

LUFTD2 - TFMS -- 5016 - 1 - 54 - 90.

STATISTISK ANALYS AV  
RADONDATA FRÅN BOSTÄDER

av

Bengt Johansson

Examensarbete på  
linjen för teknisk fysik

1990:E2



**UNIVERSITY OF LUND AND  
LUND INSTITUTE OF TECHNOLOGY  
DEPARTMENT OF MATHEMATICAL STATISTICS**

**LUND, SWEDEN**

CODEN: LUFTD2/(TFMS-5016)/1-54/(1990) ,

Lunds Tekniska högskola  
Inst för Matematisk Statistik  
Box 118  
221 00 Lund



Correct:

LUFTD2-TFMS--5016-

1-54-90.

See Refer:

STATISTISK ANALYS AV  
RADONDATA FRAN BOSTADER

av

Bengt Johansson

Examensarbete på  
linjen för teknisk fysik

1990:E2

mars 1990

<b>TYP AV DOKUMENT</b> <input type="checkbox"/> Doktorsavhandling <input checked="" type="checkbox"/> Examensarbete <input type="checkbox"/> Konpendium	<input type="checkbox"/> Ansökan <input type="checkbox"/> Reserapport <input type="checkbox"/> Delrapport <input type="checkbox"/> Slutrapport	<input type="checkbox"/> Tidskriftsartikel <input type="checkbox"/> Konferensuppsats <input type="checkbox"/> .....	<b>DOKUMENTBETECKNING/CODEN</b> CODEN:LUFTD 2/(TFMS-5016)/1-54/ (1990)
--	---	---	--

**AVDELNING/INSTITUTION** Institutionen för matematisk statistik  
 Box 118, 221 00 Lund

**FÖRFATTARE** Bengt Johansson

**DOKUMENTTITEL OCH UNDERTITEL** Statistisk analys av radondata från bostäder

**SAMMANFATTNING**

Detta examensarbete innehåller en statistisk analys av radondata från bostäder, som insamlats vid fysiska institutionen vid Lunds universitet. Med hjälp av variansanalys har undersökts på vilket sätt den uppmätta radonhalten beror av faktorer såsom rökning, mättidpunkt, olika placering av detektor i huset samt den geografiska mätplatsens inverkan. Förutom klassificeringen i risknivåer, visade sig mättidpunkten ha stor betydelse. Den statistiska analysen har utförts med hjälp av programpaketet SAS. Förutom diskussionen om olika faktorerers inverkan ges också förslag på hur mätningar i framtiden skall läggas upp.

**NYCKELORD** Radon, rökning, variansanalys, residualanalys, SAS,

**DOKUMENTTITEL OCH UNDERTITEL - SVENSK ÖVERSÄTTNING AV UTLÄNDSK ORIGINALTITEL**

**TILLÄMPNINGSOMRÅDE**

**TILLÄMPNINGSNYCKELORD**

<b>UTGIVNINGSDATUM</b> år 90   mån 03	<b>ANTAL SID</b> 50	<b>SPRÅK</b> <input checked="" type="checkbox"/> svenska <input type="checkbox"/> engelska <input type="checkbox"/> annat
--	---------------------	--

<b>ÖVRIGA BIBLIOGRAFISKA UPPGIFTER</b>	<b>ISSN</b>
	<b>ISBN</b>
	<b>PRIS</b>

I, the undersigned, being the copyright owner of the abstract, hereby grant to all reference source permission to publish and disseminate the abstract.

Date: 13/3 1990  
 Signature: Bengt Johansson

## **Förord**

**Denna skrift är resultatet av ett examensarbete i matematisk statistik på linjen för teknisk-fysik vid Lunds tekniska högskola. Den statistiska analysen har utförts på material insamlat på fysiska institutionen vid Lunds tekniska högskola.**

**Jag vill härmed tacka mina handledare professor Georg Lindgren vid institutionen för matematisk statistik, och docent Gilbert Jönsson vid fysiska institutionen, som båda på alla sätt varit mig behjälpliga. I detta tack inbegripes alla övriga, som på olika sätt hjälpt mig, ingen nämnd, ingen glömd.**

**Lund den 11/3 1990**

**Bengt Johansson**

<b>1....Inledning .....</b>	<b>1</b>
1.1..Introduktion .....	1
1.2..Radon - en fysikalisk beskrivning.....	1
1.3..Källor för radon i bostäder.....	1
1.4..Cancerrisk .....	2
1.5..Gränsvärden.....	3
<b>2....Mätning.....</b>	<b>4</b>
2.1..Metod.....	4
2.2..Mätningarnas uppläggning.....	5
<b>3....Statistisk analysmetod .....</b>	<b>6</b>
3.1..Introduktion .....	6
3.2..Variansanalys med balanserade data .....	6
3.3..Variansanalys med obalanserade data .....	12
3.4..Behandling av obalanserade data i programpaketet SAS.....	12
3.5..Radondatas utseende och anpassning till ovanstående analys.....	14
<b>4....Resultat - en enkel modell.....</b>	<b>15</b>
4.1 Introduktion.....	15
4.2..Hässleholm .....	16
4.3..Lund.....	17
4.4..Malmö.....	21
4.5..Nynäshamn.....	24
4.6..Simrishamn.....	28
4.7..Sjöbo .....	32
4.8..Trelleborg.....	36
4.9..Ystad .....	37
4.10.Totala materialet .....	40
<b>5....Resultat - modell med samspelseffekter .....</b>	<b>45</b>
5.1..Introduktion .....	45
5.2..Analys av det totala materialet.....	46
<b>6....Modellernas riktighet .....</b>	<b>47</b>
<b>7....Slutsatser.....</b>	<b>49</b>
7.1..Sammanfattning av resultatet .....	49
7.2..Förslag till förbättringar av mätmetoden .....	50
<b>Litteratur .....</b>	<b>54</b>

# 1 Inledning

## 1.1 Introduktion

Radonet har under senare år uppmärksammats relativt mycket på grund av det samband man tror finns mellan lungcancer och hög radonhalt i omgivningen. På fysiska institutionen vid Lunds tekniska högskola har man i samarbete med olika miljö- och hälsoskydds-förvaltningar gjort en serie mätningar av radonhalten i bostäder i ett försök att kartlägga radonsituationen. En noggrannare statistisk analys av dessa har hitintills saknats, vilket detta arbete delvis skall försöka råda bot på. Likaså är målet att undersöka om det finns ett bättre sätt att lägga upp mätningarna på.

## 1.2 Radon - en fysikalisk beskrivning

Radon är en ädelgas som förekommer naturligt i naturen. Den är radioaktiv och ett flertal isotoper finns, vilka ingår i olika sönderfallsserier. Den för radonhalten inomhus viktigaste är  $^{222}\text{Rn}$  (radon) vilken återfinns i sönderfallsserien för uran (Tabell 1.1). Denna isotop har nämligen en längre halveringstid än övriga isotoper och detta möjliggör för gasen att diffundera ut såväl från byggnadsmaterial som från marken.  $^{222}\text{Rn}$  faller i sin tur, med avgivande av  $\alpha$ -strålning, sönder till  $^{218}\text{Bi}$  (vismut), vilken är den första av de så kallade radondöttrarna. De övriga är  $^{214}\text{Pb}$  (bly),  $^{214}\text{Bi}$  och  $^{214}\text{Po}$  (polonium). Radon-döttrarna är fasta partiklar med relativt kort livstid. De fäster gärna på väggar och aerosol-partiklar. För radonmätningarna är det också av intresse att de fäster på ytan av de detektorer som används vid radonundersökningar.

## 1.3 Källor för radon i bostäder

Det finns tre viktiga källor som kan orsaka en hög radonhalt inomhus. Byggnadsmaterial i form av blå lättbetong kan ofta innehålla uranhaltig alunskiffer. Typiska värden på aktiviteten härrörande från  $^{226}\text{Ra}$  (radium) är 1500-2500 Bq/kg byggnadsmaterial (Bq=bequerel, d v s sönderfall/s). Kranvatten kan också vara en viktig källa för radon. Detta gäller om vattnet kommer från områden med uranhaltig berggrund eller jordart. Den tredje källan är marken under huset, som kan innehålla uran. I Sverige kan alunskiffer och vissa sorters granit vara uranhaltig.

För radonhalten i ett hus är inte bara tillförseln av radon viktig. Om man har god ventilation försvinner en stor del av radonet ut, och radonhalten blir därför högre i alltför välisolerade hus och vid alltför sällsynt vädring. Mängden radon som frigörs ur ursprungsmaterialet påverkas av temperatur, tryck och luftfuktighet. Detta medför att radonhalten är olika vid olika årstider, men även sådana faktorer som ökad luftning vid varm väderlek påverkar mängden radon i huset.

*Tabell 1.1 Urans sönderfallsserie och data för den. Streck och pil innebär övergång via andra ämnen*

Isotop	Halveringstid	Huvudsaklig strålning	Energi
$^{238}\text{U}$	$4.5 \times 10^9$ år	alfa	4.20 MeV
↓			
$^{226}\text{Ra}$	1602 år	alfa	4.78 M
↓			
$^{222}\text{Rn}$	3.82 dygn	alfa	5.49 MeV
↓			
$^{218}\text{Po}$	3.05 min	alfa	6.0 MeV
↓			
$^{214}\text{Pb}$	26.8 min	beta	
↓			
$^{214}\text{Bi}$	19.7 min	beta	
↓			
$^{214}\text{Po}$	$164 \times 10^{-6}$ s	alfa	3.69 MeV
↓			
$^{206}\text{Pb}$	stabil		

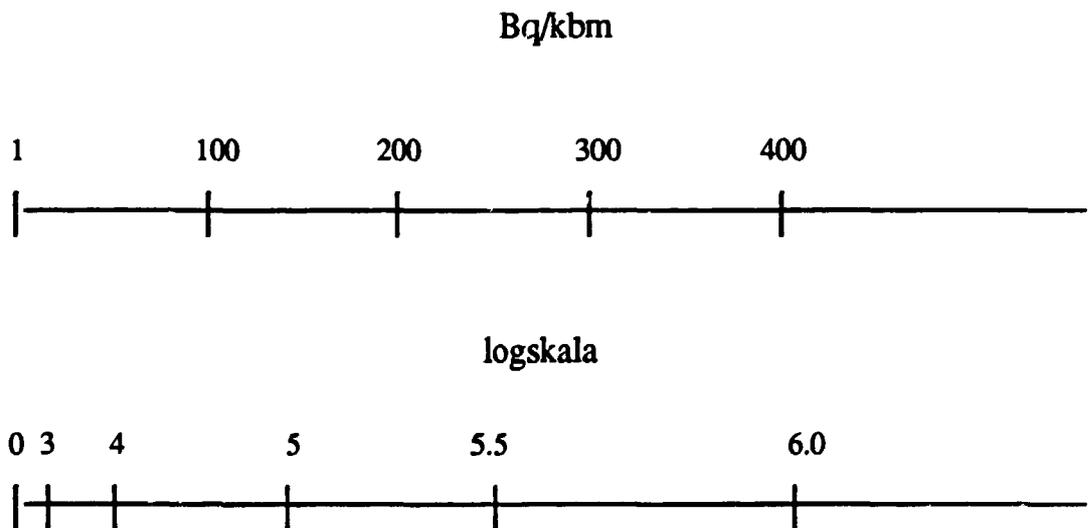
## 1.4 Cancerrisk

Ett flertal undersökningar har utförts för att undersöka om något samband råder mellan lungcancer och hög radonhalt i bostäderna. Dessa undersökningar har indikerat att ett samband finns (Edling, Kling, Axelson 1984). Dessutom tycks det även finnas en samverkan mellan rökning och hög radonhalt, som ökar risken för lungcancer (Pershagen, Damber, Falk 1984). Detta är en långsam process och vid en radondotterhalt på 400 Bq/kbm luft är tidsskalan för utveckling av lungcancer ungefär 25 år.

## 1.5 Gränsvärden

Ett antal gränsvärden för tillåten radondotterhalt finns. 400 Bq/m<sup>3</sup> är den gräns som kräver att ett hus måste åtgärdas. Efter åtgärdandet får radondotterhalten ej överstiga 200 Bq/m<sup>3</sup>. Vid nybyggnad av hus tillåts högst 70 Bq/m<sup>3</sup>. Denna gräns kan vara svår att hålla åtminstone på platser med hög markradonhalt. Var tionde bostad beräknas ha en radondotterhalt som är högre än 100 Bq/m<sup>3</sup>. Ovanstående gränsvärden har valts med hänsyn såväl till medicinska som ekonomiska intressen och hänsyn har också tagits till mättekniska möjligheter. Det finns (dec 1989) långt framskridna planer på att sänka åtgärdsgränsen till 200 Bq/m<sup>3</sup> och gränsen efter åtgärd till 70 Bq/m<sup>3</sup>. Längre fram i skriften kommer logaritmerade halter att användas vid beräkningarna och det kan därför vara av intresse att placera in dessa gränsvärden på en logskala detta görs i Figur 1.1.

*Figur 1.1 Sambandet mellan radondotterhalt och de logaritmerade värdena av densamma*



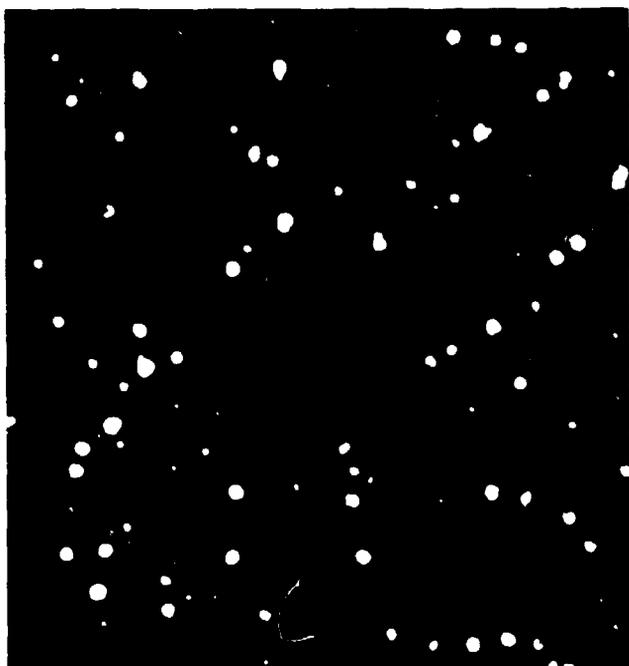


## 2 Mätning

### 2.1 Metod

Vid mätningarna har använts en spårfilm, Kodak LR115, som har exponerats för alfastrålning härrörande såväl från radon som radondöttrarna. När  $\alpha$ -partiklarna passerar genom filmen uppstår en latent skada i filmens cellulosanitrat-skikt. Skadans storlek beror på partikelns energi. Efter exponeringen etsas filmen varefter antalet hål i filmen räknas. Antalet hål kan sedan med hjälp av en kalibreringsskala omräknas till  $\text{Bq}/\text{m}^3$ . Filmen är känslig för partiklar inom energiområdet 1.7-4.1 MeV, vilket bl a medför att de radondöttrar som fäst på filmytan inte kommer att registreras på grund av för hög energi hos alfapartiklarna vid passage. Osäkerheten i mätningarna beroende på de kemiska processerna och räkningen av hålen kan uppskattas till ungefär 10 %. Dessutom tillkommer osäkerhet från själva exponeringen av filmen, elektrostatiska effekter och tillverkningsfel på filmen.

*FIG 2.1 Bilden visar plastfilmen fylld av hål orsakade av alfapartiklar*



## 2.2 Mätningarnas upplägning

För analysen har använts resultat från 813 olika filmer. Dessa har placerats ut av bostadsinnehavaren antingen på initiativ av respektive Miljö- och hälsoskyddsförvaltningar eller på eget initiativ och egen bekostnad. De mätningar som Miljö- och hälsoskyddsförvaltningarna initierat har fr a gällt hus som man misstänkt skulle ha höga radonhalter. Mätningar har gjorts i följande kommuner: Hässleholm, Lund, Malmö, Nynäshamn, Simrishamn, Sjöbo, Trelleborg och Ystad. Dessutom finns mätningar från ett stort antal andra orter. Emellertid finns bara ett fåtal mätningar från varje ort och dessa kommer framöver enbart att betecknas "övriga". (I Tabell 2.1 ges antalet mätningar samt mätningstid för respektive ort).

*Tabell 2.1 Antalet mätningar och period under vilken mätningarna utförts för respektive ort*

Kommun	Antal mätningar	Period för mätningar
Hässleholm	37	5/1 - 17/10 1989
Lund	135	4/1 1988 - 22/9 1989
Malmö	43	13/8 1987 - 1/10 1989
Nynäshamn	88	20/1 - 17/10 1988
Simrishamn	193	3/2 1988 - 12/5 1989
Sjöbo	78	24/1 - 27/8 1989
Trelleborg	42	23/3 - 24/5 1988
Ystad	78	12/8 1988 - 10/7 1989
Övriga	119	10/12 1987 - 24/9 1989

*Tabell 2.2 Klassindelning av materialet (Förutom geografisk eller tidsmässig indelning)*

Rumsplacering	Rökning	Radonkälla
0 Okänd	0 Okänt	0 okänd
1 Vardagsrum	1 Förekommer	1 Huset beläget på lågriskmark
2 Sovrum	2 Förekommer ej	2 Huset beläget på högriskmark
3 Källare	3 Förekommer ibland	3 Huset av blå lättbetong
4 Pannrum, uthus		
5 Kök		
6 Kontor, arbetsrum		
7 Övrig		

Mätningarna härrör sig från tiden från den 13/8 1987 till den 17/10 1989. Detektorerna har suttit i bostaden i en till tre månader vilket integrerar bort de största fluktuationerna i radonhalten. Bostadsinnehavaren har därvid antecknat i vilket rum mätningen har skett, om rökning förekommit, adress samt datum för uppsättande respektive nedtagande av filmen. Till denna information har lagts notering om huruvida bostaden är belägen på lågriskområde, högriskområde samt om huset är byggt av blå lättbetong. I Tabell 2.2 visas klassindelningen av materialet.

### 3 Statistisk analysmetod

#### 3.1 Introduktion

Datamaterialet innehåller som ovan angivits förutom radonhalten uppgifter om faktorer som kan tänkas ha en inverkan på denna. Det kan då vara intressant att undersöka huruvida dessa faktorer verkligen har en statistiskt säkerställd inverkan på radonhalten i bostäder. Hypoteser rörande samband mellan radonhalt och olika förklarande variabler kan prövas med hjälp av variansanalys. Därvid kan man också uppskatta storleken av de olika faktorernas effekter på radonhalten. I detta kapitel beskrivs kortfattat variansanalysens grundläggande modeller och metoder. Eftersom programpaketet SAS utnyttjas för beräkningarna beskrivs dessutom hur vissa av dess beräkningar görs.

#### 3.2 Variansanalys med balanserade data

En enkel modell för radonhaltens beroende av olika faktorer är att låta den uppmätta halten vara en summa av en konstant, oberoende av vilka förhållanden som råder, och olika additiva termer som beskriver de olika faktorernas inverkan samt deras eventuella samspel. Därtill får man lägga en term som beskriver de slumpmässiga variationer av halten som råder. En modell för fallet att man har två inverkanse faktorer A och B kan då se ut så här. Faktor A antas kunna uppträda på nivåerna  $A_1, A_2, \dots, A_p$  och faktor B på nivåerna  $B_1, B_2, \dots, B_q$ . Om  $y_{ijk}$  betecknar resultatet av den  $k$ :te mätningen på faktorkombinationen  $A_i B_j$ , så kan man skriva

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$$

där  $\mu$  är medeleffekten,  $\tau_i$  och  $\beta_j$  betecknar huvudeffekterna av faktor  $A_i$  respektive  $B_j$ ,  $(\tau\beta)_{ij}$  är effekten av samspel mellan  $A_i$  respektive  $B_j$ , medan  $\epsilon_{ijk}$  är den slumpmässiga variationen. Följande bivillkor gäller:

$$\sum_{i=1}^a \tau_i = 0$$

$$\sum_{j=1}^b \beta_j = 0$$

$$\sum_{i=1}^a (\tau\beta)_{ij} = 0$$

$$\sum_{j=1}^b (\tau\beta)_{ij} = 0$$

Här är  $y_{ijk}$  resultatet vid en viss mätning,  $\mu$  är en konstant,  $\tau_i$  och  $\beta_j$  är inverkan från faktor  $A$  respektive  $B$ ,  $(\tau\beta)_{ij}$  är inverkan från samspelet mellan faktorerna och  $\epsilon_{ijk}$  är den slumpmässiga termen.

Vid variansanalys förutsätter man vanligen att den slumpmässiga termen är normalfördelad  $N(0,\sigma)$ . Detta medför nämligen en väsentlig förenkling av analysen. Det är mycket viktigt att standardavvikelsen är den samma vid alla mätningar och vid alla faktorkombinationer.

Vid variansanalys ställer man upp hypoteser om huruvida de olika faktorerna inverkar på mätresultatet. I ovanstående exempel kan man till exempel testa följande hypoteser:

A:  $H_0: \tau_i = 0$  för alla  $i$   
mot  
 $H_1: \tau_i \neq 0$  för något  $i$

B:  $H_0: \beta_j = 0$  för alla  $j$   
mot  
 $H_1: \beta_j \neq 0$  för något  $j$

AB:  $H_0: (\tau\beta)_{ij} = 0$  för alla  $i$  och  $j$   
mot  
 $H_1: (\tau\beta)_{ij} \neq 0$  för något  $i$  och  $j$

För testen behövs två hjälpfördelningar, nämligen  $\chi^2$ -fördelningen och F-fördelningen. Dessa beskrivs kortfattat så här. En stokastisk variabel  $X$  som har täthetsfunktionen

$$f_X(x) = kx^{f/2-1}e^{-x/2} \quad x > 0$$

sägs vara  $\chi^2$ -fördelad med  $f$  frihetsgrader. Om man har  $n$  stycken oberoende variabler  $X_i$  där  $X_i$  tillhör  $N(0,1)$  gäller att

$$\sum_{i=1}^n X_i^2 \text{ tillhör } \chi^2(n) \text{ där } n \text{ är antalet frihetsgrader.}$$

Dessutom gäller att

$$\sum_{i=1}^n (X_i - \bar{X})^2 \text{ är } \chi^2(n-1) \text{ fördelat då } \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Då man ej har standardiserade normalfördelningar utan i stället  $X_i$  som tillhör  $N(m, \sigma)$  gäller istället:

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ tillhör } \chi^2(n-1).$$

En stokastisk variabel  $X$  med täthetsfunktionen

$$f_X(x) = kx^{f_1/2-1} \left(1 + \frac{f_1}{f_2}x\right)^{-(f_1+f_2)/2} \quad x > 0$$

sägs vara F-fördelad med  $f_1$  och  $f_2$  frihetsgrader.

Ett viktigt samband mellan de båda fördelningarna råder. Om  $X$  tillhör  $\chi^2(f_1)$  och  $Y$  tillhör  $\chi^2(f_2)$ , och  $X$  och  $Y$  äroberoende gäller att

$$Z = \frac{X/f_1}{Y/f_2}$$

tillhör  $F(f_1, f_2)$ .

Låt oss nu förutsätta att i försöket  $i=1, \dots, a; j=1, \dots, b; k=1, \dots, n$ . Detta innebär att man tagit  $n$  stickprov för varje faktorkombination. Inför följande beteckningar:

$$y_{i..} = \sum_{j=1}^b \sum_{k=1}^n y_{ijk}$$

$$\bar{y}_{i..} = \frac{y_{i..}}{bn}$$

$$y_{.j.} = \sum_{i=1}^a \sum_{k=1}^n y_{ijk}$$

$$\bar{y}_{.j.} = \frac{y_{.j.}}{an}$$

$$y_{ij.} = \sum_{k=1}^n y_{ijk}$$

$$\bar{y}_{ij.} = \frac{y_{ij.}}{n}$$

$$y_{...} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}$$

$$\bar{y}_{...} = \frac{y_{...}}{abn}$$

Man kan då dela upp den totala kvadratsumman

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y} \dots)^2$$

i fyra olika delsummor som uttrycker effekten av faktorerna A och B, av samspelet mellan A och B, samt den slumpmässiga variationen:

$$SS_T = bn \sum_{i=1}^a (\bar{y}_{i..} - \bar{y} \dots)^2 + an \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y} \dots)^2 + n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y} \dots)^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2$$

Detta kan symboliskt skrivas som

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$

Här är de tre första termerna ett mått på hur mycket de olika stickprovsmedelvärdena skiljer sig från varandra. Den sista termen är residualkvadratsumman, vilken är ett mått på hur mycket värdena inom ett stickprov skiljer sig från varandra. Kvadratsumman  $SS_T$  har  $abn - 1$  frihetsgrader,  $SS_A$  har  $a - 1$ ,  $SS_B$  har  $b - 1$ ,  $SS_{AB}$  har  $(a - 1)(b - 1)$  och  $SS_E$  har  $ab(n - 1)$  frihetsgrader. Vi inför också medelkvadratsummorna

$$MS_A = \frac{SS_A}{(a - 1)}, \quad MS_B = \frac{SS_B}{(b - 1)}, \quad MS_{AB} = \frac{SS_{AB}}{(a - 1)(b - 1)}, \quad MS_E = \frac{SS_E}{ab(n - 1)}$$

Nu gäller det att  $SS_E/\sigma^2$  tillhör  $\chi^2(f)$  där  $f = ab(n - 1)$ . Viktigt är också att om respektive hypotes  $H_0$  är sann gäller dessutom att  $SS_A/\sigma^2$  tillhör  $\chi^2(f)$ , där  $f = a - 1$ ,  $SS_B$  tillhör  $\chi^2(f)$  med  $f = b - 1$ , samt  $SS_{AB}$  tillhör  $\chi^2(f)$  med  $f = (a - 1)(b - 1)$ . Dessutom gäller att  $SS_E$  är oberoende av de övriga kvadratsummorna.

Detta medför att om hypoteserna är sanna är  $MS_A/MS_E$  fördelad som  $F(a - 1, ab(n - 1))$ ,  $MS_B/MS_E$  tillhör  $F(b - 1, ab(n - 1))$  samt  $MS_{AB}/MS_E$  tillhör  $F((a - 1)(b - 1), ab(n - 1))$ . För att kunna testa hypoteserna måste man känna till väntevärdena för medelkvadratsummorna. Dessa är

$$E(MS_A) = \sigma^2 + \frac{bn}{(a - 1)} \sum_{i=1}^a \tau_i^2$$

$$E(MS_B) = \sigma^2 + \frac{an}{(b - 1)} \sum_{j=1}^b \beta_j^2$$

$$E(MS_{AB}) = \sigma^2 + \frac{n}{(a-1)(b-1)} \sum_{i=1}^a \sum_{j=1}^b (\tau\beta)_{ij}^2$$

$$E(MS_E) = \sigma^2$$

Man ser att alla medelkvadratsummorna har väntevärdet  $\sigma^2$  om hypoteserna är sanna. Därför bör man förkasta hypoteserna om medelkvadratsummorna för faktorerna är tillräckligt mycket större än MSE, som ju i medeltal alltid är  $\sigma^2$ , oavsett om hypotesen är sann eller falsk. Testet utförs genom att man sätter upp teststorheter  $F_0$  enligt tabell och jämför med  $\alpha$ -kvantilen  $F_\alpha(f_1, f_2)$  ( $f_1$ =antalet frihetsgrader för faktorn som testas,  $f_2$  =antalet frihetsgrader för  $SS_E$ ). Hypotesen förkastas om  $F_0 > F_\alpha(f_1, f_2)$ . Här är  $\alpha$  signifikansnivån, d v s den sannolikhet med vilken man kommer att förkasta en korrekt hypotes.

Tabell 3.1 Exempel på variansanalystabell samt teststorheter

Faktor	Kvadratsumma	Frihetsgrader	Medelkvadrat	Teststorhet $F_0$
A	$SS_A$	$a-1$	$MS_A$	$MS_A/MS_E$
B	$SS_B$	$b-1$	$MS_B$	$MS_B/MS_E$
AB	$SS_{AB}$	$(a-1)(b-1)$	$MS_{AB}$	$MS_{AB}/MS_E$
Slumpfel	$SS_E$	$ab(n-1)$	$MS_E$	
Totalt	$SS_T$	$abn-1$		

Om man funnit att någon faktor inverkar signifikant på resultatet kan det vara intressant att veta vilken eller vilka faktornivåer som skiljer sig från de övriga. Om man bara har två faktornivåer kan man utnyttja t-testet. För detta test krävs ytterligare en hjälpfördelning nämligen t-fördelningen. Det gäller att en stokastisk variabel med täthetsfunktionen

$$k(1 + \frac{x^2}{f})^{-(f+1)/2}, \quad -\infty < x < \infty,$$

är t-fördelad med  $f$  frihetsgrader. Dessutom gäller att om  $X$  tillhör  $N(0, 1)$  och  $Y$  tillhör  $\chi^2(f)$  och  $Z = \frac{X}{\sqrt{Y/f}}$  så har  $Z$  en t-fördelning med  $f$  frihetsgrader. Detta medför att

$$\frac{\bar{y}_1 - \bar{y}_2}{d} \text{ är } t(f),$$

där  $d = s\sqrt{\frac{2}{n}}$  och  $f$  är antalet frihetsgrader för  $s$ , den skattade standardavvikelsen. På detta kan man förkasta en hypotes om att faktorerna har samma väntevärde med konfidensgraden  $\alpha$  om absolutbeloppet av ovanstående kvot blir större än  $t_{\alpha/2}(f)$ . Om det finns fler faktornivåer skulle man kunna tänka sig att utföra jämförelser på ovanstående sätt

mellan varje faktornivå. Detta kommer emellertid att ge en felaktig konfidensgrad och därför har ett flertal test tagits fram för att man skall kunna jämföra olika nivåer med en gemensam konfidensgrad (Duncan, Tukey e t c). Eftersom programpaketet SAS inte utnyttjar dessa för material med utseende som det undersökta (se kapitel 2.3) behandlas dessa inte utförligare här.

För att skatta parametrarna i ekvationen ovan används minsta kvadratmetoden d v s man minimerar uttrycket

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \mu^* - \tau_i^* - \beta_j^* - (\tau\beta)_{ij}^*)^2.$$

Detta leder till ett stort ekvationssystem som löses med hjälp av bivillkoren

$$\begin{aligned} \sum_{i=1}^a \tau_i^* &= 0 \\ \sum_{j=1}^b \beta_j^* &= 0 \\ \sum_{i=1}^a (\tau\beta)_{ij}^* &= 0 \\ \sum_{j=1}^b (\tau\beta)_{ij}^* &= 0 \end{aligned}$$

I det balanserade fallet med lika många observationer på varje faktorkombination kan lösningen anges explicit, och man får följande skattningar:

$$\begin{aligned} \mu^* &= \bar{y} \dots \\ \tau_i^* &= \bar{y}_{i..} - \bar{y} \dots \\ \beta_j^* &= \bar{y}_{.j.} - \bar{y} \dots \\ (\tau\beta)_{ij}^* &= \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y} \dots \end{aligned}$$

Dessa skattningar kan användas för att ge ett predikterat värde för en observation av en cell med A-faktorn  $A_i$  och B-faktorn  $B_j$ . Detta predikterade värde blir

$$y_{ijk}^* = \mu^* + \tau_i^* + \beta_j^* + (\tau\beta)_{ij}^*$$

För att undersökningssättet ovan skall vara möjligt krävs att vissa villkor skall vara uppfyllda. Hela ovanstående förfarande bygger på att observationerna är hämtade från en normalfördelning med konstant standardavvikelse  $\sigma$ . Man måste undersöka om detta stämmer på det undersökta materialet. Detta sker genom residualanalys. Residualerna



erhålles som skillnaden mellan observerat värde i en cell och det predikterade värdet för samma cell enligt följande formel.

$$e_{ijk} = y_{ijk} - \hat{y}_{ijk}^*$$

Normalfördelningen testas t ex genom att man för in residualerna i ett normalfördelningspapper. Dessutom plottas residualerna mot predikterade värden, mot olika faktorer etc. Härvid gäller att man anser att modellen är giltig om inget system i residualernas spridning syns. För variansanalysens giltighet är det betydligt viktigare att variansen för residualerna är oberoende av faktorkombinationer än att normalfördelningsantagandet gäller.

### 3.3 Variansanalys med obalanserade data

Vid en statistisk undersökning, som skall analyseras med variansanalys, försöker man hålla sig till det balanserade fallet, d v s man har lika många observationer i alla celler. Ofta försvinner emellertid vissa observationer, och det kan till och med finnas vissa celler som är helt tomma. Det finns ett flertal tekniker för att analysera sådana obalanserade material. Om materialet nästan är balanserat kan man föra över experimentet till ett balanserat experiment. Om det saknas några observationer i ett fåtal celler kan man tänka sig att man skattar de saknade värdena. Om det finns ett enstaka värde mer i någon cell kan man på liknande sätt utelämna detta. Emellertid blir svårigheterna stora om vissa celler saknar observationer helt och hållet och det finns stora skillnader mellan antalet observationer i övriga celler. Då måste en mera exakt metod användas. Det finns flera sådana och den som SAS använder beskrivs i följande avsnitt.

### 3.4 Behandling av obalanserade data i programpaketet SAS.

Antag att man har totalt  $m$  olika faktornivåer ( $A_1, \dots, A_p, B_1, \dots, B_q$  osv). En matris  $X$  med storlek  $m+1 \times n$  bildas, där  $n$  är antalet observationer som gjorts. För varje observation sätts en etta i kolumnen för de faktornivåer som gäller för observationen och en nolla i övriga. En kolonnvektor  $Y$  med storlek  $n \times 1$  skapas också innehållande de beroende variablerna vid varje observation. En vektor  $\beta$  införs. Den består av de parametrar som skall skattas ( $\mu, \tau_1, \dots, \tau_p$  etc.). SAS använder minsta kvadratmetoden för att skatta parametrarna. Det innebär att man vill minimera uttrycket:

$$(Y - X\beta)'(Y - X\beta)$$

Detta medför att man vill lösa följande ekvation:

$$X'X\beta = X'Y$$

Härur fås en minsta kvadratskattning av parametrarna. För varje effekt testas sedan om dess parametrar signifikant skiljer sig från 0. De kvadratsummor som används för obalanserade data motsvarar ej dem som tagits fram i kapitlet om balanserade data. Om man hade använt dessa skulle teststorheten för en effekt påverkas av när effekten införts i modellen. I stället använder man kvadratsummor som ger teststorheter som motsvarar fallet att effekten är den senast införda

I programpaketet SAS utförs för obalanserade data enbart parvisa jämförelser av effekterna. Detta medför att de konfidensgrader med vilka man särskiljer effekternas olika nivåer kan bli relativt osäkra.

Man tar fram ett marginellt medelvärde för en effekt istället för det vanliga medelvärdet. Detta marginella medelvärde kommer då att vara det förväntade medelvärdet för en effekt om man utfört ett balanserat försök. Det krävs alltså i det obalanserade fallet en korrigering av de skattade parametrarna. Att observera är att dessa medelvärden för en effekt motsvarar  $\mu + \tau_i$  i modellen i kapitel 3.3. För att få fram de marginella medelvärdena sätter man upp en matris  $Z$  efter följande premisser.

- 1) Matrisen blir av storleken  $m \times m + 1$  och kolonnerna  $a$  motsvarar de olika effekternas nivåer medan varje rad motsvarar den nivå man undersöker med avseende på medelvärde.
- 2) Betrakta alla effekter som innefattas i en undersökt effekt. Sätt alla nivåer som ingår i den undersökta nivån till 1 övriga till 0.
- 3) Betrakta den undersökta effekten. Sätt 1 på den nivå som undersöks 0 på övriga.
- 4) Betrakta de effekter som innefattar den undersökta effekten. Ge de nivåer som innehåller den undersökta nivån värdet  $1/k$  där  $k$  är antalet sådana nivåer. Ge övriga värdet 0.
- 5) Betrakta de effekter som ej undersöks. Ge nivåerna i dessa värdet  $1/j$  där  $j$  är antalet nivåer i effekten.

Ett exempel på hur man får fram en matris  $Z$  om man har två effekter med två nivåer och samverkans effekter ges nedan. Det finns i detta fall fyra samverkansnivåer. Detta ger en  $8 \times 9$  . matris enligt följande:

Undersökt faktor		A <sub>1</sub>	A <sub>2</sub>	B <sub>1</sub>	B <sub>2</sub>	A <sub>1</sub> B <sub>1</sub>	A <sub>1</sub> B <sub>2</sub>	A <sub>2</sub> B <sub>1</sub>	A <sub>2</sub> B <sub>2</sub>
A <sub>1</sub>	1	1	0	1/2	1/2	1/2	1/2	0	0
A <sub>2</sub>	1	0	1	1/2	1/2	0	0	1/2	1/2
B <sub>1</sub>	1	1/2	1/2	1	0	1/2	0	1/2	0
B <sub>2</sub>	1	1/2	1/2	0	1	0	1/2	0	1/2
A <sub>1</sub> B <sub>1</sub>	1	1	0	1	0	1	0	0	0
A <sub>1</sub> B <sub>2</sub>	1	1	0	0	1	0	1	0	0
A <sub>2</sub> B <sub>1</sub>	1	0	1	1	0	0	0	1	0
A <sub>2</sub> B <sub>2</sub>	1	0	1	0	1	0	0	0	1

En vektor  $\beta'$  med en skattning av de marginella medelvärdena erhålles ur relationen:

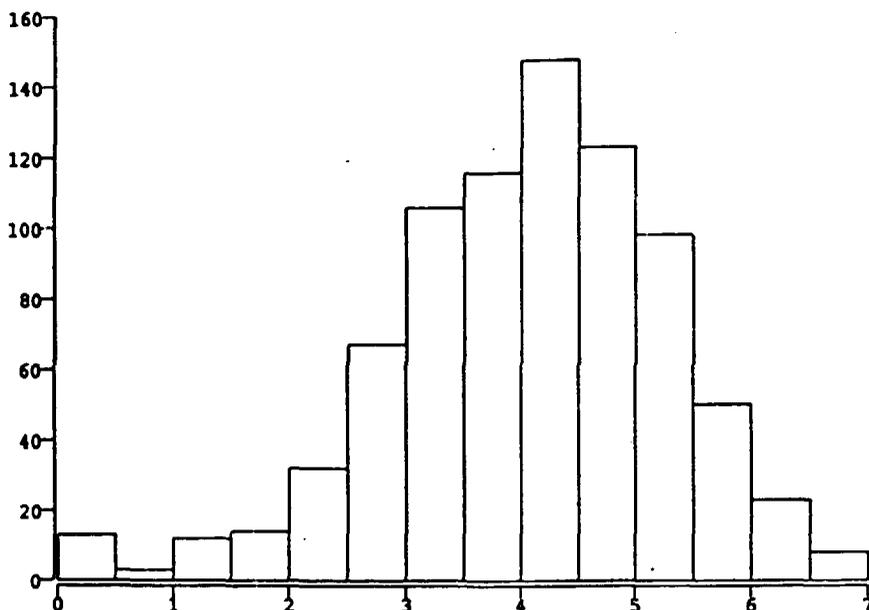
$$\beta' = Z\beta$$

### 3.5 Radondatas utseende och anpassning till ovanstående analys

Radondatas fördelning visar upp en skevhet vilket gör materialet olämpligt att direkt behandla med variansanalys. Ett vanligt sätt att angripa sådana fördelningar är att logaritmera materialet. Om observationerna  $y_{ijk}$  är lognormalfördelade erhåller man vid logaritmering  $\ln(y_{ijk})$ , som är normalfördelade. Exempel på hur logaritmerade radondata ser ut ges i Fig 3.1. Här kan man se att fördelningen är relativt lik en normalfördelning. Logaritmering tjänar också syftet att stabilisera variansen. Som nämntes i samband med residualanalysen (kap 3.2) är en noggrann överensstämmelse med normalfördelningen inte nödvändig för analysens giltighet. Bedömningen om modellens riktighet för datamaterialet behandlas utförligare i kapitel 6.

De radondata jag haft att analysera utgör ett ytterst obalanserat material. Inom vissa kommuner är t ex alla mätningar gjorda på högriskmark, i andra är alla gjorda på lågriskmark. Mätningarna är gjorda under olika månader inom olika kommuner. Inom vissa kommuner har det hänt att enbart högriskmark mätts under vissa månader medan lågriskmark mätts under andra månader. Detta medför att vissa celler blir helt utan observationer, vilket försvårar analysen och gör det omöjligt att skilja effekter av marktyp från effekten av årstiden. I detta sammanhang kan också påpekas att i de fall endast vissa faktorkombinationer finns kan man helt förhindras att särskilja olika effekters inverkan på radonhalten. Skattningarna av olika parametrar kan också påverkas.

Figur 3.1 Radondata för det totala materialet efter logaritmering



## 4 Resultat - en enkel modell

### 4.1 Introduktion

För alla orter har en enkel linjär modell testats. Detta har gjorts bl a på grund av litenheten av materialet för varje ort, samt avsaknaden av många celler, men också för att en enkel modell alltid är önskvärd. Följande bokstäver kommer i fortsättningen vara förknippade med de olika effekterna.  $\tau$  kommer att förknippas med effekten från rökning,  $\beta$  med effekten från rumsplaceringen,  $\gamma$  med effekten från klassindelningen (högriskområde, lågriskområde etc),  $\omega$  med effekten från månaden då detektorn sattes ut (observera att om endast månad nämns innebär det utsättningsmånad),  $\lambda$  är effekten av året då detektorn sattes ut,  $\alpha$  ger effekten av en ortsfaktor inom en kommun och  $\kappa$  betecknar effekten från kommunen man mätt i. I modellen för det totala materialet utnyttjar man klasseffekten och kommuneffekten, medan för den enskilda orten ortseffekterna används. Som signifikansgräns för förkastande av en effekts inverkan har valts 0.05. För de effekter som har signifikans har skattningar av det marginella medelvärdet getts.

Om man tar medelvärdet av de marginella medelvärdena för en effekt får man samma värde oavsett vilken effekt det gäller. Det beror på att detta medelvärde för alla effekter ger

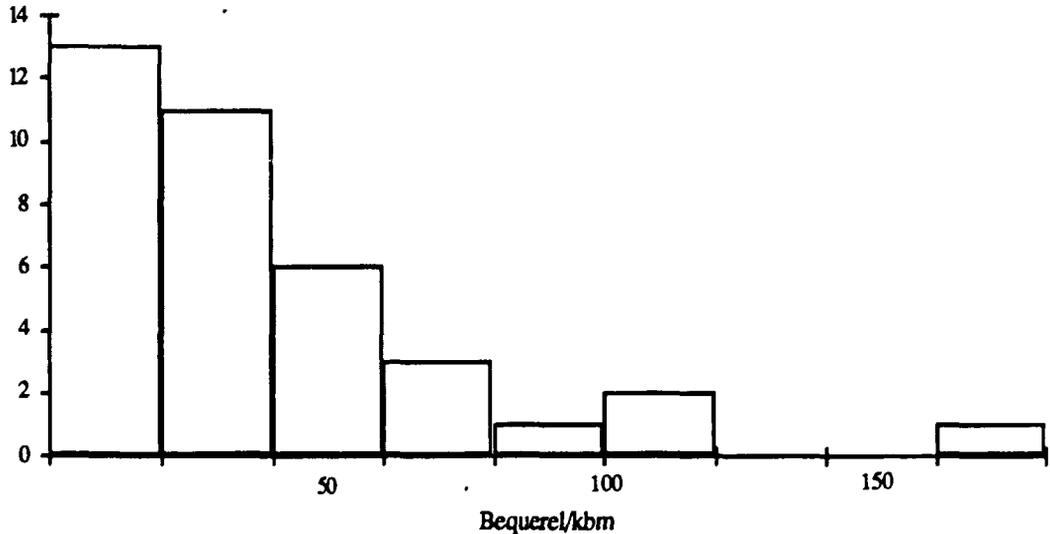
skattningen av  $\mu$ . Detta medelvärde skulle om man gjort ett balanserat försök överensstämma med medelvärdet för observationerna (D v s det medelvärde som anges i variansanalystabellerna). Då materialet är obalanserat gäller inte det längre. Om man har fler observationer från nivåer med låg radonhalt kommer skattningen av  $\mu$  bli större än det uppmätta medelvärdet. Eftersom  $\mu$  skattas lika kommer även de marginella medelvärdena för faktorer där alla observationerna är jämt fördelade på alla nivåer att öka i detta fall.

I följande avsnitt redovisas datamaterialet ort för ort, variansanalys utförs samt eventuella egenheter hos materialet diskuteras.

## 4.2 Hässleholm

Materialet från Hässleholm innehåller 37 observationer. Mätningarna är gjorda under tiden 5/1-17/10 1989 och alla mätningar är gjorda i lågriskområde. Datamaterialets utseende ges i Figur 4.1 samt Tabell 4.1.

Figur 4.1 Histogram av datamaterialet från Hässleholm



Eftersom materialet är hämtat från ett och samma år och endast utgörs av mätningar på lågnivåmark innehåller en rimlig modell rumsfaktor, rökfaktor, månadsfaktor samt ortsfaktor, d v s modellen ser ut:

$$\ln y_{ijkln} = \mu + \tau_i + \beta_j + \omega_k + \alpha_l + \varepsilon_{ijkln}$$

Denna modell testas och redovisas i Tabell 4.2.

Tabell 4.1 Datamaterialets fördelning inom olika effekter

Mätort	antal	Rum	antal	Rökning	antal	Månad	antal
Okänd	4	Vardagsr.	7	Förekommer	10	Jan	13
Hässleholm	6	Sovrum	12	Förekommer ej	27	Feb	8
Hörja	2	Källare	4			Mar	8
Vinslöv	4	Kök	3			Apr	2
Sösdala	4	Kontor	1			Maj	2
Tormestorp	4	Övrigt	10			Juni	2
Örkellunga	13					Sep	2

Tabell 4.2 Varianslystabelle, Hässleholm

$$\sqrt{MSE} = 0.917$$

MEDELVÄRDE FÖR LOGHALT 3.23

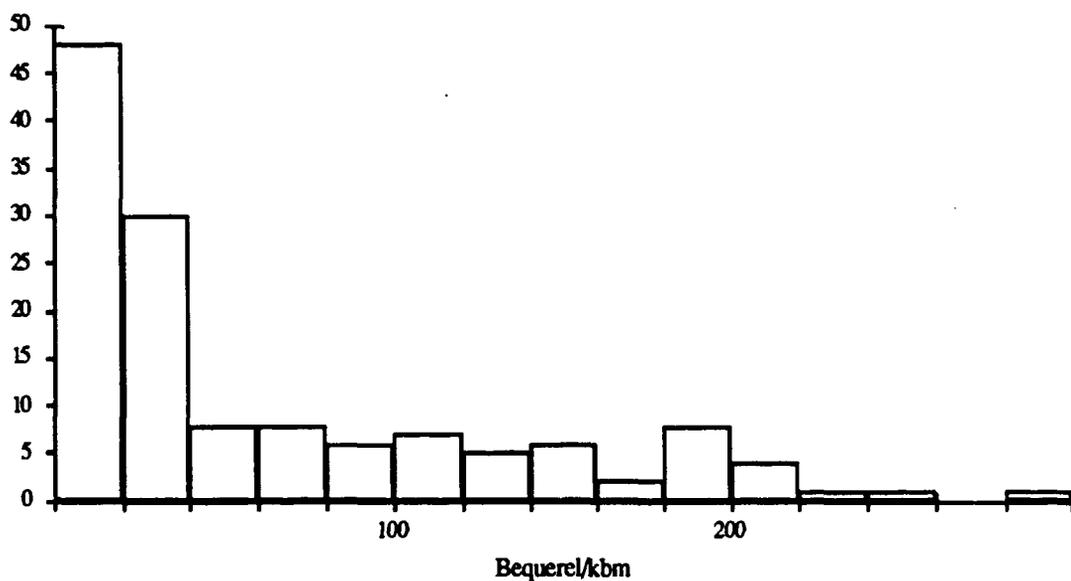
EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Ort	4	2.634	0.78	0.550
Månad	4	3.970	1.18	0.350
Rökning	1	0.182	0.22	0.647
Rum	5	3.341	0.79	0.566

Ingen effekt visar någon signifikant inverkan. Detta kan bero på att materialet är för litet för att kunna urskilja en sådan, men det kan också bero på att radonhalten är låg och ett tillskott beroende på någon faktors inverkan då blir litet.

### 4.3 Lund

Datamaterialet innehåller 135 observationer insamlade mellan 21/1 1988 och 22/9 1989. Det var 53 observationer från hus på lågriskmark, 78 från högriskmark och fyra stycken var från blå lättbetonghus. Det har ett utseende som redovisas i Figur 4.2 och Tabell 4.3.

Figur 4.2 Histogram av datamaterialet från Lund



Tabell 4.3 Datamaterialets fördelning inom olika effekter

Mätort	antal	Rum	antal	Rökning	antal	Månad	antal
Okänd	6	Vardagsr	33	Okänt	8	Jan	40
S Sandby	78	Sovrum	44	Förekommer	23	Feb	38
Hardeberga	3	Källare	8	Förekommer ej	95	Mar	21
Stångby	2	Kontor	5	Förekommer	9	Apr	8
Flyinge	2	Övrigt	29	ibland		Maj	6
Dalby	4	Okänt	16			Juni	2
Genarp	8					Sep	8
Lund	26					Nov	8
Hoby	2					Dec	4
Veberöd	2						
T. Hällestad	2						

Vi har här mätningar från olika år och det kan ge en rimlig modell innehållande rök-, rums-, Orts-, månads- och årsfaktorer:

$$\ln y_{ijklmn} = \mu + \tau_i + \beta_j + \omega_k + \lambda_l + \alpha_m + \epsilon_{ijklmn}$$

Test av modellen ges i Tabell 4.4

*Tabell 4.4 Variansanalystabell, Lund modell 1*

$\sqrt{MSE} = 1.164$  MEDELVÄRDE FÖR LOGHALT 3.39

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Ort	10	51.29	3.79	0.0002
Månad	8	35.11	3.24	0.0024
Rökning	3	22.07	5.43	0.0016
Rum	5	10.623	1.57	0.175
År	1	0.0035	0.00	0.960

Rumsfaktorn visar inte någon signifikans. Inte heller verkar årseffekten ha någon inverkan. Därför testas en ny modell, modell 2, som innehåller Orts-, rökning- samt månadseffekter. Testet redovisas i Tabell 4.5.

*Tabell 4.5 Variansanalystabell, Lund modell 2*

$\sqrt{MSE} = 1.173$  MEDELVÄRDE FÖR LOGHALT 3.39

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Ort	10	79.09	5.75	0.0001
Månad	8	41.71	3.79	0.0006
Rökning	3	26.47	6.41	0.0005

Skattningar av marginella medelvärdena i ovanstående modell ges i Tabell 4.6. I Figur 4.3 visas variationen i tiden grafiskt.

*Tabell 4.6 Skattningar av marginella medelvärden för modell 2, Lund*

Månad	medelvärde	medelfel
Januari	4.120	0.358
Februari	4.000	0.394
Mars	3.569	0.381
April	4.301	0.511
Maj	3.843	0.681
Juni	3.800	0.923
September	6.626	0.561
November	5.090	0.548
December	3.934	0.688

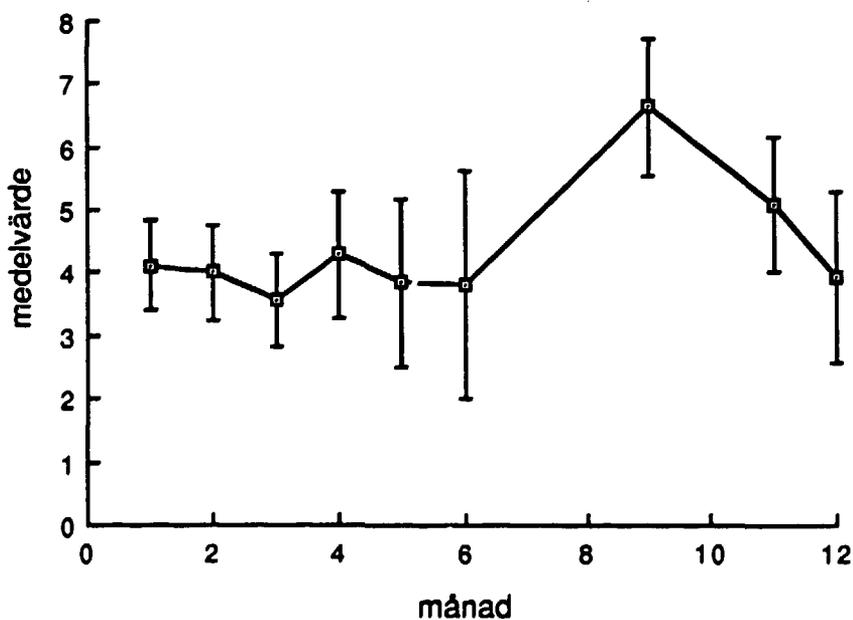


Rökning	medelvärde	medelfel
Okänt	3.844	0.556
Förekommer	4.472	0.362
Förekommer ej	3.660	0.236
Förekommer ibland	5.483	0.516

Ort	medelvärde	medelfel
Okänd	2.539	0.571
S Sandby	3.485	0.245
Hardeberga	5.382	0.792
Stångby	3.474	0.987
Flyinge	2.120	0.973
Dalby	4.358	0.740
Genarp	6.074	0.545
Lund	4.920	0.295
Hoby	5.401	0.925
Veberöd	4.281	0.925
T Hällestad	5.978	1.061

Figur 4.3 Radonhaltens medelvärde som funktion av månaden

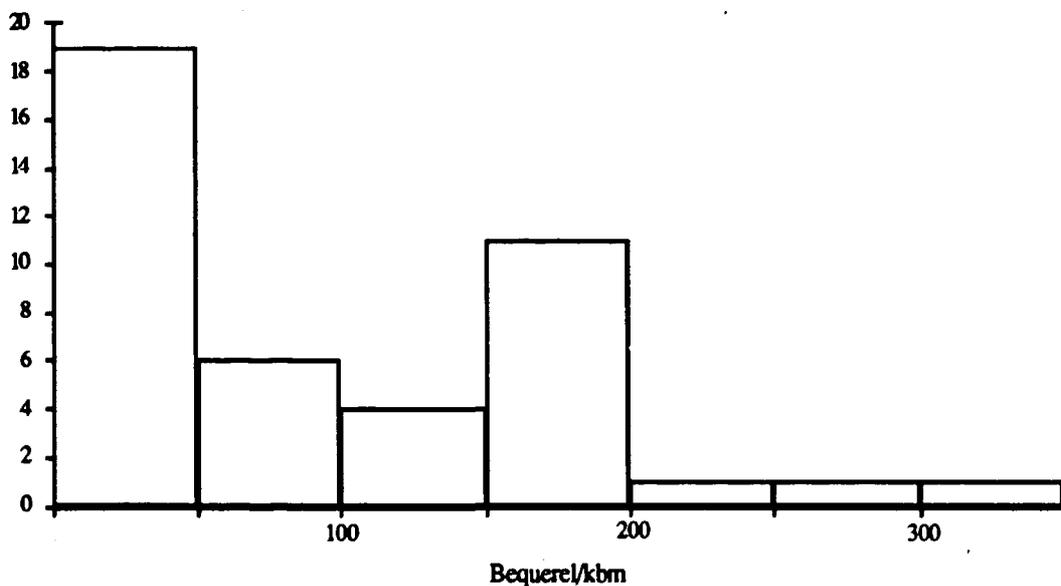


Sammanfattningsvis kan sägas att S Sandby och Flyinge verkar ha lägre radonhalt än de flesta andra. Genarp verkar å andra sidan ha högre halt. Lägg märke till att Flyinge endast har två observationer och man bör inte generalisera alltför långt utifrån dessa. De olika röningsnivåerna verkar skilja sig från varandra. Skattningen av skillnaden mellan hus där rökning förekommer och där det inte förekommer blev 0.812 med standardavvikelse på skattningen som var 0.303. Med signifikans 0.0085 skiljer sig nivåerna åt. Dessutom verkar september ha klart högre radonhalt än övriga månader.

#### 4.4 Malmö

Datamaterialet innehåller 43 observationer insamlade mellan den 13/8 1987 och den 1/10 1989. Alla mätningar gjordes på mark som klassificerats som lågnivåmark. Materialets utseende redovisas i Figur 4.4 samt Tabell 4.7.

Figur 4.4 Histogram över datamaterialet från Malmö



Eftersom vi här har mätt under olika år kan detta tänkas påverka radonhalten. En lämplig modell kan därför tänkas vara:

$$\ln y_{ijklmn} = \mu + \tau_i + \beta_j + \omega_k + \lambda_l + \alpha_m + \varepsilon_{ijklmn}$$

Test av modellen ges i Tabell 4.8

Tabell 4.7 Datamaterialets fördelning inom olika effekter

Mätort	antal	Rum	antal	Rökning	antal	Månad	antal
Okänd	2	Vardagsr	11	Förekommer	5	Jan	7
Malmö	26	Sovrum	18	Förekommer ej	34	Feb	2
Bunkeflo-		Källare	2	Ibland	4	Mar	5
strand	8	Kök	1			Apr	2
Oxie	4	Kontor	4			Maj	3
		Övrigt	7			Juni	7
						Aug	8
						Dec	7

Tabell 4.8 Variansanalystabell för Malmö

$$\sqrt{MS_E} = 0.507 \quad \text{MEDELVÄRDE FÖR LOGHALT 4.14}$$

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Ort	3	4.252	5.52	0.0056
Månad	8	8.123	3.95	0.0049
Rökning	2	1.793	3.49	0.0483
Rum	5	4.336	3.38	0.0206
År	2	4.334	8.44	0.0019

Alla effekter är således signifikanta och marginella medelvärden skattas. Dessa redovisas i Tabell 4.9. I Figur 4.5 visas variationen i tiden grafiskt.

Tabell 4.9. Skattningar av marginella medelvärden, Malmö

Månad	Medelvärde	medelfel
Januari	5.033	0.822
Februari	6.292	0.887
Mars	4.059	0.589
April	6.797	0.887
Maj	4.135	0.598
Juni	5.099	0.465
Juli	6.823	0.887
Augusti	4.917	0.420
December	6.667	0.795

Ort	Medelvärde	medelfel
Okänd	6.170	1.272
Malmö	4.673	0.320
Bunkeflostrand	4.748	0.473
Oxie	6.552	0.799

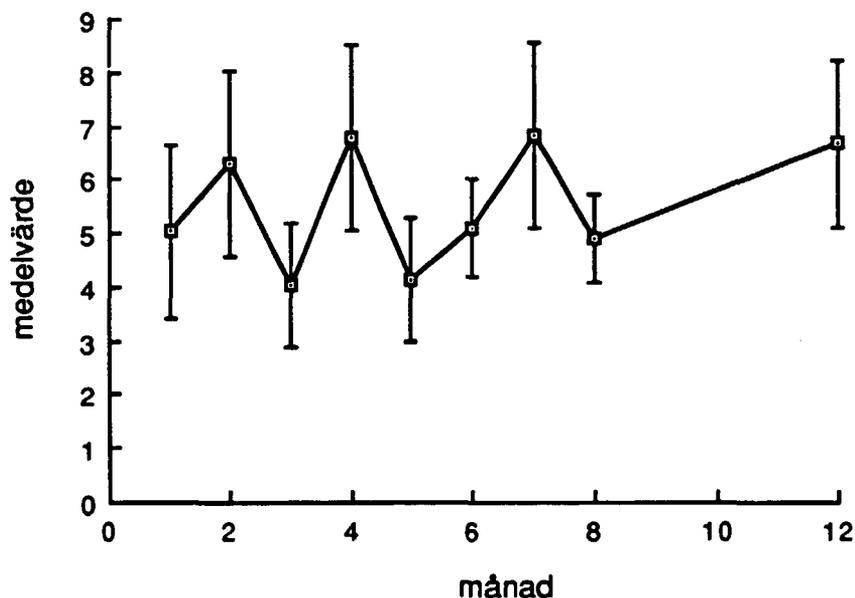
Rökning	Medelvärde	medelfel
Förekommer	5.529	0.626
Förekommer ej	4.860	0.486
Förekommer ibland	6.218	0.671

Rum	Medelvärde	medelfel
Vardagsrum	5.824	0.535
Sovrum	6.082	0.516
Källare	5.687	0.647
Kök	5.900	0.821
Kontor	4.559	0.687
Övrigt	5.160	0.546

År	Medelvärde	Standardavvikelse på medelvärdet
1987	7.046	0.860
1988	5.187	0.587
1989	4.374	0.308

Skattningen av skillnaden mellan hus där rökning förekommer och där det inte förekommer blev 0.67 med ett medelfel som var 0.593. Med signifikans 0.0321 skiljer sig nivåerna åt.

Figur 4.5 Radonhaltens marginella medelvärde som funktion av månaden



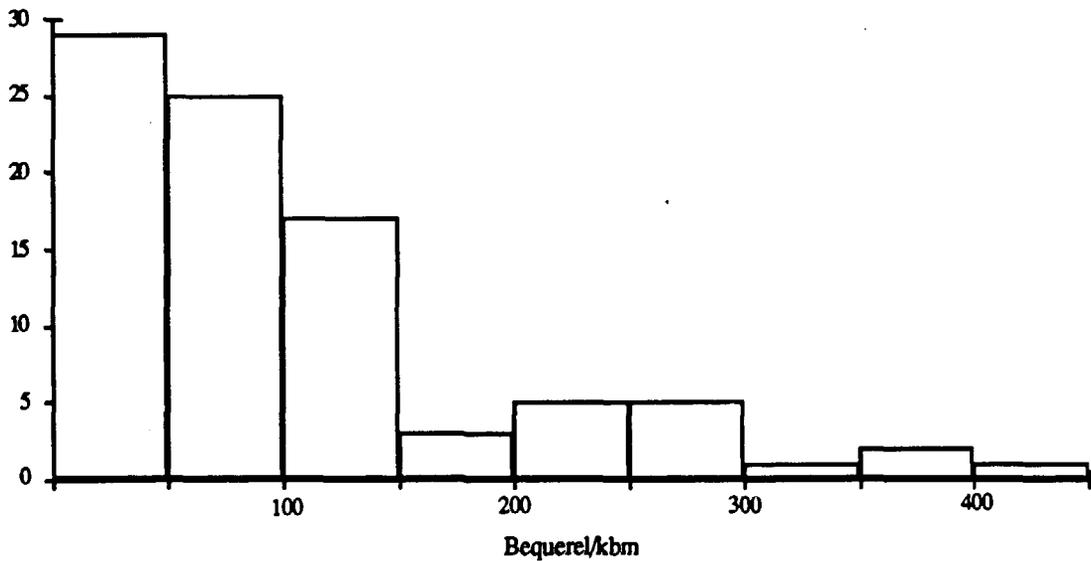
#### 4.5 Nynäshamn

Datamaterialet består av 88 observationer vilka insamlats mellan den 20/1 och den 17/10 1988. Alla observationer är tagna på högriskområde. Datamaterialets utseende visas i Figur 4.6 samt tabell 4.10.

Tabell 4.10 Datamaterialets fördelning inom olika effekter

Mätort	antal	Rum	antal	Rökning	antal	Månad	antal
Okänd	34	Vardagsr	23	Okänt	4	Jan	66
Nynäshamn	26	Sovrum	28	Förekommer	7	Feb	12
Sorunda	10	Källare	18	Förekommer ej	74	Mar	4
Ösmo	14	Pannrum	2	Ibland	3	Maj	4
St Vika	4	Kök	2			Augusti	2
Övrigt	13						
Okänt	2						

Figur 4.6 Histogram för datamaterialet från Nynäshamn



Den modell som var aktuell att utgå ifrån ser ut som nedan

$$\ln y_{ijkmn} = \mu + \tau_i + \beta_j + \omega_k + \alpha_m + \epsilon_{ijkmn}$$

Tabell för test av modellen ges i Tabell 4.11.

Tabell 4.11 Variansanalystabell för modell 1 Nynäshamn

$\sqrt{MS_E} = 1.075$  MEDELVÄRDE FÖR LOGHALT 4.17

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Ort	4	6.49	1.51	0.209
Månad	4	7.00	0.178	0.177
Rökning	3	7.96	2.47	0.0692
Rum	6	17.13	2.66	0.0223

Detta resultat tyder på att man bör testa en modell där enbart rums- och rönkningsfaktorena ingår. Detta gjordes varvid modell 2 erhöles.(Tabell 4.12).

**Tabell 4.12** *Variansanalystabell för Nynäshamn modell 2*

$\sqrt{MS_E} = 1.088$  MEDELVÄRDE FÖR LOGHALT 4.17

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Rökning	3	5.47	1.54	0.211
Rum	6	13.28	1.87	0.097

Denna modell verkade inte speciellt bra. Ingen av effekterna hade här signifikans. Däremot ger två andra modeller signifikans nämligen modell 3, som innehåller effekterna månad, rum och rökning samt modell 4 som innehåller enbart ortseffekten. Test för dessa redovisas i Tabell 4.13 respektive Tabell 4.15.

**Tabell 4.13** *Variansanalystabell för Nynäshamn modell 3*

$\sqrt{MS_E} = 1.050$  MEDELVÄRDE FÖR LOGHALT 4.17

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Rökning	3	10.96	3.31	0.025
Rum	6	20.95	3.16	0.0081
Månad	4	10.67	2.42	0.056

Skattningar av de marginella medelvärdena göres i Tabell 4.14. I Figur 4.7 visas variationen i tiden grafiskt.

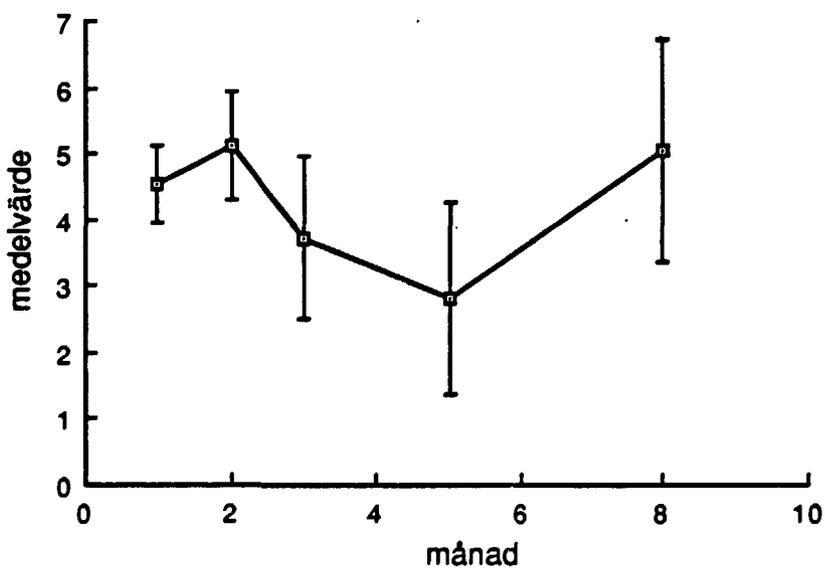
**Tabell 4.14** *Skattningar av marginella medelvärden modell 3, Nynäshamn*

Månad	Medelvärde	medelfel
Januari	4.546	0.297
Februari	5.131	0.423
Mars	3.731	0.631
Maj	2.826	0.735
Augusti	5.044	0.851
Rökning	Medelvärde	medelfel
Okänt	6.331	0.863
Förekommer	3.622	0.542
Förekommer ej	3.226	0.378
Förekommer ibland	3.842	0.726

Rum	Medelvärde	medelfel
Okänt	1.242	1.232
Vardagsrum	4.399	0.384
Sovrum	4.631	0.391
Källare	5.447	0.422
Pannrum	4.676	0.851
Kök	4.116	0.836
Övrigt	5.289	0.414

Skattningen av skillnaden mellan hus där rök förekommer och där rökning inte förekommer blev 0.397 med ett medelfel på skattningen av 0.484. Skillnaden är icke signifikant.

Figur 4.7 Radonhaltens marginella medelvärde som funktion av månaden





Tabell 4.15 Variansanalystabell för Nynäshamn modell 4

$$\sqrt{MS_E} = 1.086 \quad \text{MEDELVÄRDE FÖR LOGHALT 4.17}$$

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Ort	4	11.73	2.49	0.05

Skattningar av parametrarna ges i Tabell 4.16.

Tabell 4.16 Skattningar av marginella medelvärdena Nynäshamn

Ort	Medelvärde	medelfel
Okänd	3.786	0.186
Nynäshamn	4.263	0.213
Sorunda	4.192	0.343
Ösmo	4.754	0.290
St Vika	4.823	0.543

Båda modellerna verkar var någorlunda bra. Orsaken till att dessa modeller verkar bra kan vara det ytterst obalanserade materialet som gör det svårt att särskilja månads- och ortseffekter. Detta samband visas i Tabell 4.17.

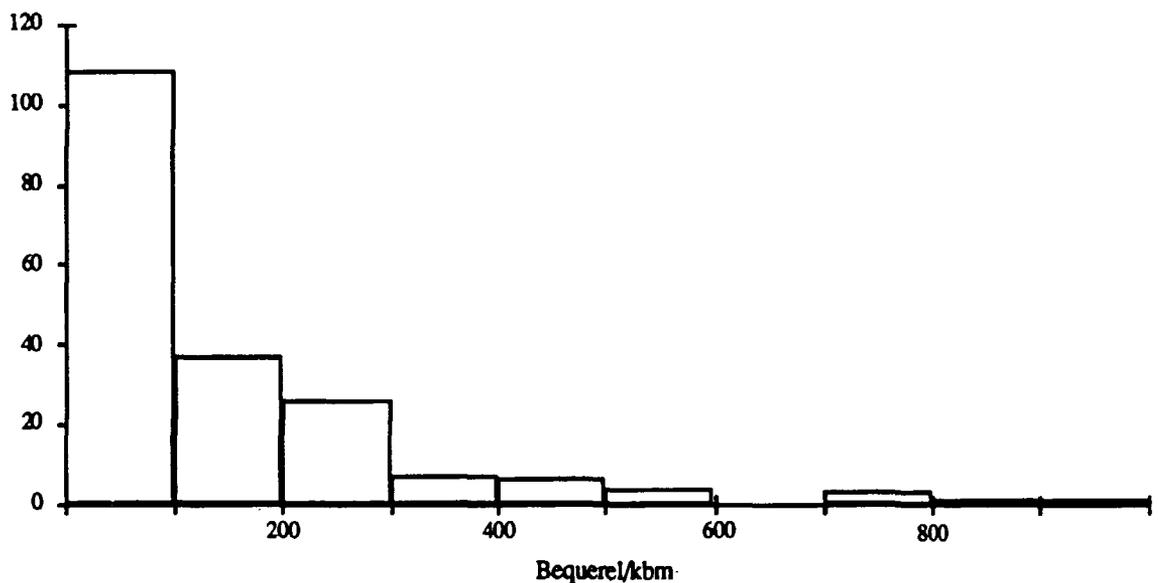
Tabell 4.17 Antal observationer för olika kombinationer av faktornivåerna månad och ort i Nynäshamn

Ort/Månad	Januari	Februari	Mars	Maj	Augusti
Okänd	26	2	4	2	
Nynäsh.	18	6	-	2	-
Sorunda	8	2	-	-	-
Ösmo	12	2	-	-	-
St Vika	2	-	-	-	2

## 4.6 Simrishamn

Datamaterialet innehåller 193 observationer insamlade mellan den 3/2 1988 och 12/5 1989. Alla mätningar är gjorda på högnivåmark. Materialets utseende visas i Figur 4.8 och Tabell 4.18.

Figur 4.8 Histogram av datamaterialet från Simrishamn



Tabell 4.18 Fördelning på olika effekter i Simrishamn

Mätort	antal	Rum	antal	Rökning	antal	Månad	antal
Okänd	17	Vardagsr	34	Okänt	22	Jan	15
Simrishamn	46	Sovrum	59	Förekommer	43	Feb	16
Kivik	18	Källare	11	Förekommer ej	116	Mar	6
Borrby	48	Pannrum	1	Ibland	12	Apr	20
Gärsnäs	14	Kök	25			Maj	2
St Olof	8	Kontor	22			Juni	2
Hammenhög	14	Övrigt	34			Juli	2
Gislöv	4	Okänt	7			Okt	82
Skillinge	10					Nov	26
Vitaby	8					Dec	22
Tommarp	6						

Den modell som testas innehåller månadsfaktorn, årsfaktorn, ortsfaktorn, rumsfaktorn samt röningsfaktorn. Den ser ut som nedan:

$$\ln y_{ijklm} = \mu + \tau_i + \beta_j + \omega_k + \lambda_l + \alpha_m + \epsilon_{ijklm}$$

Testresultatet ges i Tabell 4.19.

*Tabell 4.19 Variansanalystabell för Simrishamn modell 1*

$\sqrt{MS_E} = 1.060$  MEDELVÄRDE FÖR LOGHALT 4.38

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Ort	10	21.44	1.91	0.048
Månad	9	28.97	2.86	0.0037
Rökning	3	10.76	3.19	0.025
Rum	7	10.05	1.28	0.264
År	1	0.317	0.28	0.596

Detta resultat tyder på att rumsfaktorn och årsfaktorn inte är viktiga, och en ny modell utan dessa faktorer testas:

$$\ln y_{ikmn} = \tau_i + \omega_k + \alpha_m + \epsilon_{ikmn}$$

Denna modell testas också , och resultatet ges i Tabell 4.20

*Tabell 4.20 Variansanalystabell för Simrishamn modell 2*

$\sqrt{MS_E} = 1.064$  MEDELVÄRDE FÖR LOGHALT 4.38

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Ort	10	25.12	2.22	0.019
Månad	9	35.56	3.49	0.0006
Rökning	3	10.89	3.21	0.025

Denna modellen verkar rimlig och parametrarna skattas (Tabell 4.21 ). I Figur 4.9 visas variationen i tiden grafiskt.

*Tabell 4.21 Skattningar av marginella medelvärden, Simrishamn*

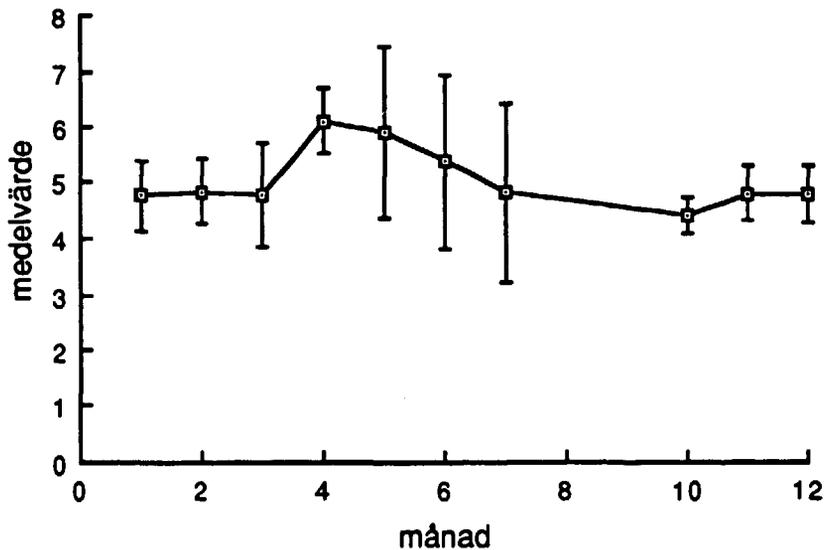
Rökning	Medelvärde	medelfel
Okänt	5.293	0.308
Förekommer	4.646	0.250
Förekommer ej	4.756	0.191
Förekommer ibland	5.582	0.364

Ort	Medelvärde	medelfel
Okänd	4.903	0.331
Simrishamn	4.784	0.236
Kivik	5.221	0.299
Borrby	4.550	0.206
Gärsnäs	4.652	0.357
St Olof	4.256	0.435
Hammenhög	4.876	0.359
Gislöv	5.493	0.572
Skillinge	5.492	0.391
Vitaby	5.364	0.437
Tommarp	6.173	0.492

Månad	Medelvärde	medelfel
Januari	4.775	0.328
Februari	4.857	0.305
Mars	4.797	0.471
April	6.103	0.297
Maj	5.923	0.786
Juni	5.391	0.797
Juli	4.815	0.815
Oktober	4.434	0.166
November	4.802	0.252
December	4.798	0.266

Man kan härur lägga märke till att det framför allt är månaden April som skiljer sig från övriga månader, samt att de som röker ibland verkar ha högre radonhalt än övriga. Skattningen av skillnaden mellan hus där rökning förekommer och där det inte förekommer skattades till -0.110 med ett medelfel på 0.209. Denna skillnad är inte signifikant.

Figur 4.9 Radonhaltens marginella medelvärde som funktion av månaden.



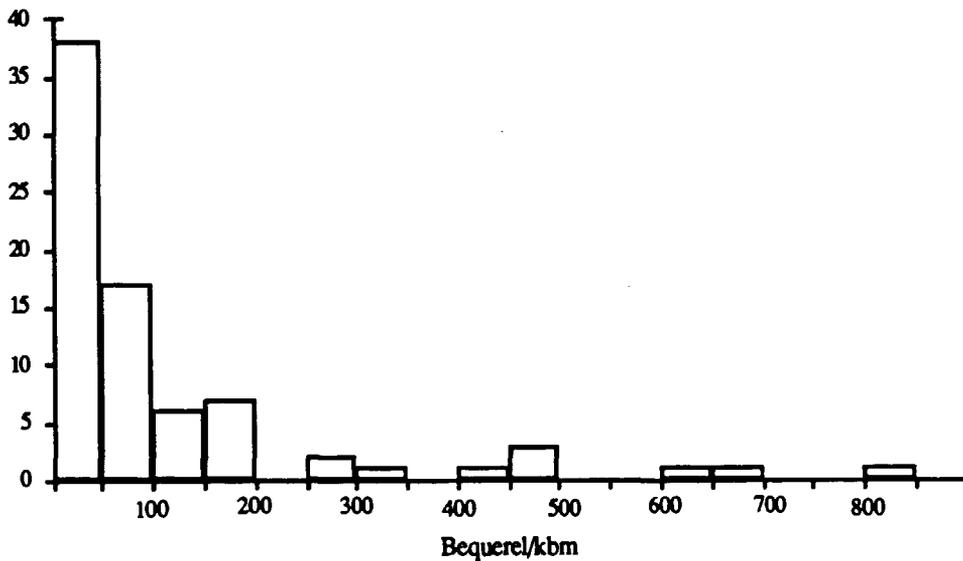
#### 4.7 Sjöbo

Datamaterialet innehåller 78 observationer insamlade mellan den 24/1 1989 och den 27/8 1989. Av mätningarna är 30 gjorda på lågnivåmark och 48 på högnivåmark. Materialets utseende visas i Figur 4.10 och Tabell 4.22.

Tabell 4.22 Fördelning på olika effekter i Sjöbo

Mätort	antal	Rum	antal	Rökning	antal	Månad	antal
Okänd	1	Vardagsr	19	Okänt	4	Jan	31
Lövestad	30	Sovrum	43	Förekommer	15	Feb	31
Hörr	6	Källare	2	Förekommer ej	55	Mar	9
Tolånga	10	Källare	2	Ibland	4	Apr	5
Vanstad	1	Kök	1			Maj	2
Rygård	1	Kontor	3				
Vollsjö	6	Övrigt	4				
Sjöbo	12	Okänt	4				
Äsperöd	1						

Figur 4.10. Histogram över datamaterialet från Sjöbo



Eftersom alla observationer var från samma år ansattes en modell innehållande månads-, Orts-, rökning- och rumseffekter:

$$\ln y_{ijkmn} = \mu + \tau_i + \beta_j + \omega_k + \alpha_m + \epsilon_{ijkmn}$$

Testet av modellen redovisas i Tabell 4.23.

Tabell 4.23 Variansanalys av material från Sjöbo modell 1

$\sqrt{MS_E} = 0.974$  MEDELVÄRDE FÖR LOGHALT 4.14

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Ort	8	13.99	1.84	0.089
Månad	4	7.50	1.97	0.111
Rökning	3	1.64	0.58	0.633
Rum	7	10.03	1.51	0.183

Ingen av effekterna inverkar signifikant men såväl ortsfaktorn som månadsfaktorn kan vara värd ytterligare undersökning. Det visar sig då att såväl månadsfaktorn som ortsfaktorn ger signifikans var för sig (Tabell 4.25 och 4.27). Om man tittar på datamaterialet och undersöker vilka kombinationer av månader och orter som mätts (Tabell

4.24) visar det sig att det troligtvis är samma effekt som ger utslag, och det är därför svårt att särskilja de olika effekternas inverkan.

*Tabell 4.24 Antal observationer för olika kombinationer av faktornivåerna månad och ort i Sjöbo*

Ort/Månad	Januari	Februari	Mars	April	Maj
Okänd	4	5	2	-	-
Lövestad	12	14	3	-	1
Hörr	3	2	1	-	-
Tolånga	-	4	1	-	-
Vanstad	-	1	-	-	-
Rygård	-	1	-	-	-
Vollsjö	2	1	-	3	-
Sjöbo	5	2	2	2	1
Äsperöd	-	1	-	-	-

*Tabell 4.25 Variansanalys för Sjöbo modell 2*

$$\sqrt{MSE} = 1.01$$

MEDELVÄRDE FÖR LOGHALT 4.14

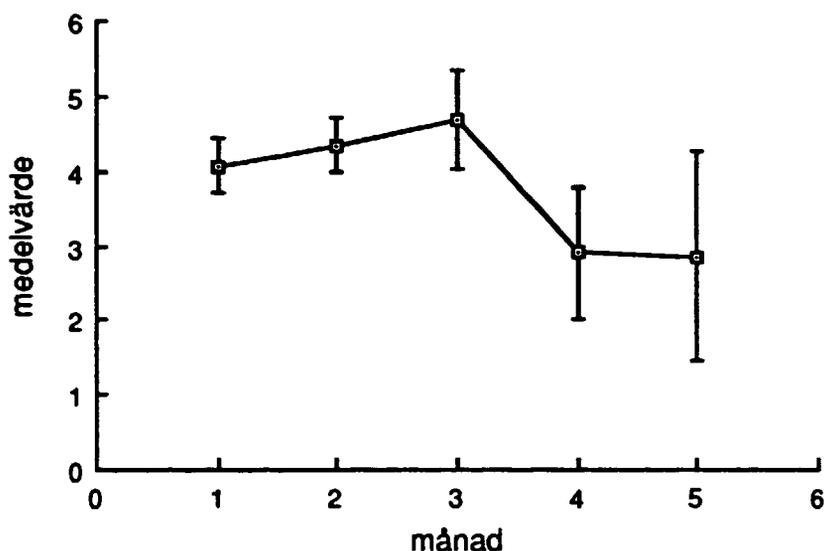
EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
månad	4	15.07	3.67	0.0089

Skattningar av parametrar ges i Tabell 4.26. I Figur 4.11 visas variationen i tiden grafiskt.

*Tabell 4.26 Skattningar av marginella medelvärden modell 2, Sjöbo*

Månad	Medelvärde	medelfel
Januari	4.069	0.182
Februari	4.346	0.182
Mars	4.678	0.338
April	2.902	0.453
Maj	2.852	0.717

Figur 4.11 Radonhaltens marginella medelvärde som funktion av månaden



Dessa jämförelser ger vid handen att april och maj skiljer sig från övriga månader. Kommentarer om detta ges i slutet av kapitel 4.7.

Tabell 4.27 Variansanalys för Sjöbo modell 3

$$\sqrt{MS_E} = 1.02$$

MEDELVÄRDE FÖR LOGHALT 4.14

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Ort	8	17.80	2.12	0.045

Skattningar av parametrarna ges i Tabell 4.28

Rygård och Hörr verkar ha högre radonhalt än övriga orter, med reservation taget för osäkerheten om samspelet mellan ort och mättid. Man ska här observera att endast en mätning är gjord i Rygård och denna mätning representerar endast sig själv.

Man vet att under de första månaderna gjorde mätningar främst i riskområden medan det i april och maj gjordes även privata mätningar inom blandade områden. Detta skulle ge vid handen att modell 2 inte är så intressant eftersom mätningarna från olika månader är från områden med skiftande förväntad radonhalt.



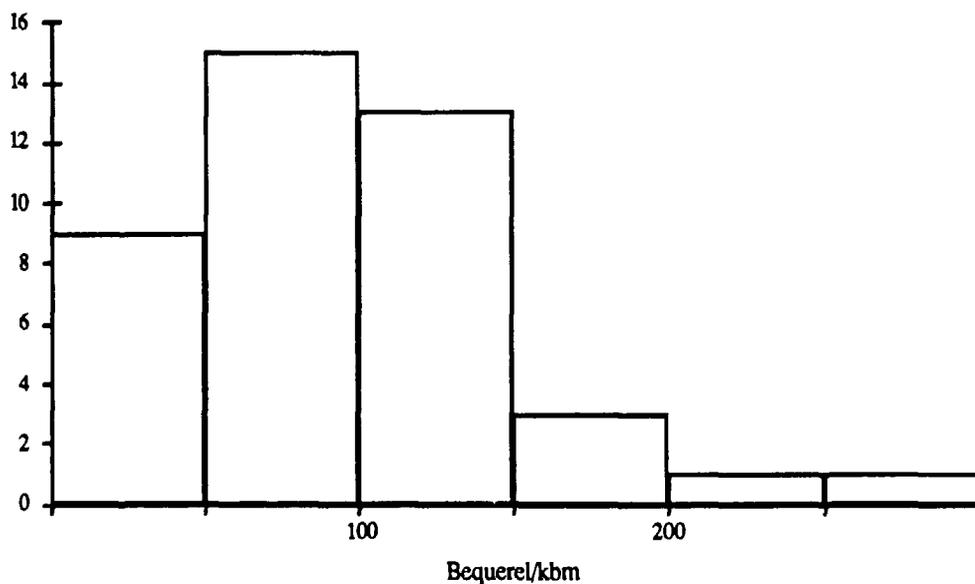
Tabell 4.28 Skatningar av marginella medelvärden. modell 3, Sjöbo

Ort	Medelvärde	medelfel
Okänd	4.371	0.309
Lövestad	4.171	0.187
Hörr	5.022	0.418
Tolånga	3.941	0.323
Vanstad	4.522	1.023
Rygård	6.739	1.023
Vollsjö	3.414	0.418
Sjöbo	3.733	0.295
Äsperöd	3.891	1.023

## 4.8 Trelleborg

Materialet består av 42 observationer insamlade mellan den 23/3 och den 24/5 1988. Alla mätningar är från centralorten och uppmätt på lågnivåmark. Alla observationer har samma utsättningsmånad. Datamaterialets utseende ges i Figur 4.12 samt i Tabell 4.29.

Figur 4.12. Histogram över datamaterialet från Trelleborg



Tabell 4.29 Datamaterialets fördelning på olika effekter

Rum	antal	Rökning	antal
Okänt	1	Okänt	1
Vardagsrum	14	Förekommer	11
Sovrum	4	Förekommer ej	24
Källare	2	Förekommer ibland	6
Kök	20		
Övrigt	1		

Den modell som kan vara aktuell här är beroende på materialets homogenitet endast rökning- och rumsfaktorerna. Modellen blir sålunda:

$$\ln y_{ijn} = \mu + \tau_i + \beta_j + \varepsilon_{ijn}$$

Testets resultat ges i Tabell 4.30.

Tabell 4.30 Variansanalystabell för Trelleborg

$$\sqrt{MS_E} = 0.95$$

MEDELVÄRDE FÖR LOGHALT 4.27

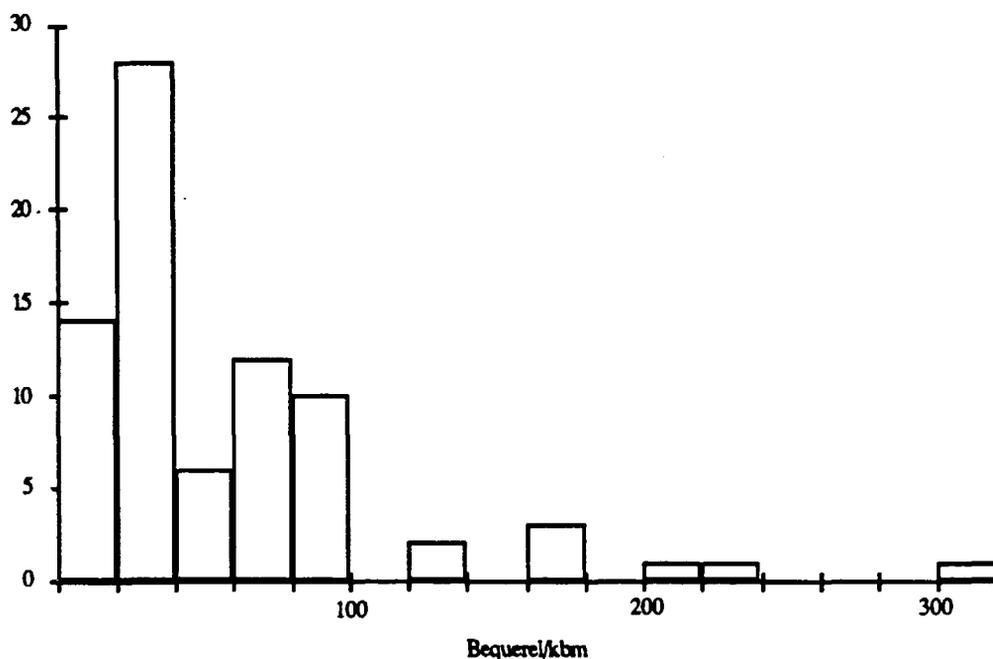
EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Rökning	3	1.03	0.38	0.767
Rum	5	5.89	1.31	0.285

Ingen av de undersökta effekterna ger någon signifikans. Detta kan bero på materialets litenhet.

## 4.9 Ystad

Datamaterialet består av 78 observationer insamlade mellan den 12/8 1988 och den 10/7 1989. Alla mätningar är gjorda på lågnivåmark. Materialets utseende visas i Figur 4.13 samt i Tabell 4.31.

Figur 4.13 Histogram över radonhalten i Ystad



Tabell 4.31 Materialets fördelning på olika effekter

Ort	Antal	Rum	Antal	Rökning	Antal	Månad	Antal
Okänd	30	Vardagsr	27	Okänt	10	Feb	4
Köpingebro	30	Sovrum	27	Förekommer	19	Mar	2
Svenstorp	2	Källare	12	Förekommer ej	43	Maj	2
Löderup	8	Kök	3	Förekommer	6	Augusti	46
Hörup	2	Kontor	3	ibland		September	22
St Köpinge	2	Övrigt	6			November	2
Ystad	2						
Glemminge- bro	2						

Den modell som testas innehåller rums- rökning- orts- samt månadsfaktorer. Resultatet ges i Tabell 4.32. Modellens utseende är som följer.

$$\ln y_{ijkmn} = \mu + \tau_i + \beta_j + \omega_k + \alpha_m + \epsilon_{ijkmn}$$

Tabell 4.32. Variansanalys för Ystad modell 1

$$\sqrt{MS_E} = 0.690 \quad \text{MEDELVÄRDE FÖR LOGHALT 3.71}$$

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Ort	7	10.86	3.26	0.0056
Månad	5	5.62	2.36	0.0514
Rökning	3	2.29	1.61	0.155
Rum	5	3.99	1.67	0.1980

Resultatet tyder på att en modell innehållande månads- och ortseffekten bör undersökas noggrannare. Test för denna ges i Tabell 4.33.

Tabell 4.33 Variansanalys för Ystad modell 2

$$\sqrt{MS_E} = 0.727 \quad \text{MEDELVÄRDE FÖR LOGHALT 3.71}$$

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Ort	7	14.96	4.04	0.0010
Månad	5	8.34	3.15	0.0131

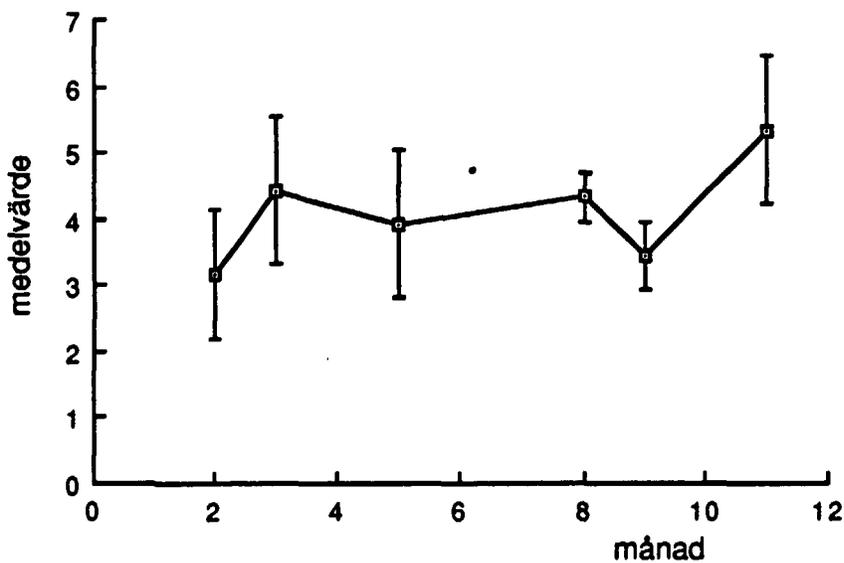
Denna modell verkar bra och skattningar av de marginella medelvärdena görs (Tabell 4.34). I Figur 4.14 visas variationen i tiden grafiskt

Tabell 4.34 Skattningar av marginella medelvärden för Ystad

Ort	medelvärde	medefel
Okänd	3.991	0.198
Köpingebro	3.418	0.272
Svenstorp	3.065	0.571
Löderup	3.679	0.330
Hörup	3.523	0.571
St Köpinge	5.041	0.571
Ystad	6.080	0.564
Glemmingebro	4.050	0.693

Månad	medelvärde	medelfel
Februari	3.176	0.501
Mars	4.438	0.564
Maj	3.925	0.564
Augusti	4.326	0.195
September	3.442	0.256
November	5.330	0.563

Figur 4.14 Radonhaltens marginella medelvärde som funktion av månaden



Såväl Ystad som St Köpinge verkar ha högre halt än övriga orter.

#### 4.10 Totala materialet

Det totala materialet innehåller 813 observationer. Dessa undersöks ej på ortsnivå utan endast på kommun eller risknivå. Modellen innehåller månadsfaktorn, årsfaktorn, rökfaktorn, rumsfaktorn samt antingen kommunfaktorn eller klassfaktorn (högrisk, lågrisk). Eftersom man för de flesta kommunerna enbart mätt antingen på högriskmark eller lågriskmark kommer dessa faktorer att i stort sett visa samma effekt. Om materialet hade varit mer balanserat hade man kunnat använda kommun och klassfaktorerna samtidigt. Vi gör två modeller, modell 1 innehåller kommunfaktorn och modell 2 innehåller klassfaktorn. I Tabell 4.35 visas variansanalystabellen för modell 1:

*Tabell 4.35 Variansanalystabell för det totala materialet modell 1, Totala materialet*

$\sqrt{MSE} = 1.111$  MEDELVÄRDE FÖR LOGHALT 4.05

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Kommun	8	121.0	12.26	0.0001
Månad	11	25.97	1.91	0.035
Rökning	3	11.52	3.11	0.026
Rum	7	34.85	4.03	0.0002
År	2	6.66	2.70	0.068

Alla faktorer är signifikanta utom årsfaktorn som ligger på gränsen till signifikans. Trots allt behålls den eftersom den visar sig vara signifikant i modell 2 och de övriga faktorerna ej får ökad signifikans om den tas bort. Skattningsarna av de olika parametrarna ges i Tabell 4.36. I Figur 4.15 visas variationen i tiden grafiskt

*Tabell 4.36 Skattningar av marginella medelvärdeni modell 1, Totala materialet*

Kommun	Antal obs	medelvärde	medelfel
Okänd	119	4.875	0.209
Hässleholm	37	3.680	0.269
Lund	135	3.841	0.207
Malmö	43	4.337	0.211
Nynäshamn	88	4.526	0.233
Simrishamn	193	5.001	0.200
Sjöbo	78	4.556	0.232
Trelleborg	42	4.676	0.286
Ystad	78	3.762	0.286

Rökning	medelvärde	medelfel
Okänt	4.494	0.236
Förekommer	4.135	0.192
Förekommer ej	4.197	0.170
Förekomer ibland	4.620	0.239

Månad	medelvärde	medelfel
Januari	4.393	0.200
Februari	4.262	0.208
Mars	4.339	0.217
April	4.590	0.242
Maj	3.995	0.268
Juni	4.469	0.274
Juli	4.604	0.408
Augusti	4.718	0.240
September	4.618	0.271
Oktober	3.921	0.248
November	4.439	0.259
December	3.991	0.265

Rum	medelvärde	medelfel
Okänt	4.328	0.264
Vardagsrum	4.401	0.181
Sovrum	4.549	0.175
Källare	4.975	0.212
Pannrum, uthus	3.829	0.492
Kök	4.439	0.229
Kontor	4.206	0.243
Övrigt	4.166	0.193

År	medelvärde	medelfel
1987	5.073	0.465
1988	3.996	0.116
1989	4.015	0.119

Man ser att Simrishamn verkar ha signifikant högre halt än övriga kommuner och källarlokalerna har högre halt än övriga rum. För rökningens del är skillnaden mellan rökning och icke rökning 0.06 vilket inte är signifikant. Däremot ger "rökning ibland" signifikant högre halt.

Modell 2 som innehåller klassfaktorn redovisas i Tabell 4.37.

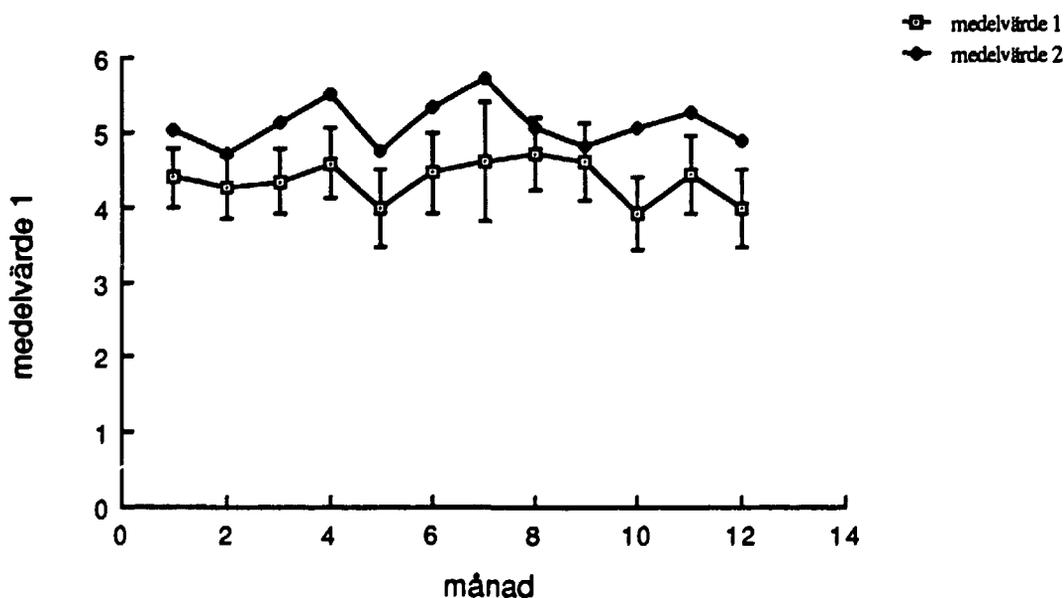
Tabell 4.37 Variansanalystabell för totala materialet för modell 2, Totala materialet

$\sqrt{MS_E} = 1.164$  MEDELVÄRDE FÖR LOGHALT 4.05

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Klass	2	18.38	6.78	0.0012
Månad	11	36.79	2.47	0.0048
Rökning	3	15.43	3.79	0.0102
Rum	7	50.12	5.28	0.0001
År	2	13.87	5.12	0.0062

Alla faktorer är signifikanta. I Tabell 4.38 redovisas skattningar av de olika parametrarna. I Figur 4.15 visas variationen i tiden grafiskt.

Figur 4.15 Radondatas marginella medelvärden som funktion av månaden för de två modellerna för det totala materialet. Medelvärde 1 är för modell 1, medelvärde 2 är för modell 2. Konfidensintervall har av tydlighetsskäl endast redovisats för modell 2





Tabell 4.38 Skattning av marginella medelvärden för modell 2, Totala materialet

Klass	medelvärde	medelfel
Lågrisknivå	4.424	0.180
Högrisknivå	4.702	0.182
Blå lättbetong	6.213	0.622
Rökning	medelvärde	medelfel
Okänt	5.338	0.310
Förekommer	4.815	0.272
Förekommer ej	4.961	0.259
Förekommer ibland	5.339	0.312
Månad	medelvärde	medelfel
Januari	5.038	0.277
Februari	4.718	0.271
Mars	5.119	0.289
April	5.505	0.389
Maj	4.751	0.337
Juni	5.348	0.340
Juli	5.731	0.460
Augusti	5.051	0.299
September	4.824	0.337
Oktober	5.075	0.296
November	5.289	0.320
December	4.906	0.320
Rum	Medelvärde	medelfel
Okänt	4.813	0.333
Vardagsrum	5.161	0.267
Sovrum	5.305	0.259
Källare	5.744	0.292
Pannrum, uthus	4.754	0.546
Kök	5.372	0.300
Kontor	4.946	0.312
Övrigt	4.808	0.268

År	Medelvärde	medelfel
1987	5.945	0.485
1988	4.635	0.223
1989	4.758	0.234

Man kan ur detta material lägga märke till att klassindelningen som gjorts är rimlig. Dessa värden skiljer sig signifikant åt. De som röker ibland skiljer sig från dem som röker och dem som inte röker. Skattningen av skillnaden mellan dem som röker och dem som inte röker är  $-0.146$  med en standardavvikelse på  $0.111$ . Differensen skiljer sig inte signifikant från  $0$ . Intressant är också att källarlokalerna verkar ha högre halt än övriga rum. I figuren kan man se att de marginella medelvärdena för månadsfaktorn systematiskt är högre i modellen med klassfaktorn. Det kan förklaras av att nivån blå lättbetong är kraftigt underrepresenterad, och samtidigt har betydligt högre radonhalt än övriga nivåer. Några så stora skillnader i radonhalt mellan nivåerna för kommunerna finns ej. Detta medför att  $\mu$  skattas högre i modellen med klassfaktorn än i modellen med kommunfaktorn. Detta medför i sin tur att även de marginella medelvärdena för månaderna måste bli större i modellen med klassfaktorn än i modellen med kommunfaktorn.

## 5 Resultat - modell med samspelseffekter

### 5.1 Introduktion

Man kan tänka sig att värdet hos en observation inte enbart beror av summan av de olika faktorernas inverkan och en slumpmässig term. Man kan också tänka sig att det finns en samverkan mellan olika faktorer och att man därför måste addera samverkanstermer för de faktorer som inverkar vid en viss observation. I nedanstående beräkningar tas hänsyn till samverkanstermer mellan två faktorer. På grund av det obalanserade materialet skulle ett hänsynstagande till samverkanstermer mellan fler termer vara meningslös då informationen i materialet helt enkelt inte räcker till. Modellen behandlar enbart det totala materialet.

## 5.2 Analys av det totala materialet.

Materialet består som tidigare av 813 observationer. En undersökning av de samverkans effekter som fysikaliskt verkar rimliga ger till slut upphov till två modeller där samverkan råder mellan år och månad, år och rökning samt månad och rökning. Modell 1 innehåller månadseffekten, årseffekten, rumseffekten, rökningseffekten samt kommuneffekten. Modell 2 innehåller samma effekter förutom kommuneffekten. I stället ingår klasseffekten. Inga skattningar av parametrarna är möjliga. Resultaten redovisas i Tabell 5.1 och Tabell 5.2.

*Tabell 5.1 Variansanalys för modell 1, samverkansmodell*

$\sqrt{MS_E} = 1.111$  MEDELVÄRDE FÖR LOGHALT 4.05

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Kommun	8	121.0	12.26	0.0001
Månad	11	10.84	0.85	0.592
Rökning	3	8.55	2.45	0.062
Rum	7	31.11	3.82	0.0004
År	2	8.72	3.75	0.024
Månad*rökn.	27	59.94	1.91	0.0038
Månad*År	12	25.63	1.84	0.0390
Rökning*år	3	16.29	4.67	0.0031

*Tabell 5.2 Variansanalys modell 2, samverkansmodell*

$\sqrt{MS_E} = 1.164$  MEDELVÄRDE FÖR LOGHALT 4.05

EFFEKT	f	SS	F-VÄRDE	KONFIDENSGRAD
Klass	2	17.99	7.03	0.0009
Månad	11	22.67	1.61	0.0909
Rökning	3	11.39	2.97	0.0313
Rum	7	40.20	4.49	0.0001
År	2	8.02	3.13	0.0443
Månad*rökn.	27	51.52	1.49	0.0529
Månad*År	12	39.99	2.60	0.0021
Rökning*år	3	12.97	3.38	0.0180

Modellerna är signifikanta och man bör alltså ha med samverkans effekter i en bra modell. Emellertid blir det i vårt fall problem p g a obalansen i materialet. Om man för in en samverkans effekt i materialet blir nämligen signifikansen för de enskilda effekterna sämre för varje samverkan som införs. Dessutom har man ingen möjlighet att undersöka hur samverkan är beskaffad då inga skattningar kan göras.

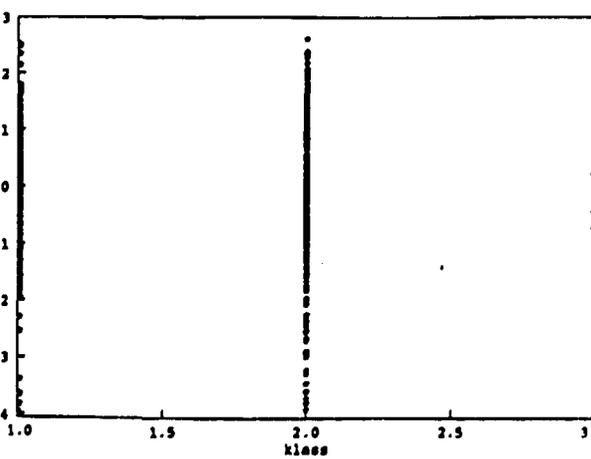
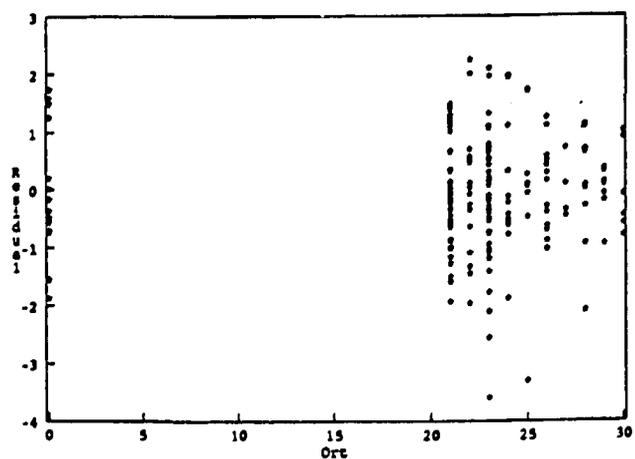
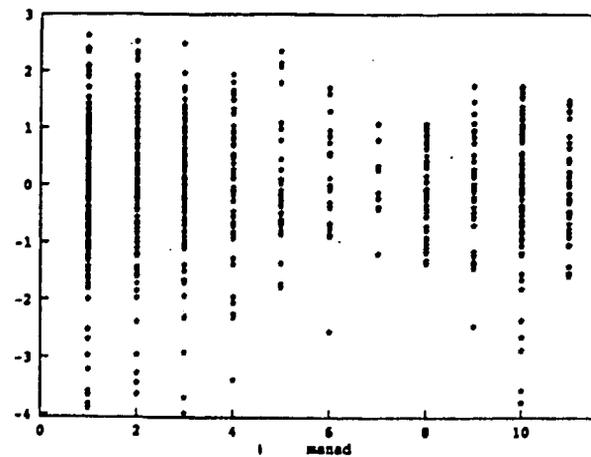
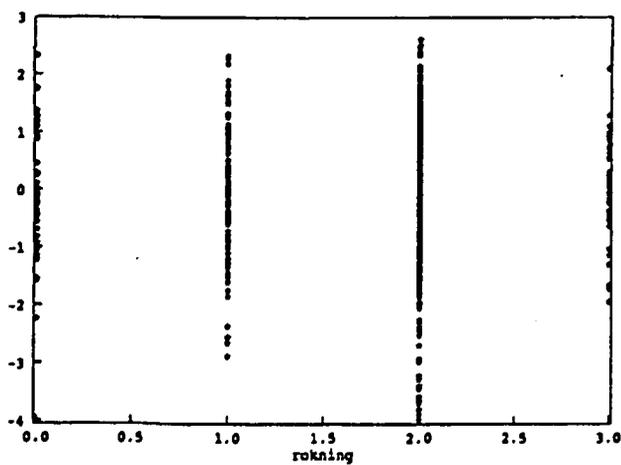
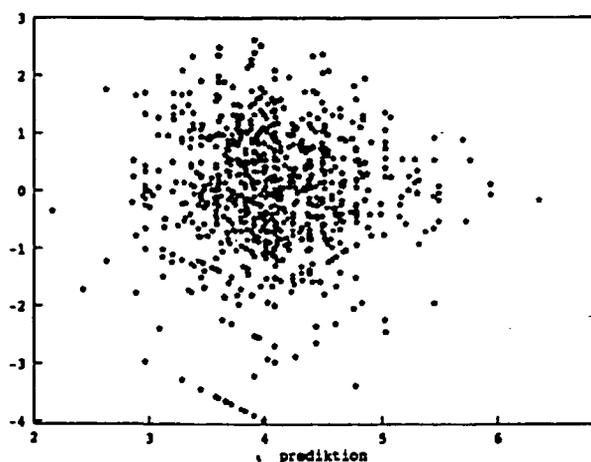
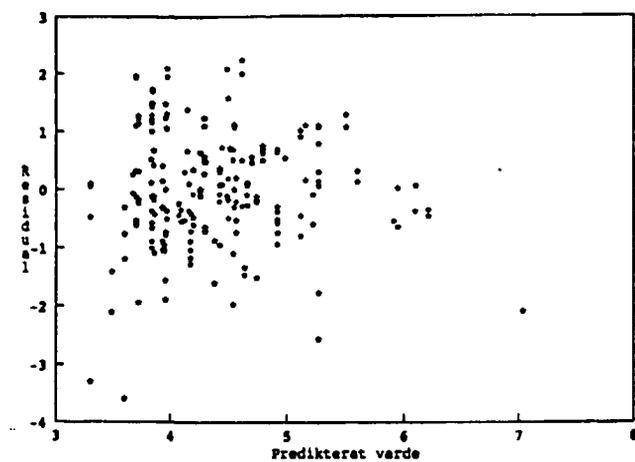
## 6 Modellernas riktighet

För att se om modellen verkar riktig gjordes analys av residualerna. För att variansanalys skall fungera skall mätningarnas varians vara lika stor oavsett vilken faktorkombination man mäter på. Dessutom skall mätningarna vara observationer från en normalfördelning. För modellens riktighet är det första villkoret viktigast. För att undersöka dessa villkor plottas residualerna mot de olika faktorerna. För att modellen då skall anses riktig bör inte någon skillnad i varians för de olika faktornivåerna kunna urskiljas. En svårighet är att kunna se om variansen är olika stor för olika nivåer om det finns fler observationer från vissa nivåer. Ju större antal observationer man har från en nivå desto sannolikare är det ju att man får en stor residual. Att sedan visuellt kunna bedöma om variansen är konstant blir därför ganska komplicerat. I Figur 6.1 plottas residualerna mot de olika faktornivåerna och mot predikterat värde för modellen från Simrishamn och modell 1 med samverkan.

Man ser inget direkt system i spridningen som skulle tyda på att variansen inte skulle vara konstant. Residualerna plottas i ett normalfördelningspapper. I Figur 6.2 visas normalplottar för samma modeller som i Figur 6.1. Man ser här viss skillnad mot normalfördelningen för små och stora värden. Skillnaden är emellertid inte större än att modellen kan användas med någorlunda noggrannhet.

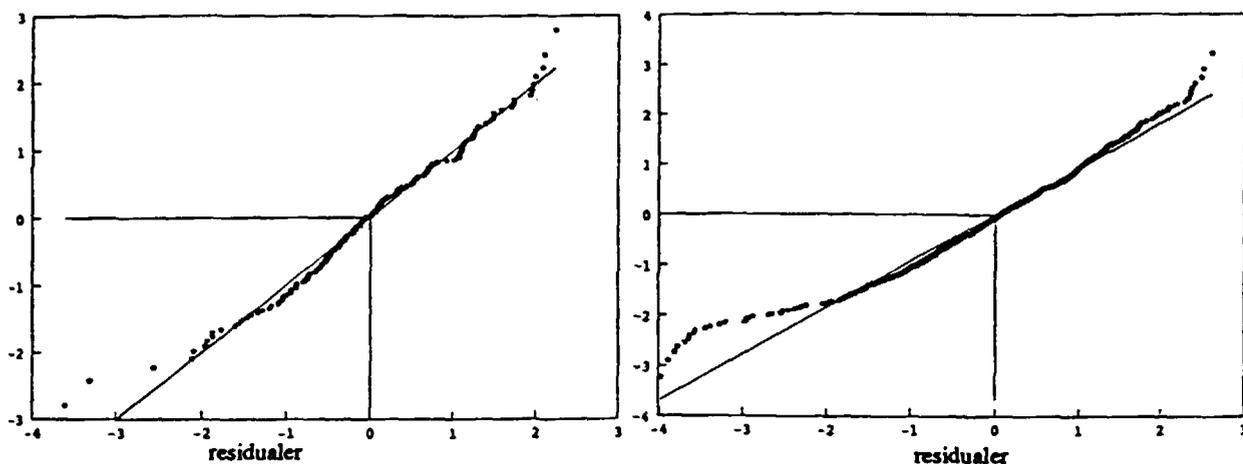
Här har endast redovisats residualanalysen för två modeller. Övriga modeller har också undersökts och resultaten är likartade de två här redovisade modellerna.

Figur 6.1 Residualerna plottade mot några olika faktorer. De tre till vänster från modellen från Simrishamn. De tre till höger från modell 1 med samverkan



Figur 6.2 Normalfördelningsplottar av residualerna för modellen från Simrishamn samt modell 1 med samverkan

Standard-  
avvikelser



## 7 Slutsatser

I detta kapitel görs ett försök att sammanfatta resultaten från analysen av radondata. Det kommer att innehålla såväl sammanfattning av resultaten som förslag till hur mätningarna i fortsättningen bör läggas upp.

### 7.1 Sammanfattning av resultatet

Analysen har visat att radonhaltens fördelning ganska väl överensstämmer med en lognormalfördelning. Detta visades genom att de logaritmerade observationerna relativt väl överensstämde med en normalfördelning. Residualanalysen visade också att variansen för de logaritmerade värdena var konstanta vilket medför att variansanalys kan användas på materialet.

Det har under analysen visat sig att såväl tidsfaktorer, rumsfaktorer som rönkningsfaktor är signifikanta. Speciellt har det betydelse under vilken tid på året man mäter. Det visar sig emellertid svårt att peka på någon tid på året som rent allmänt har högre radonhalt än övriga. Detta beror troligen på skiftande lokala förhållanden. Det kan vara såväl skiftande mark som väderförhållanden som ger hög halt för olika platser vid

olika tidpunkter. Att årsfaktorn kan ha en inverkan kan bero på att olika år har olika väderläge. Medelvärdets variation med månaden redovisas i Figur 7.1

Ortsfaktorn verkar rent allmänt ha stor betydelse och det kan förklaras av att orterna är belägna på skiftande berggrund och jordart. Variationerna beskrivs tydligare i delmodellerna för de olika orterna. Det bör än en gång poängteras att de skattade marginella medelvärdena inte utgör en skattning av väntevärdet av det medelvärde man skulle få om man slumpmässigt tog ut ett antal observationer på en ort utan är en skattning av väntevärde för det medelvärde man får om man gör ett balanserat blockförsök. Dessutom finns problem med generaliserbarheten för observationerna från de orter där observationerna är få. De skattade ortsfaktorerna i Skåne har i Figur 7.2 förts in på en karta över Skåne. För Hässleholm och Trelleborg där ingen signifikant ortsfaktor finns har de skattade värdena för kommunfaktorn angetts.

Rumsfaktorn har en mindre betydelse än de tidigare nämnda. Man kan bara få signifikans för den i vissa fall. Det är i dessa fall källaren som särskiljer sig med en högre halt. Detta kan förklaras med att markradon koncentreras i dessa lokaler som dessutom för det mesta har sämre ventilation.

Rökfaktorn har betydelse. Det visar sig att radonhalten inte skiljer sig nämnvärt mellan hus där rökning förekommer och hus där det inte förekommer. Däremot verkar hus, i vilka man angett att rökning förekommer ibland, ha högre radonhalt än övriga kategorier. Man kan anta att rökning faktorn har med skiftande vädringsförhållanden att göra. Det skulle i så fall innebära att man luftar mindre i hus där man röker ibland än i hus där man röker ofta. Detta antagande är svårt att förklara.

Vad gäller samverkans effekter visade det sig att samverkan finns mellan tidsfaktorn och rökfaktorn. Detta kan kanske förklaras av att luftningsfrekvensen ändras med årstidernas gång. Denna varierade luftning kommer naturligtvis att påverka hus med olika luftningsfrekvens olika, beroende på varierande rökningens intensitet. Dessutom råder samverkan mellan årsfaktorn och månadsfaktorn. Detta kan förklaras med att även om ett visst år har ett väder, som ger hög radonhalt kan det mycket väl finnas månader under just det året som är tvärtom. Detta kommer då att märkas som en samverkans effekt.

## 7.2 Förslag till förbättringar av mätmetoden

Ett stort problem vid analysen har varit den stora obalans som råder i materialet. En viss nivå av en faktor har förekommit tillsammans med en bestämd nivå av en annan faktor. Detta har medfört att inverkan av de olika faktorerna inte kunnat särskiljas. Ett annat problem har varit de många tomma celler som funnits. Detta har medfört att värden för dessa celler har skattats. Detta har gett ett oönskat beroende inom materialet. Man har inte

kunnat utnyttja den fina teknik som variansanalys för balanserade data innebär. De jämförande testen (Duncan, Tukey etc ) som kan användas för balanserade experiment för att kunna säga om två faktornivåer skiljer sig signifikant åt har inte kunnat användas. I stället har gjorts parvisa jämförelser av olika faktornivåer, vilket ger en sådan osäkerhet om signifikansnivåer att dessa inte angetts.

Hur mätningarna skall läggas upp i framtiden beror på vilket syfte man har med mätningarna. Dessa kan vara flera t ex

- 1) att undersöka en viss faktors inverkan på radonhalten.
- 2) att samtidigt undersöka flera faktorerers inverkan på radonhalten.
- 3) att undersöka radonhalten i en viss ort.
- 4) att undersöka radonhalten i bestämda hus.

Olika rekommendationer om hur mätningarna bör läggas upp får ges för de olika fallen.

1) Om man vill undersöka en viss faktors inverkan, t ex rökningens inverkan på radonhalten måste man se till att övriga inverkande faktorer inte systematiskt stör jämförelserna. Man kan till exempel se till att man gör alla mätningarna vid samma tidpunkt på året, och att man dessutom gör mätningarna på platser med likartade markförhållanden. Man kan också se till att man inte gör mätningarna i källaren, vilket ju gett högre radonhalt än mätningar i övriga rum. Undersökningen utförs som variansanalys, där idealet är en balanserad försöksplan.

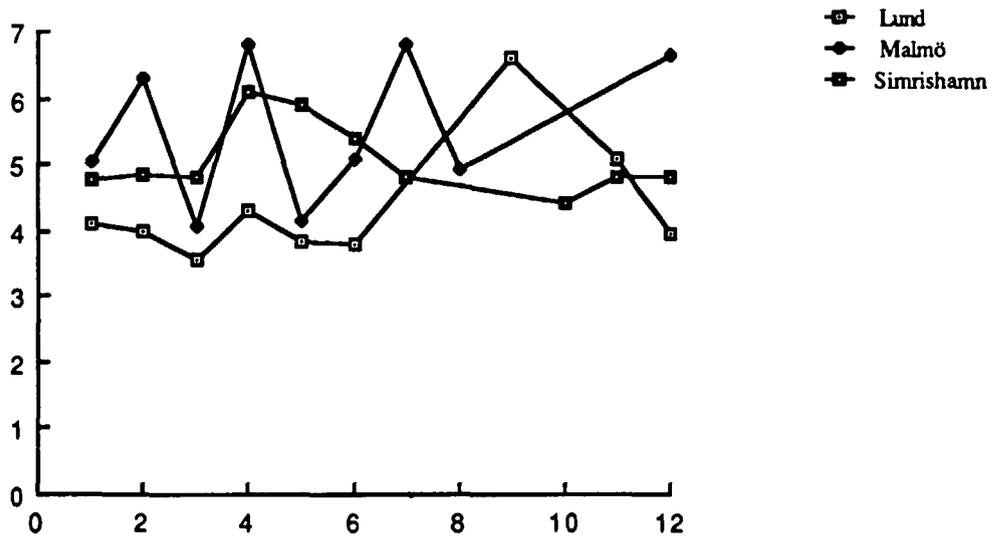
2) Vill man undersöka flera faktorer samtidigt bör man också utföra balanserad variansanalys d v s man bör ha lika antal observationer för samtliga faktorkombinationer. En sådan analys har nackdelen att många observationer krävs. Den bör därför endast göras om man verkligen är intresserad av att undersöka samtliga effekter.

3) Om man vill undersöka radonhalten i en viss ort bör husen som man undersöker väljas slumpmässigt. Det är viktigt att man inte väljer ut husen på grund av misstanke om hög radonhalt. Även här bör man på något sätt försöka göra mätningarna under likartade förhållanden till exempel genom att välja likadana mätplatser i varje hus.

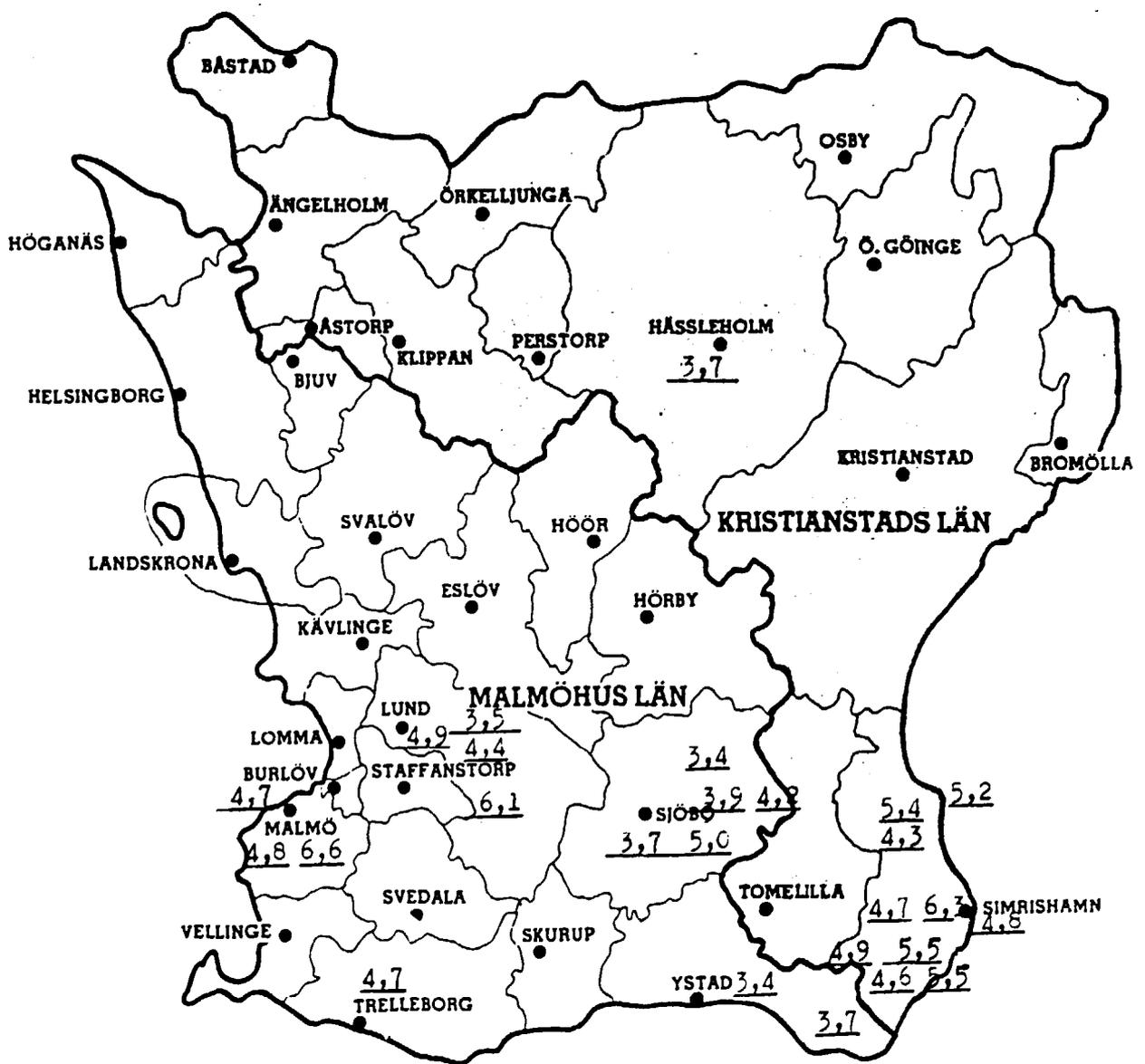
4) Om man vill undersöka radonhalten i bestämda hus är det inte längre ett statistiskt problem, men man bör ändå på något sätt, för att sedan kunna jämföra olika mätningar, även här göra alla sina mätningar under likartade förhållanden.



Figur 7.1 *Marginella* medelvärdets variation med månaden för Lund, Malmö och Simrishamn.



Figur 7.2. De skattade marginella medelvärdena för orterna inlagda i en Skånekarta. Endast orter med mer än fyra observationer har medtagits



## Litteratur

Blom, G (1984): Sannolikhets teori med tillämpningar, Studentlitteratur Lund.

Blom, G (1984): Statistikteori med tillämpningar Lund.

Edling C, Kling Hans, Axelson O (1984): Radon in homes - A possible cause of lung cancer. Scand J Work Environ Health 25-34

Jönsson, G (1987): Indoor radon gas and its detection with kodak plastic film. Nucl. Tracks Radiat Meas, 13, pp 85-91

Jönsson G (1988): Indoor  $^{222}\text{Rn}$  measurements in Sweden with the solid-state nuclear track detector technique, Health Physics Vol. 54, 271-281

Montgomery, D C (1984): Design and analysis of experiments, Wiley, New York

Pershagen, G Damber, L & Falk, R (1984): Exposure to radon in dwellings and lung-cancer: A pilot study. Ur Proceedings of the third international conference on indoor quality and climate August 20-24 vol 2 73-78

SAS users guide, SAS Institute INC 1982