

732G01/732G40
Grundläggande statistik
(7.5hp)

Grundläggande statistik, 7.5 hp

Mål:

- Kursens mål är att den studerande ska tillägna sig en översikt över centrala begrepp och betraktelsesätt inom statistik.

Innehåll:

- Beskrivande mått
- Sannolikhetslära
- Slumpvariabler (stokastiska variabler)
- Statistiska fördelningar
- Konfidensintervall
- Hypotesprövning

Grundläggande statistik, 7.5 hp

Lärare

- Annika Tillander
- E-post: annika.tillander@liu.se
- Rum: 3E:475 (B-huset, plan 2 ovanför café JAVA)

Litteratur

- Wahlin, K. *Tillämpad statistik - en grundkurs*. Andra upplagan. Bonnier utbildning. ISBN: 9789152327821

Kursupplägg 732G01/732G40

- **Föreläsningar:** teori presenteras (9 st)
- **Lektioner:** läraren räknar uppgifter på tavlan (4 st)
- **Räknestugor:** studenterna löser uppgifter i boken (4 st)
- **Datorövningar:** övningar i praktiska problem på dator (3 st)

Examination

- **Skriftlig tentamen** (värd 6 hp) kan ge betygen VG, G eller U
- **Inlämningsuppgift** (värd 1.5 hp) kan ge betygen G eller U
- Slutbetyg på kursen är detsamma som betyget på tentamen (givet att inlämningsuppgiften är godkänd)

Skriftlig tentamen

- 4 timmar
- Tillåtna hjälpmedel:
 - Dubbelsidigt handskrivet A4 med anteckningar
 - Formel- och tabellsamling (delas ut med tentan)
 - Miniräknare av valfri modell

Inlämningsuppgift

- Görs i grupper om 1-2 studenter
- Arbetet sker i grupper om högst två personer och genomförs **självständigt** inom varje grupp - samarbete mellan grupper är alltså **inte** tillåtet.
- Uppgiften ska lämnas in senast **kl. 23.59 onsdag den 13 september** och kommer bedömas som godkänd eller underkänd.
- Grupper på gränsen till godkänt erbjudas möjlighet till komplettering. Kompletteringen måste lämnas inom en vecka efter att den första versionen rättats och därefter görs en slutgiltig bedömning om godkännande.
- En slutgiltigt underkänd uppgift eller uppgift inlämnad efter deadline kan inte bli godkänd, utan uppgiften måste genomföras nästa gång kursen ges.

Kurs-hemsida

- [LISAM](#)
- Kan också googla 732G01 för generell information.

FÖRELÄSNING 1

Kapitel 2 – sid 11-46

Populationer, stickprov och variabler

Population

- Den samling enheter (exempelvis individer) som vi vill dra slutsatser om
- Populationen definieras på logisk väg med utgångspunkt från den frågeställning vi vill besvara
- Exempel:
 - Studerande vid LiU
 - Röstberättigande i Sverige
- Antalet enheter i populationen betecknas med N

Ändlig och oändlig population

- En ändlig population har ett begränsat antal enheter
 - Ex: Studenter i ett klassrum
- I en oändlig population kan man observera oändligt många enheter
 - Ex: Upprepade tärningskast
 - Leder ofta till enklare uträkningar
- Population brukar (approximativt) betraktas som oändlig om urvalet utgör mindre än 10 procent av populationsstorleken

Stickprov eller urval

- En del av enheterna ur populationen, oftast slumpmässigt utvalda
- Stickprovet ska vara **representativt** för populationen
- Antalet enheter i stickprovet betecknas med n

Variabler

Resultatet av upprepade mätningar eller observationer av ett fenomen

- **Kvalitativa** variabler:
 - Icke-numeriska mätningar
 - Ex: Nationalitet
- **Kvantitativa** variabler:
 - Numeriska mätningar
 - Ex: Ålder

Variabler

- Kvantitativa variabler kan delas upp i:
 - **Diskreta:** Antar bara heltalsvärden
 - Ex: Antal anställda på ett företag
 - **Kontinuerliga:** Decimalers noggrannhet
 - Ex: Längd

Variabler

- En variabel betecknas med X (stort)
- De värden som observeras för variabeln betecknas med x_1, x_2, \dots (små)

Nominalskala

- Förekommer bland kvalitativa variabler
- Grupper som **inte** kan rangordnas!
- Ex: Favoritfärg

Ordinalskala

- Förekommer bland både kvalitativa och kvantitativa variabler
- Grupper som **kan** rangordnas!
- Ex: Betyg

Metrisk skala

- Förekommer bland kvantitativa variabler
- Har numeriska värden (siffror)
- Ex: Daglig försäljning i en butik

Variablers fördelning

- En fördelning är en *sammanställning över vilka värden variabeln kan anta och hur ofta*
- Presenteras ofta med diagram
- Olika angreppssätt används för att beskriva
 - Kvalitativa variabler
 - Kvantitativa diskreta variabler
 - Kvantitativa kontinuerliga variabler

Exempel

- Företagshälsovården vid ett företag sänder ut en enkät där de anställda bland annat får svara på frågan

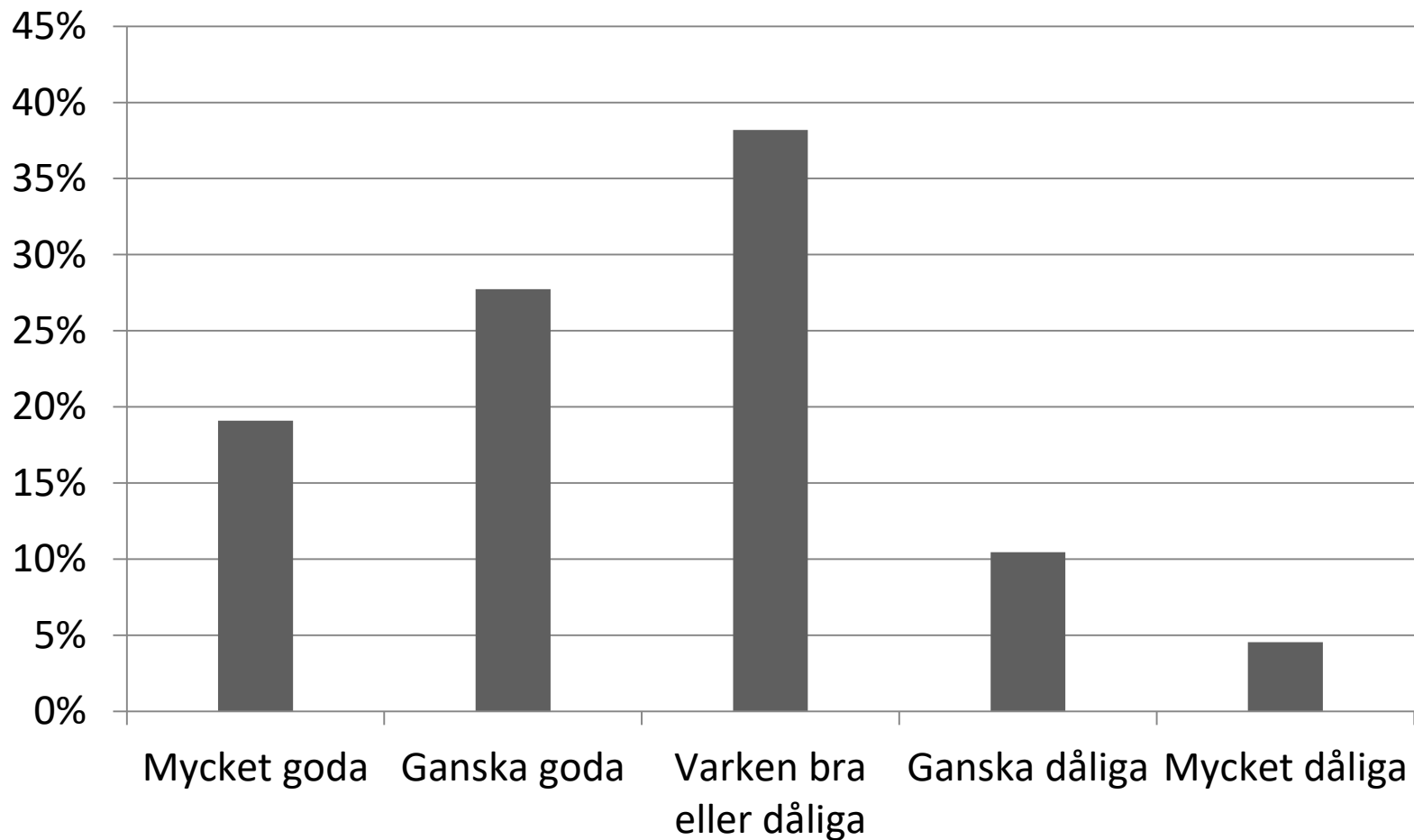
Hur bedömer Du Din närmaste chefs ledaregenskaper?

() Mycket goda () Ganska goda () Varken bra eller dåliga
() Ganska dåliga () Mycket dåliga

- Resultaten sammanställs i följande frekvenstabell
- Vilken sorts variabel är detta?

Åsikt (x)	Antal (f)
Mycket goda	42
Ganska goda	61
Varken bra eller dåliga	84
Ganska dåliga	23
Mycket dåliga	10
Totalt	220

Kvalitativ variabel: Stapeldiagram



Exempel

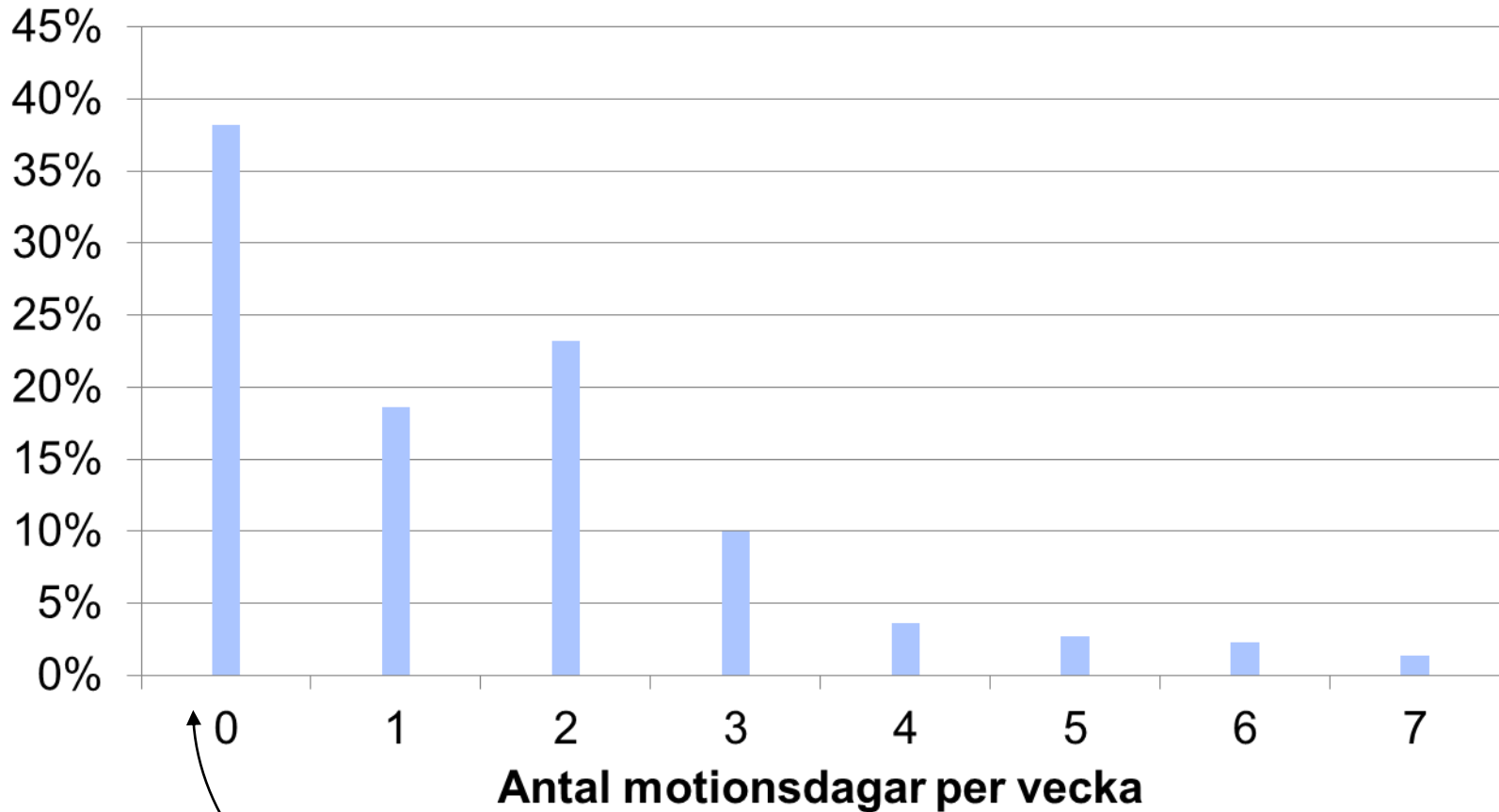
- En annan fråga på enkäten löd

Hur många dagar i veckan motionerar Du?
 Ingen 1 2 3 4 5 6 7

- Resultaten sammanställs enligt
- Vilken sorts variabel är detta?

Antal dagar (x)	Antal (f)	Andel (%)
0	84	38
1	41	19
2	51	23
3	22	10
4	8	4
5	6	3
6	5	2
7	3	1
Totalt	220	100%

Diskret kvantitativ: Stolpdiagram



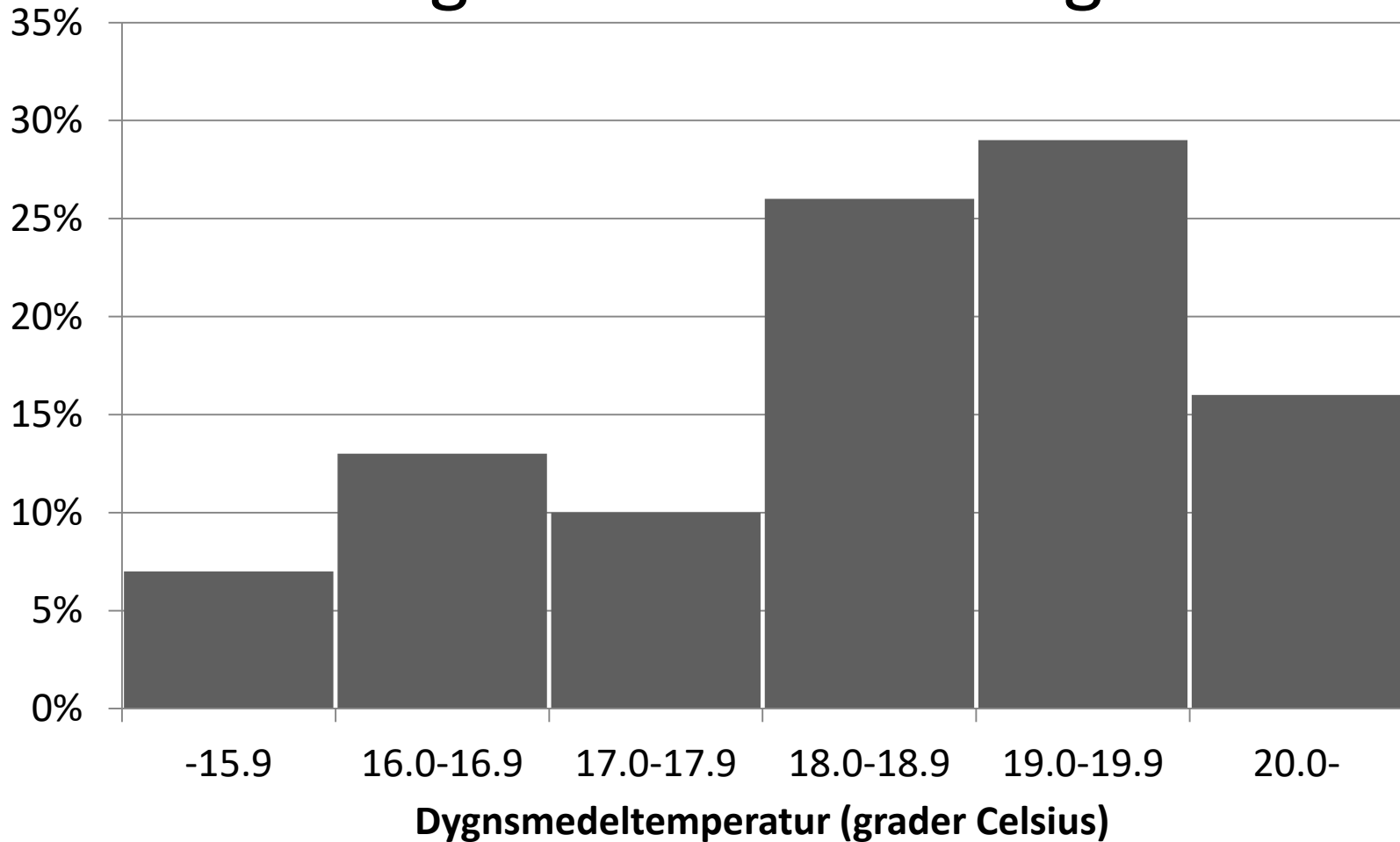
Exempel

- Dygnstemperaturen (grader Celsius) i centrala Linköping under juli månad 2015

Dag	1	2	3	4	5	6	7
Temp	20.9	20.7	19.1	16.6	18.7	19.8	19.1
Dag	8	9	10	11	12	13	14
Temp	19.2	18.6	18.4	17.3	17.8	16.0	14.7
Dag	15	16	17	18	19	20	21
Temp	16.1	16.7	18.2	15.6	18.7	19.0	18.6
Dag	22	23	24	25	26	27	28
Temp	19.7	20.1	17.0	19.1	18.4	18.4	20.8
Dag	29	30	31				
Temp	20.1	19.0	19.9				

- Variabelsort?

Kontinuerlig kvantitativ: Histogram



Beskrivande mått

- Stickprovsmedelvärde

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

- Populationsmedelvärde

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

- Ex: Längden på fem slumpmässigt utvalda personer ur en population: 165, 188, 159, 170, 198

$$\bar{x} = \frac{1}{5} \sum_{i=1}^5 x_i = \frac{1}{5} \cdot (165 + 188 + 159 + 170 + 198) = 176 \text{ cm}$$

Beskrivande mått - Frekvenstabell

- Stickprov

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

- Population

$$\mu = \frac{\sum_{i=1}^g f_i \cdot x_i}{N}$$

där g är antalet klasser i tabellen

Antal dagar (x)	Antal (f)	Andel (%)
0	84	38
1	41	19
2	51	23
3	22	10
4	8	4
5	6	3
6	5	2
7	3	1
Totalt	220	100%

Beskrivande mått

- Stickprovsstandardavvikelse (beräkningsformel)

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$$

- Population

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N}}$$

Beskrivande mått - Frekvenstabell

- Stickprov

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^g f_i \cdot (x_i - \bar{x})^2} = \sqrt{\frac{\sum_{i=1}^g f_i \cdot x_i^2 - \frac{(\sum_{i=1}^g f_i \cdot x_i)^2}{n}}{n-1}}$$

- Population

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^g f_i \cdot (x_i - \mu)^2} = \sqrt{\frac{\sum_{i=1}^g f_i \cdot x_i^2 - \frac{(\sum_{i=1}^g f_i \cdot x_i)^2}{N}}{N}}$$

Beskrivande mått - Andel

- Stickprovsandel

$$p = \frac{\text{antal enheter i stickprovet med studerad egenskap}}{n}$$

- Population

$$\pi = \frac{\text{antal enheter i populationen med studerad egenskap}}{N}$$

- Ex: Bland 550 anställda i ett stickprov uppgav 187 att de röker. Vilken andel röker bland urvalet?

$$p = \frac{187}{550} = 0.34 = 34\%$$

Beskrivande mått - Median

- Medianen är alltid det mittersta värdet i ett storleksordnat material
 - Om n (eller N) är udda: mittersta värdet
 - Om n (eller N) är jämnt: medelvärdet av de två mittersta värdena

- Ex: Längder på fem personer

159, 165, 170, 188, 198

Medianen = 170 *cm*

- Ex: Vikten av fyra personer

53, 62, 70, 85

Medianen = $\frac{62+70}{2} = 66\text{kg}$

Beskrivande mått - Kvartiler

- Första kvartilen (q_1) är mitten av första halvan av materialet
- Andra kvartilen (q_2) är medianen
- Tredje kvartilen (q_3) är mitten av andra halvan av materialet

Beskrivande mått - Typvärde

- Det vanligaste förekommande värdet i en fördelning
- Ex: Vid en tentamen har studenterna följande betyg
U, U, G, G, G, VG, VG

Typvärdet = G

När bör vi använda de olika måtten?

Kvalitativ variabel	Kvantitativ variabel
Typvärde	Median
Median	Kvartiler
Kvartiler	Medelvärde
Andelar	Standardavvikelse
	Andelar

Standardvägning

- Ibland är medelvärden missvisande
- Ex: Ett fackförbund önskar jämföra medellönen vid två företag inom samma verksamhetsområde. Följande information samlas in.

	Bolag A		Bolag B	
Befattning	Antal personer	Medellön (tkr)	Antal personer	Medellön (tkr)
Mellanchef/chef	6	36.6	15	34.5
Tjänstemän	77	20.4	34	19.8
Administrativ personal	89	17.2	21	17.1

- Jämför lönerna i de två bolagen

Exempel – Vägda medeltal

$$\bar{x}_A = \frac{6 \cdot 36.6 + 77 \cdot 20.4 + 89 \cdot 17.2}{6 + 77 + 89} = \frac{3321.2}{172} \approx 19.309$$

$$\bar{x}_B = \frac{15 \cdot 34.5 + 34 \cdot 19.8 + 21 \cdot 17.1}{15 + 34 + 21} = \frac{1549.8}{70} \approx 22.140$$

Företag B verkar ha en högre medellön. **MEN** det verkar som att det finns fler (större andel) chefer bland företag B.

Hur skulle medellönen se ut om andelen som jobbar i de olika befattningarna är densamma i de två företagen?

Exempel - Standardvägning

	Bolag A	Bolag B	
Befattning	Medellön (tkr)	Medellön (tkr)	Standard- population
Mellanchef/chef	36.6	34.5	21
Tjänstemän	20.4	19.8	111
Administrativ personal	17.2	17.1	110

$$\bar{x}_A = \frac{21 \cdot 36.6 + 111 \cdot 20.4 + 110 \cdot 17.2}{242} = \frac{4925}{242} \approx 20.351$$

$$\bar{x}_B = \frac{21 \cdot 34.5 + 111 \cdot 19.8 + 110 \cdot 17.1}{242} = \frac{4803.3}{242} \approx 19.848$$

Om befattningsfördelningen vore densamma, har bolag A högst medellön.