Stockholm University

Föreläsning 4: Samband mellan kategoriska variabler

Matias Quiroz[1]

[1]Statistiska institutionen, Stockholms universitet

VT 2023

Innehåll

- Beskriva fördelningen för två kategoriska variabler.
- Undersöka samband mellan två kategoriska variabler.
- Marginella fördelningar.
- Simultana fördelningar.
- Betingade fördelningar.
- Begränsningar vid sambandstolkningar.
- Undersöka samband mellan tre kategoriska variabler.
- Beräkningar i R.

Repetition kategoriska variabler

- En **kategorisk variabel** är en variabel vars utfall är kategorier som inte behandlas som numeriska värden.
- En frekvenstabell, alternativt en relativ frekvenstabell, ger en beskrivning av fördelningen för en endaste kategoriska variabel.
- För variabeln "Class" i Titanic datasetet:

Class	Count
First	324
Second	285
Third	710
Crew	889

Class	Percentage (%)
First	14.67
Second	12.91
Third	32.16
Crew	40.26

Figure 1: Tabell 2.2 och 2.3 i De Veaux et al. (2021). Frekvenstabell (vänster) och relativa frekvenstabell (höger).

- Ibland vill man analysera **två** kategoriska variabler. Vi kan göra en frekvenstabell för vardera variabel och studera de separat.
- Med denna ansats kan vi beskriva variablerna separat, men det ger oss **ingen information om ett eventuellt samband mellan variablerna**.

Korstabeller för att undersöka samband

- För att studera sambandet mellan variablerna behöver vi en korstabell (**contingency table** på engelska).
- Exempel: Från en dejting websida finns data med 14294 personer som äger hund och/eller katt och som idenfierar sig som män eller kvinnor. Vi undersöker om det finns ett samband mellan kön och vilka husdjur man äger.
- Första steget är att identifiera vilka variabler man vill studera.
  1. Kön (`Gender`).
  2. Husdjurspreferens (`Pets`).
- Nästa steg är att klassificera variabelns typ samt bestämma dess utfall.
  1. `Gender` är en kategorisk variabel. Utfallen antas vara: Man eller kvinna.
  2. `Pets` är en kategorisk variabel. Utfallen antas vara: Hund, katt, eller bägge.

Korstabeller för att undersöka samband, forts.

- Korstabellen för datamaterialet:

		Gender		
		Female	Male	Total
Pets	Has cats	3412	2388	5800
	Has dogs	3431	3587	7018
	Has Both	897	577	1474
	Total	7740	6552	14,292

Figure 2: Från lärarmaterialet skapat av utgivaren av De Veaux et al. (2021).

- I relativa tal (%) (dela varje cell med 14292 och $\times$ 100):

		Gender		
		Female	Male	Total
Pets	Has cats	23.9%	16.7%	40.6%
	Has dogs	24.0%	25.1%	49.1%
	Has both	6.3%	4.0%	10.3%
	Total	54.2%	45.8%	100%

Figure 3: Tabell 3.4 i De Veaux et al. (2021).

- Kan vi utläsa en relativ frekvenstabell för varje variabel separat?

Korstabeller för att undersöka samband, forts.

- Ja, kolumnsummorna ger en relativ frekvenstabell för Gender

Gender	**Percentage (%)**
Female	54.2
Male	45.8

och radsummorna ger en relativ frekvenstabell för Pets

Pets	**Percentage (%)**
Has cats	40.6
Has dogs	49.1
Has both	10.3

- En fördelning för en enstaka variabel kallas för en marginell fördelning (**marginal distribution** på engelska).
- En **fördelning summerar alltid till 100%** (stäm av ovan).
- En marginell fördelning är en fördelning för en variabel **oavsett vilket värde den andra variabeln antar**.

Korstabeller för att undersöka samband, forts.

- Betrakta korstabellen igen:

		Gender		
		Female	Male	Total
Pets	Has cats	23.9%	16.7%	40.6%
	Has dogs	24.0%	25.1%	49.1%
	Has both	6.3%	4.0%	10.3%
	Total	54.2%	45.8%	100%

Figure 4: Tabell 3.4 i De Veaux et al. (2021).

- De orange-fyllda cellerna kallas för en simultan fördelning (**joint distribution** på engelska) för variablerna `Gender` och `Pets`.
- En simultan fördelning är en **fördelning över de olika utfallen av de båda variablerna**. Exempel på olika utfall:
  1. Kvinna som har katt.
  2. Man som har både hund och katt.
- En simultan fördelning är en fördelning. En **fördelning summerar alltid till 100%** (stäm av ovan).

Korstabeller för att undersöka samband, forts.

- Betrakta återigen korstabellen:

		Gender		
		Female	Male	Total
Pets	Has cats	23.9%	16.7%	40.6%
	Has dogs	24.0%	25.1%	49.1%
	Has both	6.3%	4.0%	10.3%
	Total	54.2%	45.8%	100%

Figure 5: Tabell 3.4 i De Veaux et al. (2021).

- Jämföra om kön påverkar preferensen för katter: Jämföra siffran 16.7% mot 23.9%?
- Vad betyder 16.7% och 23.9% i korstabellen?
  - 16.7%: Andelen av alla i undersökningen som hade katt och var män.
  - 23.9%: Andelen av alla i undersökningen som hade katt och var kvinnor.
- "Andelen av alla" (oavsett kön) är inte vad vi är intresserade av.

Korstabeller för att undersöka samband, forts.

- Vi är intresserade av att jämföra:
  1. Andelen av alla män i undersökningen som hade katt.
  2. Andelen av alla kvinnor i undersökningen som hade katt.
- Ingen av dessa motsvarar en simultan fördelning eller en marginell fördelning.
- Vi behöver introducera en ny sorts fördelning, kallad betingad fördelning (**conditional distribution** på engelska).
- En betingad fördelning är en **fördelning över en av variablerna betingat ett värde på den andra**.
- Betingat ett värde betyder givet ett värde.
- Vår jämförelse:
  1. Andelen av alla män i undersökningen som hade katt. **Här betingar vi på `Gender` = male** och får en fördelning över `Pets`.
  2. Andelen av alla kvinnor i undersökningen som hade katt. **Här betingar vi på `Gender` = female** och får en fördelning över `Pets`.

Korstabeller för att undersöka samband, forts.

- Vi räknar den betingade (betingad på kön) fördelningen (höger) från korstabellen (vänster):

		Gender		
		Female	Male	Total
Pets	Has cats	3412	2388	5800
	Has dogs	3431	3587	7018
	Has both	897	577	1474
	Total	7740	6552	14,292

		Gender		
		Female	Male	Total
Pets	Has cats	44.1%	36.4%	40.6%
	Has dogs	44.3%	54.8%	49.1%
	Has both	11.6%	8.8%	10.3%
	Total	100%	100%	100%

Figure 6: Tabell 3.1 och 3.2 i De Veaux et al. (2021). Korstabell (vänster) och betingad fördelning (höger).

- Den betingade fördelningen för `Pets` betingat på:
  - `Gender` = male fås genom att dela Male kolumnen i korstabellen med 6552 (antal män).
  - `Gender` = female fås genom att dela Female kolumnen i korstabellen med 7740 (antal kvinnor).
- Kom ihåg vår fråga: Påverkar kön preferensen för katter?
- Jämför siffran 36.4% för gruppen män mot 44.41% för gruppen kvinnor (istället för 16.7% mot 23.9%).

Korstabeller för att undersöka samband, forts.

- Notera att tabellen också innehöll den marginella fördelningen (`Pets` obetingat på `Gender`) i högerkolumnen.

		Gender		
		Female	Male	Total
Pets	Has cats	3412	2388	5800
	Has dogs	3431	3587	7018
	Has both	897	577	1474
	Total	7740	6552	14,292

		Gender		
		Female	Male	Total
Pets	Has cats	44.1%	36.4%	40.6%
	Has dogs	44.3%	54.8%	49.1%
	Has both	11.6%	8.8%	10.3%
	Total	100%	100%	100%

Figure 7: Tabell 3.1 och 3.2 i De Veaux et al. (2021). Korstabell (vänster) och betingad fördelning (höger).

- En betingad fördelning är en fördelning. En **fördelning summerar alltid till 100%** (stäm av ovan).
- Förefaller det finnas ett samband mellan `Pets` och `Gender`?
- Hur hade de två betingade fördelningarna ovan tett sig **om det inte finns ett samband mellan `Pets` och `Gender`**?

Korstabeller för att undersöka samband, forts.

- De två betingade fördelningarna illustrerade med hjälp av stapeldiagram:

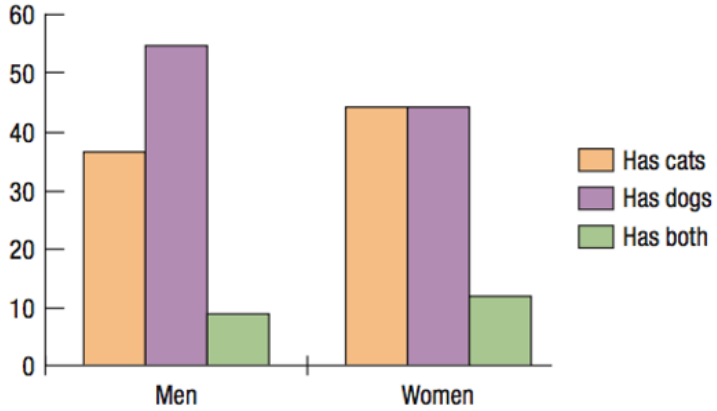

Figure 8: Figur 3.2 i De Veaux et al. (2021).

- Det föreligger ett samband. Män föredrar hundar mer än katter, medan kvinnor inte verkar ha en tydlig preferens över att enbart ha hund eller katt.
- Om det inte hade funnits ett samband hade **fördelningarna varit (ungefärliga) likadana** och **sammanfalligt (ungefärligt) med den marginella fördelningen**.

Korstabeller för att undersöka samband, forts.

- Ett vanligt problem med betingade fördelningar är att de är enkla att feltolka:

		Gender		
		Female	Male	Total
Pets	Has cats	44.1%	36.4%	40.6%
	Has dogs	44.3%	54.8%	49.1%
	Has both	11.6%	8.8%	10.3%
	Total	100%	100%	100%

Figure 9: Tabell 3.2 i De Veaux et al. (2021).

- Tabellen verkar visa att en mycket större andel män (54.8%) än kvinnor (44.3%) har enbart hundar.
- Fel! Notera att vi gör ett påstående om variabeln `Gender` (givet `Pets` = Has dogs).
- Tabellen ovan visar fördelningarna för `Pets` betingat `Gender` (dvs tvärtom det vi försöker påstå!).
- Vi kan svara på frågan genom att räkna den omvända betingade fördelningen.

Korstabeller för att undersöka samband, forts.

- Fördelningen över `Gender` betingat på Pets:

Pets	Gender		
	Female	Male	Total
Has cats	3412	2388	5800
Has dogs	3431	3587	7018
Has both	897	577	1474
Total	7740	6552	14,292

Pets	Gender		
	Female	Male	Total
Has cats	58.8%	41.2%	100%
Has dogs	48.9%	51.1%	100%
Has both	60.9%	39.1%	100%
Total	54.2%	45.8%	100%

Figure 10: Tabell 3.1 och 3.3 i De Veaux et al. (2021). Korstabell (vänster) och betingad fördelning (höger).

- Den betingade fördelningen för `Gender` betingat på:
  - `Pets` = Has cats fås genom att dela Has cats raden i korstabellen med 5800 (antal som har katter).
  - `Pets` = Has dogs fås genom att dela Has dogs raden i korstabellen med 7018 (antal som har hundar).
  - `Pets` = Has both fås genom att dela Has both raden i korstabellen med 1474 (antal som har katter och hundar).
- Vi ser nu att det är en liten andel fler män än kvinnor som enbart har hundar, 51.1% mot 48.9% (istället för 54.8% mot 44.3%).

Vad händer om vi får nya data?

- Nya data ger andra tabeller!
- Som alltid måste vi **tänka på slumpfaktorn när vi tar ett stickprov**.
- När vi jämför 51.1% mot 48.9% finns det två scenarior:
  1. I den underliggande populationen vi undersöker är det verkligen så att det är fler män än kvinnor som enbart har hundar, precis som stickprovet indikerade.
  2. Stickprovet visar att det är fler män än kvinnor som enbart har hundar på grund av slumpfakorn. I populationen stämmer inte det här påståendet.
- Del två av den här kursen behandlar **inferentiell statistik** som är ett ramverk för att kunna urskilja mellan 1. och 2..
- Del ett behandlar **deskriptiv statistik**, dvs beskriva sambanden i det faktiska stickprovet vi har.
- Låt oss göra ett tankeexperiment för att utforska slumpfaktorn i ett faktiskt experiment som utreder om det finns ett samband mellan vilken sida man föredrar att sova på och om man drömmer mardrömmar.

Vad händer om vi får nya data?, forts.

- Exempel: 63 personer tillfrågades om deras drömvanor samt vilken sida de sover på. Målet var att undersöka om det finns ett samband mellan vilken sida man sover på och drömvanor, speciellt mardrömmar.
- Första steget är att identifiera variablerna.
  1. Drömvanor (`Dreams`).
  2. Vilken sida man föredrar att sova på (`Side`).
- Klassificera variabelns typ samt bestämma dess utfall.
  1. `Dreams` en kategorisk variabel. Utfallen antas vara: mardrömmar eller drömmar.
  2. `Side` är en kategorisk variabel. Utfallen antas vara: vänster eller höger.

Vad händer om vi får nya data?, forts.

- Data från experimentet:

		Side		
		Right	Left	Totals
Dreams	Nightmares	6	9	15
	Sweet Dreams	35	13	48
	Totals	41	22	63

Figure 11: Tabell från s.100 i De Veaux et al. (2021).

- Kommentarer:
  - Av 63 personer hade 15 mardrömmar ($15/63 \approx 0.24$) och 48 drömmar ($48/63 \approx 0.76$).
  - Av de 63 sov 41 på högersida ($41/63 \approx 0.64$) och 22 på vänstersida ($22/63 \approx 0.36$).
  - Av de 41 som sov på högersida hade 6 mardrömmar ($6/41 \approx 0.15$) och 35 drömmar ($35/41 \approx 0.85$).
  - Av de 22 som sov på vänstersida hade 9 mardrömmar ($9/22 \approx 0.41$) och 13 drömmar ($13/22 \approx 0.59$).
- Det verkar som att en högre andel av de som sover på vänstersidan drömmer mardrömmar jämfört med de som sover på högersidan ($0.41 > 0.15$).

Vad händer om vi får nya data?, forts.

- Betrakta igen data från experimentet:

		Side		
		Right	Left	Totals
Dreams	Nightmares	6	9	15
	Sweet Dreams	35	13	48
	Totals	41	22	63

Figure 12: Tabell från s.100 i De Veaux et al. (2021).

- Tankeexperiment för att utreda om detta beror på slumpen eller inte: Antag att det inte hade spelat någon roll vilken sida man sover på.
- Då skulle de 15 mardrömmarna **fördelat sig slumpmässigt bland kategorierna högersida och vänstersida**.
- Notera att det finns en högre andel ($41/63 \approx 0.65$) som sover på högersida jämfört med vänstersida ($22/63 \approx 0.35$).
- Om inte sidan man sover på hade spelat roll så hade **i genomsnitt**:
  - 65% av 15 mardrömmar (9.75) varit bland de som sov på högersida.
  - 65% av 15 mardrömmar (5.25) varit bland de som sov på vänstersida.

Vad händer om vi får nya data?, forts.

- Ett hypotetiskt nytt stickprov från tankeexperimentet:

		Side		
		Right	Left	Totals
Dreams	Nightmares	11	4	15
	Sweet Dreams	30	18	48
	Totals	41	22	63

Figure 13: Tabell från s.100 i De Veaux et al. (2021).

- Jämför med våra ursprungsdata:

		Side		
		Right	Left	Totals
Dreams	Nightmares	6	9	15
	Sweet Dreams	35	13	48
	Totals	41	22	63

Figure 14: Tabell från s.100 i De Veaux et al. (2021).

- Hur skulle antalet som sover på vänstersida och drömmer mardrömmar se ut om vi slumpade ut 1000 tabeller?

Vad händer om vi får nya data?, forts.

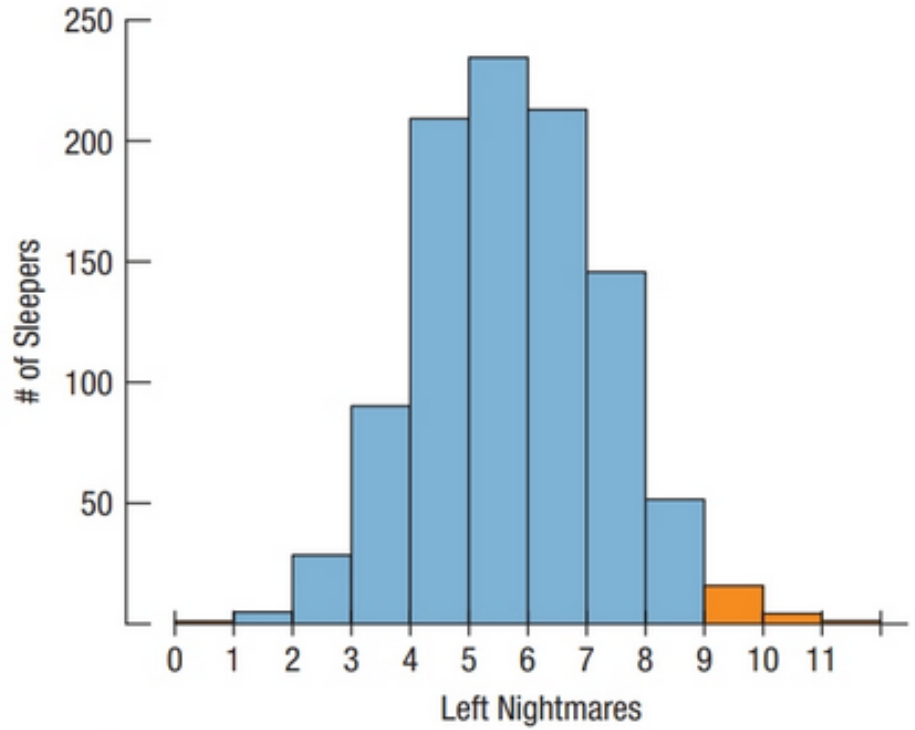

Figure 15: Figur 3.3 i De Veaux et al. (2021).

- Vårt experiment visade 9 som sov på vänstersida och hade mardrömmar.
- Vilken slutsats ger tankeexperimentet om **hypotesen** att den sidan vi sover på inte har något samband med att drömma mardrömmar?

Grafiska illustrationer av korstabeller

- En korstabell över `Class` och `Survival` (överlevnad) i Titanic datasetet:

			Class				
			First	Second	Third	Crew	Total
Survival	Alive	Count	201	119	180	212	**712**
		% of Column	62.0%	41.8%	25.4%	23.8%	**32.3%**
	Dead	Count	123	166	530	677	**1496**
		% of Column	38.0%	58.2%	74.6%	76.2%	**67.7%**
	Total	Count	**324**	**285**	**710**	**889**	**2208**
			100%	**100%**	**100%**	**100%**	**100%**

Figure 16: Från lärarmaterialet skapat av utgivaren av De Veaux et al. (2021).

- Stapeldiagram för `Survival` (betingat på `Class`):

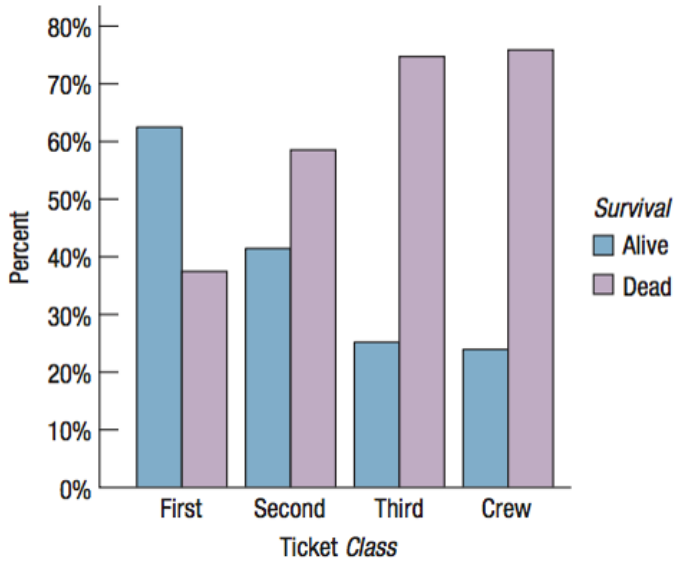

Figure 17: Figur 3.4 i De Veaux et al. (2021).

Grafiska illustrationer av korstabeller, forts.

- Ett alternativ till stapeldiagrammet är ett staplat stapeldiagram (**stacked barchart** på engelska).
- Staplat stapeldiagram för `Class` (betingat på `Survival`):

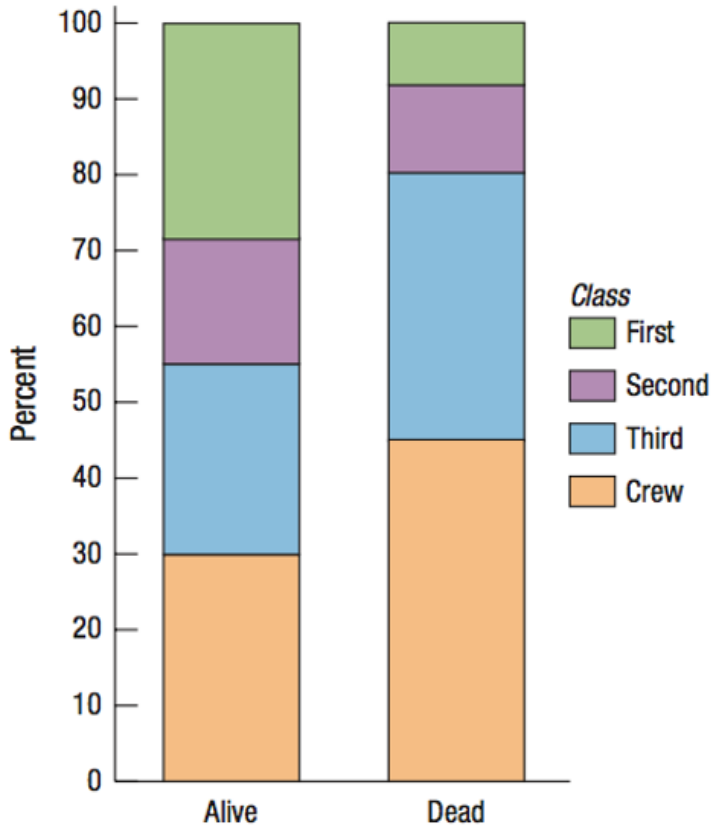

Figure 18: Figur 3.5 i De Veaux et al. (2021).

Grafiska illustrationer av korstabeller, forts.

- Innehåller ingen information om marginella fördelningen för `Survival`.
- En **mosaic plot** är ett staplat stapeldiagram som också visar marginella fördelningen för survival:

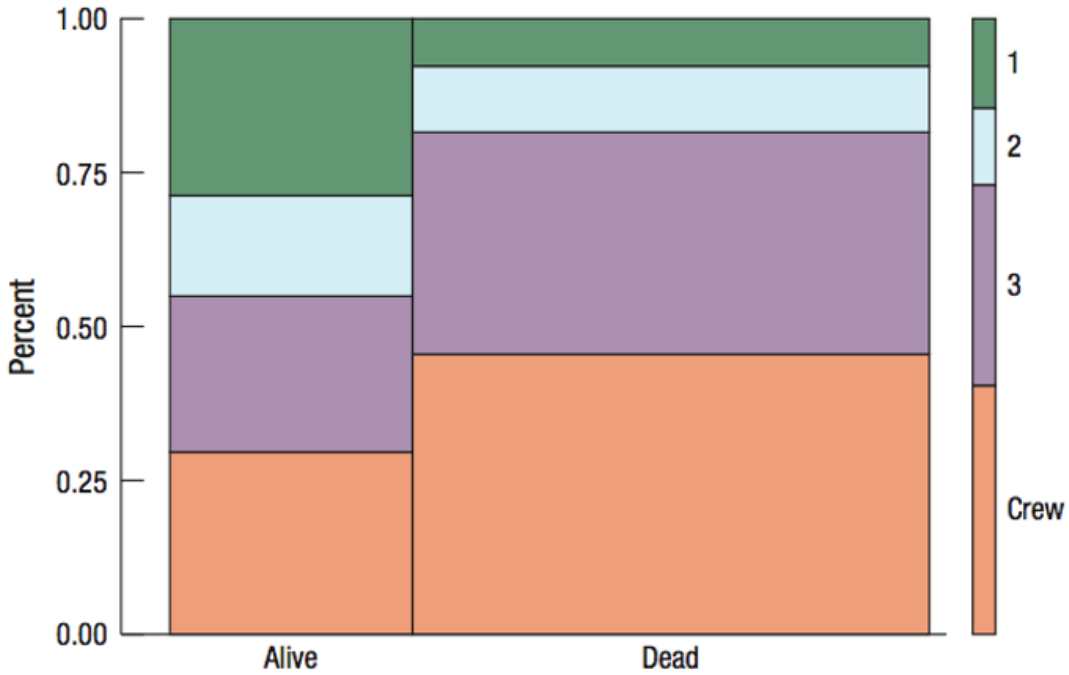

Figure 19: Figur 3.6 i De Veaux et al. (2021).

- Vi ser att ungefär 1/3 överlevde och 2/3 dog.

Utmaningar med sambandstolkningar

- När vi hittar ett samband mellan två variabler är det viktigt att förstå begränsningarna i resultatet.
- Det eviga problemet: Samband och kausalitet (**relationship** och **causality** på engelska) är inte samma sak.
- Ett tillsynes samband mellan två variabler kan istället förklaras av en tredje variabel. Kallas för **lurking variable** på engelska.
- **Simpsons paradox**: Ett samband mellan två variabler försvinner, eller omvänds, när stickprovet från populationen delas in i subgrupper.
- Ett klassiskt exempel på Simpsons paradox är en undersökning som undersökte om kön påverkar antagning till doktorandprogrammen i University of California, Berkley (UC Berkley).

Utmaningar med sambandstolkningar, forts

- Data från studien:

	Admit	Reject	%Admit
Men	1158	1493	43.7%
Women	557	1278	30.4%

Figure 20: Tabell från s.106 i De Veaux et al. (2021).

- Betingade fördelningar för antagning (betingat på kön):

Gender	**Admit (%)**	**Reject (%)**
Men	43.7	57.3
Women	30.4	69.6

- Det verkar som en tydlig diskriminering: Endast 30.4% av kvinnor antas till utbildningen jämfört med 43.7% för männen.
- Finns det någon annan variabel än kön som kan förklara skillnaden?

Utmaningar med sambandstolkningar, forts

- Vilken institution sökte man till?

School	Male Admits	Female Admits	Male%	Female%
A	512	89	62.1%	82.4%
B	313	17	60.2%	68.0%
C	120	202	36.9%	34.1%
D	138	131	33.1%	34.9%
E	53	94	27.7%	23.9%
F	22	24	5.9%	7.0%

Figure 21: Tabell från s.106 i De Veaux et al. (2021).

- Kommentarer:
  - Institution A och B är enklast att komma in i (för bägge könen).
  - Antagningsgraden för kvinnor $\approx$ män. Ibland mycket högre.
  - En hög andel män söker till institution A och B.
  - En hög andel kvinnor söker till de svårare utbildningarna.
- Slutsats: Kvinnor söker oftare till program som är mycket svårare att komma in på, och det förklarar deras lägre antagningsgrad.

Utmaningar med sambandstolkningar, forts

- Simpsons paradox i det vardagliga livet:

Figure 22: Från @PeterSweden7s Twitter profil.

- Ovaccinerade och de med endast en dos har överlag en bättre hälsa än de som tar sin fjärde dos (tillhör riskgrupper).

Samband mellan tre kategoriska variabler

- Det finns ibland anledning att undersöka samband mellan tre kategoriska variabler.
- Principen är densamma, men det är inte överskådligt med en tre dimensionell korstabell.
- Man kan göra räkningarna givet olika värden av den tredje variabeln.
- På dejting websidan så är droganvändning en annan variabel som registreras.
- Betingade fördelningar över `Pets` (betingat på `Gender` och `Drugs`):

Drugs = "No"

Pets	Gender: Female	Male	Total
Has cats	40.8%	32.3%	37.1%
Has dogs	47.0%	58.5%	52.0%
Has both	12.2%	9.18%	10.9%
Total	100%	100%	100%

Drugs = "Yes"

Pets	Gender: Female	Male	Total
Has cats	51.7%	44.2%	47.7%
Has dogs	36.3%	46.1%	41.5%
Has both	12.1%	9.70%	10.8%
Total	100%	100%	100%

Figure 23: Tabell 3.6 i De Veaux et al. (2021).

Samband mellan tre kategoriska variabler, forts.

- Betingade fördelningar för `Class` (betingat `Survival` och `Gender`):

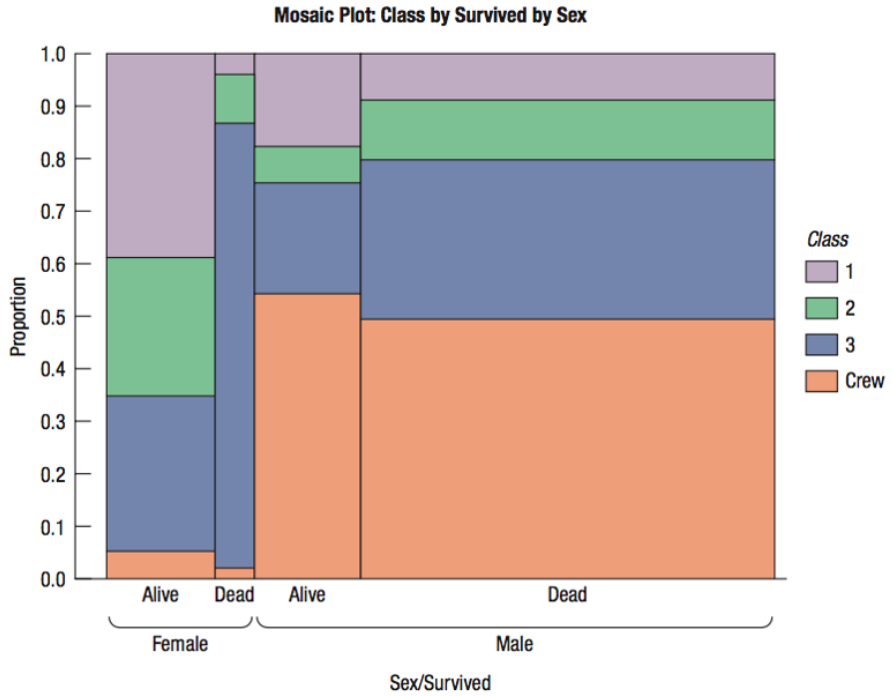

Figure 24: Tabell från s.105 i De Veaux et al. (2021).

- Kommentarer:
  - Högre andel män än kvinnor ombord.
  - Högre andel kvinnor överlevde jämfört med män.

Samband mellan tre kategoriska variabler, forts.

- Betingade fördelningar för `Class` (betingat `Survival` och `Gender`):

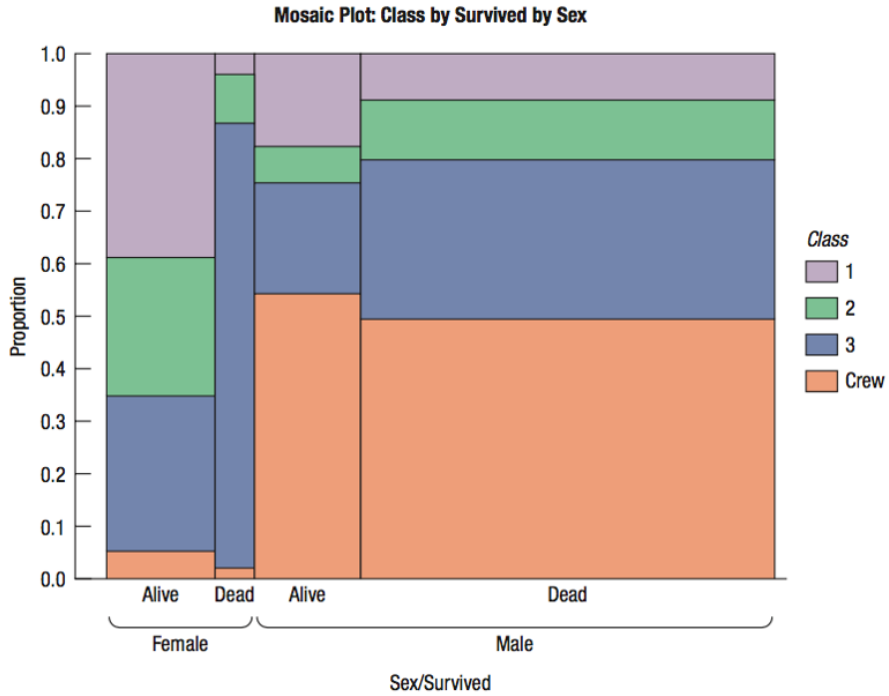

Figure 25: Tabell från s.105 i De Veaux et al. (2021).

- Kommentarer (forts.):
  - `Class` fördelningen mellan de som överlevde och dog ganska lika för män.
  - `Class` fördelningen mellan de som överlevde och dog väldigt olika för kvinnor.

Beräkningar i R

- I R kan `mosaic` paketet användas för att beräkna korstabeller och samt göra diverse plottar.
- Dataset över forcerad utandningsvolym hos 606 barn och ungdomar[1]. Kategoriska variabler rökning, kön, åldersgrupp
- Räkna tre typer av fördelningar (marginell, simultan, betingad):

```
library(mosaic)
load("Datasets/FevChildren.RData")  Smoking coded as yes = 1 and no = 0
 marginal smoking
tally(~smoking, data = FevChildren, format = "prop")
 marginal age.group
tally(~age.group, data = FevChildren, format = "prop")
 joint distribution
tally(~smoking + age.group, data = FevChildren, format = "prop")
 conditional (on age.group)
tally(~smoking | age.group, data = FevChildren, format = "prop")
 conditional (on smoking)
tally(~age.group | smoking, data = FevChildren, format = "prop")
 conditional (on age.group and gender)
tally(~smoking | age.group + gender, data = FevChildren, format = "prop")
```

[1] Skapad från https:raw.githubusercontent.com/GTPB/PSLS20/master/data/fev.txt.

Beräkningar i R, forts.

```
> head(FevChildren)
    fev height smoking gender age age.group
1 1.708 144.78       0      f   9       6-9
2 1.724 171.45       0      f   8       6-9
3 1.720 138.43       0      f   7       6-9
4 1.558 134.62       0      m   9       6-9
5 1.895 144.78       0      m   9       6-9
6 2.336 154.94       0      f   8       6-9
> tally(~ smoking, data = FevChildren, format = "prop")
smoking
        0         1
0.8993399 0.1006601
> tally(~ age.group, data = FevChildren, format = "prop")
age.group
      6-9     10-14     15-17
0.4455446 0.4884488 0.0660066
> tally(~ smoking + age.group, data = FevChildren, format = "prop")
       age.group
smoking         6-9       10-14       15-17
      0 0.443894389 0.420792079 0.034653465
      1 0.001650165 0.067656766 0.031353135
> tally(~ smoking | age.group, data = FevChildren, format = "prop")
       age.group
smoking         6-9       10-14       15-17
      0 0.996296296 0.861486486 0.525000000
      1 0.003703704 0.138513514 0.475000000
```

Figure 26: Output från R.

Beräkningar i R, forts.

```
> tally(~ age.group | smoking, data = FevChildren, format = "prop")
         smoking
age.group          0          1
    6-9   0.49357798 0.01639344
    10-14 0.46788991 0.67213115
    15-17 0.03853211 0.31147541
> tally(~ smoking | age.group + gender, data = FevChildren, format = "prop")
, , gender = f

       age.group
smoking         6-9       10-14       15-17
      0 1.000000000 0.808510638 0.411764706
      1 0.000000000 0.191489362 0.588235294

, , gender = m

       age.group
smoking         6-9       10-14       15-17
      0 0.992647059 0.909677419 0.608695652
      1 0.007352941 0.090322581 0.391304348
```

Figure 27: Output från R.

Beräkningar i R

- Lite olika figurer med `base R` samt `mosaic`:

```
library(mosaic) library(vcd)
 Pretty colors
cs <- brewer.pal(3, "Paired")
load("Datasets/FevChildren.RData")  Smoking coded as yes = 1 and no = 0
 Stacked barchart (base R)
table <- tally(~smoking | age.group, data = FevChildren, format = "prop")
barplot(table, col = c(cs[1], cs[2]), xlab = "Age.gr", ylim = c(0, 1.3), legend = c("No smoker", "Smoker"))
 Barchart (non-stacked) (mosaic)
bargraph(~smoking | age.group, data = FevChildren, type = "prop")
 Mosaic plot (vcd)
mosaic(~smoking | age.group, data = FevChildren)
```

- Lättare att göra stacked barcharts i `base R`.
- `bargraph()` är från `mosaic` paketet.
- Mosaicplotten är från `vcd` paketet.

Beräkningar i R, forts.

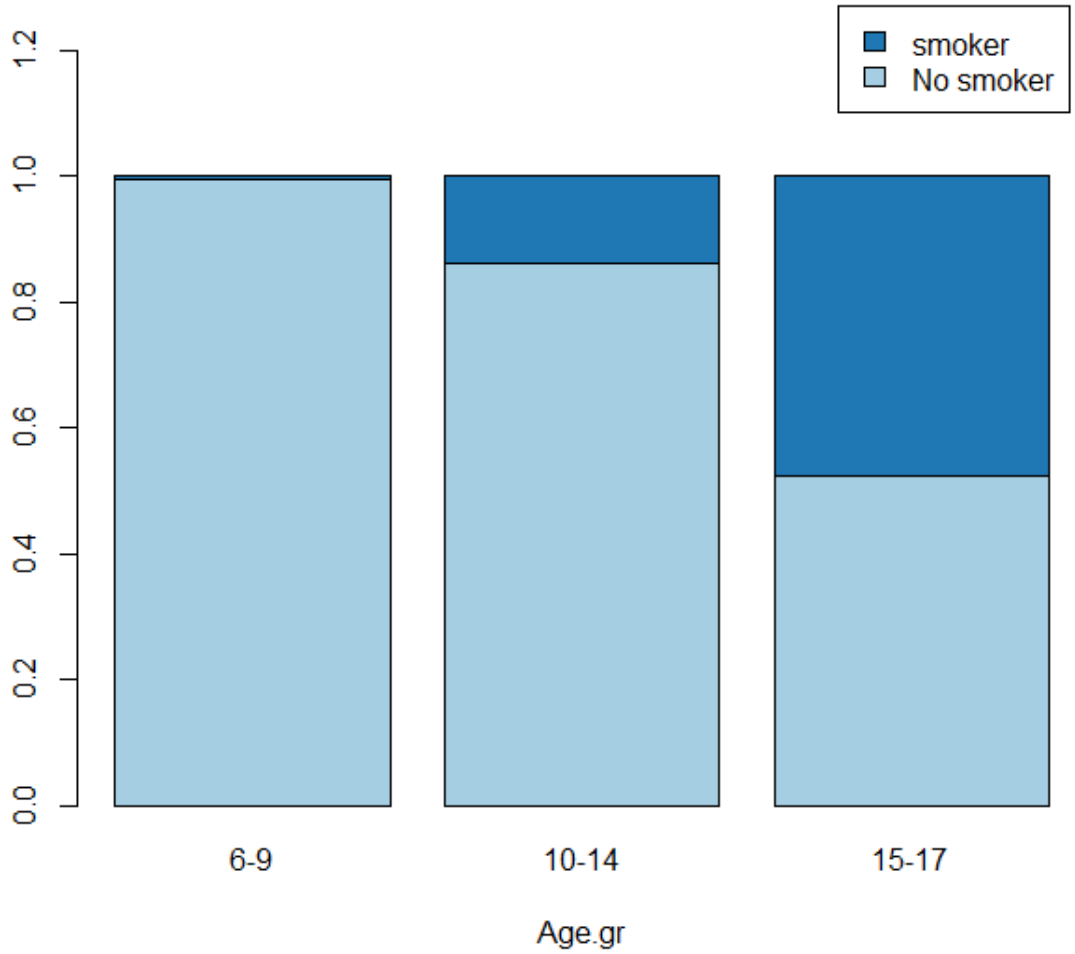

Figure 28: Stacked barchart för `smoking` betingat på `age.gr`.

Beräkningar i R, forts.

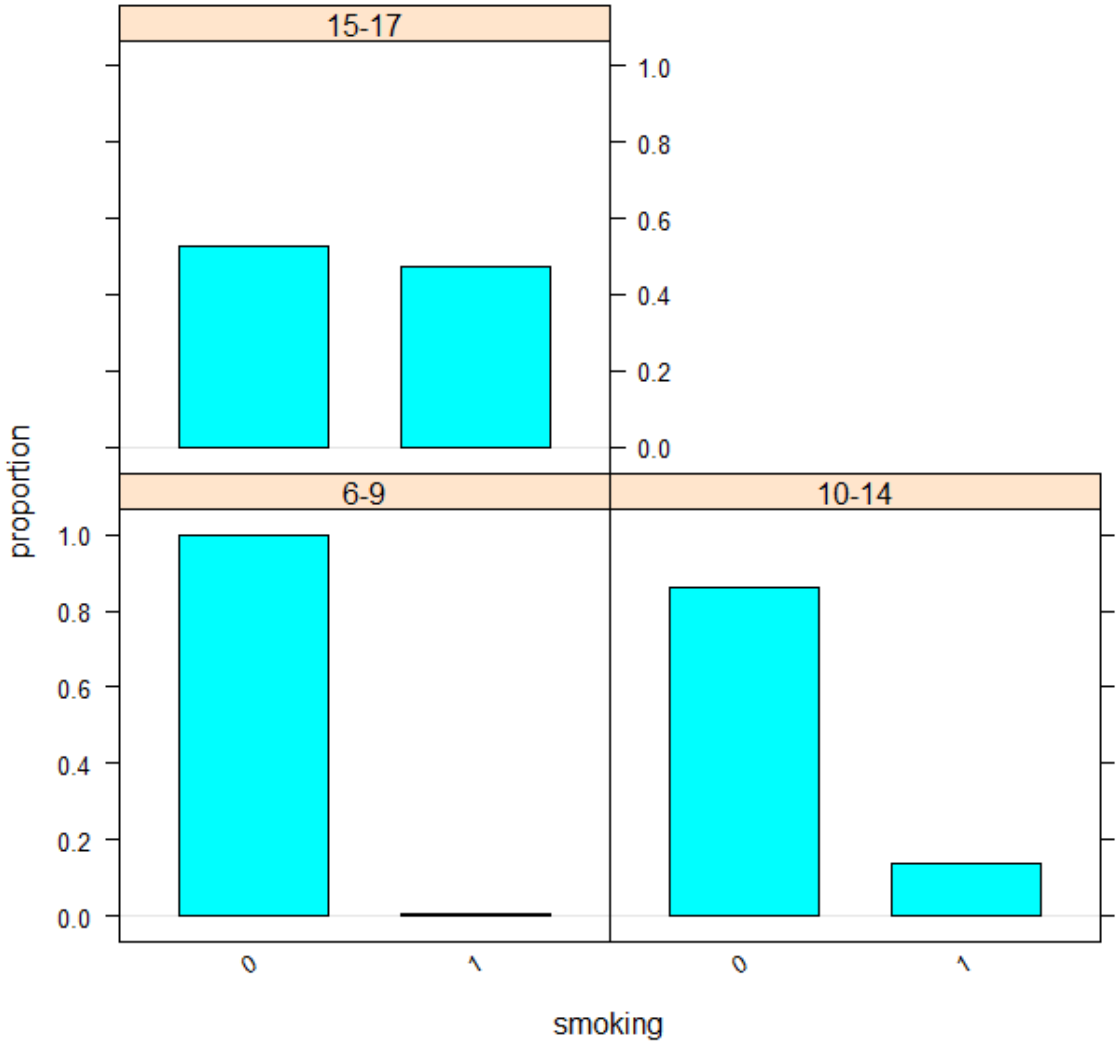

Figure 29: Barchart för `smoking` betingat på `age.gr`.

Beräkningar i R, forts.

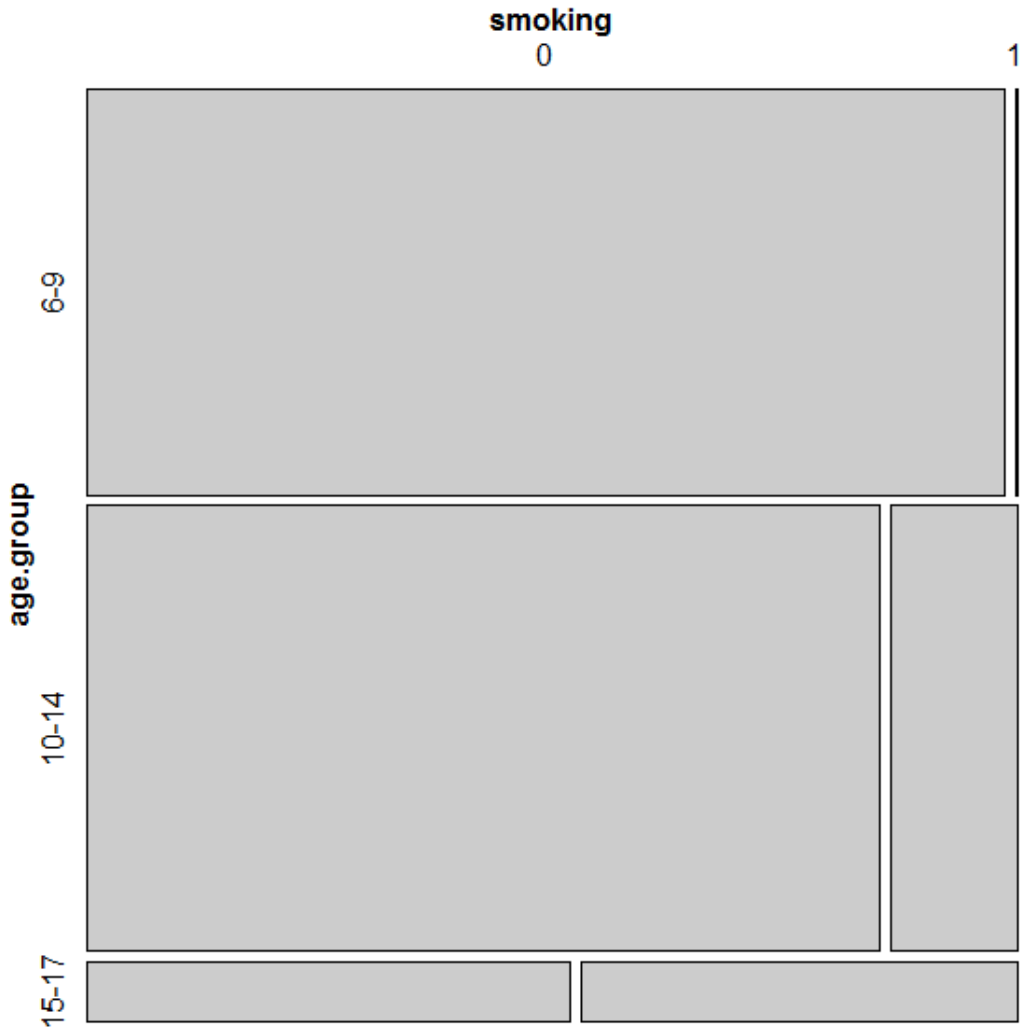

Figure 30: Mosaicplot för `smoking` betingat på `age.gr`.

References I

De Veaux, R. D., Velleman, P., and Bock, D. (2021). *Stats: Data and Models*. Pearson, Harlow, United Kingdom, fifth edition.