

Statistik och Dataanalys I

Föreläsning 22 - Chi2-test och beslut under osäkerhet

Mattias Villani



Statistiska institutionen
Stockholms universitet



mattiasvillani.com



@matvil



@matvil



mattiasvillani

- Chi2-test för goodness of fit
- Beslutsfattande under osäkerhet

Kortkampanj (Uppgift 22.2 i SDM)

- Bank har tre sorters kreditkort: Silver, Gold och Platinum.
- Marknadsföringskampanj. Skillnad i vilken kortklass kunder ansöker om?
- Undersöker $n = 200$ personers ansökningar efter kampanj.

Korttyp	Historiskt	Efter kampanj	Observerat antal	Förväntat antal
Silver	60%	55%	111	$200 \cdot 0.6 = 120$
Gold	30%	29.5%	59	$200 \cdot 0.3 = 60$
Platinum	10%	15%	30	$200 \cdot 0.1 = 20$

Chi2-test

- **Räknedata.** Antal.
- Hypotestest:
 - ▶ H_0 : räknedata följer fördelning med sannolikhet p_k i cell k .
 - ▶ H_A : räknedata följer annan fördelning.
- Totalt antal i hela tabellen: n .
- **Förväntat antal** i cell k : $\text{Exp}_k = n \cdot p_k$
- **Observerat antal** i cell k : Obs_k
- Chi2 (χ^2) test för tabell med N celler - **teststatistika**

$$\chi^2 = \sum_{k=1}^K \frac{(\text{Obs}_k - \text{Exp}_k)^2}{\text{Exp}_k} = \sum_{\text{all cells}} \frac{(\text{Obs} - \text{Exp})^2}{\text{Exp}}$$

- **Under H_0 - Chi2-fördelning** med $K - 1$ frihetsgrader

$$\chi^2 \sim \chi_{K-1}^2$$

Chi2-fördelningen

Frihetsgrader, ν

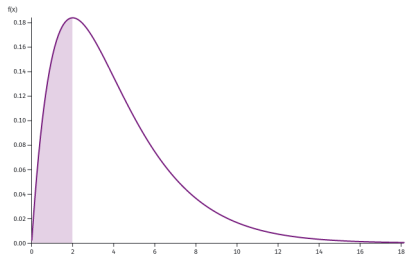
Kvantil:

Om $X \sim \chi^2(4)$ så gäller att

$$E(X) = \nu = 4.0$$

$$Var(X) = 2\nu = 8.0$$

$$P(X \leq 2) = 0.2642$$



■ Teststatistika

$$\chi_{obs}^2 = \sum_{\text{all cells}} \frac{(\text{Obs} - \text{Exp})^2}{\text{Exp}} = \frac{(111 - 120)^2}{120} + \frac{(59 - 60)^2}{60} + \frac{(30 - 20)^2}{20} = 5.6917$$

■ Under H_0 - **Chi2-fördelning** med $3 - 1 = 2$ frihetsgrader

■ Kritiskt värde på signifikansnivå 5% från χ^2 -tabell: 5.991.

■ Eftersom $\chi_{obs}^2 < \chi_{crit}^2$ kan vi inte förkasta H_0 .

■ Finns inte stöd för att kampanjen har ändrat fördelningen över olika kortklasser.





Beslut under osäkerhet

- Vi behöver ofta **fatta beslut** i en miljö med **osäkerhet**.
 - ▶ **Beslut**: Ska jag ta med ett paraply när jag går ut?
 - ▶ **Osäkerhet**: kommer det att regna?
 - ▶ **Beslut**: ska jag investera i aktier eller spara på banken?
 - ▶ **Osäkerhet**: börsens och inflationens utveckling under min placeringshorisont.
 - ▶ **Beslut**: Ska Sverige satsa på snabbtåg?
 - ▶ **Osäkerhet**: hur kommer elbilar utvecklas? klimatet? vad kommer det kosta? etc etc

Beslut och statistik

- Ett fattat beslut har **konsekvenser**.
- **konsekvenserna beror på osäkra faktorer** som vi inte vet när vi fattar beslutet.
- Vi behöver **sannolikhetsfördelningen** för de osäkra kvantiteterna.
- Modellerar **osäker kvantitet** i form av en **slumpvariabel X** .
- Använder **data** (och expertkunskap) för att beräkna dessa sannolikheter. **Statistik!**

Beslut + Utfall = Konsekvens

		Väder	
		Regn	Sol
Beslut	Paraply		
	Inget paraply		

■ Beslutsprocess:

- ▶ Du **fattar beslutet** a .
- ▶ X **realiseras** som x .
- ▶ Kombinationen a och x ger dig viss **nytta** (eng. **utility**):

$$U(a, x)$$

- Ibland: **förlust** $L(a, x)$ - vilket bara är negativ nytta

$$L(a, x) = U(a, x)$$

		Väder	
		Regn	Sol
Beslut	Paraply	0	50
	Inget paraply	-100	100

Maximin - en pessimistisk beslutsregel

- **Maximin**: välj beslut a som maximerar den minimala nyttan.
- **Garderar mot det värsta** som kan hända (pessimist).

		Väder	
		Regn	Sol
Beslut	Paraply	0	50
	Inget paraply	-100	100

- Maximin ignorerar hur sannolika utfallen är.

		Väder	
		0.2 Regn	0.8 Sol
Beslut	Paraply	0	50
	Inget paraply	-100	100

		Väder	
		0.01 Regn	0.99 Sol
Beslut	Paraply	0	50
	Inget paraply	-100	100

- I **spelteori** med intelligent **motståndare** är maximin optimal.

Maximera förväntad nytta

- Beslutsregel välj beslut a som maximerar förväntade nytta

$$EU(a) = \sum_{\text{alla } x} U(a, x) \cdot P(X = x)$$

- Paraply-beslutet:

$$a_1 = \text{Paraply} : EU(a) = 0.2 \cdot 0 + 0.8 \cdot 50 = 40$$

$$a_2 = \text{Inget paraply} : EU(a) = 0.2 \cdot (-100) + 0.8 \cdot 100 = 60$$

- Optimalt beslut: ta inte med paraply.

		Väder	
		0.2 Regn	0.8 Sol
Beslut	Paraply	0	50
	Inget paraply	-100	100

Maximera förväntad nytta

■ Paraply-beslutet i Bergen:

$$a_1 = \text{Paraply} : \quad EU(a) = 0.7 \cdot 0 + 0.3 \cdot 50 = 15$$

$$a_2 = \text{Inget paraply} : EU(a) = 0.7 \cdot (-100) + 0.3 \cdot 100 = -40$$

■ Optimal beslut i Bergen: Paraply!

		Väder	
		0.7 Regn	0.3 Sol
Beslut	Paraply	0	50
	Inget paraply	-100	100