavku

da je nulta hipoteza istinita ispunjeno $\dfrac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$ ~t(n-1).

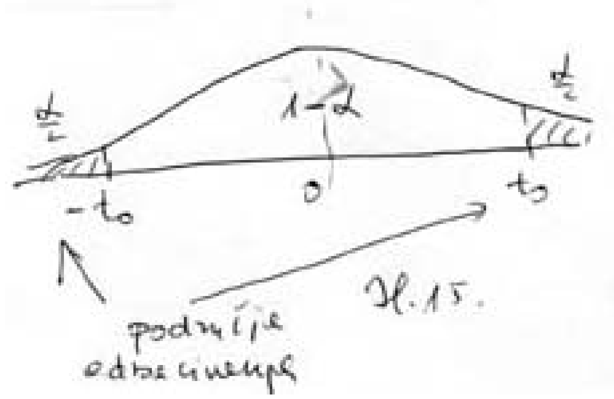
Zato broj $\dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ možemo interpretirati kao slučajnu vrijednost slučajne varijable t(n-1).
Postupak opisujemo uz kontrahipotezu $\mu \neq \mu_0$, dakle imamo:

H_0: $\mu = \mu_0$
H_a: $\mu \neq \mu_0$

1. Računamo $|t_{\exp}| = \dfrac{|\bar{x} - \mu_0|}{\frac{s}{\sqrt{n}}}$.
2. U tablici t-razdiobe određujemo kritičnu vrijednost t_0 (analogno kao i prije, ovisno o broju stupnjeva slobode k=n-1, nivou signifikantnosti što je obično 0.05 i kontrahipotezi koja je, ako drukčije ne specificiramo $\mu \neq \mu_0$)
3. Ako je $|t_{exp}| < t_0$ hipotezu prihvaćamo, inače je odbacujemo (sl.14.).



Napomena o razini značajnosti i području odbacivanja.
Za razliku od testiranja varijance gdje se područje odbacivanja sastoji od jednog dijela, ovdje područje odbacivanja ima dva simetrična dijela, svaki površine $\frac{\alpha}{2}$, gdje je α nivo signifikntnosti (sl.15.). To je zato što je kontrahipoteza oblika $\mu \neq \mu_0$, pa se dopuštaju otkloni na obje strane. Dakle, u slučaju α =0.05, broj t_0, označava broj iza kojega je ispod grafa t-razdiobe površina jednaka 0.025.



Primjer 4. Proizvođač kemikalija je deklarirao na svojim proizvodima da sadrže 1 litru kemikalije uz maksimalnu pogrješku ± 0.09 litara. Kupac mjerenjem uzorka od 12 posuda ustanovio prosječni rezultat 0.97 uz standardno odstupanje 0.04. Jesu li rezultati u skladu s deklaracijom?

Tu je, prema pravilu «tri sigme», σ_0 =0.03, jer je 3·0.03 = 0.09. Zato je:
$\mu_0 = 1.00, \quad \sigma_0^2 = 0.03^2, \quad n = 12, \quad \bar{x} = 0.97, \quad s^2 = 0.04^2.$

Prvo treba testirati hipotezu o jednakosti varijanca:
H_0': $\sigma^2 = \sigma_0^2$.
Dobivamo:
k=n-1 = 11

$$W_{\exp} = k\frac{s^2}{\sigma_0^2}$$
=19.5556

U tablici “hikvadrat” razdiobe za k=11, i nivo signifikantnosti 0.05 dobivamo pripadajuću kritičnu vrijednost 19.6751.
Kako je 19.5556 < 19.6751
hipotezu o jednakosti varajanca prihvaćamo (ali jedva).

Sad prelazimo na testiranje očekivanja.

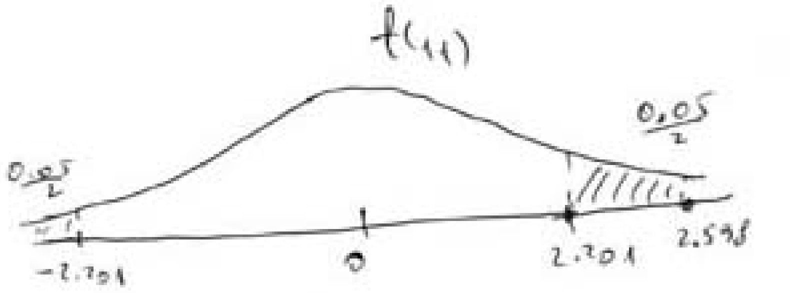
H_0: $\mu = \mu_0$

H_a: $\mu \neq \mu_0$

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

$= 2.598$

Pripadna kritična vrijednost u t-razdiobi (za kontrahipotezu $\mu \neq \mu_0$, uz k=11 i nivo signifikantnosti $\alpha = 0.05$)
t_0 =2.201 (sl.16.).



sl.16.

Kako je 2.598 > 2.201,
hipotezu o jednakosti očekivanja odbacujemo (tj. smatramo da se one bitno razlikuju).
Tako smo odbacili deklaraciju.

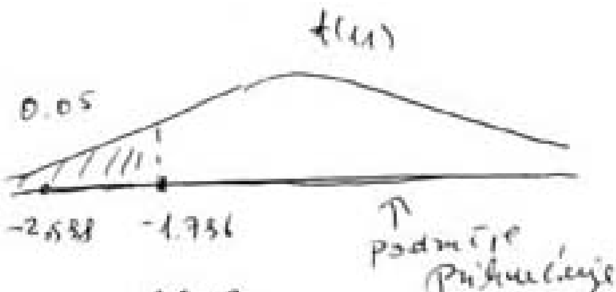
Napomene.
1. Da smo umjesto kontrahipoteze $\mu \neq \mu_0$, uzeli kontrahipotezu $\mu < \mu_0$
(što bismo napravili da su, na primjer, svi rezultati mjerenja ili gotovo svi, bili manji od deklarirane, što ovdje vjerojatno nije slučaj), hipotezu o jednakosti bismo još uvjerljivije odbacili jer bi nam kritična vrijednost ispala 1.796, jer je P(t>1.796)=0.05
Naime, tada bismo imali:

H_0: $\mu = \mu_0$

H_a: $\mu < \mu_0$

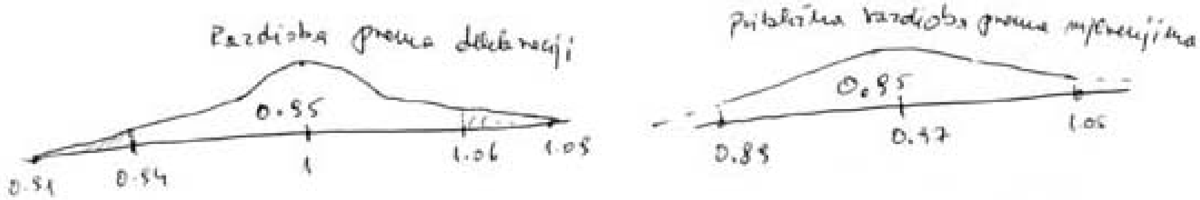
pa bismo gledali (sad bez apsolutne vrijednosti)

$t_{\exp:=} \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = -2.598$, što je u kritičnom području (sl.17.).



sl.17.

2. Uz pretpostavku normalne distribucije sadržaja posuda (što je prirodna predpostavka i već smo je prihvatili), prema pravilu “tri sigme”:
prema deklaraciji je sadržaj između 0.91 i 1.09 (između 0.94 i 1.06 uz vjer. 0.95)
prema mjerenjima je sadržaj (približno jer nije riječ o normalnoj razdiobi) između 0.85 i 1.09 (između 0.89 i 1.05 uz vjer. 0.95)
odakle možemo dobiti intuitivnu predodžbu o tome zašto smo odbacili hipotezu, ali i o tome da smo je umalo prihvatili. Vidimo da je nismo prihvatili jer je vrijednost $\mu_0=1$ ispala izvan intervala pouzdanosti uz vjerojatnost 0.95 koji je 0.97 ± 0.0254 (sl.18.).



sl.18.

4. U deklaraciji bi pisalo da je sadržaj posude 1 ± 0.09 (odakle smo zaključili da je standardno odstupanje, prema pravilu *tri sigme* trećina od 0.09, tj. 0.03). Napominjemo da se u deklaricijama u pravilu koristi pravilo «dvije sigme», pa bi, ako bi tako nešto prihvatili trebali uzeti $\sigma_0=0.045$.

Primjer 5. Možemo li prihvatiti da je sadržaj posude u Primjeru 1 jednak 1 ± 0.015?

U primjeru je bilo n = 10, $\bar{x} = 0.997$, $s = 0.014944$, a prema deklaraciji je
$\mu_0 = 1$, $\sigma_0 = 0.005$ (opet smo išli prema pravilu «tri sigme»)
$W_{exp} = 80.396 > 16.919$ pa se varijance bitno razlikuju. Zato odbacujemo deklaraciju.
Razlog ovog drastičnog odbacivanja jest u tome što početni podatci nisu bili približno normalno distribuirani (što je pretpostavka za važenje testa).

Testiranje hipoteze $\mu_1 = \mu_2$ (t-test).
Tom testu u pravilu predhodi F-test. Nakon što taj prođe nastavlja se s t-testom (testiranju očekivanja), tj. s testiranjem hipoteze:

H_0: $\mu_1 = \mu_2$ (nulta hipoteza)

Hipoteza se, primjenom t-testa, provodi ovako:

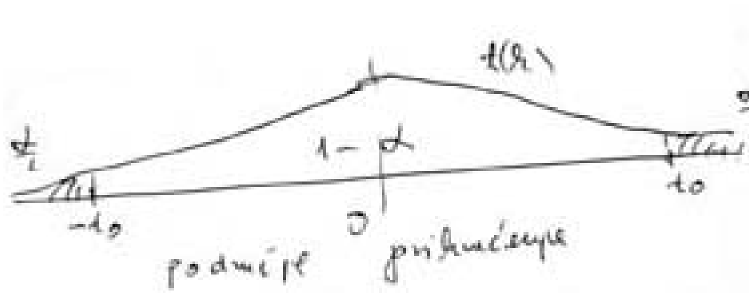
1. Izračuna se:

$$t_{\exp} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{(n_1 - 1)\ s_1^2 + (n_2 - 1)\ s_2^2}{n_1 + n_2 - 2}}\sqrt{\dfrac{n_1 + n_2}{n_1 n_2}}}$$

gdje obično označavamo: $s_d = \sqrt{\dfrac{(n_1 - 1)\ s_1^2 + (n_2 - 1)\ s_2^2}{n_1 + n_2 - 2}}\sqrt{\dfrac{n_1 + n_2}{n_1 n_2}}$

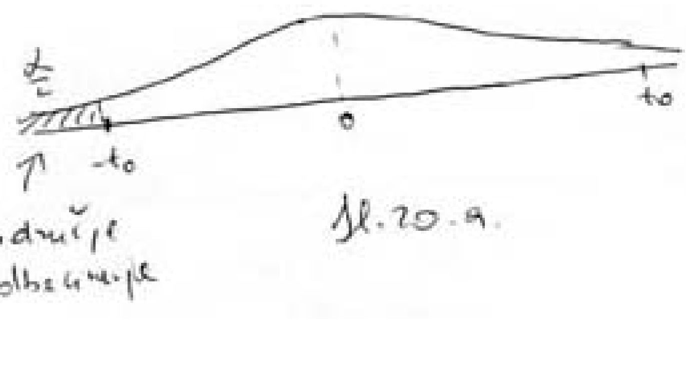
2. Odredi se broj stupnjeva slobode $k=n_1+n_2-2$.
3. Prihvati se neki nivo signifikantnosti α (obično α =0.05, ali može i α =0.01 ili α =0.1) Smisao nivoa signifikantnosti u testiranju je sljedeći:
   P(Postavljena se hipoteza odbacuje| postavljena je hipoteza istinita) = α .
4. Iz tablica t-razdiobe izračuna se kritična vrijednost pomoću koje odredjujemo upada li izračunata vrijednost t_{exp} u kritično područje. Kritična vrijednost ovisi o nivou signifikantnosti α , o broju stupnjeva slobode (dakle o broju mjerenja), ali i o našoj kontrahipotezi koja može biti:

a) $\mu_1 \neq \mu_2$ (kad testiramo jesu li te dvije veličine jednake ili različite). Tada kritična vrijednost t_0 ima značenje: $P(|t|>t_0) = \alpha$ (sl.19.), gdje t označava Studentovu (t-razdiobu).
Hipotezu prihvaćamo ako je $|t_{exp}|<t_0$ (inače je odbacujemo).
Ako izričito drukčije ne kažemo uvijek smatramo da je kontrahipoteza takva.



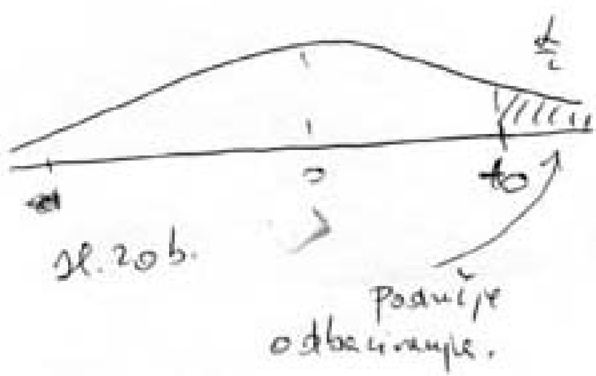
Sl.19.

b) $\mu_1 > \mu_2$ (koja ima smisla samo ako je $\bar{x}_1 > \bar{x}_2$) .
Tada kritična vrijednost t_0 ima značenje: $P(t>t_0) = \alpha$ (t_0 je drukčiji od onog iz a)).
Hipotezu prihvaćamo ako je $t_{exp}<t_0$, inače je odbacujemo (sl.20a).



Sl.20.a.

c) $\mu_1 < \mu_2$ (koja ima smisla samo ako je $\bar{x}_1 < \bar{x}_2$).
Tada kritična vrijednost t_0 takodjer ima značenje: $P(t>t_0) = \alpha$.
Hipotezu prihvaćamo ako je $t_{exp} > - t_0$, inače je odbacujemo (sl.20b).



Da se bolje uvidi razlika između a) , b) i c), neka je $\alpha = 0.05$; k = 8.
Tada je u a) $t_0 = 2.306$, a u b) i c) $t_0 = 1.860$.

Primjer 6. Neka je iz 8 mjerenja neke normalne slučajne veličine dobiven prosjek 12.56 uz standardno odstupanje 1.36; a iz 11 mjerenja druge normalne slučajne veličine prosjek 13.01 uz standardno odstupanje 0.84. Razlikuju li se bitno te veličine?

Podatci se mogu zapisati ovako:

$n_1 = 8, \quad \bar{x}_1 = 12.56, \quad s_1 = 1.36$

$n_2 = 11, \quad \bar{x}_2 = 13.56, \quad s_2 = 0.84$

$k_1 = 7, \quad k_2 = 10, \quad k = 17.$

1. F-test.

H'_0: $\sigma_1^2 = \sigma_2^2$

$$F_{\exp} = \frac{s_1^2}{s_2^2}$$

$$=2.6213$$

$F_{0.05}(k_1, k_2)$=3.14.

Kako je 2.6213 < 3.14 hipotezu prihvaćamo, tj. smatramo da varijance tih slučajnih veličina nisu bitno različite.

2. t-test. Testiramo:

H_0: $\mu_1 = \mu_2$

H_a: $\mu_1 \neq \mu_2$

Dobijemo:

$s_d = 0.504035$

$t_{exp} = - 1.984$, $|t_{exp}| = 1.984$

Kritična vrijednost (za nivo signifikantnosti 0.05 i za k=17) je $t_0 = 2.110$, jer je $P(t>2.110)=0.025$ (polovica od vjerojatnosti 0.05).

Kako je 1.984 < 2.110, hipoteza se prihvaća pa se smatra da se dvije mjerene veličine bitno ne razlikuju.

Sljedeće napomene upozoravaju na relativnost zaključka pri testiranju u odnosu na male promjene podataka ili na odabir kontrahipoteze i razine značajnosti.

Napomena 1. Da smo u podatcima imali $\bar{x}_2 = 13.66$, a da su ostali podatci ostali isti, sve bi bilo isto osim završnog rezultata t_{exp}. Naime, bilo bi:

t_{exp}= - 2.1824, $|t_{exp}|$= 2.1824

a kako je 2.2419 > 2.110, hipotezu o jednakosti očekivanja bismo odbacili.

2. Da smo u izvornom zadatku odabrali kontrahipotezu H_a: $\mu_1 < \mu_2$

(što načelno ima smisla jer je $\bar{x}_1 < \bar{x}_2$), hipotezu o jednakosti očekivanja takodjer bismo odbacili. Naime, tada bi kritična vrijednost bila t_0 = 1.740, jer je, za k=17, P(t>1.740)=0.05 . Kako je 1.984>1.740 nula hipotezu bismo odbacili.

3. Da smo imali sve kao u izvornom zadatku i izvornom rješenju, sam da smo odabrali razinu značajnosti α = 0.1, tada bismo hipotezu takodjer odbacili, jer bi tada kritična vrijednost bila kao i u 2., tj. bilo bi t_0 = 1.740.

4. Testiranje teoretskih razdioba (χ^2 - test)

Jedno od najčešćih pitanja u statistici jest ponašaju li se mjereni podatci prema nekom teoretskom zakonu (razdiobi) ili se bitno od njega razlikuju.

Primjer 7. Registriranjem broja poruka na nekoj adresi u fiksiranom vremenskom intervalu, dobiveni su sljedeći podatci:

0	1	2	3	4	5 ili više
16	16	36	15	10	7

Dakle, u 16 mjerenja nije bila ni jedna poruka, u 16 mjerenja točno jedna, u 36 mjerenja točno 2 itd. Formulacija u zadnjem stubcu je takva jer je, možda bilo i 6 ili 7 poziva koji put, pa smo to skupili u jedan podatak. Ukupno je bilo n=100 mjerenja, koje smo svrstali u L=6 grupa.

Postavlja se pitanje ponašaju li se ti podatci prema Poissonovu zakonu ili, možda, bitno odudaraju od njega. O tome je zaista teško odgovoriti samo uvidom u podatke.

Odgovor na pitanje pomoću χ^2-testa (predloženog Karlom Pearsonom 1900) zasniva se na sljedećem razmišljanju.

Brojevi u drugom redku tablice zovu se **eksperimentalne frekvencije** f_i, dakle:

f_0=16, f_1=16, f_2=36, f_3=15, f_4=10, f_5=7.

Prosječan broj poruka, dobije se kao:

$$a=\frac{0\bullet 16+1\bullet 16+2\bullet 36+3\bullet 15+4\bullet 10+5\bullet 7}{100}=2.08$$

Jasno je da je a procjena za očekivanje slučajne varijable X koja registrira broj poruka u fiksiranom vremenskom intervalu, a ako se podatci zaista ponašaju prema Poissonovu zakonu onda je, približno, *X ~ P(a)*, tj. *X ~ P(2.08).*

U nastavku ćemo izračunati pripadne **teoretske vjerojatnosti** p_i, i=0,1,2,... te razdiobe, prema formuli

$$p_i := e^{-a}\frac{a^i}{i!}, \quad \text{tj.} \quad p_i := e^{-a}\frac{2.08^i}{i!}$$

i pripadne **teoretske frekvencije** prema formuli

$f_{ti} := n \cdot p_i$, tj. $f_{ti} := 100 \cdot p_i$. Imamo, dakle:

$p_0 = e^{-2.08} =$ 0.124930 $\qquad f_{t0} = 12.4930$

$p_1 = e^{-2.08}\dfrac{2.08}{1!} = 0.259855$ $\qquad f_{t1} = 25.9855$

$p_2 =$ 0.270249 $\qquad f_{t2} = 27.0249$

$p_3 =$ 0.187373 $\qquad f_{t3} = 18.7373$

$p_4 =$ 0.097434 $\qquad f_{t4} = 9.7434$

$p_5 := 1 - (p_0 + p_1 + p_2 + p_3 + p_4)$
$= 0.060159$ $\qquad f_{t5} = 6.0159$

Tu smo, umjesto pravog p_5 stavili zbroj svih vjerojatnosti od pete na dalje, tako da ukupan zbroj vjerojatnosti bude 1; slično tako smo dobili da je ukupan zbroj teoretskih frekvencija jednak 100.
Sljedeći je korak uvodjenje **mjere udaljenosti** eksperimentalnih i teoretskih frekvencija:

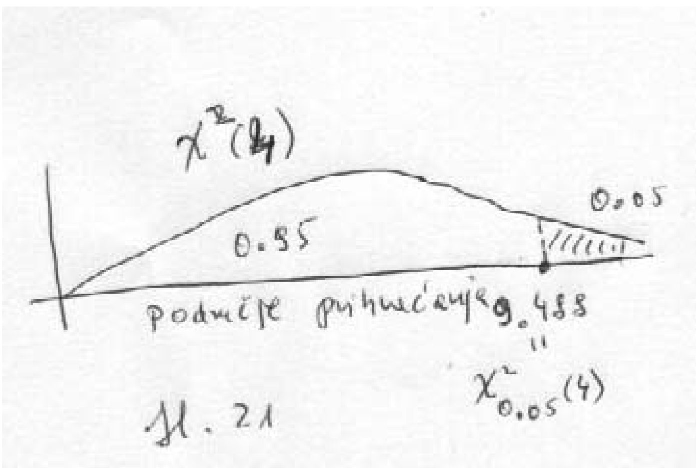
$$\chi^2_{\exp} := \frac{(f_0 - f_{t0})^2}{f_{t0}} + \frac{(f_1 - f_{t1})^2}{f_{t1}} + \frac{(f_2 - f_{t2})^2}{f_{t2}} + \frac{(f_3 - f_{t3})^2}{f_{t3}} + \frac{(f_4 - f_{t4})^2}{f_{t4}} + \frac{(f_5 - f_{t5})^2}{f_{t5}}$$

$$+ \frac{(16 - 12.4930)^2}{12.4930} + \frac{(16 - 25.9855)^2}{25.9855} + \frac{(36 - 27.0249)^2}{27.0249} + \frac{(15 - 18.7373)^2}{18.7373}$$

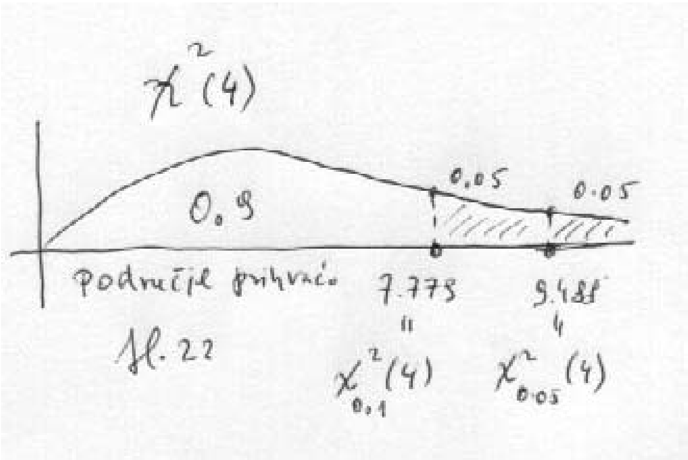
$$+ \frac{(10 - 9.7434)^2}{9.7434} + \frac{(7 - 6.0159)^2}{6.0159}$$

$= 8.715$

Završni je korak prihvaćanje ili odbacivanje hipoteze o Poissonovoj razdiobi. Taj se kriterij zasniva na činjenici, da je, **ako je ispunjena nulta hipoteza**:
H_0 : podatci se ponašaju prema Poissonovoj razdiobi
onda broj $\chi^2_{\exp}$ možemo interpretirati kao slučajnu vrijednost slučajne varijable koja **približno** ima hi-kvadrat razdiobu s $k := L - 2 = 6 - 2 = 4$ stupnjeva slobode (sl.21.).



Iz tablica vidimo da je $\chi^2_{0.05}(4) = 9.488$, što je veće od broja $\chi^2_{\exp}$, pa, uz razinu značajnosti α =0.05, hipotezu o Poissonovoj razdiobi prihvaćamo (iako ne sasvim uvjerljivo).
Takodjer vidimo da je $\chi^2_{0.1}(4) = 7.779$, pa na razini značajnosti α =0.1 tu hipotezu odbacujemo, tj. smatrama da postoji bitno odstupanje od Poissonove razdiobe (sl.22.).



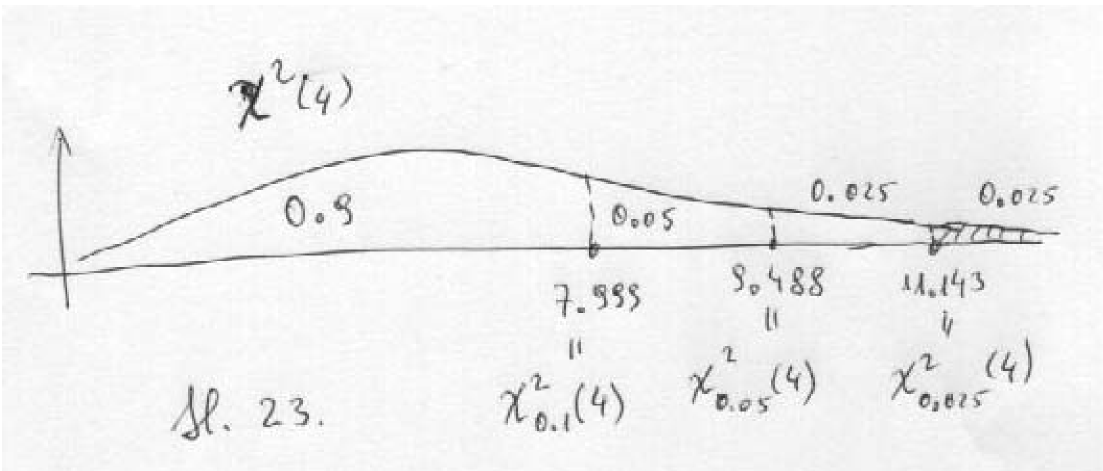
U sljedećem ćemo primjeru dodatno ilustrirati zašto je spomenuta razdioba rubno Poissonova, tako što ćemo samo malo promijeniti podatke.

Primjer 8. Registriranjem broja poruka na nekoj adresi u fiksiranom vremenskom intervalu, dobiveni su sljedeći podatci:

0	1	2	3	4	5 ili više
16	16	37	14	9	8

Treba testirati predpostavku o Poissonovoj razdiobi.

Lako se vidi da je tu, kao i u Primjeru 7. ispunjeno:
n=100, L=6, k=4, a=2.08 pa su i odgovarajuće teoretske frekvencije jednake. Medjutim, tu je $\chi^2_{\exp}$=10.412, pa hipotezu o Poissonovoj razdiobi odbacujemo na razini značajnosti α =0.05.
Treba napomenuti da bismo predpostavku prihvatili na razini značajnosti α =0.025, jer je $\chi^2_{0.025}(4) = 11.143$ (sl.23.).



Općenito, a ne samo za Poissonovu razdiobu, imamo:

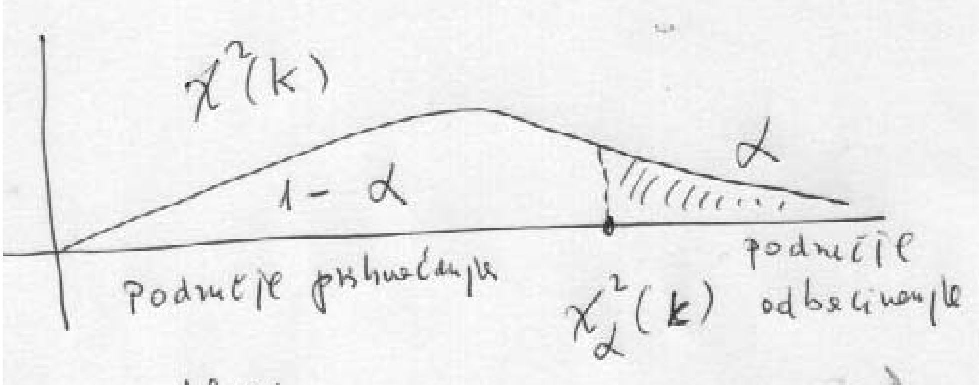
$\chi^2_{\exp} := \sum \frac{(f_i - f_{ti})^2}{f_{ti}}$, $k := L\text{-}l\text{-}1$, gdje je l broj parametara o kojima ovisi teoretska razdioba, tj. l=2 za normalnu i binomnu

l=1 za Poissonovu i eksponencijalnu

l=0 za jednoliku.

Hipotezu o suglasnosti s teoretskom razdiobom prihvaćamo na razini značajnosti α (u pravilu je α =0.05)

ako je $\chi^2_{\exp} < \chi^2_{\alpha}(k)$, inače je odbacujemo (sl.24.).



sl.24.