

## Uppgift 1

### Vikt

*Vikt* är en variabel på kvotskalan. Det gör att vi kan räkna med aritmetiskt medelvärde ( $m$ ) som centralmått (Djurefeldt, 2003:59). Medelvärdet är 35,85 kg. Det saknas värden för två observationer, vilket sannolikt påverkar medelvärdet, men inte i speciellt stor grad. Om vi fått två vikter på minimivärdet på 26 kg skulle vi ha fått ett medelvärde på 35,76. Med tanke på standardavvikelsen på 5,45 kg kan denna skillnad på 0,09 kg anses försumbar. Med två vikter på maxvärdet 49 kg skulle medelvärdet påverkas med 0,14 kg uppåt till 35,99. Samma tolkning här, i stort sett en försumbar påverkan.

**Descriptive Statistics**

|                    | N   | Minimum | Maximum | Mean  | Std. Deviation |
|--------------------|-----|---------|---------|-------|----------------|
| vikt i kg          | 198 | 26      | 49      | 35,85 | 5,450          |
| Valid N (listwise) | 198 |         |         |       |                |

Med en kvotvariabel som *vikt* kan vi också använda median som centralmått (Djurefeldt 2003:59) och med hjälp av SPSS finner vi att medianen är 35 kg. Den ligger nära det aritmetiska medelvärdet, vilket tyder på att medelvärdet visar ett rättvist värde.

**Statistics**

vikt i kg

|             |         |       |
|-------------|---------|-------|
| N           | Valid   | 198   |
|             | Missing | 2     |
| Median      |         | 35,00 |
| Percentiles | 25      | 32,00 |
|             | 50      | 35,00 |
|             | 75      | 39,00 |

### **Totalt antal kr per månad**

*Totalt antal kr per månad* är en variabel på kvotskalan. Vi använder medelvärdet (m) som är 108,46 kr per månad. Det saknas värde för två observationer\*, men medelvärdet skulle påverkas ganska lite, även av extremvärden.

**Descriptive Statistics**

|                          | N   | Minimum | Maximum | Mean   | Std. Deviation |
|--------------------------|-----|---------|---------|--------|----------------|
| total antal kr per månad | 198 | 50      | 250     | 109,01 | 39,587         |
| Valid N (listwise)       | 198 |         |         |        |                |

\*I datafilen återfinns ett värde 0 men det är för en observation där vi saknar angivelse för både vecko- eller månadspeng och angiven summa per utbetalningstillfälle. Efter justering av 0-värdet till ett saknat värde får vi sammanställningen ovan.

Medianen för *Totalt antal kr per månad* är 100 kr och pekar på att medelvärdet är lite högt. Med en blick i histogrammet i uppgift 2 (se sidan 7) ser vi att det stämmer, och att det aritmetiska medelvärdet dras upp på grund av en mängd höga observationer, men att de flesta fördelar sig kring medianvärdet.

**Statistics**

total antal kr per månad

|             |         |        |
|-------------|---------|--------|
| N           | Valid   | 198    |
|             | Missing | 2      |
| Median      |         | 100,00 |
| Percentiles | 25      | 80,00  |
|             | 50      | 100,00 |
|             | 75      | 130,00 |

### **Vem som bäddar sängen**

*Vem som bäddar sängen* är en nominalvariabel. Det går endast att använda typvärde som centralmått (Djurfeldt,2003:47). Typvärdet är *mamma* som bäddar sängen 70 gånger av de observerade 198 gångerna. Det saknas värde för två observationer, men de påverkar inte typvärdet alls. Anmärkningsvärt är att *pappa* förekommer färre gånger än *själv*, vilket jag i min egen roll som pappa finner både lustigt och skrämmande.

**vem som bäddar sängen**

|         |        | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|--------|-----------|---------|---------------|--------------------|
| Valid   | Själv  | 60        | 30,0    | 30,3          | 30,3               |
|         | Mamma  | 70        | 35,0    | 35,4          | 65,7               |
|         | Pappa  | 58        | 29,0    | 29,3          | 94,9               |
|         | Ingen  | 10        | 5,0     | 5,1           | 100,0              |
|         | Total  | 198       | 99,0    | 100,0         |                    |
| Missing | System | 2         | 1,0     |               |                    |
| Total   |        | 200       | 100,0   |               |                    |

## Uppgift 2

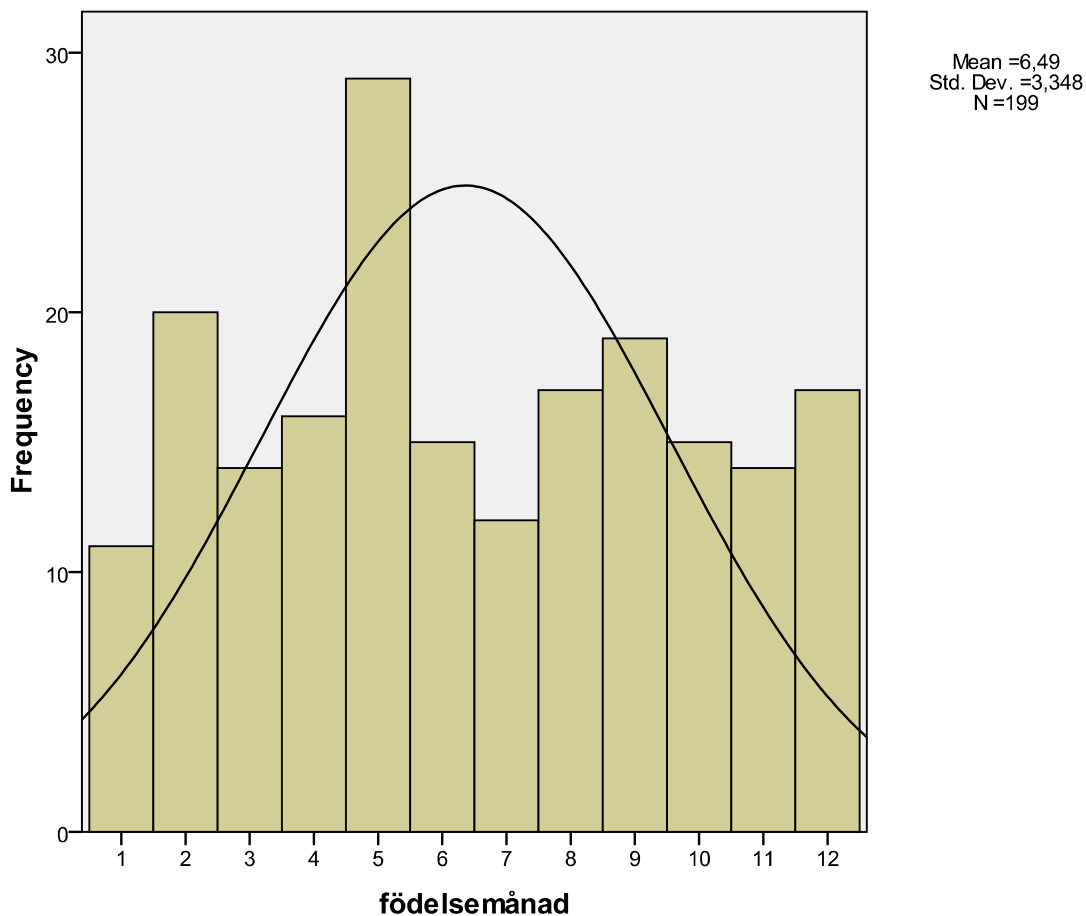
### **Födelsemånad**

*Födelsemånad* är en variabel på ordinalskalan men som har en sådan struktur att ett centralmått som median har svårt att ge någon information alls om mätningen. Typvärdet är såklart intressant när det gäller en tidsorienterad mätning. Bland alla månader är maj i detta exempel med 29 av 199 observationer, eller 14,6% av alla angivna födelsemånader den vanligaste födelsemånaden.

### **Histogram**

Men mer intressant i fallet med födelsemånad är hur utfallet fördelar sig över året. Det är exakt tolv olika värden som variabeln kan ikläda sig och de är dessutom väldigt lätta att förstå intuitivt, årets månader lär vi ju oss tidigt som barn.

Av histogrammet ser vi att födelsemånaderna är någorlunda jämnt fördelade över året. Vi har dock tre månader som sticker ut ur mätningen; januari, maj och juli.



Maj står för en stor andel medan januari och juli för en ovanligt liten del av alla födselar vilka med det sunda förnuftet borde vara jämnt fördelade över året. Om vi kikar på frekvensfördelningen ser vi mer i detalj hur utfallet fördelar sig.

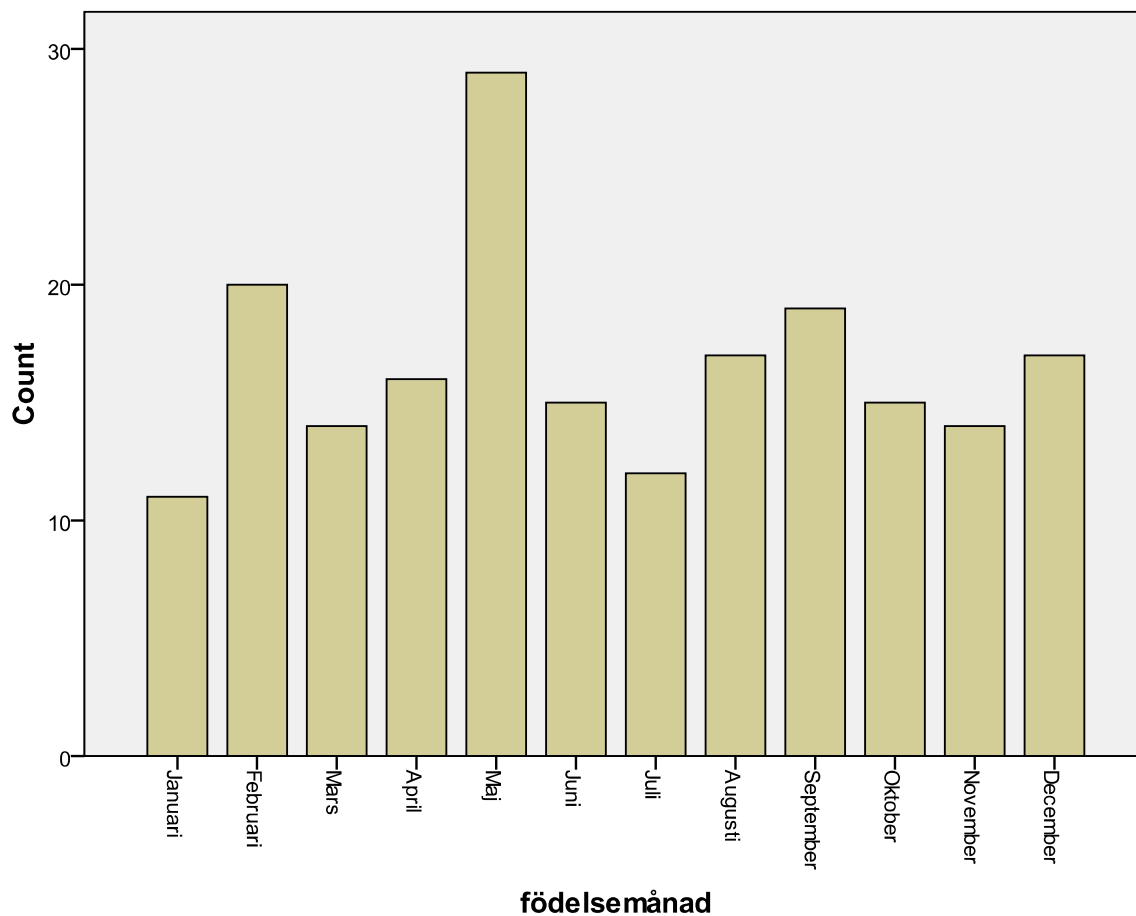
|         |           | födelsemånad |         |               |                    |
|---------|-----------|--------------|---------|---------------|--------------------|
|         |           | Frequency    | Percent | Valid Percent | Cumulative Percent |
| Valid   | Januari   | 11           | 5,5     | 5,5           | 5,5                |
|         | Februari  | 20           | 10,0    | 10,1          | 15,6               |
|         | Mars      | 14           | 7,0     | 7,0           | 22,6               |
|         | April     | 16           | 8,0     | 8,0           | 30,7               |
|         | Maj       | 29           | 14,5    | 14,6          | 45,2               |
|         | Juni      | 15           | 7,5     | 7,5           | 52,8               |
|         | Juli      | 12           | 6,0     | 6,0           | 58,8               |
|         | Augusti   | 17           | 8,5     | 8,5           | 67,3               |
|         | September | 19           | 9,5     | 9,5           | 76,9               |
|         | Oktober   | 15           | 7,5     | 7,5           | 84,4               |
|         | November  | 14           | 7,0     | 7,0           | 91,5               |
|         | December  | 17           | 8,5     | 8,5           | 100,0              |
|         | Total     | 199          | 99,5    | 100,0         |                    |
| Missing | System    | 1            | ,5      |               |                    |
| Total   |           | 200          | 100,0   |               |                    |

Då ser vi att Maj har 14,6% av alla giltiga värden för födelsemånad. Det sunda förnuftet säger att ett normalutfall borde ligga kring 8,25% för alla månader. Nu är det säkert inte helt jämnt fördelat, men det är nog viktigt att kontrollera sannolikheten i detta utfall. På samma sätt är Januari med 5,5% och juli med 6,0% av alla barn i underkant och bör ställas under en extra analys av utfallet.

### Stapeldiagram

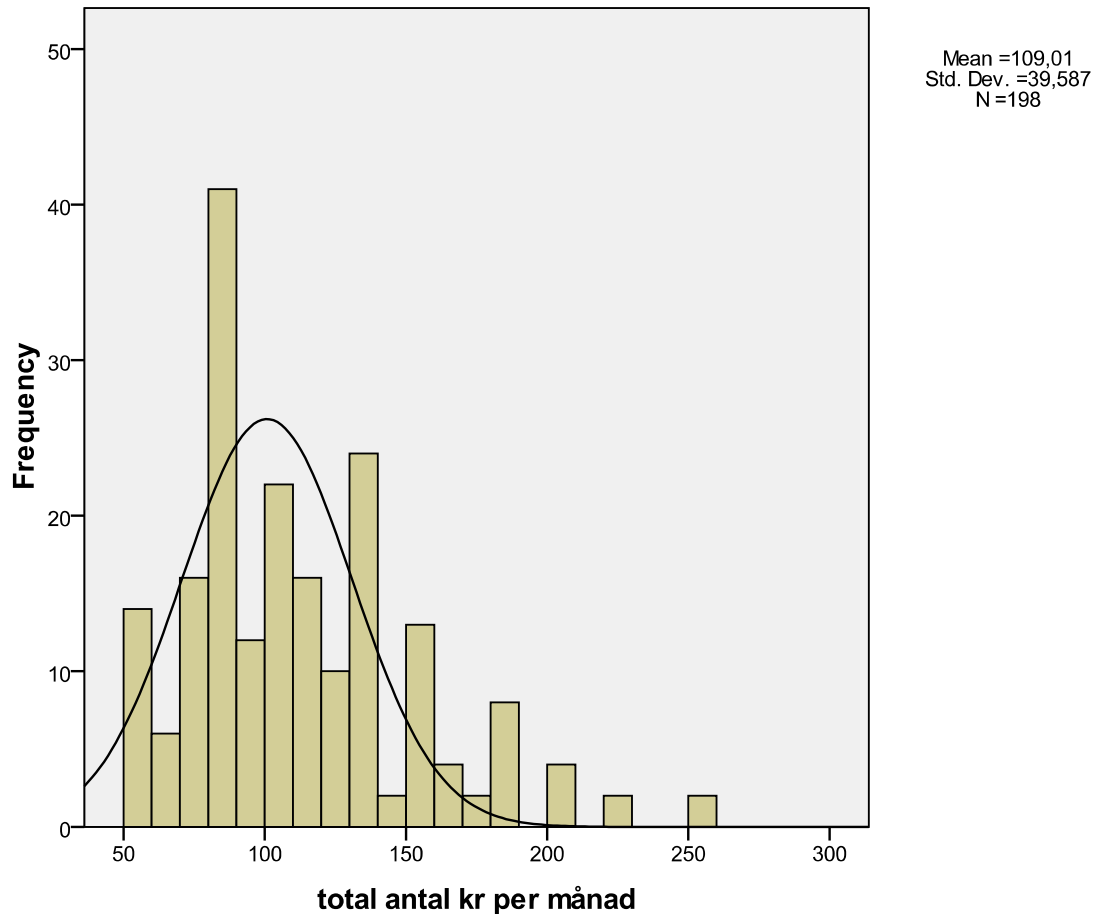
Med tanke på att *födelsemånad* är en ordinalskalevariabel är histogram ett något missvisande sätt att beskriva materialet. Ett stolpdiagram (Djurfeldt,2003:51), som tydligare visar att

variabeln har ett antal diskreta värden att anta, är mer rättvist. Det visar SPSS också genom att automatiskt presentera varje stolpe med månadsnamnet.



### **Totalt antal kr per månad**

*Totalt antal kr per månad* är en variabel på kvotskalan och passar bra att visa med histogram. Vi ser att den stora mängden observationer återfinns kring medianen som är 100 kr vilket analysen i uppgift 1 visar (se sidan 2). Histogrammet visar tydligt hur förskjuten fördelningen är mellan olika månadspengar. Det är några värden som sticker iväg och frågan är hur giltiga de allra högsta observationerna är. Finns det andra behov som styr registreringen av värden över 200 kronor, eller finns det brister i omformningen av veckopeng till månadspeng? Det är frågor som väcks när histogrammet studeras.



### Uppgift 3

Variabeln *Antal besökta länder* har inget bortfall så vi har alltså 200 mätvärden. De fördelar sig från 0 besökta länder till 7 besökta länder. Fördelningen är ojämn. Variabeln hör till kvotskalan men ändå är typvärdet ett av de starkaste måtten i en snabb analys.

|         | Frequency | Percent |
|---------|-----------|---------|
| Valid 0 | 16        | 8,0     |
| 1       | 69        | 34,5    |
| 2       | 44        | 22,0    |
| 3       | 32        | 16,0    |
| 4       | 19        | 9,5     |
| 5       | 12        | 6,0     |
| 6       | 4         | 2,0     |
| 7       | 4         | 2,0     |
| Total   | 200       | 100,0   |

#### Centralmått

Typvärdet, dvs ”1 resa” har modalvärdet 34,5% av alla observationer vilket är mer än 1 av 3. Om vi gör en grov förenkling av vår modell och låter resultatet beskrivas genom värdena 0 aldrig, 1-2 resor, 3 resor eller fler, ser vi att värdet 1-2 resor står för 56,5% av alla svar.

#### Kumulativ analys

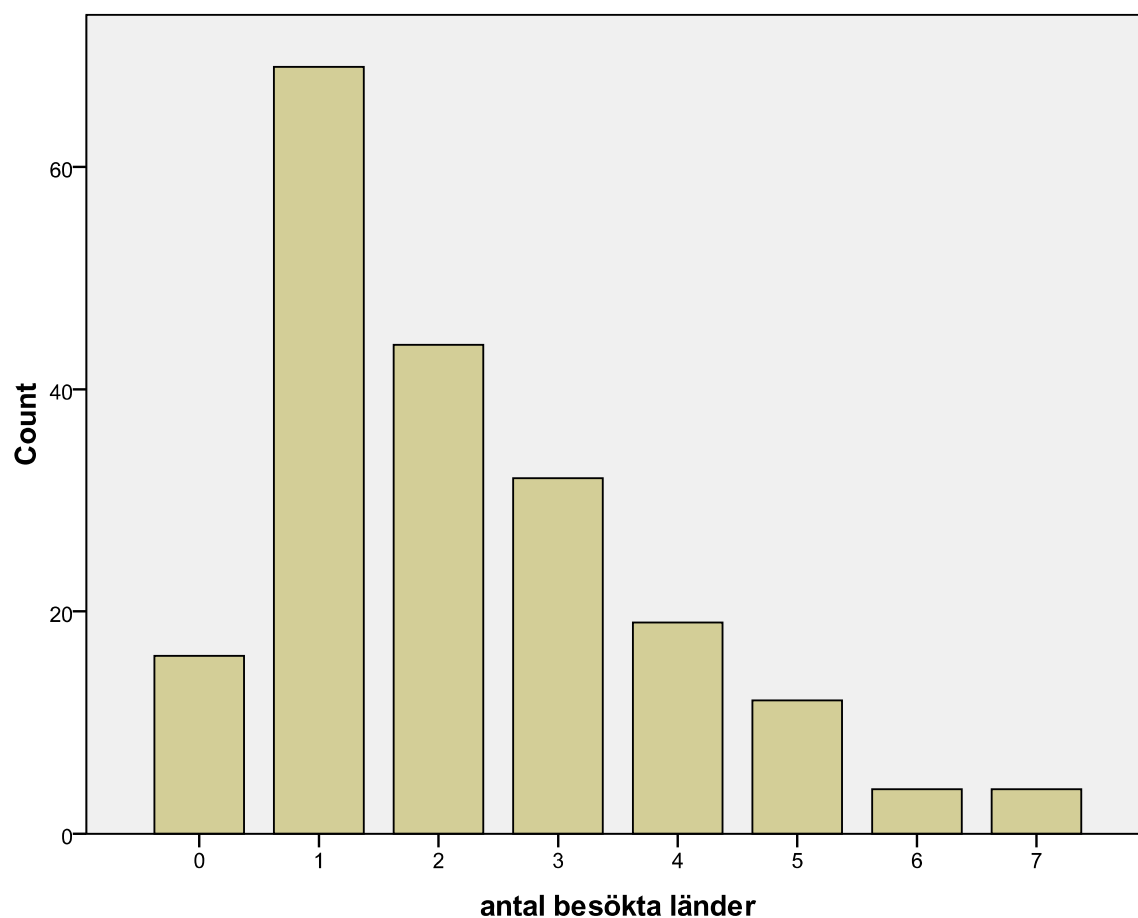
Om vi gör en analys av den kumulativa fördelningen kan vi se ytterligare en rätt kraftig signal. Det visar sig att 92% av alla elever har någon gång besökt ett annat land.



**antal besökta länder**

|         | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|-----------|---------|---------------|--------------------|
| Valid 7 | 4         | 2,0     | 2,0           | 2,0                |
| 6       | 4         | 2,0     | 2,0           | 4,0                |
| 5       | 12        | 6,0     | 6,0           | 10,0               |
| 4       | 19        | 9,5     | 9,5           | 19,5               |
| 3       | 32        | 16,0    | 16,0          | 35,5               |
| 2       | 44        | 22,0    | 22,0          | 57,5               |
| 1       | 69        | 34,5    | 34,5          | 92,0               |
| 0       | 16        | 8,0     | 8,0           | 100,0              |
| Total   | 200       | 100,0   | 100,0         |                    |

Det är vid en första anblick en väldigt hög siffra, och nästan var femte barn har varit i fyra eller fler länder. Vi kan fortsätta och se på spridningen av länderna.



Vi ser det vi redan visste, att 1 och 2 besök i andra länder dominerar. Spridningen upp till maxvärdet sker med färre och färre observationer. Om vi minns diagrammet över månadspeng (se histogram uppgift 1 på sidan 7) som liknar spridningen av resor till andra länder, väcks en hypotes om att variablerna har ett samband.

---

Djurfeldt Göran, Larsson Rolf, Stjärnhagen Ola (2003). *Statistisk verktygslåda*.  
Studentlitteratur