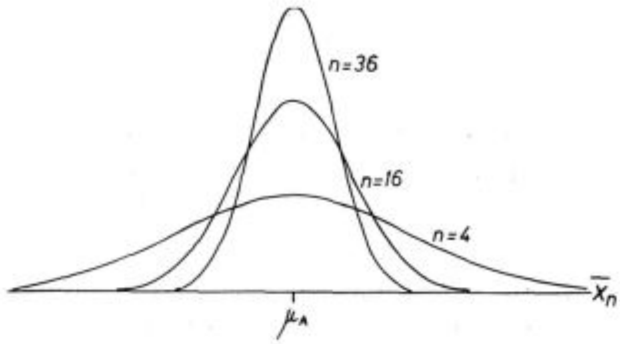
Božidar Brudar, dipl. inž.
Železarna Jesenice

DK: 519
ASM/SLA: U 4 j

Preverjanje statističnih hipotez s pomočjo operacijskih karakteristik

V članku je opisano, kako po predhodni izbiri dopustnih verjetnosti za napako prvega in drugega reda določimo velikost vzorca s pomočjo operacijskih karakteristik (OC curve — operating characteristic curve).

Kot primer je navedeno preskušanje hipoteze $\mu = \mu_0$ v primeru, ko sta standardni deviaciji obeh distribucij znani in enaki.



Slika 1

VZOREC IN POPULACIJA

Pri raziskovalnem delu pogosto skušamo priti do informacij o večji množici podatkov ene vrste (populacija) tako, da analiziramo manjši vzorec. Pri tem pa navadno ne moremo najti točne vrednosti parametra, ki je karakterističen za celo populacijo. Najdemo le interval, v katerem lahko z neko določeno verjetnostjo pričakujemo vrednost parametra.

Zaradi lažjega razumevanja si mislimo populacijo A sestavljeno iz velike množice vrednosti, ki so porazdeljene okrog srednje vrednosti μ_A s standardno deviacijo σ_A. Vzemimo iz te populacije slučajnostni vzorec n vrednosti in izračunajmo aritmetično srednjo vrednost vzorca $\overline{X}_n^{(1)}$. Ponovimo tak poskus in spet povsem slučajno izberimo n vrednosti! Srednja vrednost novega vzorca bo na primer $\overline{X}_n^{(2)}$. Če tak poskus večkrat ponovimo, dobimo množico srednjih vrednosti vzorcev $\overline{X}_n^{(1)}$, $\overline{X}_n^{(2)}, \ldots$, za katere velja, da se porazdeljujejo po normalni distribuciji s srednjo vrednostjo, ki je enaka μ_A in standardno deviacijo $\frac{\sigma_A}{\sqrt{n}}$. Tri take distribucije z različnimi n so narisane na sliki 1.

Očitno je, da so aritmetične srednje vrednosti večjih vzorcev porazdeljene v ožjem pasu okrog vrednosti μ_A.

Na podoben način bi lahko prišli tudi do zaključkov o vrednosti drugih parametrov. Tako bi lahko ocenili standardno deviacijo σ_A iz vrednosti standardnega odklona v vzorcu.

Kadar želimo primerjati karakteristični vrednosti dveh populacij, si tudi pomagamo z manjšimi vzorci. Ko se predhodno odločimo za dopustno napako prve in druge vrste, si velikost vzorca izberemo s pomočjo familije operacijskih karakteristik.

NAPAKE PRVE IN DRUGE VRSTE PRI TESTIRANJU HIPOTEZ

Pri primerjanju karakterističnih vrednosti dveh populacij lahko neko v naprej postavljeno trditev zavržemo ali pa sprejmemo.

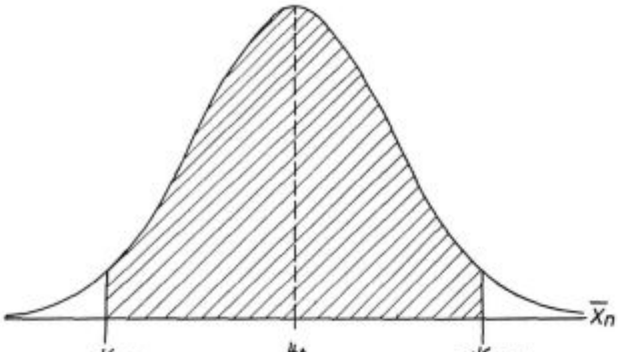
Oglejmo si, na primer, kako postopamo pri preverjanju hipoteze, da je $\mu_B = \mu_A$, pri pogoju, da je $\sigma_B = \sigma_A$, če je μ_B srednja vrednost neke populacije B.

Najprej se vprašajmo, kaj bi lahko pričakovali za srednjo vrednost vzorcev n vrednosti, ki bi jih vzeli iz populacije A s srednjo vrednostjo μ_A in s končno varianco σ_A^2. Ker je pri tem porazdelitev srednjih vrednosti vzorcev normalna, lahko najdemo dve taki vrednosti $-K_{\frac{\alpha}{2}}$ in $+K_{\frac{\alpha}{2}}$ med katerima leži $(1 - \alpha)$. 100 % vseh srednjih vrednosti vzorcev*. Pri tem si mislimo, da smo naše merilo premaknili tako, da je $\mu_A = 0$.

Če si bomo izbrali $\alpha = 0,05$, bomo rekli, da pričakujemo med $-K_{0,025}$ in $+K_{0,025}$ 95 % vseh možnih srednjih vrednosti vzorcev z n podatki. Za srednjo vrednost enega vzorca pa lahko rečemo, da je 95 % verjetno, da bo ležala na intervalu

* Glej članek B. Rodeta v Železarskem zborniku št. 3, letnik 1967, str. 189.

$(-K_{0,025}, +K_{0,025})$. Na sliki 2 je narisana porazdelitvena krivulja za srednje vrednosti vzorcev z n podatki. Šrafirana ploskev predstavlja 95 % vse ploščine med krivuljo in absciso.



Slika 2

Kolikšna mora biti aritmetična srednja vrednost vzorca n vrednosti iz populacije B, da bomo lahko trdili, da je $\mu_B = \mu_A$?

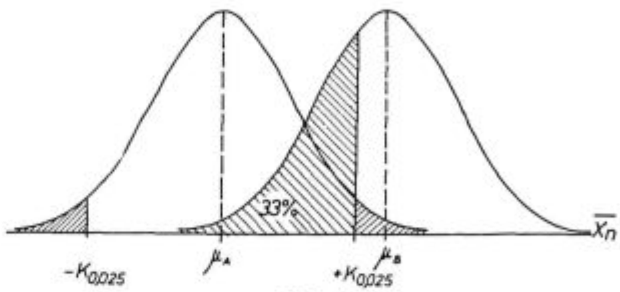
Če je μ_B v **resnici** enak μ_A, lahko pričakujemo na primer s 95 % gotovostjo, da bo $\overline{X}_n$ vzorca n vrednosti iz populacije B ležala med $-K_{0,025}$ in $+K_{0,025}$.

Lahko se odločimo, da bomo sprejeli hipotezo, da je $\mu_B = \mu_A$ vedno takrat, kadar bo srednja vrednost vzorca n vrednosti iz populacije B ležala na intervalu $(-K_{\frac{\alpha}{2}}, +K_{\frac{\alpha}{2}})$. Pri takem sklepanju pa se lahko zgodi, da napravimo tako imenovano napako prve vrste: $\alpha \cdot 100$ % je namreč verjetno, da leži aritmetična srednja vrednost vzorca B izven intervala $(-K_{\frac{\alpha}{2}}, +K_{\frac{\alpha}{2}})$. V takem primeru bi pa zavrgli hipotezo, ki je pravilna. Če bi hoteli zmanjšati možnost za napako prve vrste bi morali interval $(-K_{\frac{\alpha}{2}}, +K_{\frac{\alpha}{2}})$ razširiti.

Zastavimo pa si še drugo vprašanje! Če smo že sprejeli hipotezo, da je $\mu_B = \mu_A$, ali nismo mogoče sprejeli napačne hipoteze? Vse gornje izvajanje velja namreč le za primer, ko je naša hipoteza **v resnici** pravilna.

Če pa bi bila srednja vrednost populacije μ_B enaka vrednosti μ_1, ki je nekoliko **večja** od μ_A, bi imeli opravka z distribucijo, ki je glede na prvo premaknjena nekoliko v desno. Ako bi bila pri tem standardna deviacija σ_B enaka σ_A, bi to lahko ponazorili s sliko 3.

Verjetnost, da bi zdaj srednja vrednost vzorca ležala na intervalu $(-K_{0,025}, +K_{0,025})$ ni več 95 %, ampak manjša. Nikakor pa ni zanemarljivo majna, proporcionalna je srednji šrafirani ploskvi na sliki 3.



Slika 3

Če bi se torej še vedno držali prvotnega predpisa, da bi namreč sprejeli hipotezo $\mu_B = \mu_A$, kadar bi srednja vrednost vzorca padla v interval $(-K_{\frac{\alpha}{2}}, +K_{\frac{\alpha}{2}})$, bi pri tem lahko naredili pomembno napako: lahko bi sprejeli hipotezo, ki v resnici ni pravilna. To imenujemo napako druge vrste. Na sliki 3 smo za vzorec izbrali 16 meritev, za razliko $\mu_1 - \mu_A$ pa $0,6\,\sigma_A$. V takem slučaju je verjetnost za napako druge vrste kar precejšnja, saj bi na primer v povprečju pri eni tretjini vseh vzorcev trdili, da je $\mu_B = \mu_A$, ker bi jih toliko ležalo v intervalu $(-K_{0,025}, +K_{0,025})$, to pa ne bi bilc prav.

Napaka druge vrste je tem večja, čim manjša je razlika med μ_A in μ_B in čim večja je standardna deviacija teh srednjih vrednosti vzorcev. Ta pa se spreminja z velikostjo vzorca.

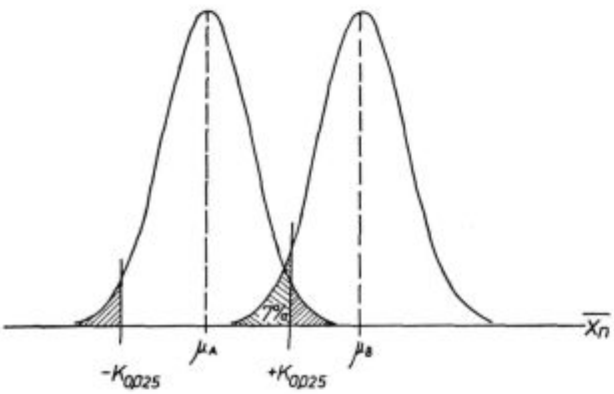
Če bi imeli v gornjem primeru v vzorcu 36 vrednosti namesto 16, bi bila verjetnost, da bi srednja vrednost tako velikega vzorca padla v interval $(-K_{0,025}, +K_{0,025})$ bistveno manjša. Znašala bi le približno 7 % in takšna bi bila tudi verjetnost za to, da bi sprejeli napačno hipotezo $\mu_B = \mu_A$. Na sliki 4 sta narisani porazdelitveni krivulji za primer, če bi imeli po 36 vrednosti v vzorcu. Pri tem pa je razlika $\mu_1 - \mu_A$ prav tako $0,6\,\sigma_A$.

Srednja šrafirana ploskev predstavlja 7 % ploščine med porazdelitveno krivuljo in absciso.

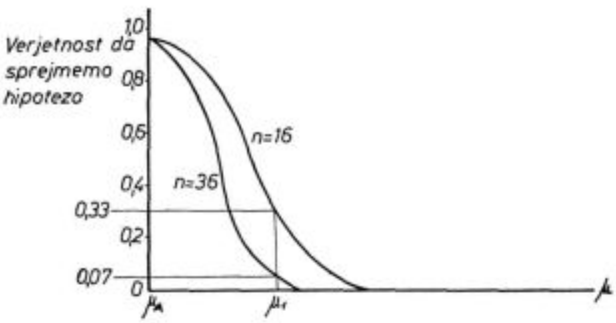
OPERACIJSKA KARAKTERISTIKA

Lahko konstruiramo krivuljo, ki predstavlja verjetnost, da sprejmemo hipotezo $\mu_B = \mu$, če se držimo gornjega kriterija. Na sliki 5 sta narisani dve operacijski karakteristiki — za večji (n = 36) in manjši vzorec (n = 16). Za α smo si izbrali spet vrednost 0,05. Razlika med μ_1 in μ_A pa znaša spet $0,6\,\sigma_A$.

Če bi bil n. pr. naš vzorec sestavljen iz 16 vrednosti, bi bilo 33 % verjetno, da smo s tem, ko smo sprejeli hipotezo, da je $\mu_B = \mu_A$ naredili napako:



Slika 4



Slika 5

ker smo namreč rekli, da je $\mu_B = \mu_A$, ko je pa v resnici morda $\mu_B = \mu_1$. Očitno je verjetnost, da bi zgrešili večjo razliko manjša, za manjšo razliko pa večja, nikakor pa ne gre brez neke določene napake.

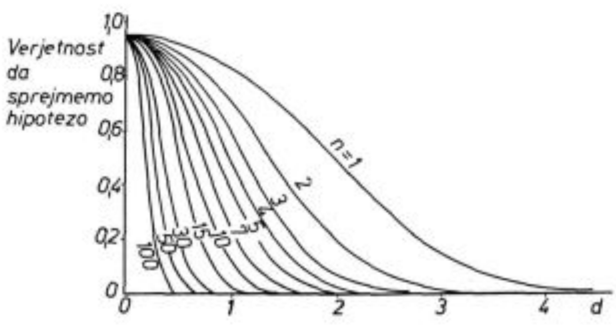
To nam pa ne sme vzeti zaupanja v tak način ocenjevanja hipotez. Vsak raziskovalec namreč ve, kolikšna razlika je pri določeni analizi še pomembna.

Predno se sploh lotimo meritev in zbiranja podatkov, se odločimo za napako prve vrste. Nato v naprej določimo, kolikšno razliko smo še pripravljeni z neko določeno verjetnostjo (β) tolerirati pri preverjanju hipoteze. Nato pa iz familije operacijskih krivulj določimo število potrebnih meritev v vzorcu. Na raziskovalnem oddelku Železarne Jesenice navadno izbiramo za napako prve vrste vrednost $\alpha = 0{,}05$ ali pa $\alpha = 0{,}01$.

Na sliki 6 je narisana familija operacijskih karakteristik za različne velikosti vzorcev ($\alpha = 0{,}05$). Na abscisi je brezdimenzijska spremenljivka $d = \frac{|\mu - \mu_o|}{\sigma}$. Kako si lahko pomagamo s to sliko, si oglejmo na posebnem primeru.

PRIMER ZA UPORABO OPERACIJSKIH KARAKTERISTIK

Radi bi preverili, če se je zaradi spremenjenega postopka pri valjanju spremenila povprečna trdnost žice od prejšnjih 45 kpmm^{-2} (μ_o). Znano je, da znaša standardna deviacija 2,5 kpmm^{-2}. Pred-



Slika 6

postavljamo, da se ta ni spremenila zaradi spremenjenega načina valjanja.

Želimo torej ugotoviti ali se trdnost ni spremenila ali se je spremenila. Do tega zaključka želimo priti s 95 % gotovostjo ($\alpha = 0{,}05$). Obenem smo se pa odločili, da sme biti verjetnost za napako druge vrste pri razliki 2 kp/mm^2 le 10 % ($\beta = 0{,}10$).

Najprej izračunamo d:

$$d = \frac{|\mu - \mu_o|}{\sigma} = \frac{2 \ \text{kpmm}^{-2}}{2{,}5 \ \text{kpmm}^{-2}} = 0{,}8$$

Nato poiščemo točko s koordinatama (0,8; 0,1) v diagramu na sliki 6. Ta točka leži v bližini krivulje, ki pripada vrednosti n = 15. Odločili smo se za 16 meritev v vzorcu in dobili naslednji rezultat:

$n = 16$

$\overline{X}_{16} = 46$ kpm/mm^2

Ker so srednje vrednosti vzorca z n podatki normalno porazdeljene okrog μ_o s standardno deviacijo $\frac{\sigma}{\sqrt{n}}$, bomo hipotezo $\mu = \mu_o$ sprejeli, če bo srednja vrednost vzorca $\overline{X}_n$ ležala znotraj intervala $(-K_{0,025}, +K_{0,025})$.

Za standardno normalno distribucijo ($\mu_o = 0$, $\sigma = 1$) pa velja, da je $K_{0,025} = 1{,}96$. Naša porazdelitev srednjih vrednosti vzorcev je pa tudi normalna s srednjo vrednostjo μ_o in standardno deviacijo $\frac{\sigma}{\sqrt{n}}$. Zato izračunamo parameter K, za katerega velja standardna normalna distribucija:

$$K = \frac{(X_n - \mu_o)\ \sqrt{n}}{\sigma} = \frac{1 \ \text{kpmm}^{-2} \sqrt{16}}{2{,}5 \ \text{kpmm}^{-2}} = 1{,}6$$

Ker je $-K_{0'025} < K < +K_{0'025}$, bomo sprejeli hipotezo, da je $\mu = \mu_o$ oziroma bomo trdili, da se pod navedenimi pogoji srednja trdnost žice ni spremenila.

Do prav takega zaključka bi tudi prišli, ako bi na primer namesto 16 napravili le 5 meritev in ako bi dobili enak rezultat za $\overline{X}_5$. V takem primeru bi bil K enak $\frac{1 \cdot \sqrt{5}}{2{,}5}$, kar je še manj kot 1,6.

Verjetnost za napako druge vrste bi bila pa večja. S slike 6 se vidi, da bi v takem primeru ($d = 0{,}8$) bila verjetnost, da smo sprejeli napačno hipotezo (ko bi se resnični μ razlikoval od μ_o za 2 kp/mm²) približno 50 %. 10 % verjetnost za napako drugega reda bi bila šele pri $d \cong 1{,}5$, kar bi pomenilo, da je $\mu - \mu_o = 1{,}5 \,.\, 2{,}5\ \text{kp/mm}^2 = 3{,}75\ \text{kp/mm}^2$.

Na podlagi petih vrednosti bi sicer tudi sprejeli hipotezo, napaka druge vrste pa bi bila prevelika, tako da se na tako majhen vzorec ne bi mogli zanesti. V tem primeru bi namreč pri sprejeti hipotezi, da je $\mu = \mu_o$ dopuščali tudi možnost, da se resnični μ razlikuje od μ_o za več kot 2 kp/mm², kar pa za nas ne bi bilo več nepomembno.

V mislih predpostavimo še tretjo možnost: da bi namesto 16 meritev naredili 100 meritev in da bi bila srednja vrednost vzorca spet enaka kot prej $\overline{X}_{100} = 46\ \text{kpm/mm}^2$.

V takem primeru bi bila vrednost

$$K = \frac{1\ \text{kpmm}^{-2}\ \sqrt{100}}{2{,}5\ \text{kpmm}^{-2}} = 4$$

kar bi padlo daleč izven intervala, v katerem bi bili pripravljeni sprejeti hipotezo $\mu = \mu_o$. Hipotezo bi torej zavrgli: rekli bi, da se srednja vrednost trdnosti nove žice razlikuje od predpisane vrednosti μ_o. Ako bi torej pri vzorcu 100 vrednosti dobili takšen rezultat ($\overline{X}_{100} = 46\ \text{kp/mm}^2$), bi bili sigurni, da gre za razliko med μ in μ_o. Z diagrama na sliki 6 pa je razvidno, da bi bila v takem primeru 10 % verjetnost za napako druge vrste pri $d \cong 0{,}3$ oziroma pri razliki $(\mu - \mu_o) = 0{,}75\ \text{kp/mm}^2$.

Vprašanje pa je, ali bi bilo zares potrebno narediti 100 meritev, ko pa že vnaprej vemo, da za nas razlika v trdnosti do 2 kp/mm² ni tako pomembna.

V tem je tudi prednost uporabe operacijskih karakteristik. Čemu bi namreč delali veliko število meritev in na primer ugotavljali z veliko verjetnostjo, da gre za razlike v srednjih vrednostih populacij, ko pa te razlike niso pomembne. S tem lahko prihranimo dosti časa in denarja.

Omenjeni primer za uporabo operacijskih karakteristik je le eden od mnogih načinov za testiranje raznih statističnih hipotez.

Pri vseh si pomagamo s podobnimi operacijskimi karakteristikami, ki pa so seveda za različne statistične parametre različne, saj gre navadno pri tem tudi za druge vrste distribucij.

Na raziskovalnem oddelku Železarne Jesenice preverjamo na primer tudi hipoteze $\mu = \mu_o$, ko standardna deviacija ni znana in jo šele ocenimo iz standardnega odklona vzorca. Preverjamo tudi hipoteze o srednjih vrednostih in standardnih deviacijah dveh populacij, ko imamo na razpolago le vzorca in v naprej ne poznamo nobenega parametra ene od populacij. Zelo pogosto preverjamo tudi hipoteze tipa $\mu_1 \leqq \mu_2$ (enostransko preverjanje). Pri tem je postopek praktično povsem enak, le da se operacijske karakteristike nekoliko razlikujejo od tistih, pri katerih testiramo trditve tipa $\mu = \mu_o$ (dvostransko preverjanje).

V splošnem je z uporabo operacijskih karakteristik dana možnost hitrejšega in učinkovitejšega dela na vseh področjih merjenja. Obenem pa posredno navaja raziskovalca na to, da si vedno skuša odgovoriti na vprašanje, kaj lahko z dobljenimi rezultati ugotovi in česa ne more. Tako se mu veča tudi zaupanje v pravilnost trditev in odločitev, ki bazirajo na statističnih analizah.

ZAKLJUČEK

Pri preskušanju hipotez o karakterističnih parametrih populacije, si pogosto zastavljamo vprašanje, koliko meritev je treba narediti, da bo naša odločitev pravilna, oziroma da verjetnost za napačno odločitev ne bo presegala neke določene vrednosti.

Pri tem se je treba pred samo meritvijo odločiti, kolikšna naj bo verjetnost za napako prve vrste (α) in kolikšna sme biti še dopustna verjetnost za napako druge vrste (β). Nato pa s pomočjo familije operacijskih karakteristik določimo potrebno velikost vzorca.

Opisan je primer preskušanja hipoteze $\mu = \mu_o$ pri pogoju, da je $\sigma = \sigma_o$. Za napako prve vrste smo izbrali 5 %, obenem pa smo si izbrali tudi napako druge vrste. S pomočjo operacijskih karakteristik smo določili velikost vzorca (16) in določili rezultat, ki je potrdil hipotezo. Predpostavili smo tudi možnost, da bi enak rezultat dobili pri manjšem (5) in večjem vzorcu (100 meritev). V prvem primeru bi sprejeli isto hipotezo $\mu = \mu_o$, v drugem primeru pa bi lahko hipotezo zavrgli, če bi se držali opisanih pravil.

Literatura

1. A. H. Bowker, G. J. Lieberman: Engineering Statistics Prentice Hall, Inc. (1959)
2. B. Ostle: Statistics in Research, Iowa State University Press (1969)

ZUSAMMENFASSUNG

Bei der Überprüfung der Hypothesen über die charakteristischen Parametern einer Population wird oft die Frage gestellt, wieviele Messungen nötig sind, um einen richtigen Entschluss zu fassen bzw., dass die Wahrscheinlichkeit für einen falschen Entschluss einen bestimmten Wert nicht überschreiten wird.

Dabei muss vor der eigentlichen Prüfung die Wahrscheinlichkeit für den Fehler ersten Ranges (α) und die Wahrscheinlichkeit für den Fehler zweiten Ranges (β) festgestellt werden. Danach wird mit Hilfe einer Familie der Operationscharakteristiken die nötige Probenzahl bestimmt.

Ein Beispiel der Hypothesen-überprüfung $\mu = \mu_0$ bei der Bedingung $\sigma = \sigma_0$ ist in diesem Artikel beschrieben. Für den Fehler zweiten Ranges sind 5 % gewählt worden. Auch der Fehler zweiten Ranges ist bestimmt worden. Mit Hilfe der Operationscharakteristiken ist die Probenzahl (16) und das Ergebniss, welches die Hypothese bestätigt, bestimmt worden.

Bei der Voraussetzung der Möglichkeit dasselbe Ergebniss auch bei einer kleineren (5) und einer grösseren (100) Probenzahl zu bekommen, hat sich herausgestellt, dass im ersten Fall die Hypothese $\mu = \mu_0$ angenommen werden kann, im zweiten Fall müssten wir jedoch die Hypothese bei Einhaltung der beschriebenen Regeln verwerfen.

SUMMARY

By testing of hypotheses about characteristic parameters of some population we often wish to know the sample size. This enables us to make the correct decision, so that the probability of the wrong decision does not exceed some defined value.

Before performing the experiment one must choose the probability of making the Type I. error (α) and the probability of making the error of Type II. (β). Then we find the necessary sample size by means of the family of OC curves — (Operating characteristic curves).

An example of testing the hypothesis $\mu = \mu_0$ when $\sigma = \sigma_0$ is described in the article. The Type I error was fixed 5 % and the error of Type II. was chosen too. By means of OC curves we found the necessary sample size (16). Then we made the measurements and we got such a result that we could accept the hypothesis.

We also supposed that we might get the same result in some smaller (5) and in a bigger sample (100). In the first case we should accept the same hypothesis but in the second one we should reject it using the criteria for rejection.

ЗАКЛЮЧЕНИЕ

При исследовании гипотез характеристичных параметров часто наступает вопрос, сколько измерений необходимо выполнить для того, чтобы решение было правильным т.е. чтобы вероятность ошибочных решений не превышало определённой предельной величины.

Прежде чем начать измерение надо определить вероятность ошибки первой группы (α). После этога также надо определить допустимую вероятность ошибки второй группы (β). Наконец, при помощи семейства характеристик, определяем необходимую величину образца.

В статье описан пример исследования гипотезы $\mu = \mu_0$ при условии когда $\sigma = \sigma_0$. Для ошибки первой группы выбрали 5 %. Определили также ошибку второй группы. Наконец, на основании операционных данных определили количество образцов (16) и получили результат который доказал что гипотеза правильная. Предположили также возможность одинаковога результата при небольшом количестве образцов (5) и при измерении количества 100 образцов. В первом случае могли бы воспринять гипотезу $\mu = \mu_0$; во втором случае опровергнуть, в случае если бы придерживались описанных правил,