

NÄKÖKOHTIA TILASTOTIETEELLISTEN TUTKIMUSMENETELMIEN KÄYTÖSTÄ MAATALOUSEKONOMIASSA

LAURI KETTUNEN

Maatalouden taloudellinen tutkimuslaitos

Saapunut 12. 4. 1966

Tilastotieteen käytön yleistymisen maataloudellisissa tutkimuksissa on tuonut esille eräitä vaikeuksia, jotka ovat ennen muuta tilastotieteellisten tutkimusmenetelmien luonnetta koskevia. Itse menetelmien teknillinen soveltaminen — siis laskelmien suorittaminen — on useimmiten suhteellisen helppoa tai se voidaan jättää tietokonekeskusten tehtäväksi. Sen sijaan tilastollisia menetelmiä voidaan soveltaa väärin sen tähden, että kiinnitetään liian vähän huomiota niiden käyttöön liittyviin olettamuksiin. Tästä voi olla seurauksena, että hankittu perusaineisto ei täytä tilastolliseen tutkimukseen kelpaavan näytteen vaatimuksia ja/tai että saatuja tuloksia ei tarkastella tarpeeksi kriittisesti pitäen mielessä ne rajoitukset, jotka liittyvät käytettyyn aineistoon ja menetelmään. Tilastollisia tutkimusmenetelmiä voidaan myös soveltaa epämielekkäällä tavalla esimerkiksi triviaalisten seikkojen tutkimiseen tai päätyä niitä sovellettaessa epäloogisiin tuloksiin, joiden syntymistä tilastolliset menetelmät eivät läheskään aina paljasta.

Tässä yhteydessä ei tietenkään voida ryhtyä tarkastelemaan kaikkia tilastotieteellisiä tutkimusmenetelmiä. Seuraavassa onkin rajoitettu vain esittämään muutamia näkökohtia tutkimusaineiston edustavuudesta lähinnä kirjanpitoiltoilta saatavaa tilastoaineistoa ajatellen, syy-yhteyksien analysoinnista korrelaatio- ja regressioanalyysin avulla sekä hypoteesien testaamisesta. Tarkastelun kohteet edustavat kuitenkin melko suurta osaa niistä tilastotieteellisistä menetelmistä, joita joudutaan käyttämään maatalouden tutkimuksissa.

Kirjanpitoaineiston edustavuudesta

Tilastotieteen tehtävänä on luoda menetelmiä, joiden avulla saadaan tietoja (tunnuslukuja) jostakin kokonaisaineistosta, joka on liian suuri tutkittavaksi kokonaan. Populaatio voi olla käytännöllisesti katsoen äärettömän suuri, joten jokai-

sen yksikön havaitseminen on mahdotonta. Mutta vaikka jokainen yksikkö olisikin havaittavissa (mitattavissa), koko populaation tutkimisesta on usein luovuttava liian suurten kustannusten takia. Otantamenetelmät tarjoavat tällöin tieteellisesti hyväksyttävän keinon tutkimuksen suorittamiseksi. Niiden avulla ei kuitenkaan saada ehdottoman oikeita tuloksia, sillä sattumasta johtuen parametrien tai tunnuslukujen estimaatit useimmiten poikkeavat todellisista arvoista. Otantateorian tehtävänä onkin tutkia, minkälainen ja miten suuri on kyseinen poikkeama, toisin sanoen on löydettävä estimaatin jakautumafunktio. Kun se tunnetaan, voidaan laskea ns. luotettavuusvälit, joille estimaateista osuu annettu prosenttimäärä. Esimerkiksi 95 %:n luotettavuusväli todellisen keskiarvon kahden puolen tarkoittaa sitä, että 95 % kaikista otannalla saaduista keskiarvon estimaateista osuu kyseiselle välille (vert. esim. 3, p. 110 – 111 tai 5, p. 209 – 220)¹.

Tutkimalla koko populaatio ei välttämättä saada täysin oikeita tuloksia, sillä aineiston keruuseen koulutetun työvoiman puutteen, ylimalkaisemman tutkimuksen yms. seikkojen takia havaintovirheet tulevat sitä suuremmiksi mitä laajempi on tutkimuskohde. Näissä ns. peruslaskentatuloksissa on siis myös virheitä. Voi olla jopa niin, että otannalla saadaan tarkempia arvioita populaation tunnusluvuista (kuten esimerkiksi keskiarvoista) kuin tutkimalla kaikki perusjoukon yksilöt, sillä koska otantamenetelmiä käytettäessä tyydytään suhteellisen pieneen aineistoon, havaintovirheet pienenevät tarkemman tutkimuksen ja kontrollin avulla.

Ne analyysimenetelmät, joita tilastotiede on kehittänyt, perustuvat oletukselle, että tutkittava aineisto on ns. satunnaisnäyte, jonka saamiseen on käytettävissä useita eri menetelmiä (vrt. 5, p. 657). Tutkijan on siis varmistauduttava siitä, että kyseinen vaatimus on täytetty, sillä jollakin muulla tavalla saadun aineiston jakautumafunktiosta on vaikea tehdä mitään olettamusta, minkä takia todennäköisyyskriteerien mukaisten johtopäätösten teko vaikeutuu, ellei tule aivan mahdottomaksi.

Kirjanpitotilat muodostavat yksityiskohtaisimman maataloja koskevan tilastoaineiston lähteen. On siis luonnollista, että useita tutkimuksia on tehty ja tehdään edelleenkin siltä pohjalta. Kirjanpitoaineiston heikkoutena on kuitenkin, että se ei ole satunnaisnäyte maamme tiloista, vaan pikemminkin se on erikoisryhmä, joka erottuu muista maataloista mm. siten, että keskimääräinen kokonaistuotto on näillä tiloilla ollut korkeampi kuin kaikilla maamme tiloilla. Sen tähden kirjanpitotiloilta saatuja tuloksia ei voida yleistää kaikkia tiloja käsittäviksi lukuunottamatta muutamia yksityiskohtia, joita esitetään jäljempänä. Kirjanpitoaineiston käyttö rajoittaa siis tutkimustulokset vain kirjanpitotiloja koskeviksi. Tutkittava populaatio on lisäksi äärellinen ja niin pieni, että se on useinkin mahdollista tutkia kokonaan. Kun usein tutkimus rajoitetaan vielä jonkin alueen kirjanpitotiloihin, populaatio on sitäkin pienempi. Jos kuitenkin perusaineisto voidaan tutkia kokonaan, eikä havaintovirheitä oleteta syntyvän, ei tilastotieteellisiä menetelmiä tarvita, koska mitään sattumasta johtuvaa virhettä ei lasketuissa suureissa ole. Se on kuitenkin tutkimukselle vain eduksi, sillä tilastolliseen analyysiin turvautuminen vaikeuttaa

¹Artikkelissa on tyydytty antamaan vain muutamia tunnetuimpia lähdeoteoksia, sillä esitetyt teoreettiset kohdat löytyvät yleensä kaikista tilastotieteen oppikirjoista.

ehdottomien johtopäätösten tekoa. Tosin kirjanpitoaineistosta voidaan ottaa otoksia, kun koko aineistoa ei katsota voitavan tutkia, mutta tällöin on kuitenkin otettava huomioon korjausterminä otoksen ja populaation suuruuksien suhde (vrt. esim. 5, p. 146).

Vaikka kirjanpitotilat poikkeavat taloudelliselta tulokseltaan kaikista tiloista, ei se välttämättä merkitse sitä, että ero olisi havaittavissa myös kaikkien tuotannon tekijäin kohdalla. Kyseisten tilojen keskimääräinen taloudellinen paremmuus johdukin ilmeisesti inhimillisistä tekijöistä kuten ammattitaidosta ja yleisestä aktiivisuudesta, jota juuri liittyminen kirjanpito toimintaan ilmentää ja ehkä myös vahvistaa. Eräiden tekijäin kuten maan suhteen kirjanpitotilat lienevät verrattavissa koko maan tiloihin. Mikäli näin todella olisi, tarjoaisi se mahdollisuuden, eräin varauksin, tehdä kirjanpitoaineiston pohjalta yleistettäviä johtopäätöksiä. Esimerkkinä siitä, kuinka tällaisia yleistyksiä voidaan tehdä, on seuraavassa tarkasteltu tuotantofunktioiden estimointia kirjanpitoaineiston pohjalta. Korostettakoon kuitenkin, että tarkastelua ei ole tarkoitettu ehdottomaksi todistukseksi, vaan ainoastaan suuntaa antavaksi, sillä käytännössä on tietenkin todennettava seuraavassa esitetty olettamukset.

Funktioiden soveltamista ajatellen on tärkeä, että parametreja estimoidessa muuttujien vaihtelualue on laaja, sillä yleensä estimoituja funktioita ei tulisi soveltaa muuttujien vaihtelualan ulkopuolella (1, p. 322). Toiseksi muuttujien mittauksen tulee olla yksikäsitteinen: esimerkiksi kahdella eri tilalla tehdyt havainnot samanmittaisia (lannoitteet samaa laatua, samoin väkirehut, maan luokitus sama jne.). Mikäli nämä kaksi olettamusta, muuttujien yksikäsitteisyys ja vaihtelualan laajuus, ovat voimassa myös kirjanpitoaineistossa, antavat siitä estimoidut tuotantofunktiot melko hyvän yleiskuvan vastaavasta koko maan tuotantofunktiosta.

Tuotantofunktioiden tärkeimpinä ja ehkä ainoina tekijöinä voitaneen pitää maan laatua ja lannoitusta sekä kotieläinten laatua ja ruokintaa. Kasvinviljelytuotteen vaikuttavan maaperän laadun vaihtelujen voidaan olettaa olevan samaa suuruusluokkaa kirjanpitotiloilla kuin koko maassa. Jos siis maaperä voidaan luokitella tuotantofunktioita varten, kuten toivottavaa on, on kyseinen tuotantofunktio tältä osin yleistettävissä koko maahan. Tosin tällöinkin on tutkijan selvitettävä, onko sovellettava havaintoarvo estimointiin käytettyjen havaintoarvojen vaihtelualan sisäpuolella. Lannoitteiden osalta yleistettävyyden on selvä, sillä kirjanpitoiloilla käytetään samoja lannoitteita kuin muillakin tiloilla ja lannoituksen määrässä on huomattavia vaihteluita.

Kotieläintuotantofunktion osalta tarkastelu on edellisen kaltainen. Eläinten laatu olisi otettava erääksi tekijäksi tuotantofunktioon. Toinen rajatuotokseen vaikuttava tekijä on ruokinta. Jos vain molemmissa tekijöissä on tarpeeksi vaihtelua kirjanpitotiloilla, ovat niiden aineistosta lasketut tuotantofunktiot yleistettävissä koko maassa.

Syy-yhteyksien analysointi korrelaatio- ja regressioanalyysin avulla

Kun aineiston keruussa otantamenetelmiä käytettäessä on otettava huomioon todennäköisyyslait, on korrelaatio- ja regressioanalyysia sovellettaessa ennen muuta

kiinnitettävä huomiota asetettujen hypoteesien ja mallien loogisuuteen. Tutkijan ratkaistavaksi jää kokonaan riippuvuussuhteen suunta, mikäli sen voidaan olettaa olevan yksikäsitteinen sekä primääristen tekijöiden valinta esimerkiksi regressioanalyysillä estimoitaviin funktioihin. Varsinaisten kysymykseen tulevien estimaattien, korrelaatio- ja regressiokertoimien sekä niihin liittyvien standardipoikkeamien laskeminen on sen jälkeen vain mekaaninen tehtävä. Parametrejä tulkittaessa ja jo muuttujia valittaessa on tietenkin otettava huomioon *ceteris paribus* -olettamuksen asettamat rajoitukset (vrt. I, p. 176).

Riippuvuussuhteita tutkittaessa käytetään usein yksinkertaista Pearsonin korrelaatiokerrointa r . Sitä sovellettaessa empiiriseen aineistoon on kuitenkin muistettava *ceteris paribus* -olettamus. Se merkitsee sitä, että korrelaatiokerroin ilmaisee tekijöiden (muuttujien) välisen riippuvuuden asteen olettaen, että muut tekijät pysyvät vakioina. Tällainen tilanne on todellisuudessa varmaan hyvin harvinainen, sillä normaalisti tarkasteltavat ilmiöt ovat useiden tekijöiden tuloksia. Vaikka jokin niistä olisikin selvästi voimakkain, voivat muut tekijät silti vaikeuttaa sen vaikutuksen selvittämistä tarkasteltavassa ilmiössä. Ainoastaan silloin, kun muiden tekijöiden vaikutus voidaan eliminoida esimerkiksi järjestettyjen kokeiden avulla, ovat yksinkertaisen korrelaatiokertoimen laskemiseen vaadittavat olettamukset voimassa. Tällaiseen on esimerkiksi taloudellisten ilmiöiden osalta mahdollisuus hyvin harvoin jos ollenkaan.

Varsinkin teknillistä transformaatiota mutta myös luonnonilmiöitä kuvaavia funktioita estimoitaessa saadaan helposti korkeita korrelaatiokertoimia. On kuitenkin harkittava, onko kyseinen tulos jo *a priori* odotettavissa. Jos näin on, ei tulokselle ole syytä antaa liikaa painoa. Voi olla jopa epämielekästä käyttää korrelaatiokerrointa tällöin muussa mielessä kuin osoittamaan estimoidun funktion selvityskykyä. Sen sijaan riippuvuussuhteen määrittäminen funktion muodossa voi olla tällöin hyvinkin mielenkiintoista.

Mainittakoon lisäksi tässä yhteydessä, vaikka jäljempänä tarkastellaan testien käyttöä, että korrelaatiokertoimen merkitsevyys riippuu näytteen koosta. Mitä pienempi se on, sitä suuremman tulee olla korrelaatiokertoimen, ennenkuin sen perusteella voidaan väittää tutkittavien ilmiöiden välillä olevan tilastollisessa mielessä merkitsevän riippuvuussuhteen. Esimerkiksi 1 %:n riskitasolla korrelaatiokertoimen tulee olla $= 0.71$, jos otoksen suuruus $n = 10$, tai $= 0.54$, kun $n = 20$ (5, p. 771) olettaen tietenkin lisäksi, että molemmat muuttujat ovat normaalisti jakautuneita.

Käytettäessä yksinkertaista yhden muuttujan regressioanalyysia estimoitavan funktion mielekkyys tulee merkitseväksi. Tilastotiede ei tässä suhteessa voi antaa mitään kriteerejä. Tutkijalle jää täysi vastuu siitä, että estimoitava funktio on looginen. Loogisuuteen ei kuulu ainoastaan vaikutussuunnan määrittäminen, toisin sanoen, onko y x :n vain x y :n funktio, vaan myös se että kyseisessä funktiossa on ilmiöiden välinen syy-yhteys suora. Voihan olla, että kun y on x :n funktio, x korreloi z :n kanssa niin voimakkaasti, että y näyttää olevan mahdollista selvittää myös z :n avulla, vaikka niiden välillä ei olekaan mitään loogista riippuvuussuhdetta. Tekijöitä koskevat **m i t t a u s t u l o k s e t** ovatkin usein voimakkaasti keskenään korreloituneita, mutta tekijöiden itsensä ei välttämättä tarvitse olla **t o i**

sistaan riippuvia. Aikasarjat ovat hyviä esimerkkejä tästä. Melkein minkätähansa kahden taloudellista toimintaa kuvaavan aikasarjan välillä on voimakas korrelaatio, mutta se ei merkitse sitä, että kyseiset ilmiöt olisivat syy-yhteydessä toisiinsa. Samoin maataloudesta löytyy ilmiöitä, joiden välillä on näennäisen korkea korrelaatio, mutta ei mitään todellista suoraa vaikutusyhteyttä. Kyseisten ilmiöiden esittäminen funktion muodossa on ymmärrettävästi epämielekästä. Tämä ei tietenkään ole tilastotieteellinen vaan sen hypoteesin asettamiseen liittyvä ongelma, joka halutaan tilastotieteellisin menetelmin osoittaa yhtäpitäväksi tilaston kanssa.

Myös useamman muuttujan funktiota käytettäessä on tärkeää, että käytetyt selittävät muuttujat ovat primäärisiä tekijöitä eikä jonkin muun tekijän kautta vaikuttavia tai sen kanssa korreloituneita. Yleensäkin on vältettävä liittämästä funktioon vain korkean korrelaation toivossa muuttujia, joiden vaikutus on epäselvä tai kokonaan perustelematon. Maatalouden kokonaistuotantofunktioihin on sisällytetty mm. rakennuspääoma, joka on tietenkin välttämätön tiettyyn minimiin saakka tuotantosuunnasta riippuen, mutta sen ylittäessä kyseisen minimin se on tuskin enää tuotantoa lisäävä.

Usean muuttujan funktion avulla voidaan periaatteessa tutkia kunkin tekijän vaikutus erikseen. Regressiokertoimethan ovat nytkin tulkittavissa *ceris paribus* eli muiden tekijöiden pysyessä vakiona: kukin regressiokerroin ilmaisee kyseisen muuttujan rajavaikutussuhteen. Multikollineaarisuuden eli selittävien muuttujien välisen korrelaation vuoksi voi kuitenkin olla vaikeaa saada esille yksityisen muuttujan regressiokerrointa vaan ainoastaan kyseisten korreloituneiden muuttujien yhteisvaikutus. Tämä ongelma on yleinen käytettäessä aikasarjoja. Mikäli mallia sovelletaan vain ennustamiseen (siis tekijöiden yhteisvaikutuksen määrittämiseen) ei siitä yleensä ole haittaa (JOHNSTON 1963, s. 207), mutta yksityisiä parametrien estimaatteja on tulkittava varoen. Yleensäkin on otettava huomioon, että estimoidut regressiokertoimet ovat vain keskimääräisiä, sillä niiden ei todellisuudessa tarvitse olla vakioita kuten tavallisesti oletetaan, vaan esim. muuttujista riippuvaisia. Lineaarinen funktiohan antaa vakiorajasuureita (rajatuottoja, -kustannuksia, jne.), logaritmfunktio sisältää olettamuksen vakiojoustoista, jne. Malleja on kuitenkin yleensä pakko yksinkertaistaa niiden tulkittavuuden parantamiseksi sekä muuttujien että myös parametrien suhteen.

Hypoteesien testaamisesta

Edellä on jo viitattu otannalla saatujen estimaattien tarkasteluun ja vertailuun eli ns. hypoteesien testaamiseen. Tilastotieteellisen analyysin lopputulos on tunnusluku, jonka laskemisella analyysi alkaa ja josta tehtäviin johtopäätöksiin se päättyy. Johtopäätösten tekoa vaikeuttaa kuitenkin se, että jonkin tunnusluvun otannalla saatu estimaatti tavallisesti poikkeaa sattuman takia todellisesta arvosta. Tilastotieteen kehittämin menetelmin voidaan kuitenkin estimatille laskea ns. luotettavuusväli, joka annettulla todennäköisyydellä (95 %, 99 % jne) sisältää todellisen arvon. Näitä luotettavuusvälejä apuna käyttäen voi-

daan esimerkiksi kahden populaation tunnuslukuja verrata toisiinsa otoksen estimaattien avulla. Jos kyseiset estimaatit ovat yhtä suuret, voidaan heti tehdä johtopäätös, että myös populaatioiden tunnusluvut ovat yhtä suuret. Mikäli tunnuslukujen estimaatit ovat eri suuret, on tutkittava, mikä on se todennäköisyys, että vaikka todelliset tunnusluvut ovat yhtäsuuret, estimaatit poikkeavat toisistaan niiden erotuksen suuruisella määrällä. Todennäköisyyden ollessa suurempi kuin hyväksymämme (subjektiivinen) raja, esim. 0.95 tai 0.99, me hylkäämme asettamamme ns. nollahypoteesin, että todelliset populaatioiden tunnusluvut ovat todella yhtä suuret. Päinvastaisessa tapauksessa me hyväksymme sen. Riski, että hylkäämme nollahypoteesin eli että tunnusluvut ovat todella yhtä suuret, on tällöin 5 % (tai vastaavasti 1 %). Tämä on ns. I asteen virhe (vrt. esim. 4, p. 391). Voi kuitenkin sattua, että vaikka todelliset arvot eroavatkin toisistaan, estimaatit eivät teekään niin, ja me teemme johtopäätöksen, että tunnusluvut ovat yhtä suuret. Tällöin hyväksymme väärän hypoteesin eli teemme ns. II asteen virheen, jota kuitenkin ei normaalisti tuoda testin yhteydessä esille. Toistettakoon vielä, että tavallisesti otetaan tietty riski, että oikea hypoteesi hylätään. Otetaan esim. 5 % riski, että vaikka 2 eri populaatioista saatua estimaattia poikkeavatkin enemmän kuin asettamamme raja ($= 1.96 \times$ erotuksen standardipoikkeama), todelliset arvot ovatkin yhtä suuria.

Huomattakoon kuitenkin, että tilastollinen testi ei paljasta mitään todellisten tunnuslukujen absoluuttisesta erosta. Otannalla saatujen estimaattien luotettavuusvälihän pienenee otoksen kasvaessa. Näin ollen otosten koon kasvaessa yhä pienemmät todelliset absoluuttiset erot tulevat tilastollisesti merkitseviksi. Sitä vastoin otoksen koon ollessa pieni suuriakaan todellisia absoluuttisia eroja ei ehkä saada tilastollisesti merkitseviksi. Toisin sanoen vaikka estimaattien ero olisikin suhteellisen suuri, niiden standardipoikkeamat ovat otoksen koon pienuuden takia suuria, joten tilastollista merkitsevyyttä ei saada esille. Mikäli tilastollinen eron merkitsevyys on todettu ja otosten tunnuslukuja käytetään vastaavien populaation tunnuslukujen arvioina, tutkijalle jää vielä hänen omien subjektiivisten kriteeriensä avulla päätettäväksi, miten merkittävä kyseinen ero on taloudellisilta yms. seuraamuksiltaan.

Tilastotieteellisiä testejä ei tarvita silloin, kun tunnusluvut on saatu tutkimalla koko populaatio. Tällöinhän kyseisiin tunnuslukuihin ei liity sitä satunnaista virhettä, joka otanta käytettäessä niihin sattuman takia tulee. Tutkimuksen kohteen ollessa esimerkiksi koko maan tai jonkin alueen kirjanpitoilat, tunnusluvut (useimmiten keskiarvot) saadaan lasketuksi koko aineistosta, ja silloin niissä periaatteessa ei ole satunnaisvirheitä. Kahden tunnusluvun ero on tällöin tilastollisessa mielessä 0 %:n riskillä merkitsevä, olipa se kuinka pieni tahansa. Käytännöllisiltä seuraamuksiltaan ero sen sijaan voi olla merkityksetön.

Tiivistelmä

Edellä olevassa korostetaan tilastollisiin tutkimusmenetelmiin liittyviä todennäköisyysteoreettisia olettamuksia. Tältä kannalta tarkastellaan kirjanpitoaineiston

soveltuvuutta tutkimuskohteeksi ja todetaan, että siitä saatavia tuloksia ei tavallisesti voida yleistää koko maata koskeviksi, sillä kyseinen aineisto ei ole edustava näyte koko maan tiloista. Eräänä poikkeuksena esitetään kuitenkin tietyin varauksin mahdollisuus estimoida kirjanpitoaineistosta yleistettäviä kokonaistuotantofunktioita. Toiseksi artikkelissa painotetaan korrelaatio- ja regressiokertoimien estimointiin liittyvää *ceteris paribus*-ehtoa ja asetettujen mallien loogisuutta: vaikutussuunnan selvittämistä ja primääristen tekijöiden valitsemista estimoitaviin funktioihin. Korrelaatiokertoimen liiallista korostamista kehoitetaan välttämään silloin, kun voimakas riippuvuussuhde on tiedossa jo *a priori*. Lopuksi esitetään muutamia näkökohtia hypoteesien testaamisesta ja kiinnitetään huomiota ennen muuta siihen, että vaikka tilastolliset testit osoittavatkin todennäköisyyslakien mukaisia eroja, kyseisten erojen käytännöllinen merkitys jää tutkijan subjektiivisen harkinnan varaan.

KIRJALLISUUTTA

- (1) EZEKIEL, MORDECAI & FOX, KARL A. 1959. *Methods of Correlation and Regression Analysis*. London, 548 p.
- (2) JOHNSTON, J. 1963. *Econometric methods*. New York, 300 p.
- (3) MATTILA, SAKARI. 1965. *Tilastotiede I*. Kauppakorkeakoulu, 150 p. Helsinki.
- (4) — 1965. *Tilastotiede II*. *Ibid.*, 174 p.
- (5) MILLS, FREDERICK C. 1955. *Statistical Methods*. London, 842 p.

S U M M A R Y:

VIEWS ON THE USE OF STATISTICAL METHODS IN AGRICULTURAL ECONOMICS

LAURI KETTUNEN

Agricultural Economics Research Institute, Helsinki

The article deals with the problems arising in the application of statistical research methods to agricultural economics. As the available statistics that are often used in research work do not meet all the requirements set for statistical inference the results of the research cannot usually be generalized. The article calls attention to the logical content of function models in the use of regression analysis where the first difficulty lies in determining the causal ordering of the variables and in selecting the primary variables of the function to be estimated. Secondly, the interpretation of the results is difficult either because the *ceteris paribus* clause is not valid or because, though it may be valid theoretically, the correlation between the observed values of independent variables makes the estimates of parameters unreliable. Finally the writer of the article presents some views on the testing of hypotheses and calls attention to the fact that although statistical tests show differences that are significant, the importance of the differences, for instance in regard to their economic and other consequences, depends on the subjective judgement of the scholar.