

Forskartjänsternas regler och anvisningar

Innehåll

| | |
|---|----|
| 1. Inledning..... | 1 |
| 2. Sekretessskyldighet | 1 |
| 3. Regler för distansanvändningssystemet..... | 2 |
| 3.1 Att logga in i systemet..... | 2 |
| 3.2 Att använda systemet | 2 |
| 3.3 Resurser | 3 |
| 3.4 Underhåll..... | 3 |
| 4. Forskningslaboratoriets regler | 3 |
| 4.1 Användning av personkort och passernyckel | 3 |
| 4.2. Att arbeta i forskningslaboratoriet | 3 |
| 5. Skydd av uppgifter och kontroll av utskrifter | 4 |
| 5.1 Krav på datasekretess för utskrifter | 4 |
| 5.2 Kontroll av undersökningsresultaten i praktiken | 6 |
| 6. Publicering av resultat | 7 |
| 7. När projektet är slutfört | 7 |
| 8. Sanktioner | 8 |
| 9. Ansvar/uppgifter | 8 |
| Bilagor | 10 |

Kontaktinformation

Forskartjänster: tutkijapalvelut@stat.fi

Mikrosimulering: mikrosimulointi@stat.fi

1. Inledning

Statistikcentralens forskartjänster tillhandahåller material på enhetsnivå för vetenskaplig forskning och statistiska utredningar. Materialet kan användas via Statistikcentralens distansanvändningstjänst (FIONA) och i forskningslaboratoriet eller utanför Statistikcentralen som anonymiserad urvalsmaterial. Statistikcentralen ställer som villkor för tillståndet att använda data på enhetsnivå, att användarna följer forskartjänstens regler och anvisningar. Regler och anvisningar förknippade med sekretessskyldigheten, skydd av uppgifter, publicering av resultat och slutförandet av projektet gäller all användning av material på enhetsnivå för forskningsändamål. När det gäller distansanvändning och användning av forskningsmaterial i forskningslaboratoriet följer man ett kontrollförfarande när det gäller undersökningsresultaten, se kapitel 5. Praxisen för forskartjänstens distansanvändning gäller också distansanvändningen av SISU-mikrosimuleringsmodellen. Ett undantag vid mikrosimuleringen är att man följer en annan praxis vid granskning av undersökningsresultaten.

2. Sekretessskyldighet

Användning av Statistikcentralens material på enhetsnivå kräver tillstånd. Uppgifterna får användas endast av en person som fått tillstånd och endast för det ändamål som godkänts i beslutet. Olovlig identifiering av enskilda enheter i materialet är förbjudet. Genom att underteckna ett sekretessavtal förbinder sig forskaren att inte

lämna uppgifter om eller för egen vinning använda sådan information som han eller hon fått kännedom om i samband med forskningen och som är sekretessbelagd (personuppgifter och företagsuppgifter på enhetsnivå som ingår i undersökningsmaterialet). Tystnadsplikten gäller också dataprogram som används inom statistikproduktionen samt uppgifter som gäller regelverk vilka äventyrar datasekretessen ifall de avslöjas (statistiklagen 280/2004: 24 §). Tystnadsplikten gäller också efter att tillståndet för det enskilda projektet gått ut.

3. Regler för distansanvändningssystemet

3.1 Att logga in i systemet

Fjärråtkomst från forskarens arbetsstation till Statistikcentralens server öppnas via en webbtjänst (Bilaga B). Användarna loggar in i webbtjänsten med personliga användaridentifikationer och lösenord som ska förvaras omsorgsfullt. Det är viktigt att användaridentifikationerna och lösenorden förvaras separat.

Användaren identifieras dessutom med en sms-kod. Systemet sänder ett s.k. flashmeddelande till användarens mobiltelefonnummer. Telefoner och SIM-kort som är kopplade till detta nummer ska förvaras omsorgsfullt.

Distansanvändningen ska ske från den kunds lokaliteter som har ingått ett avtal om öppnande av förbindelsen ifall det inte finns separat överenskommelse om annat. Fjärråtkomst får bara upprättas från de IP-adresser som kunden angett. Statistikcentralen kontrollerar att de följer användningsvillkoren.

Datorn, som används vid inloggningen till distansanvändningssystemet, skall ha uppdaterad datasäkerhet samt virus- och brandvägsskydd.

3.2 Att använda systemet

För forskningsprojekten reserveras en arbetskatalog där projektets material förvaras. Rätt att läsa materialkataloger ges enligt det tillstånd som beviljats för projektet eller SISU-modellen. I SISU-modellen reserveras en egen arbetskatalog för varje användare och användaren kan om han eller hon önskar ge rättigheterna till andra som använder modellen. Användare av mikrosimuleringsmodellen kan också via en gemensam Forum-katalog dela filer bland andra användare i fjärråtkomstmiljön.

Systemet ska bara användas för det syfte som angetts i tillståndet och dess funktion får inte på något sätt äventyras. Logguppgifter om användningen av systemet samlas in för fakturering samt underhåll och övervakning av systemet¹. Användningen av systemet faktureras enligt prislistan för forskartjänster.

Distansförbindelsen är avsedd bara för den forskare som kunden angett och som har tillstånd till datan. Därför ska maskinen som används för distansförbindelsen låsas om man avlägsnar sig och maskinen lämnas oövervakad. När man inte längre arbetar via distansförbindelsen ska man logga ut ur systemet.

Materialet ska hanteras så att den information som skyddas inte hamnar hos utomstående. Säkerheten kan förbättras bl.a. med hjälp av anpassning av arbetsutrymmen samt skydd av skärmen. Material som kräver användningstillstånd får inte på något sätt avslöjas för personer som inte har tillstånd att använda materialet.

¹ Registerbeskrivning TK-00-473-08

Program- och resultatfilerna går endast via Statistikcentralens personal per e-post på adressen tutkijapalvelut@stat.fi. Undersökningsresultaten granskas för att säkra datasekretessen (se kapitel 5). Utskrifter från mikrosimuleringen har en avvikande praxis.

3.3 Resurser

För varje forskningsprojekt reserveras som utgångsställning 300 GB diskutrymme. För dem som använder SISU-mikrosimuleringsmodellen har man reserverat 40 GB/användare. Om projektet eller användaren av SISU-modellen behöver mera diskutrymme skall forskarna göra en separat överenskommelse om detta. Resurserna för kalkylering och program som kan användas i systemet är begränsade och tillgången på dem beror på antalet andra inloggade användare. Det rekommenderas att man inte lagrar överlappande material och temporära filer. Samtidiga tunga körningar bör undvikas. För enskilda kunder eller projekt kan man erbjuda egna resurser för kalkylering och program enligt ett separat avtal.

3.4 Underhåll

Statistikcentralen ansvarar för underhållet av distansanvändningssystemet under tjänstetid. Vänligen meddela om problem per e-post till adressen tutkijapalvelut@stat.fi. Vid problemsituationer bör forskaren i förstahand kontakta kontaktpersonen som utgetts i avtalet om distansanvändning i den egna institutionen. Statistikcentralen svarar inte för programmens användarstöd.

Statistikcentralen har rätt att stänga systemet på grund av underhåll. Information om driftsavbrott förknippade med underhåll ges på ingångssidan för distansanvändningen samt via ett e-postmeddelande till kontaktpersonerna i organisationerna.

4. Forskningslaboratoriets regler

4.1 Användning av personkort och passernyckel

Forskare som arbetar i forskningslaboratoriet får ett personkort med foto som alltid ska bäras synligt då man befinner sig i Statistikcentralens utrymmen. Om kortet saknas ska man kontakta Statistikcentralens personal.

I forskningslaboratoriet kan man arbeta endast då Statistikcentralens bibliotek är öppen mellan kl. 9.00–16.00. Forskare ska boka plats per e-post från forskartjänsten för var dag forskaren önskar använda utrymmet (tutkijapalvelut@stat.fi).

4.2. Att arbeta i forskningslaboratoriet

Forskarna som arbetar i forskningslaboratoriet loggar in till FIONA-systemet från datorn i forskningslaboratoriet. De tilldelas en personliga användaridentifikation och lösenord som ska förvaras omsorgsfullt. För forskarna i forskningslaboratoriet gäller samma anvisningar som för forskare som loggar in från sina arbetsplatser utanför Statistikcentralen (se kapitel 3).

Arbetsstationen ska låsas om man tillfälligt avlägsnar sig och maskinen ska stängas av när arbetsdagen är slut. I undantagsfall kan man låta arbetsstationen göra körningar enligt ett separat avtal. För forskarplatserna debiteras enligt prislistan för forskartjänster.

Program- och resultatfilerna går endast via forskartjänstens personal per e-post på adressen tutkijapalvelut@stat.fi. Undersökningsresultaten granskas för att säkra datasekretessen (se kapitel 5).

5. Skydd av uppgifter och kontroll av utskrifter

Enligt sekretesskyldigheten ska forskaren se till att det inte finns några uppgifter på enhetsnivå i undersökningsresultaten eller några möjligheter att de avslöjas. Forskartjänsterna har tillgång till kontroll av undersökningsresultaten och med hjälp av kontrollen säkerställer man datasekretessen när det gäller utskrifter som produceras utgående från forskarens material. Forskaren ska se till att utskrifterna som skickas för kontroll uppfyller de krav om datasekretess som nämns i kapitel 5.1. Utskrifterna ska vara lättolkade. I tabellerna ska antalet cellspecifika observationer synas, vilket också gäller antalet observationer som används vid beräkning av estimat och nyckeltal. Vid behov ger Statistikcentralens personal mera information om skydd av uppgifter.

5.1 Krav på datasekretess för utskrifter

I det följande beskrivs kraven på datasekretess för olika typer av utskrifter.

Frekvens- och mängdtabeller

De utskrifter som forskaren publicerar ska för datasekretessens del följa Statistikcentralens riktlinjer för skydd av uppgifter i tabellformat (bilaga A). Huvudregeln vid skydd av uppgifter om företag är att varje cell eller grupp ska innehålla minst 3 (oviktade) observationer. När det gäller GVC/international sourcing-frågematerialet samt FLOWN-ägare-materialet är gränsen undantagsvis 5 observationer. Även dominansregeln² (1,75) skall tillämpas på färskt företagsuppgifter (mindre än 15 månader från referenstidpunkten). För material om produkters utrikeshandel tillämpas alltid dominansregeln. Vid skydd av uppgifter på arbetsställenivå ska man också säkerställa skyddet på företagsnivå, dvs. i varje cell ska det finnas arbetsställen inom minst tre olika företag. På samma sätt ska skyddet på koncernnivå beaktas i företagsmaterialet som innehåller uppgifter om koncernförhållanden. Vid produktinformation (information från statistiken över industriproduktion, råvaror och förnödenheter) är företagets antal konfidentiell i alla produktionsklasser.

Vid skydd av personuppgifter följs ett cellspecifikt tröskelvärde 3 och man fäster speciell vikt vid sensitiviteten hos variabeln som ska tabelleras. När det gäller sammanlaget anställd-arbetsgivare-material ska både person- och företagsnivån skyddas, dvs. i varje tabellcell ska det finnas anställda från minst tre olika företag. När det gäller uppgifter om yrkesutövare i företagsstatistik i tabellformat tillämpas samma säkerhetspraxis som i fråga om andra företagsuppgifter.

² Uppgifter om dominansregeln finns i bilaga A i kapitel 3.

Även andra riktlinjer för skydd av data kan tillkomma för specifika datamaterial. Dessa framgår i tillståndet forskare erhåller för användandet av datamaterialet eller i lagar och förordningar om användandet av materialet.

Olika fördelningsnyckeltal

Maximi och minimi är vanligen förknippade med en observation. Om det är möjligt att identifiera denna observation kan inte maxvärdet eller minvärdet publiceras.

Fördelningspunkterna (exkl. minimi och maximi) utgör ett specialfall av tabellen där antalet observationer som blir mellan fördelningspunkterna motsvarar cellfrekvenserna. Ifall antalen överskrider gränsvärdet 3 kan fördelningspunkterna publiceras.

Typvärdet kan publiceras ifall (nästan) alla observationer inte får samma värde.

Medeltalet, andra relationstal och de högre momenten i fördelningsnyckeltalen (t.ex. varians) kan publiceras om man vid beräkningen av dem har använt minst tre observationer.

Andra numeriska typer av utskrifter

Indextal, korrelationskoefficienter och teststorheter (t, F, X^2 , o.d.) kan vanligen publiceras om man vid beräkningen har använt ett tillräckligt antal (minst 10) observationer.

Regressionsmodeller kan publiceras i sin helhet om det ligger tillräckligt med observationer bakom modellen och modellen inte beskriver tidsserien på basis av observationer av ett företag/en person. Enskilda koefficienter i modellen kan i allmänhet alltid publiceras.

Figurer

Liksom när det gäller numeriska utskrifter får inte heller figurer avslöja uppgifter om en enskild observationsenhet. Figurer som utarbetats utgående från materialet är tillåtna om en enskild figurpunkt inte kan avslöja en enskild observation som ligger bakom figurpunkten.

Figurerna tas till granskning på samma sätt som tabellerna, klart och tydligt dokumenterat. Figurformat som lämpar sig för granskning är t.ex.:

Bitkarteformat

- PNG (Portable Networks Graphics)
- BMP (Bitmap)
- JPEG (Joint Photographic Experts Group)
- TIFF (Tagged Image File Format)

Vektorformat

- EPS (Encapsulated PostScript)
- PS (PostScript)
- PDF (Portable Document Format)
- SVG (Scalable Vector Graphics)
- WMF/EMF (Windows Metafile)

I programmet Stata kan ovannämnda figurformat skapas med kommandot `graph export`. I programmet SPSS kan figurformatet väljas i funktionen `Export output`. I R-programmet får man information om ritfunktionen med kommandot `help(grDevices)`. Vissa figurtyper, som t.ex. `gph`-filer i Stata, sparar i regel det material som använts när figuren ritats och därför lämpar de sig nödvändigtvis inte för extern överföring.

Identifiering av uppgifter om enskilda observationsenheter är svårare om figuren baserar sig på urvalsmaterial. Flera typer av företagsmaterial är ändå totalmaterial, varvid enskilda företags uppgifter lättare kan identifieras utgående från figurer t.ex. på basis av extremvärden. Även om något material inte är totalmaterial av alla företag i Finland, kan det vara totalmaterial för en snävare delpopulation, t.ex. en viss näringsgren.

Stapeldiagram och andra figurer för presentation av klassificerat material är vanligen tillåtna att publiceras så länge det finns en tillräcklig mängd observationer i varje klass. Denna typ av material kan vanligen presenteras också i tabellformat och man kan då direkt tillämpa Statistikcentralens anvisningar för skydd av tabellmaterial (bilaga A).

Fördelningsbilderna innehåller ibland avvikande observationer eller extremvärden som kan avslöja uppgifter om observationsenheten. Indelningar, histogram och kumulativa fördelningsfunktioner som har utjämnats eller presenterats på en tillräckligt grov nivå är tillåtna. Programmets ritfunktioner märker ofta automatiskt ut bl.a. i låddiagram (box plot) avvikande observationer som vanligen inte ska inkluderas i de figurer som publiceras.

Spridningsdiagram används vanligen för presentation av värden hos två kontinuerliga variabler. Med tanke på datasekretessen är detta svårare att genomföra än när det gäller ovan nämnda diagram. När det gäller spridningsdiagram bör man fästa speciell uppmärksamhet vid materialets karaktär, bl.a. urvalsstorlek, känsliga uppgifter och förekomsten av avvikande observationer.

Klart förbjudna figurtyper är bl.a. figurer som presenterar värden hos avvikande observationsenheter eller spridningsdiagram utifrån vilka man kan få kännedom om t.ex. uppgifter om det största företaget inom en näringsgren.

5.2 Kontroll av undersökningsresultaten i praktiken

För forskartjänstens distansanvändning och distansanvändning av mikrosimuleringsmodellen följer man olika praxis när det gäller kontroll. Undersökningsresultat som producerats vid fjärråtkomst och i forskningslaboratoriet kontrolleras innan uppgifterna lämnas ut. Från fjärrmiljön kan man inte heller själv överföra filer till den egna arbetsstationen utan dataöverföringen sker efter att man skickat en separat begäran via e-post. Vid SISU-mikrosimuleringen överför forskaren däremot filer som innehåller forskningsresultat direkt till den egna arbetsstationen utan någon förhandskontroll.

Kontrollförfarandet inom forskningsprojekt

Undersökningsresultat som producerats vid fjärråtkomst eller i forskningslaboratoriet kontrolleras för att säkra datasekretessen. Alla filer som ska genomgå kontroll ska uppfylla samma kriterier som gäller för tabeller och figurer som publiceras.

Reviderade 27.8.2019

T.ex. när det gäller log-filer ska man be att bara de nödvändigaste delarna eller delarna som är avsedda att publiceras ska genomgå kontroll. Filerna förs till kontroll genom att kopiera dem till en katalog som granskas (...\\out) och skicka en begäran om kontroll till e-postadressen tutkijapalvelut@stat.fi. Antalet och storleken på resultatfilerna ska vara måttliga (högst 2 Mb).

Efter att datasekretessen har kontrollerats skickas resultaten till forskarens e-postadress. För kontroll reserveras 1–2 arbetsdagar.

Kontrollförfarande vid mikrosimulering

Det är tillåtet att ta ut resultat från mikrosimuleringens fjärrmiljö utan någon förhandskontroll. På så sätt kan forskaren överföra filer med forskningsresultat från fjärrmiljön till sin egen arbetsstation.

I fjärrmiljön har varje användare en egen personlig Mail e-postkatalog via vilken man kan överföra filer till användarens egen arbetsstation. Överföringen sker genom att man kopierar önskade filer (från User-, Forum- eller Admin-katalogerna) till användarens personliga Mail-katalog. Ungefär två minuter efter kopieringen överförs filen automatiskt både till användarens personliga e-post och till mikrosimuleringens e-post vid Statistikcentralen (mikrosimulointi@stat.fi). Det kommer ett separat e-postmeddelande om varje fil som kopierats till Mail-katalogen med den kopierade filen som bilaga. Av e-postmeddelandet framgår namnet på filen som överförs och sänd datum. Storleken på filer som sänds får vara högst 1 megabyte (Mb).

Statistikcentralen granskar i efterhand de överförda filerna i mikrosimuleringens e-post. Forskaren är förpliktad att följa anvisningarna och reglerna för forskartjänsternas fjärrmiljö när det gäller de uppgifter som överförs. Sådana är bl.a. skydd av överförbara forskningsresultat, begränsningar som gäller filernas storlek samt andra begränsningar, möjlighet att tolka uppgifterna och publicering av uppgifterna (kapitel 6). *Särskild vikt ska fästas vid att uppgifter som överförs från mikrosimuleringens miljö inte av misstag innehåller uppgifter på individnivå eller att det finns en möjlighet att de avslöjas.*

6. Publicering av resultat

Forskaren förbinder sig att publicera forskningsresultaten bara i sådan form som inte möjliggör identifiering av uppgifter om ett enskilt företag eller en enskild person. För att säkerställa detta kan man kräva att forskningsrapporter och -publikationer ska granskas innan resultaten publiceras. Forskaren bör komma ihåg att reservera tillräckligt med tid för kontroll av att datasekretessen beaktas (ungefär 1–2 veckor). Det lönar sig att i förväg komma överens med forskartjänstens personal om granskning av forskningsresultaten.

När resultaten offentliggörs ska Statistikcentralen uppges som uppgiftskälla.

7. När projektet är slutfört

Det material som framställts inom distansanvändnings- och forskningslaboratorieprojekt raderas när projektets eller SISU-modellens användarrätt gått ut, ifall inget

separat kontrakt om sparandet av materialet ingåtts. Det datamaterial som överlåtit till någon utanför Statistikcentralen ska returneras/förstöras när tillståndets giltighetstid går ut. Alla kopior på materialet och arbetsfiler ska också förstöras. Statistikcentralen ska informeras om att materialet förstörts.

8. Sanktioner

Ifall forskaren eller kunden bryter mot avtalet om eller anvisningarna för distansanvändning avbryts distansanvändningsförbindelsen. Förbindelsen öppnas på nytt om kunden uppvisar en godkännbar skriftlig förklaring av orsakerna till förseelsen och åtgärder med vilka förseelser i fortsättningen ska förhindras.

Brott mot tystnadsplikt bestraffas enligt kap. 38 § 1 och 2 i strafflagen. Tystnadsplikten gäller också forskare. Straffet är böter eller fängelse i högst ett år.

9. Ansvar/uppgifter

I varje institution som ingått avtal om distansanvändning bör det finnas namngivna kontaktpersoner för tekniska och administrativa ärenden. Det hör till dessa att stöda nya distansanvändare i institutionen, följa med Statistikcentralens information om distansanvändning och delta i skolningstillfällen samt tillsammans med Statistikcentralen söka lösningar i problemsituationer. Kontaktpersonerna instruerar forskare som hanterar material i fjärrmiljön, för att säkerställa att de har tillräckliga tekniska färdigheter för hantering av material och att de inte oavsiktligt äventyrar tjänstens funktion eller dataskyddet när det gäller materialet. Kontaktpersonerna ska informera Statistikcentralen om alla sådana ändringar inom organisationen som väsentligt påverkar datasäkerheten eller dataskyddet med tanke på distansanvändning.

Personen som skriver under avtalet om distansanvändning för institutionens del är ansvarig för de personer som använder FIONA-systemet samt har ansvar för att de arbetslokaler som står till forskarnas förfogande lämpar sig för distansanvändning och att det är möjligt att hantera material där på ett sådant sätt att dataskyddet inte äventyras. Denna ansvarsperson godkänner nya distansanvändare inom den egna institutionen genom att underteckna deras distansanvändningsöverenskommelser. Personen skall informera Statistikcentralen då institutionen byter ut ansvars- eller kontaktpersoner.

Forskaren ansvarar för att följa sekretessförbindelsen samt direktiven i detta dokument. Forskarna skall informera forskartjänsten då de vill avsluta arbete i FIONA-systemet eller byter arbetsplats. Vid byte av arbetsplats bör en ny distansanvändningsöverenskommelse skickas till forskartjänsten.

Bilagor

BILAGA A: Anvisningar för skydd av tabellmaterial som bildats utgående från forskningsmaterial.

BILAGA B: Anvisningar för inloggning och arbete i distansanvändningssystemet

Bilaga A: Anvisningar för skydd av tabellmaterial som utarbetats utgående från forskningsmaterial

1. Syftet med anvisningarna

Dessa anvisningar har utarbetats på basis av Statistikcentralens egna anvisningar för skydd av företags- och personuppgifter i tabellformat (TK-00-1503-17 och TK-00-348-19). Med hjälp av dessa anvisningar vill Statistikcentralen främja ansvarsfullhet inom datasekretess och underlätta tillämpningen av lagar och statistiketiska principer vid publiceringen av statistiktabel som framställts utgående från forskningsmaterial.

Varje forskare som använder Statistikcentralens forskningsmaterial och som publicerar företags- eller personuppgifter i tabellformat ska läsa igenom dessa anvisningar samt de lagar och statistiketiska principer som ligger bakom anvisningarna.

2. Att tillämpa anvisningarna

I paragraf 13 i statistiklagen (280/2004) stadgas om överlåtning av uppgifter som samlats in för statistikföring för forskningsändamål. I motiveringarna i paragrafen konstateras följande³:

"När uppgifter lämnas ut ska skyddet för personuppgifter samt för uppgifter om affärs- och yrkeshemligheter säkras från fall till fall genom praktiska åtgärder, till exempel genom att kräva tillräckliga förfaranden för informationssäkerhet samt genom att se till att det finns tillräckliga arrangemang för behövlig övervakning och uppföljning av användningen av uppgifterna. – – Eftersom resultaten av vetenskaplig forskning vanligtvis är offentliga bör det i samband med att de offentliggörs alltid separat säkerställas att de enskilda statistiska objekt som legat till grund för forskningen inte kan identifieras utifrån de offentliga resultaten."

Med stöd av detta ska man alltså se till att uppgiftslämnarnas integritetsskydd samt affärs- och yrkeshemligheter inte äventyras i slutresultaten i de vetenskapliga undersökningarna där det material som har använts är uppgifter som Statistikcentralen lämnat ut. Forskaren ska beakta dessa skyddsanvisningar vid planering och utarbetande av publikationer i tabellformat utgående från undersökningsresultaten. Med hjälp av kontrollförfarandet ser Statistikcentralen för sin del till att datasekretessen säkerställs.

3. Begrepp och metoder förknippade med datasekretess gällande material i tabellform

Material i tabellform avser statistik där material på enhetsnivå är aggregerat och arrangerat i tabellform. Statistikenheten för personstatistik i tabellform är t.ex. en privatperson, en familj, ett hushåll eller ett bostadshushåll. Statistikenheten för företagsstatistik i tabellform är t.ex. ett företag, ett arbetsställe eller en koncern. När det gäller uppgifter om yrkesutövare i företagsstatistik i tabellform tillämpas samma säkerhetspraxis som i fråga om andra företagsuppgifter.

³ Regeringens proposition till riksdagen med förslag till lagar om ändring av statistiklagen och av 2 och 3 § i lagen om landsbygdsnäringsstatistik (RP 154/2012 rd).

Tabellmaterialet kan vara antingen

- en **frekvenstabell**, där cellvärdena är antalet statistikenheter som hör till cellen, eller
- en **mängdtabell**, där cellvärdena är summor, medeltal eller andra motsvarande nyckeltal av en variabel (t.ex. omsättning) som tabelleras, eller
- en kombination av ovan nämnda där både cellfrekvenser och mängduppgifter visas.

Mängdtabeller används betydligt oftare när man presenterar företagsuppgifter, medan personuppgifter oftast visas i form av frekvenstabeller.

När det gäller tabellmaterial finns det en **risk för röjande** om det finns risk för att en statistisk enhet som förekommer i tabellen avslöjas. Vid definieringen av cellspecifik röjanderisk används **känslighetsregler**. De vanligaste känslighetsreglerna är

- **tröskelvärdesregeln** enligt vilken en cell är känslig om den innehåller färre statistiska enheter än de enheter som ingår i tröskelvärdet som fastställts på förhand.
- **dominansregeln**, dvs. **(n,k)-regeln** enligt vilken en cell är känslig ifall dess n största statistiska enheter innehåller minst k procent av cellens totalvärde. Vid användning av dominansregeln ska parametrarna n och k definieras numeriskt för att säkerställa att företag som ska statistikföras bemöts likvärdigt.
- **p%-regeln**, enligt vilken en cell är känslig ifall estimatet för den största statistiska enheten som beräknats på basis av cellens totalvärde avviker med högst p procent från det rätta värdet.

Man kan också använda flera känslighetsregler parallellt. För att en cell ska vara känslig räcker det att den är känslig enligt en av känslighetsreglerna.

Som metod för skydd av tabellerna används vanligen **täckning** eller **ändring av klassificering**.

- Till täckning hör primär täckning av celler som ligger i risk för att röjas och sekundär täckning. Med sekundär täckning säkerställs att man inte med hjälp av tabellens rad- eller kolumntotaler kan avslöja de cellvärden som täckts primärt.

Täckningen kan också göras radvis. Om en radtotal i tabellen innehåller bara en liten mängd statistiska enheter (färre än tröskelvärdet), täcks denna rad i sin helhet utan att beakta antalet statistiska enheter i de olika cellerna.

- Genom att ändra klassificeringen försöker man eliminera de celler som är i risk för att röjas genom att sammanslå klasserna som innehåller sådana celler med andra klasser i tabellen. Att ändra klassificeringen betyder ofta i praktiken att hela klassificeringen görs grövre.

En metod för att skydda tabellerna kan också vara att man ändrar värden hos de celler som är i risk för att röjas. Exempel på sådana metoder är avrundning och ersättande av ursprungligt cellvärde med ungefärliga slumptal.

4. Rekommendationer för skydd av personuppgifter i tabellform

Alltid när man gör tabeller och innan uppgifterna publiceras ska man bedöma hur stor röjanderisken är för tabelluppgifter om personer, enskilda familjer eller bostadshushåll. Dataskyddsåtgärderna ska riktas så att röjanderisken är tillräckligt liten och att materialet inte tappar onödigt mycket information i och med skyddandet. Först och främst ska man beakta statistikenhetens rätt till datasekretess, men man ska också

samtidig komma ihåg att samhället och medborgarna har rätt till tillförlitlig statistisk information för beslutsfattandet och planeringen i samhället.

4.1 Bedömning av röjanderisk i tabeller

Röjanderisk i en tabell definierar hur nödvändiga dataskyddsåtgärderna är. Vid definiering av risken för röjande beaktas

- **variabler som innehåller känsliga uppgifter i tabellen**
- **små cell- och klassfrekvenser (tröskelvärdesregeln),**
- populationens storlek,
- antalet variabler samt
- statistikområdets storlek och positionsexakthet.

Vid bedömning av behovet av skydd av uppgifterna kan det också vara viktigt att tänka på om tabellvärdena är relativa eller absoluta, gäller uppgifterna ett år eller är uppgifterna summor eller medeltal av flera år och utgörs populationen för framställningen av statistiken av en viss befolkningsgrupp (t.ex. utlänningar, brottslingar, poliser, arbetslösa eller höginkomsttagare).

Vid bedömning av röjanderisken bör man fästa särskild vikt vid känsliga uppgifter, för t.ex. de i EU:s allmänna dataskyddsförordning nämnda speciella persondatagrupper samt individuppgifter i samband med uppgifter om brottsstraff och företeelser dvs uppgifter⁴ som beskriver någons

- ras eller etniskt ursprung,
- politiska åsikter
- religiös eller filosofisk övertygelse
- medlemskap i en fackförening
- hälsa
- sexuella beteende eller inriktning eller
- fällande domar i brottmål och överträdelser eller därmed sammanhängande säkerhetsåtgärder,

och andra uppgifter som kräver särskilt skydd så som.

- dödsorsak,
- vårdåtgärder eller därmed jämförbara åtgärder som gäller en individ,
- behov av socialvård eller de socialvårdsåtgärder, stödåtgärder och andra förmåner inom socialvården som någon erhållit
- språk, nationalitet, härkomst eller födelseland,
- inkomster, skulder och förmögenhet samt
- huvudsaklig verksamhet, yrkesställning, ovanligt yrke eller någon annan variabel som beskriver socioekonomisk ställning.

Om det finns en risk att uppgifter kan röjas ska personuppgifterna i tabellen skyddas tillräckligt.

4.2 Rekommendationer för skydd av uppgifter

Röjanderisken i en tabell hänför sig alltid till vissa celler. Genom att definiera dessa celler eller klasserna som innehåller cellerna och genom att skydda dem med lämpliga skyddsåtgärder kan röjanderisken i tabellen sänkas till en godtagbar nivå.

⁴ EU: s allmänna dataskyddsförordning (2016/679)

Användning av tröskelvärde

De s.k. känsliga cellerna i tabellen, som omfattas av røjanderisk, definieras med hjälp av tröskelvärdet. Vid exaktare definiering av tröskelvärdet ska man beakta samma faktorer som vid bedömning av røjanderisken. Som tröskelvärde rekommenderas minst 10 statistikenheter vid tillämpning radvis och minst 3 statistikenheter vid tillämpning cellvis.

Skyddsmetoder

Man kan skydda celler (och samtidigt tabellen) genom att ändra tabellens struktur, täcka enskilda cellvärden eller hela rader, eller ändra cellvärden t.ex. genom avrundning.

Om tabellen innehåller många känsliga celler eller om de känsliga cellerna är koncentrerade till enskilda klasser lönar det sig att skydda uppgifterna genom att ändra tabellens struktur. Man kan ändra tabellens struktur genom att ändra variabelernas klassificering eller genom att kontrollera antalet variabler.

Antalet variabler i tabellen bör vara färre ju mindre områdets statistik det är fråga om. När det gäller småområdesstatistik ska man också undvika korstabulering av flera variabler och helst publicera enkla frekvensfördelningar.

Om tabellen innehåller bara få känsliga celler rekommenderar man att uppgifterna skyddas genom täckning. Täckningen kan göras antingen cellvis eller radvis, dvs. genom att täcka alla celler i raden.

Kompletterande rekommendationer

Flerdimensionella tabeller. Om en tabell med personuppgifter har 3 eller fler variabler av vilka minst en är känslig eller om områdesnivån är mindre än landskap, är røjanderisken mycket sannolik.

Statistikområdets storlek. I fråga om statistik som gäller större områden, t.ex. kommun- eller storområdesstatistik, behövs sällan några datasekretessåtgärder. Också i fråga om sådan statistik ska populationen och klassificeringarna ändå väljas så att tabellen inte innehåller onödigt många små cellfrekvenser.

När statistikområdets storlek minskar minskar även antalet enheter som statistikförs. Det kan också finnas en røjanderisk för kommunbaserad statistik ifall områdets population är liten. Vid statistikföring av mindre områden än kommun finns det alltid en røjanderisk.

Urvalsstatistik. Urvalet inverkar på røjanderisken. Risken är större om statistikuppgifterna har framställts med hela populationen som material än om man utgående från urvalsmaterialet med hjälp av designvikter har producerat estimat som gäller hela populationen. Ändå ska røjanderisken i persontabellerna, som framställts på basis av urvalsmaterial bedömas. För att säkerställa kvaliteten på datasekretessen och estimat ska man i statistik som baserar sig på urval använda tröskelvärde vid definiering av cellernas känslighet. Tröskelvärdet kan dock vara mindre än i motsvarande statistik som baserar sig på uppgifter om hela populationen.

Ett ärende som statistikenhet. Vid statistikföring av olika ärenden (t.ex. brottsfall, trafikolyckor) är det inte nödvändigtvis fråga om personstatistik. Vid ärendestatistik ska man dock tillämpa skyddsrekommendationerna för personstatistik, ifall det utifrån statistiken är möjligt att identifiera en enskild person eller avslöja hans eller hennes egenskaper.

5. Rekommendationer för skydd av företagsuppgifter i tabellform

Det finns olika sätt att skydda företagsstatistik i tabellform beroende på olika faktorer förknippade med framställningen av statistik. Här presenteras en gradering i tre steg för att tillämpa datasekretessen:

1. I situationer där bara ett exakt röjande är känsligt räcker det med att tillämpa **tröskelvärdesregeln**. Tröskelvärdesregeln är standardregeln. Tröskelvärdet ska alltid vara minst 3.
2. När röjande av ungefärliga företagsuppgifter är känsligt ska man som känslighetsregel använda **dominansregeln** eller **p%-regeln**. Användningen av dominans- eller p%-regeln ska dock begränsas till bara nya statistiska uppgifter och man ska också parallellt alltid använda tröskelvärdesregeln. En uppgift är ny så länge som röjande av uppgiften inverkar på marknadsläget eller ett enskilt företags verksamhet. Tidsgränsen för hur länge en uppgift är ny och för användningen av dominansregeln är 15 månader från referenstidpunkten. När det gäller äldre uppgifter ska man använda tröskelvärdesregeln.
3. Om det är möjligt att skydda uppgifterna genom att **hemlighålla uppgiftslämnarnas identitet och antal**, är detta att rekommendera. Exempel på detta är estimat som beräknas utgående från urvalsmaterial. I samband med estimaten publiceras inga uppgifter om de statistikenheter som ingår i urvalet.

Statistiktabellen behöver inte skyddas om det inte föreligger någon röjanderisk eller om uppgifterna i tabellen är offentliga enligt lagen. T.ex. uppgifter som beskriver statliga eller kommunala myndigheters verksamhet samt produktion av offentliga tjänster är i huvudsak offentliga.⁵ Över 25 år gamla uppgifter om företag är också offentliga.⁶

Kompletterande rekommendationer

Tidsdimension. När man fattar beslut om metoder för skydd av företagsuppgifter ska man beakta tidsdimensionen. Företagsuppgifternas relevans minskar avsevärt ju längre tid det har gått från statistikens referenstidpunkt till publiceringstidpunkten. Mycket aktuell kortfristig statistik ska skyddas också för ungefärligt röjande.

Cellfrekvenser. Utan att äventyra datasekretessen kan även små cellfrekvenser i mängdtabellerna publiceras även om det egentliga cellvärdet är skyddat. Å andra sidan är skydd av antalet statistikenheter i lämpliga situationer ett sätt att offentliggöra uppgifterna om själva variabeln som tabelleras.

Statistik över arbetsställen. Vid skydd av statistik över arbetsställen ska man också säkerställa skyddet av företagsnivån. Vid definiering av röjanderisken för en cell ska man fästa uppmärksamhet vid både antalet arbetsställen i cellen och antalet olika företag till vilka arbetsställena hör.

Urvalsbaserad statistik. Urvalet är nödvändigtvis inte i sig själv en tillräcklig skyddsmetod, eftersom statistiken som baserar sig på urval vanligen omfattar de största företagen och oftast hänför sig det största intresset och den största röjanderisken till storföretag. Röjanderisken ska också bedömas i företagstabeller som utarbetats på basis av urvalsmaterial. Estimat och indextal som kan möjliggöra röjande ska skyddas. Ett estimat eller indextal kan möjliggöra röjande om man vid beräkningen

⁵ Statistiklagen (280/2004), 12§.

⁶ Lagen om offentlighet i myndigheternas verksamhet (621/1999), 31 §. Om det är fråga om en enskild näringsidkare (personuppgift) gäller sekretessen i 50 år från det att personen avled.

Reviderade 27.8.2019


av estimatet eller indextalet använt uppgifter om bara några företag. Då drabbas också estimatets tillförlitlighet.

Bilaga B: Anvisningar för inloggning och arbete i distansanvändningssystemet

Detta dokument om distansanvändningssystemet beskriver hur man loggar in i FIONA fjärrmiljön för forskningsmaterial och hur man använder tjänsten. Inloggning i fjärrmiljön sker på olika sätt beroende på vilken webbläsare som används. Anvisningar för användningen av SISU-mikrosimuleringsmodellen kan fås från modellens handbok.

1. Att logga in i fjärrmiljön med IE eller Firefox webbläsare

1. Upprätta förbindelse till webbsidan för Statistikcentralens forskartjänster (Statistikcentralen -> Produkter och tjänster -> Undersökningsmaterial -> Mikro-material -> Distansanvändning) http://www.stat.fi/tup/mikroaineistot/eta-kaytto_sv.html.
2. Välj inloggning till distansanvändningstjänsten
- FIONA, CSC-miljö. Alla nya projekt finns på FIONA-systemet
3. Inloggningen kan också göras direkt till adressen <https://fiona.stat.fi>.



Käyttäjätunnus:



Salasana:

[Ajankohtaista](#)

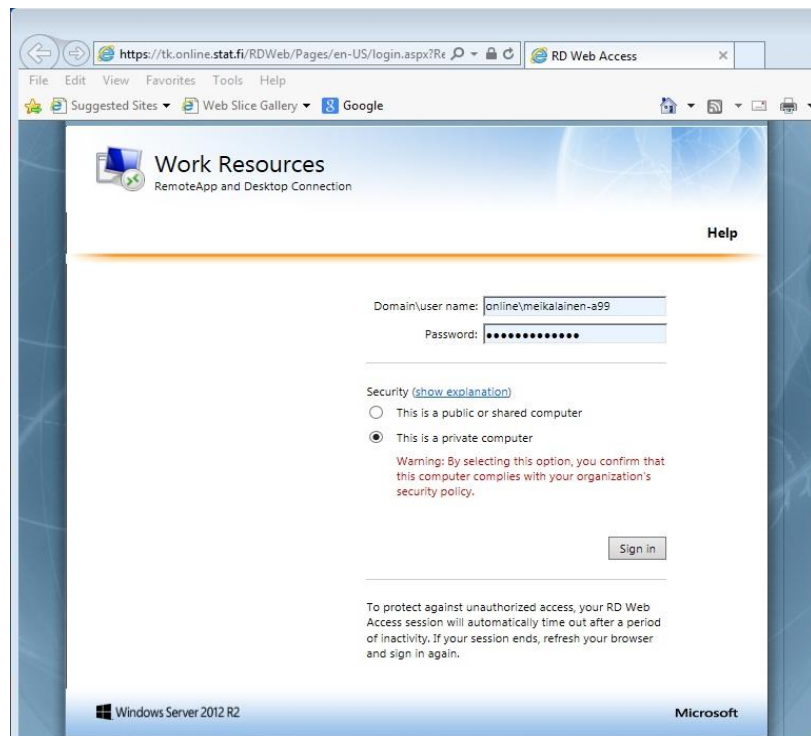
Ohjeet ja säännöt
Regler och anvisningar
Rules and instructions

Vuoden 2016 huoltokatkot:

| |
|------------|
| 9.3.2016 |
| 13.4.2016 |
| 11.5.2016 |
| 15.6.2016 |
| 13.7.2016 |
| 10.8.2016 |
| 14.9.2016 |
| 12.10.2016 |
| 9.11.2016 |
| 14.12.2016 |

 CSC-TIETEEN TIETOTEKNIIKAN KESKUS  Tilastokeskus

4. Första gången du upprättar förbindelse skall du godkänna (Allow) webbläsarens tillägg (Add-ons). ”This webpage wants to run the following add-on: Microsoft Remote Desktop Services Web Access Control...”

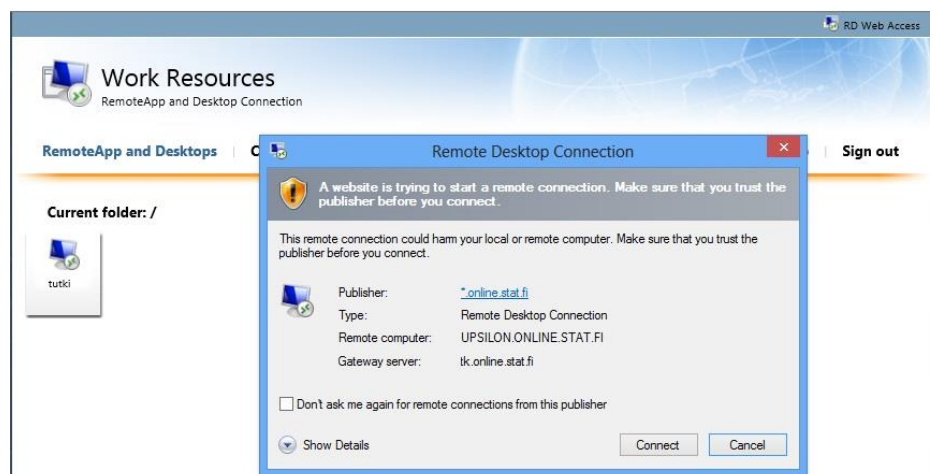


5. Vid inloggningen används endast ett användarnamn. Du loggar in genom att skriva domän-namnet (online) samt ditt användarnamn. Till användarnamnet här projektnumret, t.ex. Online\meikalainen-a99. Kryssa för "This is a private computer" innan du trycker på Sign in -knappen.
6. Work Resources -sidan visar de olika servergrupperna. Tutki-ikonen för till distansanvändningsservern Alfa, Beta, Gamma, Delta, Zeta och Eta. Dessa har 4 CPU/server.

För dem som har anmält att de använder SAS-programmet syns även Tutki-SAS -ikonen. Genom att välja den öppnas möjligheten att använda de server som innehåller SAS 9.4 (Penta, Heksa, Hepta och Okta. 4 cpu/server).

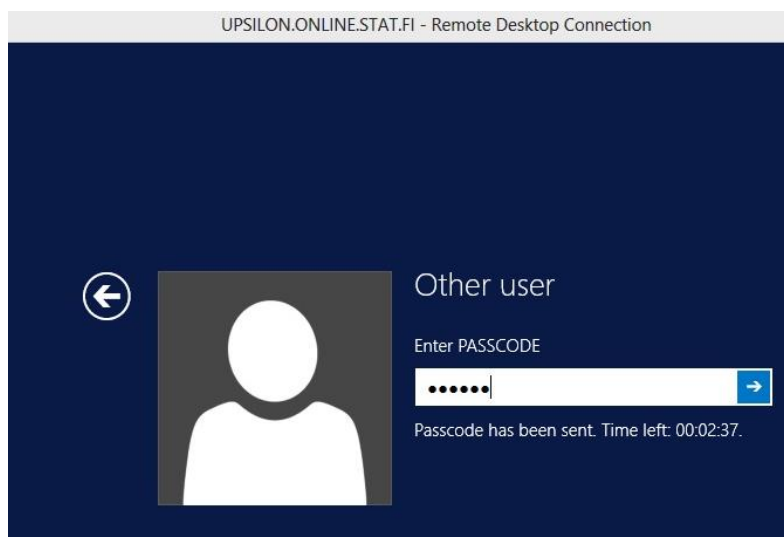
Från Mikrosim -ikonen kan du logga in på fjärrbordet för mikrosimulering.

Det kommer upp en varning om den okända servern Upsilon, som utför lastfördelningen av sessionerna. Du kan kryssa för Don't ask me again... och trycka Connect.

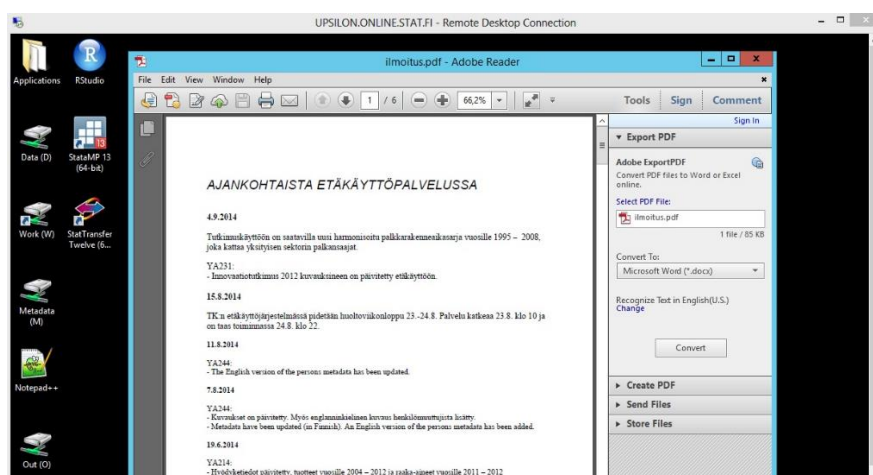


Reviderade 27.8.2019

7. Ifall du använder Firefox-läsaren, får du den första gången du loggar in en tilläggsfråga. .
” Du är i färd att öppna filen cpub-tutki-tutki-CmsRdsh.rdp. Vad skall göras med filen? ”Välj ”Öppna med hjälp av: Remote Desktop Connection (standard)” och kryssa för ”Gör detta som standard för denna typs filer”
8. För Firefox användare kommer i detta skede en ny förfrågan efter användarnamnet. Använd samma projektnummer som i punkt 4.
9. Nu följer SMS Passcode –autentisering. Till mobiltelefonen skickas ett flash-meddelande med en kod som skrivs in i Passcode-fönstret.



10. Fjärrskrivbordet öppnas



2. Allmänna anvisningar

Distansanvändningssystemet fungerar på tio Windows-serverar. Vid inloggningen styrs användarna till den server som är minst belastad vid inloggningstidpunkten. Man kan arbeta på samma sätt som på en vanlig Windows-arbetsstation. Projekten har löpande numrering med början från exempelprojektet a01.

Reviderade 27.8.2019

Allt material förvaras i den aktuella katalogen W:\a01. En säkerhetskopia tas en gång i dygnet och på detta sätt försöker man säkra att uppgifterna inte förstörs om diskettstationen skulle gå sönder. Användarna ser följande diskar:

D: Data innehåller färdigt material i forskartjänsten, läsringen beror på användningstillstånd.

M: Metadata innehåller beskrivningar av datamaterial

O:\a01 Out katalog för överföring av resultat

W:\a01 Users aktuell katalog

Om systemet belastas kan det störa andras arbete. Servern blir betydligt långsammare om systemets minne tar slut (på grund av användningen av swap-minnet). Därför ska användarna undvika att använda minnet i onödan.

Vid problemsituationer kontakta kontaktpersonen i den egna organisationen. Kontaktpersonerna svarar för programmets användarstöd. Statistikcentralens stöd svarar bara för funktionsdugligheten av program och användar-ID.

Material och program

Undersökningsmaterialen finns i SAS 7-format (.sas7bdat). Material kan överföras i önskat format med programmet Stat/Transfer. Följande program finns att tillgå i systemet:

R

Python Anaconda + komponenter

Stata

SAS

SPSS Statistics (begränsat antal användare)

Octave

Stat/Transfer

Graphviz

Kdiff

Libre Office (textbehandling och tabellkalkylering)

Lyx, Latex, MikTeX

Notepad++

Pnet, Pajek

Rstan, Rtools, Rstudio, TramineR (hör till R-programmet)

Strawberry

QGIS

Statistikcentralen erbjuder inget stöd för dessa tillämpningsprogram.

Stata-program

Programmen Stata 12, Stata 13 och Stata 14 har installerats i systemet. Programmets 64-bitarsversion betyder att det kan uppta omfattande material i minnet. Om minnet som reserverats är för stort leder det dock till att systemet blir långsammare, och därför får minnet som reserverats på Stata vara högst 4 000 MB.

När du slutar arbeta

Log Off

Reviderade 27.8.2019

När du slutar arbeta för dagen loggar du ut ur systemet. När du loggar ut blir systemets resurser lediga.

Klicka på Logoff-genvägen på skrivbordet (eller i start-menyn). Stäng webbläsarfönstret som visar inloggningsidan.

Disconnect

Om du håller en kort paus kan fjärrförbindelsen avbrytas utan utloggning. Du kan fortsätta med samma session genom att logga in på nytt. Förbindelsen avbryts genom att stänga distansförbindelsefönstret (eller via Shut Down och Disconnect i start-menyn). Disconnect lämnar sessionen öppen och inga resurser blir lediga (programlicenser och minnet som reserverats för användaren). Därför ska man alltid komma ihåg att klicka på Log Off när arbetsdagen är slut.

Fjärrskrivbordet låser sig om den inte har använts på 10 minuter. Förbindelsen avbryts om den inte har använts på 30 minuter. Inga program stängs av. Genom att öppna förbindelsen på nytt kan man fortsätta att arbeta.

Överföring av resultat

Alla forskningsresultat överförs från systemet av forskartjänstens personal. Man bör iaktta särskilt övervägande när uppgifterna överförs till granskning. Man bör undvika att överföra resultaten till granskning i många små delar. Till granskning överförda tabeller och grafer skall vara i sådan format att de kan publiceras. Se till att du följer reglerna för tabulering (se reglerna inkl. bilagor).

Dokumentera omsorgsfullt tabuleringen av resultaten i filerna som ska granskas. Varje tabell ska vara självförklarande på samma sätt som när det gäller journalartiklar. Antalet observationer ska synas i varje grupp.

Flytta filerna till katalog O:\a01 (exempelprojekt).

Skicka en begäran om överföring av resultaten till tutkijapalvelut@stat.fi (nämna projektets ordningsnummer). Efter att resultaten granskats skickas de till forskarens e-postadress.