Begrepp

* Population – N: Samtliga enheter som man vill dra slutsatser om.
* Stickprov – n: Ett antal enheter i populationen.
* Värde på ett stickprov – Xi
* Värde ifrån frekvenstabell – fi
* Kvalitativ variabel: Antar inte numeriska värden.
* Kvantitativ variabel: Antar numeriska värden.
* Diskreta kvantitativa värden: Antar endast helvärden.
* Kontinuerliga kvantitativa värden: Kan mätas med flera decimalers noggrannhet.
* Nominalskala: Kvalitativ variabel som ej går att rangordna efter variabelns värde.
* Ordinalskala: Variabelns värde har en inbördes storleksordning och kan därmed rangordnas.
* Metrisk skala: Kvantitativa variabler där avstånden är densamma mellan variabelns värden.
* Typvärde: Den variabel där det råder störst andel.
* Sannolikhet – Pr(A)
* Snitt 🡪 A∩B: Tillhör flera av de händelser som är definierade.
* Union 🡪 A∪B: Tillhör minst en av de definierade händelserna.
* Disjunkta händelser: Händelserna har ej några gemensamma element.
* Disjunkta 🡪 ≠ oberoende
* Oberoende händelser: Sannolikheten för en händelse påverkas inte av en händelse som redan inträffat.

Pr(A∩B) = Pr(A)xPr(B)

Pr(AΙB) = Pr(A)

* Utfallsrum – S: T.ex ett tärningskast 🡪 S = [1,2,3,4,5,6]
* Komplementmängd - AC: De elementutfall som ej är önskade.
* Betingad sannolikhet - Pr(AΙB): Sannolikheten för att händelsen A ska inträffa givet att händelse B redan inträffat.
* Permutation: Kombinationer där ordningen har en betydelse.
* Kombinationer: Kombinationer där ordningen inte har någon betydelse
* Slumpvariabel: Det är slumpen som bestämmer utfallet och hur ofta respektive värde i utfallsrummet antas.
* Förväntat värde – E(X)
* Väntevärde/populationsmedelvärde – μ
* Kontinuerlig sannolikhetsfördelning: Fördelningar av variabler som kan mätas med en mängd decimalers noggrannhet.
* Obundet slumpmässigt urval – OSU: Alla enheter i populationen har samma sannolikhet att bli utvalda i stickprovet.
* Stratifierat urval: En population kan delas upp i grupper med avseende på någon egenskap som vi tror har betydelse för det som studeras.
* Övertäckning: Enheter i rampopulationen som egentligen inte tillhör målpopulationen
* Undertäckning: Enheter i rampopulationen som saknas i målpopulationen.
* Replikat: Samma enhet ingår flera gånger i rampopulationen.
* Populationsparameter: Beskrivande mått för populationen.
* Stickprovsstatistika: Skattningar av populationsparametrarna och baseras på stickprovet som har dragits 🡪 slumpvariabler.
* Effektiv skattning: Skattningen som har lägst varians eller medelfel.
* De stora talens lag: Ju större stickprov som dras, desto mer lika varandra blir stickprovsstatistikorna och populationsparametrarna.
* Bortfall: När somliga personer/enheter inte svarar på en enkät eller svarar felaktigt.
* Normalfördelning: Är symmetrisk och definieras av väntevärdet och standardavvikelsen.
* Normalfördelningsapproximation: Stickprovet är större än 30 och för stort för att räkna på och därför approximeras talet för att underlätta uträkningen.
* Konfidensintervall: Med en viss säkerhet täcks den okända populationsparametern av intervallet.
* Konfidensnivå: Mått på konfidentintervallets säkerhet.
* Signifikansnivå - α: Mått på risken som finns i konfidensintervallet.
* Bredd: Differensen mellan konfidensintervallets övre- och undre gräns.
* Enkelsidigt konfidensintervall: Innehåller endast en övre- eller undre gräns.
* Dubbelsidigt konfidensintervall: Innehåller både en övre- och undre gräns.
* Hypotesprövning: Hur sannolika hypoteser är med hjälp av insamlad data.
* P-värde: Sannolikheten att testvariabeln skall anta ett värde som är observerat eller ännu längre ifrån μ0.
* Typ 1 – Fel: Risken att förkasta H0 fastän H0 är sann.
* Typ 2 – Fel: Risken att inte förkasta H0 fastän Ha är sann.
* När signifikansnivån ökar så ökar risken för typ 1–fel medan risken för typ 2-fel minskar.
* Parvisa jämförelser: Stickprov ifrån två olika populationer och man undersöker om det finns skillnader i medelvärden eller andelar. Det uppstår ett beroende mellan mätvärdena för respektive enhet eftersom det är samma enheter som undersöks före- och efter en undersökning.
* Weight: En sammanvägning av stickprovsmedelvärdena från respektive stratum.
* Allokering: Hur många enheter som skall väljas ur respektive stratum.
* Lika allokering: Lika stort stickprov ur respektive stratum.
* Proportionell allokering: Man drar lika stor andel ur respektive stratum.
* Neymanallokering: Vi har förhandsinformation om standardavvikelsen i strata. Tanken är att dra större OSU ur stratum med stor standardavvikelse.
* Optimal allokering: Tar hänsyn till kostnaden vid allokering av urvalet.
* Korstabell: Tabell som sammanfattar information om två kvalitativa variabler.
* Positivt samband: Punktsvärmen lutar uppåt. När den förklarade variabeln ökar så ökar även responsvariabeln.
* Negativt samband: Punktsvärmen lutar nedåt. När den förklarade variabeln ökar så minskar även responsvariabeln.
* Korrelationskoefficient: Graden av linjärt samband mellan två kvantitativa variabler.
* Enkel linjär regression: En rät linje i svärmen av observationer och man kan därmed få reda på hur mycket y-variabeln förändras när x-variabeln ökar/minskas en enhet.
* Prognosticering: En punkskattning av det förväntade värdet på y när x har värdet x\*
* Förklaringsgrad – r2: Ett mått på hur stor andel av variationen i y-variabeln som förklaras av x-variabeln som finns med i modellen.
* Residualanalys: Utreder hur välanpassad regressionslinjen är till datamaterialet.
* Residualer: Feltermer

Krav för beräkningar

* Regler för sannolikheter
* En sannolikhet ligger alltid mellan 0 och 1 🡪 0 ≤ Pr(A) ≤ 1
* Sannolikheten för disjunkta händelser som ingår i utfallsrummet kommer tillsammans att summera till 1
* Diskreta sannolikhetsfördelningar
* Bernoullifördelning/tvåpunktsfördelning: Varje delförsök kan endast anta ett av två olika utfall.
* Binomialfördelning: Alla delförsök är oberoende av varandra och bernoullifördelat. Stickprov med återläggning eller utgör mindre än 10% av populationen.
* Hypergeometrisk fördelning: Stickprov utan återläggning och utgör mer än 10% av populationen.
* Poissonfödelning: Ett stort antal oberoende delförsök och sannolikheten för varje delförsök ska inträffa är väldigt liten 🡪 approximering.
* Geometrisk fördelning: Man fortsätter med ett delförsök ända tills man lyckas första gången.
* Normalfördelningsapproximation av binomialfördelning
* Givet att: nπ(1-π) > 5
* Approximering av Poissonfördelning
* När n är mycket stort och π är väldigt litet.
* n > 20 och π < 0,05
* Samplingfördelningar
* Samplingsfördelningen för summor eller medelvärden anses som approximativt fördelad om stickprovsstorleken är minst 30 enheter.
* Samplingsfördelningen för stickprovsandel anses som approximativt normalfördelad om: np(1-p) > 5
* Konfidensintervall för populationsmedelvärde:
* Stickprovet ska vara draget som ett OSU 🡪 Oberoende mellan observationerna
* Samplingsfördelningen för stickprovsmedelvärdet går att betrakta som normalfördelat.
* Konfidensintervall för populationsandel:
* Stickprovet skall vara draget som ett OSU 🡪 Oberoende mellan observationerna
* np(1-p) > 5
* Hypotesprövning för populationsmedelvärde:
* Stickprovet skall vara draget som ett OSU 🡪 Oberoende mellan observationerna
* Samplingsfördelningen för stickprovsmedelvärdet går att betrakta som normalfördelat.
* Hypotesprövning för populationsandel:
* Stickprovet skall vara draget som ett OSU 🡪 Oberoende mellan observationerna
* np(1-p) > 5
* Känd standardavvikelse för populationen
* Populationsstandardavvikelsen är känd
* Stickprovet är draget som OSU
* Samplingfördelningen för stickprovsmedelvärdet går att betrakta som normalfördelad
* Jämförelse mellan p-värde och signifikansnivå
* Om P < α 🡪 H0 förkastas
* Relation mellan konfidensintervall och hypotesprövning
* För en dubbelsidig mothypotes krävs ett dubbelsidigt konfidensintervall.
* För enkelsidiga mothypoteser krävs rätt val av uppåt- eller nedåt begränsade konfidensintervall.
* Ha: μ < μ0 🡪 uppåt begränsat konfidensintervall
* Ha: μ > μ0 🡪 nedåt begränsat konfidensintervall
* Ha: μ ≠ μ0 🡪 Dubbelsidigt konfidensintervall
* Bildande av konfidensintervall medelvärdet i två populationer
* Vi har dragit två OSU 🡪 Detta garanterar oberoende mellan observationerna men också mellan stickproven
* Samplingfördelningarna för de två stickprovsmedelvärdena kan betraktas som normalfördelade.
* Hypotesprövning för jämförelse av populationsmedelvärden
* Vi har dragit två OSU 🡪 Detta garanterar oberoende mellan observationerna men också mellan stickproven
* Samplingfördelningarna för de två stickprovsmedelvärdena kan betraktas som normalfördelade.
* Hypotesprövning för jämförelse av populationsandelar
* Vi har dragit två OSU 🡪 Detta garanterar oberoende mellan observationerna men också mellan stickproven
* np(1-p) > 5
* Relation mellan konfidensintervall och hypotesprövning (två populationer)
* Ha: μ1 – μ2 < d0 🡪 Uppåt begränsat konfidensintervall
* Ha: μ1 – μ2 < d0 🡪 Nedåt begränsat konfidensintervall
* Ha: μ1 – μ2 ≠ d0 🡪 dubbelsidigt konfidensintervall
* Ändliga och oändliga populationer
* Oändlig population givet att: n/N < 10%
* Ändlig population givet att: n/N > 10%
* Konfidensintervall för totalmängd
* Populationsmedelvärde: N×μ bestäms enligt N×X(tak)
* Populationsandel: N×π bestäms enligt N×p
* Chitvå-test för korstabeller
* Det råder oberoende mellan cellerna. Samma element/person får inte förekomma i flera celler i tabellen.
* Max 20% av de förväntade frekvenserna är mindre än 5.
* Alla förväntade frekvenser är större än 1.
* Korrelationskoefficient

r Samband

>0,85 Mycket starkt

0.65-0.85 Starkt

0.35-0.65 Måttligt

0.20-0.35 Svagt

< 0.20 Mycket svagt

* Formler för kvantitativa variabler

r =$\frac{∑xy-( \frac{∑x × ∑y}{n})}{\sqrt{\left(∑x^{2}- \frac{\left(∑x\right)^{2}}{n}\right)}×\left(∑y^{2}- \frac{\left(∑y\right)^{2}}{n}\right)}$

b1 = $\frac{∑xy-( \frac{∑x × ∑y}{n})}{\left(∑x^{2}- \frac{\left(∑x\right)^{2}}{n}\right)}$

t = $\frac{b1}{\frac{s}{\sqrt{\left(∑x^{2}- \frac{\left(∑x\right)^{2}}{n}\right)}}}$

$∑(x$i – x )2 = $\left(∑x^{2}- \frac{\left(∑x\right)^{2}}{n}\right)$