

!!STATISTIK

Flera typer

Deskriptiv statistik (dvs metoder för att organisera, summera och presentera data på ett informativt sätt)

Inferentiell statistik (dvs metoder för att dra slutsatser om en population med utgångspunkt från ett stickprov)

Klassificering

Experimentella undersökningar

Kan upprepas så många gånger som behövs

Icke-experimentell undersökningar

Konstatera vad som hänt i efterhand och sedan kartlägga orsaken

Genomförande av en undersökning

Vem skall undersökas?

Hur skall undersökningen göras?

Vad skall undersökas?

Population och stickprov

Stickprov från en ändlig population

Stickprov från en oändlig population

Population = sannolighetsfördelning för den slumpvariabel som observeras

Stickprov = den följd av värden som observeras på slumpvariabeln

Slumpen bestämmer vilka variabelvärden som genereras.

Ram = företeckning över populationens element

Total- eller urvalsundersökning

Totalundersökning

Om det är omotiverat att göra en urvalsundersökning p.g.a. att det är en så pass liten population.

Risken att vissa intressant grupper försvinner vid en urvalsundersökning.

Urvalsundersökning

Stor population

Ett test är destruktivt

Vad vi kan påverka

Hur skattningen bestäms

Hur stort urval som görs

Hur observationen samlas in

Urvalsmetoder

Icke-sannolikhetsval

Ingen känd slumpmekanism. Ger dålig skattning. Ex. de 30 första du möter på gatan.

Sannorlikhetsval

Varje element i populationen har en känd sannorlikhet som är större än noll att komma med. Man vet hur bra en skattning är.

Vanliga metoder

Obundet slumpmässigt val (OSU)

Fördel: Enkelt att göra

Nackdel: Man kan missa grupper med avvikande värden. Ingen känd relevant apriori information om individerna.

Stratifierat urval

Delar in populationen i grupper med avseende på någon egenskap.

Fördel: Bättre precision i skattningen av populationen.

Nackdelen: Krångligare att utföra

Systematiskt utval

Populationen delas in i intervall ($a = N/n$). En slumpning. Sedan tas den a:te individen.

Förel: Enekel

Nackel: Cyklisk trend kan ge skeva skattningar

Gruppurval (klusterurval)

Populations individer grupperas i delpopulationer (ex geografisk spridning).

Ett OSU dras av delpopulationen. Varje individ i delpopulationen undersöks.

Fördel: Ingen urvalsram behövs

Nackdel: Stor geografisk spridning kan bli dyrt.

Insamlingsmetoder

Primärdataundersökning (data samlas in för första gången)

Sekundärdataundersökning (redan insamlad data används)

Variabeltyper

Kvantitativa (numeriskt mätbara)

Diskreta (antar distinkta värden)

Kontinuerliga (antar värden inom intervall)

Kvalitativa (kategori eller icke-numeriska värden)

Skaltyper

Nominalskala

Räknar och klassificerar frekvenser för variabelvärden. Inte möjligt att rangordna variabelvärdena,

Ordinalskala

Det är möjligt att rangordna variabelvärdena. Mätvärdenas summa eller differens ger ingen värdefull information.

Intervallskala

Kan ange skillnad mellan variabelvärden. Summa och differensen ger värdefull information. Absolut *nollpunkt* saknas dock.

Kvotskala

Variabelvärdena har en absolut nollpunkt. Vi kan bilda kvoten av variabelvärden.

Kumulativ frekvensfördelning

Kumulativa frekvenser erhålls genom att stegvis addera frekvenserna från det lägsta variabelvärdet till det högsta.

Lägesmått

Typvärdet

Motsvarar fördelningens maximum. M.a.o det vanligaste värdet/värdet med störst frekvens.

Medianen

Motsvarar mittobservationen

Percentiler

Delar in observationerna i 100 lika stora delar. Medianen hittas vid det 50:onde percentilen.

Aritmetiska medelvärdet

Medelvärdet är summan av de observerade värdena dividerat med antalet observationer.

Viktat medelvärde

Se anteckningar

Spridningsmått

Varians och standardavvikelse

Se anteckningar

Explorativ Data Analys (EDA)

Lådagram (boxplot)

Spridningsdiagram

Antag två variabler. För varje individ har vi två mätvärden. Sambandet kan presenteras i ett spridningsdiagram.