Statistik och dataanalys I

F3: Samband mellan kategoriska variabler

Valentin Zulj

Statistiska institutionen

Stockholm University

Dagens föreläsning

- Frekvenstabeller
- Korstabeller
  - Simultana-, marginella-, relativa-, och betingade fördelningar
- Kategoriska variabler och samband
  - Kausalitet och Simpsons paradox

Frekvenstabeller

En variabel

- En frekvenstabell redovisar antalet observationer i varje kategori
- En relativ frekvenstabell visar andel istället för antal

Class	Count
First	324
Second	285
Third	710
Crew	889

Table 2.2
A frequency table of the *Titanic* passengers.

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

Table 2.3
A relative frequency table for the same data.

- Har vi flera kategoriska variabler kan vi göra en frekvenstabell per variabel
- Tabeller över enskilda variabler visar dock **inte samband** mellan variabler

Korstabeller

- En **korstabell** *(en: contingency table/crosstable)* visar två variabler samtidigt
- Korstabellen nedan visar resultat ur en svensk studie som undersökte om det finns samband mellan prostatacancer och hur ofta en person äter fisk

Fish Consumption	Prostate Cancer: No	Prostate Cancer: Yes	Total
Never/Seldom	110	14	124 (2.0%)
Small Part of Diet	2420	201	2621 (41.8%)
Moderate Part	2769	209	2978 (47.5%)
Large Part	507	42	549 (8.8%)
Total	5806 (92.6%)	466 (7.4%)	6272 (100%)

Två kategoriska variabler

Fish Consumption	Prostate Cancer: No	Prostate Cancer: Yes	Total
Never/Seldom	110	14	124 (2.0%)
Small Part of Diet	2420	201	2621 (41.8%)
Moderate Part	2769	209	2978 (47.5%)
Large Part	507	42	549 (8.8%)
Total	5806 (92.6%)	466 (7.4%)	6272 (100%)

- En variabel delar in deltagare i **kategorier baserat på hur ofta de äter fisk:**
  - Sällan/adrig, lite, måttligt, mycket
- Den andra variabeln delar in deltagarna i **två kategorier:**
  - Diagnosticerades/diagonsticerades inte med prostatacancer under studieperioden

Två kategoriska variabler – Korstabeller i R

- För att skapa en korstabell i `R` använder vi funktionen `tally()`
- Skillnaden mot en frekvenstabell är att vi lägger till en extra variabel efter ett plustecken

```
> tally(~ diet + cancer, data=fish, format="count") # lägg till cancer med "+"
>           cancer
> diet         No  Yes
>   Never     110   14
>   Small    2420  201
>   Moderate 2769  209
>   Large     507   42
```

Fish Consumption	Prostate Cancer: No	Prostate Cancer: Yes	Total
Never/Seldom	110	14	124 (2.0%)
Small Part of Diet	2420	201	2621 (41.8%)
Moderate Part	2769	209	2978 (47.5%)
Large Part	507	42	549 (8.8%)
Total	5806 (92.6%)	466 (7.4%)	6272 (100%)

- I det **gula fältet** ser vi en **simultanfördelning** *(en: joint distribution)*
- En simultanfördelning delar in observationerna i grupper, baserat på två eller fler variabler
- Om en variabel har 4 kategorier och en annan variabel har 2 kategorier får vi sammanlagt $4 \cdot 2 = 8$ kategorier, en för varje möjlig kombination
- Den simultana fördelningen visar exempelvis att det fanns det 110 deltagare i studien som sällan/aldrig åt fisk **och** som inte fick prostatacancer
- Om vi lägger samman talen i det gula fältet blir summan 6272, som är det totala antalet observationer (dvs antalet deltagare i studien)

Fish Consumption	Prostate Cancer: No	Prostate Cancer: Yes	Total
Never/Seldom	110	14	124 (2.0%)
Small Part of Diet	2420	201	2621 (41.8%)
Moderate Part	2769	209	2978 (47.5%)
Large Part	507	42	549 (8.8%)
Total	5806 (92.6%)	466 (7.4%)	6272 (100%)

- **Marginalfördelningen** för **den ena variabeln** visar antalet observationer per kategori, utan att vi tar någon hänsyn till den andra variabeln
- Tabellens högermarignal ger marginalfördelningen för `Fish Consumption`
- Vi ser t.ex. att totalt 124 deltagare sällan/aldrig åt fisk, och tar ingen hänsyn till `Prostate Cancer`

Fish Consumption	Prostate Cancer: No	Prostate Cancer: Yes	Total
Never/Seldom	110	14	**124 (2.0%)**
Small Part of Diet	2420	201	**2621 (41.8%)**
Moderate Part	2769	209	**2978 (47.5%)**
Large Part	507	42	**549 (8.8%)**
Total	**5806 (92.6%)**	**466 (7.4%)**	**6272 (100%)**

- Tabellens bottenmarginal visar marginalfördelningen för variabeln `Prostate Cancer`
- Totalt 466 testdeltagare diagnostiserades med prostatacaner under studien
- Varje marginalfördelning summerar till det totala antalet observationer, enligt nedan

$$124 + 2621 + 2978 + 549 = 5806 + 466 = 6272$$

- I `R` lägger vi till marginaler i en korstabell genom att sätta `margins = TRUE`, enligt nedan

```
> tally(~ diet + cancer, data=fish, format="count", margins=TRUE)
>           cancer
> diet         No  Yes Total
>   Never     110   14   124
>   Small    2420  201  2621
>   Moderate 2769  209  2978
>   Large     507   42   549
>   Total    5806  466  6272
```

- Om vi jämför marginalfördelningarna med enskilda frekvenstabeller för vardera variabel, vad kommer vi att se då?
- Vi kommer se att varje frekvenstabell ger marginalförelningen för respektive variabel
- Frekvenstabeller visar antalet observationer per kategori, utan hänsyn till andra variabler
- Detta var vår definition av marginalfördelningar!

```
> ### Marginalfördelningar i en korstabell
>
> tally(~ diet + cancer, data=fish, format="count", margins=TRUE)
>           cancer
> diet          No  Yes Total
>   Never      110   14   124
>   Small     2420  201  2621
>   Moderate  2769  209  2978
>   Large      507   42   549
>   Total     5806  466  6272
```

```
> ### Separata frekvenstabeller
>
> tally(~ diet, data=fish, format="count", margins=TRUE)
> diet
>    Never    Small Moderate    Large    Total
>      124     2621     2978      549     6272
> tally(~ cancer, data=fish, format="count", margins=TRUE)
> cancer
>    No   Yes Total
>  5806   466  6272
```

- Genom att dela varje frekvens med det totala antalet observationer, och sedan multiplicera med 100, kan vi få en **relativ frekvenstabell**
- Det var t.ex. 507 som *åt mycket fisk* och som *inte fick cancer*, vilket motsvarar

$$\frac{507}{6272} \cdot 100 = 8.0835\%$$

- En korstabell med sådana procentsatser beskriver den **relativa fördelningen**, och vi kan skapa en sådan genom att lägga till `format = "percent"` i `tally()`

```
> tally(~ diet + cancer, data=fish, format="percent", margins=TRUE)
>           cancer
> diet                No         Yes       Total
>   Never      1.7538265   0.2232143   1.9770408
>   Small     38.5841837   3.2047194  41.7889031
>   Moderate  44.1485969   3.3322704  47.4808673
>   Large      8.0835459   0.6696429   8.7531888
>   Total     92.5701531   7.4298469 100.0000000
```

Samband mellan kategoriska variabler

- Vi har pratat om en mängd olika tabeller som beskriver olika typer av fördelningar, men ingen av dem är riktigt bra för att illustrera **samband** mellan variabler
- Vi tänker t.ex. att vi vill veta hur andelen som fick prostatacancer skiljer sig mellan de olika dietgrupperna
- En relativ korstabell innehåller andelar med hänsyn till **alla** observationer
- Om vi jämför andelen som fått cancer i grupperna `Never` och `Small` ser vi att det är 0.22% i den förra, och 3.2% i den senare
- Skillnaden blir ungefär 3 procentenheter, men detta är som andel av alla observationer
- Vi tar inte hänsyn till att det finns olika många observationer i varje dietkategori, och detta snedvrider resultaten!
- För att hantera denna typ av problem använder vi **betingade fördelningar**

Betingade fördelningar

- För att kunna se samband mellan variablerna introducerar vi begreppet **betingad fördelning** (*en: conditional distribution*)
- En betingad fördelning är fördelningen av en variabel, **givet** ett värde på en annan variabel
- Vi kan exempelvis vilja undersöka sambandet mellan prostatacancer och fisk i dieten genom att ställa frågorna:
  - Hur stor andel av de som *aldrig/sällan* åt fisk fick prostatacancer?
  - Hur stor andel av de som åt *lite fisk* fick prostatacancer?
  - Hur stor andel av de som åt *måttligt med fisk* fick prostatacancer?
  - Hur stor andel av de som åt *mycket fisk* fick prostatacancer?

Betingade fördelningar

- När vi ställer de här frågorna är vi intresserade av risken för prostatacancer **betingat på** en viss diet
- För att få svaren tittar vi på en diet-kategori i taget, där en diet-kategori definieras av hur ofta en person äter fisk
- När vi gör detta tar vi hänsyn till att de olika dietgrupperna är olika stora
- Till slut kan vi jämföra personer som tillhör olika dietgrupper för att se om det finns samband

Betingade fördelningar

- När vi räknar ut betingade fördelningar intresserar vi oss bara för den kategori som vi betingar på
- Vill vi ha fördelningen av variabeln prostatacancer betingat på att individen aldrig/sällan äter fisk, ignorerar vi de individer som inte tillhör den kategorin

		Prostate Cancer		
		No	Yes	Total
Fish Consumption	Never/Seldom	110	14	**124 (2.0%)**
	Small Part of Diet	2420	201	**2621 (41.8%)**
	Moderate Part	2769	209	**2978 (47.5%)**
	Large Part	507	42	**549 (8.8%)**
	Total	**5806 (92.6%)**	**466 (7.4%)**	**6272 (100%)**

Betingade fördelningar

- När vi räknar ut betingade fördelningar intresserar vi oss bara för den kategori som vi betingar på
- Vill vi ha fördelningen av variabeln prostatacancer betingat på att individen aldrig/sällan äter fisk, ignorerar vi de individer som inte tillhör den kategorin

		Prostate Cancer		
		No	Yes	Total
Fish Consumption	Never/Seldom	110	14	124 (2.0%)
	[redacted]			

- 14 deltagare fick prostatacancer, och 110 fick inte det
- Vi tar fram den relativa frekvensen på samma sätt som när vi bara har en variabel
- Andelen med cancer är $14/124 = 0.1129$ och andelen utan cancer är $110/124 = 0.8871$

Betingade fördelningar

- I `R` kan vi skapa en tabell som visar den betingade fördelningen för varje kategori
- I `tally()` skriver vi `cancer|diet`, som uttalas *cancer givet/betingat på diet*
- Tabellen nedan anger fördelningen i procent, vilket är mycket vanligt

```
> # Tillägget |> t() vänder på tabellen. Testa både med och utan.
>
> tally(~cancer|diet,data=fish,format="percent",margins=TRUE) |> t()
>           cancer
> diet               No        Yes      Total
>   Never     88.709677  11.290323 100.000000
>   Small     92.331171   7.668829 100.000000
>   Moderate  92.981867   7.018133 100.000000
>   Large     92.349727   7.650273 100.000000
```

```
>           cancer
> diet                No         Yes       Total
>   Never     88.709677  11.290323 100.000000
>   Small     92.331171   7.668829 100.000000
>   Moderate  92.981867   7.018133 100.000000
>   Large     92.349727   7.650273 100.000000
```

- Notera att **varje rad** summerar till 100 procent, eftersom varje rad representerar en av de kategorier som vi betingar fördelningen på
- Varje rad i tabellen kan alltså ses som en egen självständig fördelning
- Med denna tabell kan vi visa skillnader mellan grupperna på ett nytt sätt
  - Bland dem som aldrig eller sällan åt fisk hade drygt 11 procent diagnosticerats med prostatacancer
  - Bland övriga grupper är motsvarande siffra lite över 7 procent
- Jämför själva med den relativa fördelningen – är slutsatserna desamma?

Avskalat exempel

- Antag nu att vi har två olika stora datamaterial, som i tabellerna nedan

	Ej Diagnos	Diagnos	
Fisk	50 (25%)	50 (25%)	
Ej Fisk	80 (40%)	20 (10%)	
			200

	Ej Diagnos	Diagnos	
Fisk	50 (16.7%)	50 (16.7%)	
Ej Fisk	160 (53.3%)	40 (13.3%)	
			300

- Utifrån dessa drar vi **felaktigt** slutsatsen att sambanden är olika i de två datamaterialen
- Vi beräknar de betingade fördelningarna, och får

	Ej Diagnos	Diagnos	
Fisk	50%	50%	100%
Ej Fisk	80%	20%	100%

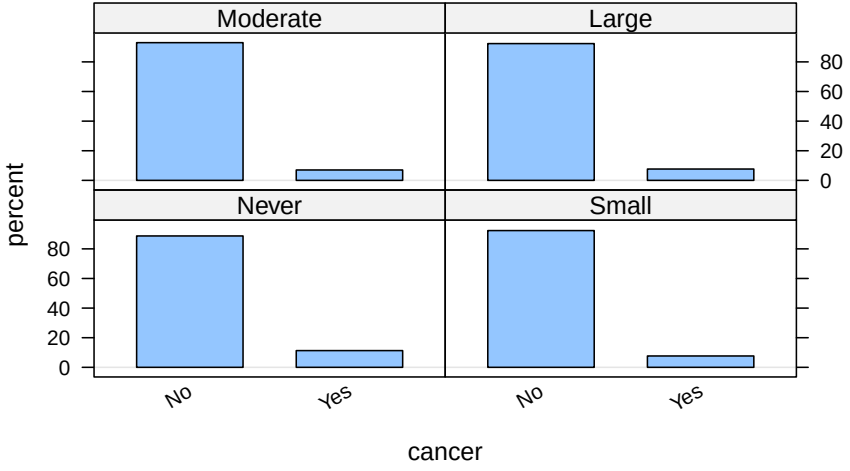
	Ej Diagnos	Diagnos	
Fisk	50%	50%	100%
Ej Fisk	80%	20%	100%

- Dessa ger en annan bild: sambanden är likadana i båda stickproven!
- Det är de olika stora stickproven som stökar till det här, och detta är en av anledningarna till att vi alltid vill använda betingade fördelningar i situationer som dessa

Figurer och samband

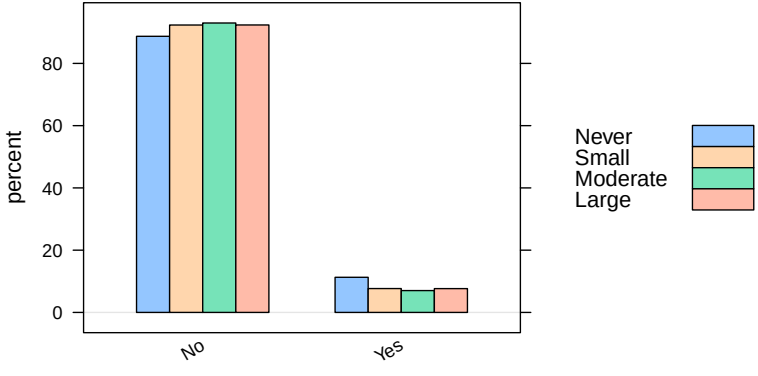
- Vi kan ge en snabbare överblick av samma information med en graf

```
> bargraph(~cancer|diet, data=fish, type="percent")
```



- Vi kan också gruppera efter variabeln cancer
- Färgerna indikerar diet, och två staplar av samma färg summerar till 100%
- Vi ser att cancer förekom något mer bland deltagare som aldrig/sällan åt fisk

```
> bargraph(~cancer, groups=diet, data=fish, type="percent")
```



- Givet en viss diet kan vi nu säga hur stor andel som har diagnosticerats med cancer
- Betyder det att vi även kan säga hur stor andel av dem som diagnosticerat med cancer som har en viss diet?
- **Nej, inte utan att räkna ut nya betingade värden**
- Den här gången vill vi veta hur ofta personer äter fisk
  - betingat på att en person har diagnosticerats med cancer
  - betingat på att en person inte har diagnosticerats med cancer

- Vi tar den här gången fram en tabell som är betingad på cancer-variabeln
- Säg att vi vill räkna ut andelen som aldrig/sällan åt fisk betingat på att de fick cancer, vi gör då på liknande sätt som tidigare:
  - Vi *ignorerar* då alla deltagare som *inte* fick cancer
  - Antalet deltagare som fick prostatacancer var 466, och av dem åt 14 aldrig/sällan åt fisk
  - Bland dem som fick cancer var den procentuella andelen som aldrig/sällan åt fisk

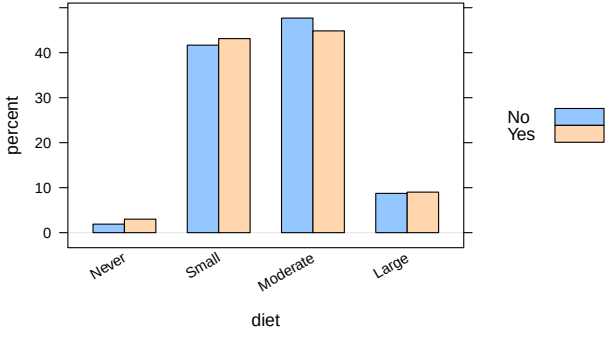
$$\frac{14}{466} \cdot 100 = 3.004292\%$$

```
> tally(~ diet | cancer, data=fish, format="percent", margins=TRUE)
>           cancer
> diet                No        Yes
>   Never       1.894592   3.004292
>   Small      41.681020  43.133047
>   Moderate   47.692043  44.849785
>   Large       8.732346   9.012876
>   Total     100.000000 100.000000
```

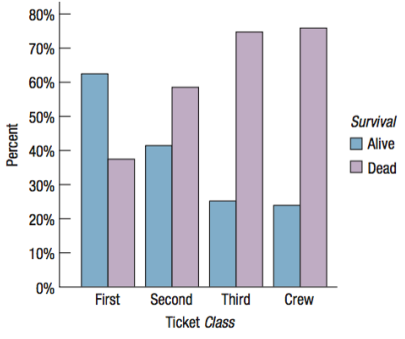
- För att göra motsvarande beräkning i `R` sätter vi `diet|cancer` i `tally()`
- Notera att det nu är **kolumnerna** som summerar till 100, då vi nu har gjort en separat frekvenstabell för varje kolumn
- Något som framgår här, och som vi inte såg när vi betingade fördelningen på diet, är att det är få av deltagarna som aldrig eller sällan äter fisk

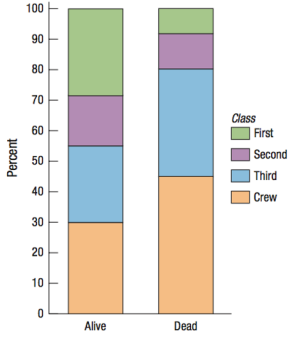
- Vi gör ett stapeldiagram grupperat på cancer-variabeln
- De blå staplarna summerar till 100 procent, och staplarna i beige summerar också till 100 procent
- Vi ser att bland de som aldrig eller sällan äter fisk är cancerdiagnoserna något överrepresenterade

```
> bargraph(~diet, groups=cancer, data=fish, type="percent")
```



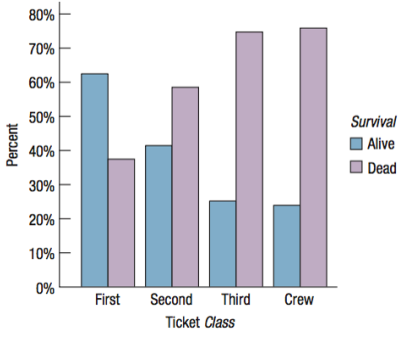
- Stapeldiagrammet till vänster är betingat på variabeln *Class*, och hjälper oss att besvara frågan om hur stor andel som överlevde inom varje biljettklass
- Stapeldiagrammet till höger är betingat på variabeln *Survived*, och hjälper oss att besvara frågan om hur stor andel överlevande/ej överlevande som reste i en viss klass

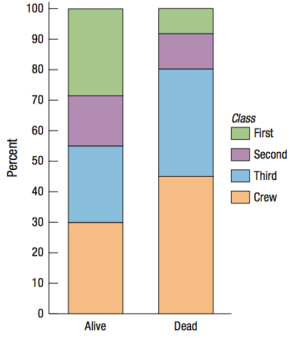




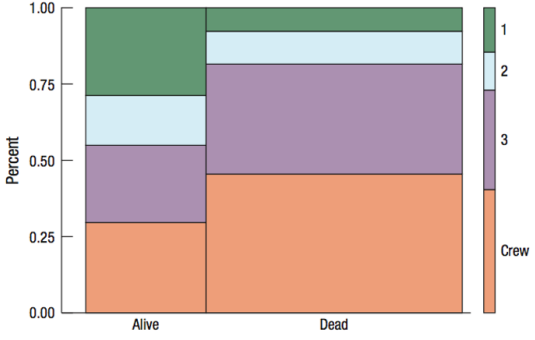
Fler exempel med Titanic

- Stapeldiagrammet till vänster ger en bild av andelen som överlevde katastrofen, men ingen information om hur många som ingick i varje klass
- Stapeldiagrammet till höger ger en bild av hur många som reste i respektive klass, men ingen information om andelen som överlevde

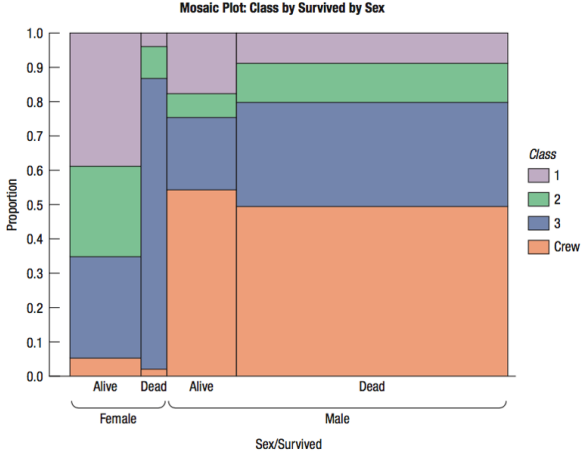




- En **mosaic plot** ger en lite utvecklad bild av en simultan fördelning än ett vanligt stapeldiagram
- Varje ruta har en **area** som motsvarar andelen observationer som rutan representerar



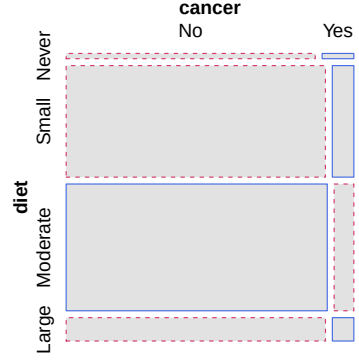
- En mosaic plot kan dessutom visa fler än två variabler i samma bild
- Grafen nedan visar variablerna *Class*, *Survived* och *Gender*



Kategoriska variabler och samband - mosaic plot

- I `R` skapar vi en mosaic-plot med funktionen `mosaic()` i paketet `vcd`
- Ett exempel med studien om fisk och prostatacancer ges nedan, och ni kommer arbeta mer med detta i datorlab 3

```
> mosaic(~diet + cancer, data=fish, shade=TRUE,
+        gp=shading_Friendly2, legend=FALSE) #Kräver paketet vcd
```

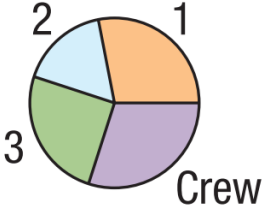


Kategoriska variabler och samband - fisk och cancer

- Det är tydligt att deltagare i studien som diagnosticerades med cancer var överrepresenterade bland dem som aldrig åt fisk
- Betyder det att vi har hittat ett samband? Ja och nej.
  - Ja, **bland deltagarna** i studien finns ett samband
  - Nej, vi kan inte utan vidare säga att det finns ett samband som gäller för **hela befolkningen**
- Det var bara 14 deltagare i studien som fick prostatacancer och som aldrig/sällan äter fisk, ett litet underlag om vi vill dra slutsatser gällande hela befolkningen
- Det kan vara slumpen som gör att vi ser ut att ha ett samband mellan två variabler

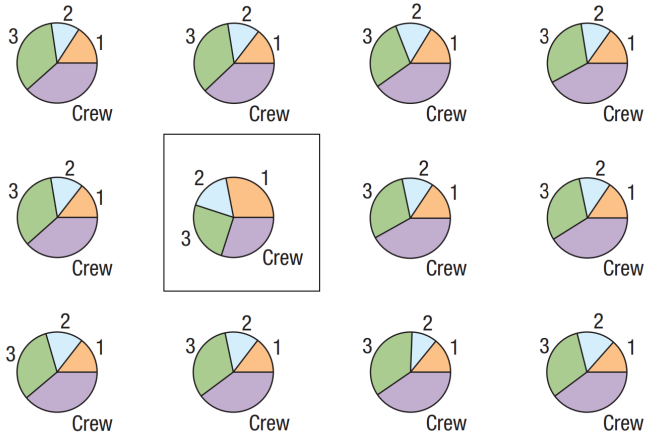
- Vi vill avgöra om sambandet vi ser i vårt stickprov beror på slumpen, eller om det faktiskt finns ett samband i populationen
- För att göra detta ger vi oss in i den gren av statistiken som kallas för **inferens**
- Formella metoder för inferens ingår i del 2 av kursen, men redan nu kan vi ge en intuitiv bild av vad det är

- En forskare ställer frågan: *Finns det ett samband mellan vilka som överlevde Titanic-katastrofen och vilket typ av biljett de reste med?*
- I vårt datamaterial är andelen överlevande större i första klass än i tredje klass (se figur längst ner)
  - Kan vi då säga att någon som reste i första klass hade större möjligheter att överleva?
  - Kan det lika gärna ha varit slumpen som gjorde att fler klarade sig i första klass?
- Kursboken (s 48-49) beskriver en metod som skulle kunna användas för att ge svar på frågan

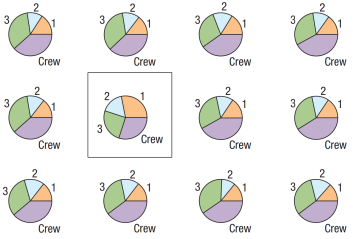


Kategoriska variabler och samband - Titanic

- För att avgöra om det finns ett verkligt samband mellan två variabler är det vanligt att ställa upp en **hypotes** som säger att det **inte** finns något samband
- I det här fallet skulle hypotesen vara att det saknas samband mellan biljettyp och överlevnad
- Om hypotesen stämmer borde de 712 livbåtsplatserna vara fördelade slumpvis mellan de 2208 passagerarna
- Vi kan själva testa att fördela livbåtsplatser slumpmässigt till alla individer på fartyget, och jämföra med den faktiska fördelningen
- Vi kan då se om det vi faktiskt observerat liknar det vi skulle förvänta oss om det **inte** fanns något samband
- Om vi upprerar vår egen uppdelning många gånger får vi ett gäng fiktiva scenarier att jämföra med, och tar samtidigt hänsyn till att slumpen kan ge många olika resultat



- Den verkliga fördelningen är markerad med en kvadrat
- Ser det ut som att den är resultatet av samma slumpvisa process som använts för att skapa de övriga fördelningarna?



- De slumpgenererade fördelningarna är i regel relativt lika, med små skillnader
- Den verkliga fördelningen **avviker mycket** från det "generella slumputseendet"
- Det kan fortfarande vara så att det faktiska resultatet bara beror på slump, men sannolikheten att helt slumpmässigt få vårt utfall verkar vara liten
- Vi lutar åt att det finns ett samband! Vi är fortfarande lite osäkra, men våra observationer pekar åt det hållet
- **Statistisk inferens** grundar sig i denna typ av tankesätt (mer om det i del 2)

Kategoriska variabler och samband - kausalitet

- Antag att vi har konstaterat att det finns ett samband mellan biljettyp och överlevnad
- Betyder det att en förstaklassbiljett per automatik **medförde** att du hade en bättre chans att få en plats i en livbåt?
- **Nej!** Bara för att sambandet existerar behöver det inte vara ett **kausalt samband** (orsakssamband)
- Det kan finnas **andra variabler som orsakar sambandet**: Var social ställning en underliggande faktor? Prioriterades kvinnor och barn på 1910-talet?
- I fallet med cancer och fisk: äter människor som är hälsomedvetna mer fisk, samtidigt som de äter mer grönsaker och ägnar sig mer åt fysisk aktivitet?

- **Att fundera över:**
  - Om det finns ett samband mellan *antalet tv-apparater* och *medellivslängd* i ett land, beror det på att tv-tittande är nyttigt eller finns det någon annan bakomliggande variabel som spelar in?
- När du tittar på ett dataset, var medveten om att det alltid finns **dolda variabler** (*en: lurking variables*)
- Dolda variabler är variabler som inte är inkluderade i datamaterialet, men som kan vara högst relevanta i kontexten

Ett (relativt) aktuellt exempel

- En artikel på svt.se lyder som följer

> För att komma in på Handelshögskolans prestigefyllda utbildningar krävs toppbetyg från gymnasiet. I år behövde de sökande också ha gjort högskoleprovet och fått ett resultat på minst 1,25. Orsaken var att skolan tyckte att det fanns omfattande problem med glädjebetyg på vissa gymnasieskolor.
>
> – Då vi inte fullt ut litade på gymnasiebetygen ville vi ”kontrollbesiktiga” dem, säger Lars Strannegård, Handelshögskolans rektor.
>
> **Fick negativa effekt på mångfalden**
>
> När Universitets- och högskolerådet (UHR) analyserade det nya behörighetskravet framkom att det fått negativa effekter på mångfalden. Andelen kvinnor, bland de antagna studenterna, hade minskat från 39 procent till 29 procent.

Ett (relativt) aktuellt exempel

- Vissa tolkade artikeln illvilligt, och menade att kraven sänktes för att kvotera in kvinnor som egentligen inte förtjänar en plats på utbildningen
- Finns det någon som kan kontextualisera?
- Många har påpekat att färre kvinnor än män skriver HP alls, eftersom kvinnor i genomsnitt har högre gymnasiebetyg (och därför inte behöver HP)
- Den “minskade mångfalden” beror så inte på att kvinnor skriver dåliga HP, utan att många kvinnor inte behövt skriva HP (och därför ej klarar kravet)
- Finns det någon som kan identifiera en dold (och högst relevant) variabel?
- Ett givet exempel skulle kunna vara gymnasiebetyg

- **Simpson's paradox** innebär att ett samband mellan två variabler kan försvinna när datamaterialet **delas in i olika grupper**
- På sid 106 i kursboken hittar vi en korstabell som tycks peka på att män hade lättare än kvinnor att bli antagna som doktorander på UC Berkeley

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

- Över 40 procent av männen som sökte blev antagna, men bara 30 procent av kvinnorna, och det ser ut som att kvinnor blev negativt särbehandlade

- När vi ser samma siffror nedbrutna per fakultet (school) blir bilden en annan

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%

- På fyra av de sex fakulteterna var andelen antagna större bland kvinnor än bland män
- Hur går det ihop?

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%

- Fler män sökte till school A och B, där det var lättare att komma in
- Fler kvinnor sökte till school E och F, där det var svårare att komma in
- Kvinnor hade lägre antagningsgrad på grund av att de sökte program som var svårare att komma in på
- Här är det alltså viktigt att ta hänsyn till den dolda variabeln som anger vilka som sökt till vilken skola!

Tack för idag!!