(1)

boken
SF1922
200318

Föreläsning 1

Beskrivande statistik: hur man tabulerar, grafiskt representerar och numeriskt bearbetar **data**

Tabulering

a) Ogrupperande data
En mängd rådata som inte blivit sorterad.

b) Grupperande data
Data av samma storlek sammanförs.
Kan representeras i t.ex. en frekvenstabell.

c) Klassindelande data
Data av ungefär samma storlek sammanförs.

Klasser

Klassindelning kan ske på olika sätt, men består dock alltid av:

a) Klassgränser
En klassgräns utgörs av ett infimum g_i och ett supremum g_{i+1} för en klass i

i) Klassbredd
Klassbredden $h_i = g_{i+1} - g_i$ bör vara konstant för alla klasser s.a. $h_i = h \ \forall i$

ii) Intervall
Ange intervallen för en klassgräns så alla värden entydligt kan föras till en klass. t.ex. $2.30 \leq x < 2.32$

iii) Öppna klasser
Bör undvikas, t.ex. "alla observationer som är mindre än 2.32"

b) Klassmitt
Klassmitten är medelvärdet av klassgränserna

Frekvenstabeller

Gör stor mängd data lättöverskådlig.
Brukar innehålla följande:

a) Olika förekommande värden
Datapunkter som grupperas/klass-indelas beroende på deras storlek, betecknat y_i

b) Klass
Värden y_i kan delas in i "i" olika klasser, där $i = 1, 2, \ldots$ och 1:a klassen representerar det (oftast) lägsta förekommande värdet

c) Absolut frekvens
Antalet datapunkter i en klass, betecknat f_i

d) Relativ frekvens
Absolut frekvens per totala antalet datapunkter (n), betecknat p_i
$\therefore \ p_i = f_i / n$ och anges ofta i %

Grafisk presentation

En grafisk presentation av en mängd data kan göras med hjälp av:

- Stolpdiagram
- Histogram
- Boxplott / Lådogram

→

Läges- och spridningsmått (Ogrupperade data)

a) Lägesmått

Anger hur stor en storhet är i medeltal. Kan använda t.ex.

i) Geometriskt medelvärde, G

Medelvärde för normaliserade resultat, används vid uträkning av genomsnittlig ränta

$$\bar{x} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

ii) Harmoniskt medelvärde, H

Inverterade värdet av det aritmetiska medelvärdet, används vid tillväxtfenomen

$$\bar{x} = \left(\frac{1}{n} \cdot \sum_{i=1}^{n} x_i^{-1}\right)^{-1}$$

iii) Kvadratiskt medelvärde, Q (RMS)

Anger variationerna hos en storhets belopp, ett generaliserat medelvärde, används inom elektrotekniken

$$x_{rms} = \sqrt{\frac{1}{n}\left(x_1^2 + x_2^2 + \ldots + x_n^2\right)}$$

överkurs för now

iv) Aritmetiska medelvärde, A

Genomsnittliga värdet av en datamängd, används oftast som lägesmått

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

← WE LIKE

v) Median

Används för att minska avvikande värdens påverkan

$\bar{x}_Q \leq \bar{x}_A \leq \bar{x}_G \leq \bar{x}_H$

b) Spridningsmått

Anger hur mycket en storhet skiljer sig åt

i) (Stickprovs) varians

Används ibland pga önskvärda egenskaper (bokens words not mine)

$$s^2 = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})^2$$

ii) Standardavvikelse

Kvadratroten av variansen

$$s = \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})^2}$$

iii) Kvartiler (medianuppdelning)

Beräkna undre kvartilen $\tilde{x}_{undre}$ och övre kvartilen $\tilde{x}_{övre}$ och ta medianen ur resp. kvartil

Kvartilavstånd: $\tilde{x}_{övre} - \tilde{x}_{undre}$

Kvartilsintervall: $(\tilde{x}_{undre}, \tilde{x}_{övre})$

iv) Variationsbredd

Används vid små datamängder, där variationsbredden R är differansen mellan det största och minsta värdet, x_{max} resp. x_{min}

$$R = x_{max} - x_{min}$$

Variationsintervallet: (x_{min}, x_{max})

v) Kvadratsumman kring $\bar{x}_A$

$$Q = (n-1)s^2 = \sum_{j=1}^{n} (x_j - \bar{x})^2 = \sum x_j^2 - \frac{1}{n}\left(\sum x_j\right)^2$$

vi) Standardavvikelse i % av $\bar{x}$

Variationskoefficienten $100 \cdot s/\bar{x}$ %

Läges- och spridningsmått fortsättning

För grupperade data kan man behandla den som för ogrupperade data, dock är det mer praktiskt att använda informationen från en frekvenstabell.

a) Lägesmått: medelvärde

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i y_i$$

där
- k = antal grupper
- f_i = absolut frekvens
- y_i = data av viss storlek

b) Spridningsmått: kvadratsumma

$$Q = \sum_{i=1}^{k} f_i (y_i - \bar{x})^2 = \sum f_i y_i^2 - \frac{1}{n}\left(\sum f_i y_i\right)^2$$

För klassindelade data approximeras denna till en grupperad datamängd där varje värde x_i ersätts med klassmitten y_i i den klass värdet hamnat i.

Korrelation

Undersöker sambandet mellan parade data och anger ett mått på sambandet. De vanligaste måttsambanden är:

a) Kovarians
Mellan x- och y-värden i en datamängd är kovariansen:

$$c_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

b) Korrelationskoefficienten
Då s_x och s_y är standardavvikelsen för x- resp. y-värde är korrelationskoefficienten:

$$r = \frac{c_{xy}}{s_x s_y}$$

Dessa kan endast definieras för numerisk data!

Def (utfall): Resultatet av ett slumpmässigt försök kallas ett utfall $(\omega_1, \omega_2, \ldots)$.

Def (utfallsrum): Mängden av alla möjliga utfall kallas utfallsrummet (Ω)

Def (händelse): En händelse är en specifik samling av utfall $(A, B, C, \ldots)$

Def (diskret/ändligt/kontinuerligt utfallsrum): Om antalet utfall är ändligt eller uppräkneligt oändligt sägs Ω vara ett diskret utfallsrum. Om antalet är ändligt sägs Ω vara ett ändligt utfallsrum. Om antalet är varken eller sägs Ω vara ett kontinuerligt utfallsrum.

Notation $(\cup/\cap)$: Minst en av händelserna inträffar:
$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \ldots \cup A_n$$

Alla händelserna inträffar:
$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \ldots \cap A_n$$

Def (parvis oförenliga, disjunkta): Händelserna $A_1, \ldots, A_n$ kan sägas vara parvis oförenliga om alla par A_i och A_j är disjunkta (= oförenliga), dvs det är omöjligt att händelserna inträffar samtidigt

Def (komplement): Komplementet till en händelse A är $A^* = \Omega \setminus A$, kan också betecknas $\complement A$.

→

boken
SF1922
20 03 18

Sats (De Morgans lagar): Låt A, B och C vara händelser. Då gäller följande likheter:

(i) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
(ii) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
(iii) $\left(\bigcup_{i=1}^{n} A_i\right)^* = \bigcap_{i=1}^{n} A_i^*$
(iv) $\left(\bigcap_{i=1}^{n} A_i\right)^* = \bigcup_{i=1}^{n} A_i^*$

Def (sannolikhet): För en händelse A finns det ett tal $P(A)$ kallat sannolikheten för A. $P(A)$ väljs s.a. den relativa frekvensen vid ett stort antal försök hamnar nära $P(A)$. Typ konvergerar. $P(\cdot)$ kallas sannolikhetsmåttet.

Def (frekvenstolkning): För $P(A) = a$ där $a \in [0, 1]$ blir frekvenstolkningen: vid ett stort antal försök blir den relativa frekvensen av A nog ungefär lika med a.

Sats (Kolmogrovs axiomsystem): För $P(\cdot)$ ska följande axiom vara uppfyllda:

(i) För varje händelse A gäller att $0 \leq P(A) \leq 1$
(ii) För hela utfallsrummet Ω gäller att $P(\Omega) = 1$
(iii) Additionsformeln: Om $A_1, A_2, \ldots$ är en diskret följd av parvis oförenliga händelser gäller att $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$

Sats (komplementsatsen): För komplementet A^* till A gäller att $P(A^*) = 1 - P(A)$

Sats (additionssatsen för två händelser): För två händelser A och B gäller att $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Sats (Booles olikhet): För två händelser A och B gäller att $P(A \cup B) \leq P(A) + P(B)$

Def (likformigt sannorlikhetsmått): Om $P(\omega_i) = \frac{1}{m}$ då $i = 1, \ldots, m$ föreligger ett likformigt sannolikhetsm

Sats (den klassiska sannolikhetsdefinitionen): Vid likformigt sannolikhetsmått är sannolikheten för en viss händelse lika med kvoten mellan antalet gynsamma fall g för händelsen och antalet möjliga fall m; $\therefore P(A) = g/m$.

Sats (multiplikationsprincipen): Om åtgärd 1 kan utföras på a_1 olika sätt och åtgärd 2 på a_2 olika sätt finns det $a_1 a_2$ olika sätt att utföra båda på.

Sats (dragning med återlämning): Dragning med återlämning av k element ur n med hänsyn till ordning kan ske på n^k olika sätt.

Sats (dragning utan återlämning): Dragning utan återlämning av k element ur n med hänsyn till ordning kan ske på $n(n-1) \cdot \ldots \cdot (n-k-1)$ olika sätt.

$P(A) = \frac{n!}{(n-k)!\, n^k}$

Antalet sätt n element kan ordnas på är lika med $n(n-1) \cdot \ldots \cdot 2 \cdot 1 = n!$, kallat n-fakultet →

boken
SF1922
20 03 18

Sats (dragning utan återläggning): Dragning utan återlämning av k element ur n utan hänsyn till ordning kan ske på $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ olika sätt.

Sats (biominaltheoremet): För varje positivt heltal n och godtyckliga tal x och y gäller att

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Uppdatering: mängdlära

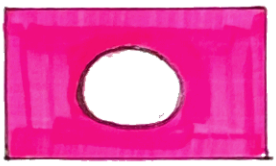
Utfallsrummet

Grundmängden

Händelsen A: A inträffar

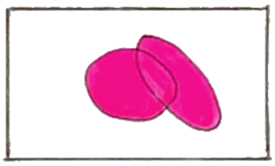
Delmängden A

Komplementära händelsen till A: A inträffar inte



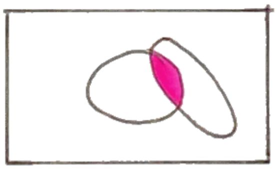
Komplementet A* till A

Unionhändelse A∪B: A eller B eller båda inträffar



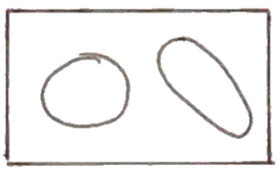
Unionen $A \cup B$

Snitthändelsen A∩B: både A och B inträffar



Snittet $A \cap B$

A och B oförenliga händelser: A och B kan inte inträffa samtidigt



A och B disjunkta
