Assignment 1 Data Science

Reproduserbarhet

Sofie Brynjelsen, Silje Marie Danielsen

# Innledning

I denne oppgaven skal vi ta for oss hva reproduserbarhet og replikerbarhet er. Samt problemet med reprodusering..

I dagens samfunn er vitenskapen i utvikling, og det vil være et stort behov for at man har tillit til forskningen som blir gjort. For å skape tillit er det vesentlig at funnen som er blitt gjort, kan reproduseres. Det vil si at når en gjennomfører samme analyse basert på samme data, vil man endre opp med de eksakt samme svarene. I forbindelse med reproduserbarhet vil det være vesentlig å nevne replikerbarhet. Replikerbarhet handler om at man får samme konklusjon, når man gjennomfører samme undersøkelse med nye data. Dette omtales også som gullstandarden når det gjelder vitenskap. vi kan se på reproduserbarhet som nødvendig, men ikke tilstrekkelig for replikerbarhet.

Forskere har funnet ut at det til tider er vanskelig med reproduserbarhet. I denne forbindelse nevner (Peng, 2011) at reproduserbarhet bær være et minstekrav for at en artikkel skal produseres.

Robus og pålitelig forskning er grunnlaget for vitesnakpelig utvikling og fremgang. Dette avhenger av at forskernes evne til å samle inn data fra tidligere arbeid. “robust and reliable sciene” omhandler at forskningen skal være reproduserbar, replikerbar og generaliserbar. Når det kommer til defineringen av disse uttrykkene, har det oppstått forvirring og misforståelser. I denne fobindelse vil vi velge å definere disse på engelsk. (Bollen et al., 2015)

## Definisjoner

*"****Replicability*** *refers to: the ability of a researcher to duplicate the results of a prior study if the same procedures are followed but the new data are collected.*"

*"****Generalizability*** *refers to: wheter the results of a study apply in other contexts or populations that differ from the originals.*"

*"****Reproducibility*** *refers to: the ability of a researcher to duplicate the results of a prior study using the same materials and prcedures used by the original investigator.*"

*Reproduserbarhet* kan også deles inn i tre hovedkategorier; *methods reproducibility*, *results reproducibility* og *robustness and generalizability*.

Methods reproducibility: “Methods reproducibility refers to the provision of enough detail about study procedures and data so the same procedures could, in theory or in actuality, be exactly repeated”

Results reproducibility: “Results reproducibility (previously descrived as replicability) refers to obtaining the same results from the conduct of an independent study whose procedure are as closely matched to the original experiment as possible.”

Robustness and generallzability: “We briefly introduce these terms because they are sometimes used in lieu of the term reproducibility. Robustness refers to the stability of experimental conclusjon to varaitions in either baseline assumptions or experimental procedures. It is somewhat related to the concept of generalizability (also known as transportability), which refers to the persistence of an effect in settings different from and outside of an experimental framework”.

##Problemer Hoveddel

# **Publication bias**

Et annet problem knyttet til forskning er publiserings skeivhet også kalt for publication bias. Dette omhandler at sannsynligheten for at en studie skal bli publisert avhenger av konklusjonen til studiet. Etter å ha undersøkt er det mange som har funnet ut at det er veldig få studier som har en negativ konklusjon. Det vil si at man ikke forkaster vår null hypotese. dette kan være et problem knyttet til at forskningen kan vise at det ikke har noen effekt, noe som videre kan fære til at man går glipp av viktig kunnskap. Et “worst case scenario” er at de vitenskapelige skriftene blir en stor samling av type 1 error.

# **Type 1 error**

Type 1 error vil si at man forkaster H0 når den i virkeligheten skal beholdes. Et av problemene knyttet til dette tema er hvis man for eksempel ufører like 100 studier som finner ingen effekt, og vi har en som finner en effekt, vil det være en sjanse for at man finner en studie i litteraturen som viser et feil resultat. Dette blir ofte omtalt som “the File Drawer Problem” (Rosenthal, 1979). Det vil si at de studiene som viser at man ikke kan forkaste H0, ikke blir publiserte. Studier som viser at man kan forkaste H0, blir publiserte. (Simmons et al., 2011) hevder at dette er en svært kostbar feil å gjøre, da de blir liggende i literaturen lenge. Dette vil også medføre at det blir mindre muligheter for andre å reprodusere studiene for å se om de samsvarer. Falske resultater vil overleve over en lang periode. et annet problem knyttet til dette er at man kan bruke de falske positve svarene som utgangspunkt for nye undersøkelser. Man vil da bruke store ressurser på å arbeide med noe som ikke er riktig. Videre kan dette føre til politiske følger og kostbare reformer som da i sin helhet er begrunnet ut ifra vitenskap som ikke gjelder. Alt i alt vil dette svekke troverdigheten.

# **Publication bias and meta-analysis**

publiserings skeivhet kan videre forplante seg i metaanalyser, hvor man tar for seg mange artikler innenfor et fagområde. hensikteten med meta-analysis er å finne det generelle svaret ved å sammenfatte resultatene fra den tidligere forskningen. Dersom man da bare tar med de falske positive artiklene som blir publisert, vil dette ha stor betydning for hvilen konklusjon man ender opp med.(Young et al., 2008)

# **The replication crisis**

Et annet problem som angår dette tema er det som blir kalt for the replication crisis. Dette har sine røtter innen psykologi men etter hvert er det flere fagområder, som for eksmepel økonomi som har vist interesse for tema. (Simmons et al., 2011) beviste at man vil føle seg yngre ved å høre på “When I’m sixty four” av The beatles. Dette er forskning som ikke stemmer, og etter dette utfallet ville de finne en løsning på problemet når det kommer til publisering av forskning. De innførte dermed seks ulike krav til forfatterne og fire guidelines til redaktørene. Hovedformålet med disse kravene og guidlinesene er at man skal kunne unngå å få slike resultat, som den tidligere nevnte forskningen.

# Løsning

Frisch (1933) var en av de første som ønsket en løsning på problemet med replikkering og reprodusering. Han kom med et forslag om at de statistike eller andre nummerisk arbeid, som skal presenteres i den ledende tiddskriften Econometrica, må ha med originale rådata og at de blir publiserte sammen med artiklene, med mindre de er omfattene og store. Dette ble et stort problem ettersom regresjon var et stort om omfattende arbeid og de ble vanskelige å flytte fra en maskin til en annen. Det vil si at replikering og reproduksjon ble umulig. JMCB gjennomførte også et prosjekt hvor de ville kjøre en replikasjon på innsendte dataset som følgte med artikler som var blitt publiserte. De undersøkte mer enn 70 arbeid og klarte å direkte reprodusere to av dem. Mange av de publiserte arbeidene var umulige å gjenprodusere på grunn av mangel på data.

Et forslag til å løse problemene med reprodusering av data, var å få arkiver i journalene som ble brukt. Man akriverer koder for å regne ut tester og modeller, noe som har vært brukt over lang tid. Senere fant McCullough et al. (2008) ut at dette ikke fikset problemet vårt med reprodusering av den grunn at det ikke finnes noe data. Senere i 2013, fant de en annen mer teoretisk løsning på problemet. Ved å lage et stort felles register der man registrerer dataene som er samlet inn.

En annen løsning på problemet er kalt for *computable documents*. Dette omhandler at koden er en del av artikkelen som har blitt publisert. Den vil da ligge inne i de opprinnelige dokumentene som er blitt tilsendt til til tidskriftene. Man vil dermed ha en kode til å lese inn data, en kode for å regne ut og behandle dem, samt koder til å teste dem og til å rapportere resultatene. Ideen er inspirert av Knuth (1992) som ville forbedre kvaliteten på dataprogrammer og kvaliteten på kodene. Ramsey (udatert) laget et program kalt for Noweb, som skulle være med på å gjøre det enklere løsning som skulle brukes på mange forskjellige programmeringsspråk. Dette kan bidra til å øke kvaliteten.

Gentleman og Lang (2007) var opptatt av at det er essensielt å intregere beregningene og kodene som ble brukt i dataanalysen, med selve dokumentet og beskrivelsene av funnene. De innførte dermed et kompenndium “computable documents” som skal være en samling av ulike elementer, tekst, kode dataene, i tillegg til at de skal kunne pakkes i ein enhet slik at det skal være enkelt å repdorusere resultatene. Får en til dette med disse kompendiene er en godt på vei til repdouserbar forskning.

***Dynamisk dokument*** i kompmendiene er en ordnet sammensettning av chuncks stykker med kode og stykker med tekst. Og at disse her kodestykkene ikke nødvnedigvis trenger å bli uført sekvensielt. Disse kan ha avanserte strukturer. I 2005 gjennomførte gentleman en case sudy for å vise at dette var gjennomførtbart.

# Avslutning

## Referanser

Bollen, K., Cacioppo, J. T., Krosnick, J. A., Olds, J. L., og Kaplan, R. M. (2015). *Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science* (Report of the Subcommittee on Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences). NSF.

Frisch, R. (1933). Editor’s Note. *Econometrica*, *1*(1), 1–4.

Gentleman, R., og Lang, D. T. (2007). Statistical Analyses and Reproducible Research. *Journal of Computational and Graphical Statistics*, *16*(1), 1–23. <https://doi.org/10.1198/106186007X178663>

Knuth, D. E. (1992). *Literate Programming*. Cambridge University Press.

McCullough, B. D., McGeary, K. A., og Harrison, T. D. (2008). Do Economics Journal Archives Promote Replicable Research? *Canadian Journal of Economics/Revue canadienne d’économique*, *41*(4), 1406–1420. <https://doi.org/10.1111/j.1540-5982.2008.00509.x>

Peng, R. D. (2011). Reproducible Research in Computational Science. *Science*, *334*(6060), 1226–1227. <https://doi.org/10.1126/science.1213847>

Ramsey, N. (udatert). *Noweb Home Page*.

Rosenthal, R. (1979). *The File Drawer Problem and Tolerance for Null Results.* *86*, 638–641.

Simmons, J. P., Nelson, L. D., og Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>

Young, N. S., Ioannidis, J. P. A., og Al-Ubaydli, O. (2008). Why Current Publication Practices May Distort Science. *PLoS Medicine*, *5*(10), e201. <https://doi.org/10.1371/journal.pmed.0050201>

## Appendiks