

F9-11

Statistikens grunder 2 dagtid

HT 2012

Repetition

F9 Hypotesprövning

- *Behöver ni komma ihåg alla formler?*
- Nej, kolla formelbladet
- *Men vilka som behövs eller inte beror på situationen*
- Det som ska läras in är när ni behöver **Z** eller **T** och hur man läser och tolkar formlerna.
- *Motivera och fastställa kritiska gränser ska ni kunna.*
- Förstå strukturen för test.

p-värden

- Antag att vi får en punktskattning för någon parameter i fokus.
- Antag att vi har en nollhypotes också.
- Men vi har inte bestämt någon signifikansnivå α .
- Antag nu att den kritiska gränsen likställs med punktskattningen.
- *Vad skulle det motsvara för signifikansnivå dvs. α ?*

Övning 1 från F8

Tenta 2010-10-27, uppgift 2.

b) Hypotesprövning om π

- $H_0 : \pi = 0,42$
- $H_1 : \pi < 0,42$
- Signifikansnivå: vi valde $\alpha = 0,05$
- Kritiskt område $z_{obs} < -1,96 = z_{0,95}$
dvs. $P(Z \leq -1,96) = 0,05$

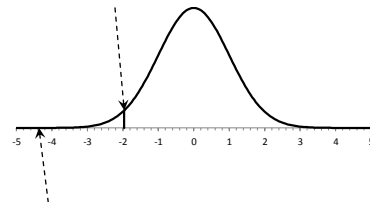
$$z_{obs} = \frac{\frac{519}{1430} - 0,42}{\sqrt{0,42 \cdot 0,58/1430}} = -4,37$$

Övning 1, forts.

- Antag att den kritiska gränsen istället var just **-4,37**
- Detta motsvarar en signifikansnivå på
 $P(Z \leq -4,37) = 0,00000621$
- **Man säger att det observerade värdet på testvariabeln Z har ett p-värde på 0,00000621**
- Resultatet skulle bli signifikant på 0,000621 % nivån
- Man behöver inte alltid bestämma α i förväg, jämför resultatet (här z_{obs}) mot olika val av α

Övning 1, forts.

$$P(Z \leq -1,96) = 0,05$$



$$P(Z \leq -4,37) = 0,00000621$$

= p-värdet

Övning 2 från F8

Tenta 2010-10-27, uppgift 5.

Hypotestest med parvisa differenser;

- $H_0 : \mu_D = 0$
- $H_1 : \mu_D \neq 0$
- Signifikansnivå: $\alpha = 0,05$
- Testvariabel och dess fördelning:

$$T = \frac{\bar{D}}{S_D / \sqrt{n}} \sim t(7)$$

- Kritiskt område: förkasta H_0 om
 $|t_{obs}| > t_{0,025}(7) = 2,365$

Övning 2, forts.

- Beräkning och insättning gav

$$|t_{obs}| = \frac{|-0,375|}{1,407886 / \sqrt{8}} = 0,75332$$

- Om det kritiska området istället hade varit

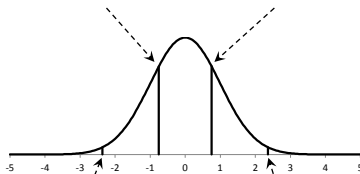
$$|t_{obs}| > 0,75332$$

hade detta motsvarat en signifikansnivå enligt

$$\begin{aligned} P(|T| > 0,75332) &= \\ &= P(T < -0,75332) + P(T > 0,75332) = \\ &= 2 \cdot 0,238 = 0,476 = \text{p-värdet} \end{aligned}$$

Övning 2, forts.

$$P(Z \leq -0,75332) = 0,238 \quad P(Z > 0,75332) = 0,238$$



$$P(Z \leq -2,365) = 0,025 \quad P(Z > 2,365) = 0,025$$

Måste komma ihåg att det är ett dubbelsidigt test när man beräknar p-värdet!

$$p\text{-värde} = 2 \cdot 0,238 = 0,476$$

Kap 18 χ^2 -test

Test hittills:

- medelvärden μ eller andelar π
- enskilda eller jämförelser
- Storheter på kvotskala

Nominal- eller ordinalskala då?

- Med dessa skaltyper hamnar observationer i en av flera möjliga **kategorier**.
- Man räknar **antal** i varje kategori
- **Möjliggör en annan typ av test**

Definitioner

n stycken observationer X_i , $i = 1, \dots, n$

$$X_i = \begin{cases} \text{kategori 1} \\ \vdots \\ \text{kategori } j \\ \vdots \\ \text{kategori } K \end{cases} \quad \begin{array}{l} \text{Nominal- eller} \\ \text{ordinalskala} \\ \text{Men, kan även} \\ \text{vara räkneta} \end{array}$$

$\pi_j =$ sannolikheten att en observation X_i ska ligga i kategori j

$N_j =$ antalet av n observationer som hamnar i kategori j (s.v.)

$n_j =$ observerat **antal** i kategori j

χ^2 -test, forts.

χ^2 -test bygger på **jämförelser** mellan **förväntat antal** observationer (under H_0) och **observerat antal** observationer

- Förväntat antal: $E(N_j) = n\pi_j$
- Jämför: $n_j - n\pi_j$
- Kvadrera: $(n_j - n\pi_j)^2$ Notera
- Relativ avvikelse: $(n_j - n\pi_j)/n\pi_j$
- Summera över alla kategorier:

$$\chi^2 = \sum_{j=1}^K \frac{(n_j - n\pi_j)^2}{n\pi_j}$$

χ^2 -test, forts.

Ett annat sätt att skriva som kanske är lättare att komma ihåg:

$$\chi^2 = \sum_{j=1}^K \frac{(\text{Observerat}_j - \text{Förväntat}_j)^2}{\text{Förväntat}_j}$$

Egenskaper:

- χ^2 är **approximativt χ^2 -fördelad** med ett visst antal frihetsgrader
- Antalet **frihetsgrader** bestäms lite olika beroende på typ av test
- Om man observerar ett **för stort värde** på χ^2 så är avvikelserna för stora och man **förkastar H_0**

χ^2 -test, forts.

Tumregel: För att det ska fungera ska **förväntat antal i varje cell** vara minst lika med 5, dvs.

$$E(N_j) = n\pi_j \geq 5 \text{ för alla } j.$$

Om det inte är fallet kan man slå ihop kategorier (kollapsa). T.ex. om man har

Kat. j	1	2	3	4	5	$\Sigma = n$
$E(N_j)$	20,9	51,6	15,2	8,2	4,1	100

kan man slå ihop kategori 4 och 5 till

Kat. j	1	2	3	4&5	$\Sigma = n$
$E(N_j)$	20,9	51,6	15,2	12,3	100

Goodness-of-fit test 1

Avsnitt 18.2

Man **antar** att kategorierna följer en viss given **sannolikhetsfördelning**, t.ex.

Kategori	A	B	C	D	E	Σ
π_j	0,35	0,25	0,20	0,10	0,10	1,00

Man observerar en **empirisk fördelning**, t.ex. av $n = 200$ observationer

Kategori	A	B	C	D	E	$\Sigma = n$
n_j	65	49	44	29	13	200

Kan man påstå att data stöder antagandet om fördelning?

Goodness-of-fit test 2

Jämför mot det förväntade antalet under antagandet (H_0):

Kategori	A	B	C	D	E	Σ
n_j	65	49	44	19	23	200
π_j	0,35	0,25	0,20	0,10	0,10	1,00
$n\pi_j$	70	50	40	20	20	200
$(n_j - n\pi_j)$	-5	-1	4	-1	3	0
$(n_j - n\pi_j)^2$	25	1	16	1	9	
$diff_j^2/n\pi_j$	25/70	1/50	16/40	1/30	9/10	1,28

$$\chi_{obs}^2 = \sum_{j=1}^K \frac{(n_j - n\pi_j)^2}{n\pi_j} = 1,277143$$

Goodness-of-fit test 3

H_0 : fördelningen för X_i enligt det givna

H_1 : ej fördelat enligt det givna

Testvariabel:

$$\chi^2 = \sum_{j=1}^K \frac{(n_j - n\pi_j)^2}{n\pi_j} \sim \chi^2(K-1)$$

dvs. med $K-1$ frihetsgrader. Varför?

- $K-1$ av sannolikheterna $\pi_1, \dots, \pi_K$ kan bestämmas (relativt) fritt
- **Så fort man har bestämt $K-1$ av dem, är den sista bestämd, t.ex. enligt**

$$\pi_K = 1 - \sum_{j=1}^{K-1} \pi_j \quad \text{Vi tappar en frihetsgrad!}$$

Goodness-of-fit test 4

I detta fall ger 5 stycken sannolikheter att vi får $5 - 1 = 4$ frihetsgrader

Signifikansnivå: säg 5 %

Kritiskt område: H_0 förkastas om

$$\chi_{obs}^2 > \chi_{0,05}^2(4) = [\text{enl. tab.}] = 9,488$$

Slutsats: H_0 kan inte förkastas, finns inget i data som säger att H_0 inte skulle vara sant på 5 % nivån.

Obs! Detta är ett enkelsidigt test, varför?

Ett värde nära noll betyder små avvikelser, dvs. bra anpassning och stöd till H_0 .

Goodness-of-fit: variant

Antag att man tror att en uppsättning observationer följer en Poissonfördelning

Men man vet inte vilken, dvs. λ är okänt och måste skattas först

Exempel: vi har observerat följande:

x	0	1	2	3	4	Σ
Frekvens	35	25	20	15	5	100

H_0 : X är Poissonfördelat

H_1 : X ej Poissonfördelat

Goodness-of-fit: variant forts.

Testvariabel:
$$\chi^2 = \sum_{j=1}^K \frac{(n_j - n\pi_j)^2}{n\pi_j}$$

λ skattas till

$$\hat{\lambda} = \bar{x} = \frac{35 \cdot 0 + 1 \cdot 25 + 2 \cdot 20 + 3 \cdot 15 + 4 \cdot 5}{100} = 1,30$$

Beräkna sannolikheterna $P(X = x)$ för $x = 0, \dots, 3$ samt $P(X \geq 4)$ med $\lambda = 1,3$ och sedan förväntat antal osv.:

x	0	1	2	3	≥ 4	Σ
$P(X = x)$	,2725	,3543	,2303	,0998	,0431	1,00
$E(N_x)$	27,25	35,43	23,03	9,98	4,31	100
n_x	35	25	20	15	5	100
$\text{Diff}_x^2/E(N_x)$	2,20	3,07	0,40	2,53	0,11	8,31

< 5, egentligen för litet!

Goodness-of-fit: variant forts.

Testvariabelns fördelning:

$$\chi^2 \sim \chi^2(K-1-1) \text{ dvs } \chi^2(3)$$

- fem kategorier totalt
- en frihetsgrad förloras för att sannolikheterna ska summera till 1
- en frihetsgrad förloras för att skatta λ

Signifikansnivå: $\alpha = 5\%$

Kritiskt område: $\chi_{obs}^2 > \chi_{0,05}^2(3) = 7,815$

Slutsats: Då $\chi_{obs}^2 = 8,31 > 7,815$ förkastas H_0 på 5 % nivån, data pekar på att X inte är Poissonfördelad.

Avsnitt 18.4

Oberoendetest 1

- Att pröva en hypotes om att två s.v. är statistiskt oberoende
- Antag två att man har två **kategoriska** s.v. X och Y
 - ex. nominal eller ordinalvariabler men fungerar även på **kategoriserade** intervall- och kvotskalevariabler
- Hypoteser:
 - H_0 : X och Y är oberoende
 - H_1 : X och Y är ej oberoende

Oberoendetest 2

Exempel: antag följande tvåvägstabell med sannolikheter:

Sannol.	Y = 1	2	3	Σ
X = 1	π_{11}	π_{12}	π_{13}	$\pi_{1\cdot}$
2	π_{21}	π_{22}	π_{23}	$\pi_{2\cdot}$
3	π_{31}	π_{32}	π_{33}	$\pi_{3\cdot}$
4	π_{41}	π_{42}	π_{43}	$\pi_{4\cdot}$
Σ	$\pi_{\cdot 1}$	$\pi_{\cdot 2}$	$\pi_{\cdot 3}$	1

Om två s.v. är statistiskt **oberoende** gäller

$$P(X = x \cap Y = y) = P(X = x) \cdot P(Y = y)$$

dvs. $\pi_{xy} = \pi_{x\cdot} \pi_{\cdot y}$ för varje x och y .

Oberoendetest 3

En tvåvägstabell med frekvenser för X och Y observeras (eng. *contingency table*):

Antal	Y = 1	2	3	Σ
X = 1	n_{11}	n_{12}	n_{13}	$n_{1\cdot}$
2	n_{21}	n_{22}	n_{23}	$n_{2\cdot}$
3	n_{31}	n_{32}	n_{33}	$n_{3\cdot}$
4	n_{41}	n_{42}	n_{43}	$n_{4\cdot}$
Σ	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	n

Skattningar av **marginalsannolikheterna**:

$$\hat{\pi}_{x\cdot} = \frac{n_{x\cdot}}{n} \quad \hat{\pi}_{\cdot y} = \frac{n_{\cdot y}}{n}$$

Oberoendetest 4

Av totalt n stycken par av X och Y kan en skattning av det förväntade antalet i varje cell skrivas

$$E(N_{xy}) = n\hat{\pi}_{x\cdot}\hat{\pi}_{\cdot y} = n \cdot \frac{n_{x\cdot}}{n} \cdot \frac{n_{\cdot y}}{n} = \frac{n_{x\cdot} \cdot n_{\cdot y}}{n}$$

Testvariabel är

$$\chi^2 = \sum_{x=1}^R \sum_{y=1}^C \frac{(n_{xy} - n\hat{\pi}_{x\cdot}\hat{\pi}_{\cdot y})^2}{n\hat{\pi}_{x\cdot}\hat{\pi}_{\cdot y}}$$

som är χ^2 -fördelad med $(R-1)(C-1)$ frihetsgrader.

Dvs. $(\# \text{ rader} - 1) \times (\# \text{ kolumner} - 1)$ exklusive marginaler.

Övning oberoendetest

En sammanställning av trivselnivån på arbetet bland säljarna på ett läkemedelsföretag fördelat på kön ges nedan:

<i>Antal</i>	Y = Låg	Medium	Hög	Σ
X = Man	46	61	53	160
Kvinna	8	10	12	30
Σ	54	71	65	190

H_0 : trivsel och kön är oberoende

H_1 : trivsel och kön är ej oberoende

Testvariabel:
$$\chi^2 = \sum_{x=1}^2 \sum_{y=1}^3 \frac{(n_{xy} - n\hat{\pi}_{x\cdot}\hat{\pi}_{\cdot y})^2}{n\hat{\pi}_{x\cdot}\hat{\pi}_{\cdot y}}$$

Övning, forts.

Testvariabel:

$$\chi^2 = \sum_{x=1}^2 \sum_{y=1}^3 \frac{(n_{xy} - n\hat{\pi}_{x\cdot}\hat{\pi}_{\cdot y})^2}{n\hat{\pi}_{x\cdot}\hat{\pi}_{\cdot y}}$$

$$= \sum_{x=1}^2 \sum_{y=1}^3 \frac{(n_{xy} - n_{x\cdot} \cdot n_{\cdot y} / n)^2}{n_{x\cdot} \cdot n_{\cdot y} / n}$$

<i>Antal</i>	Y = Låg	Medium	Hög	Σ
X = Man	46	61	53	160
Kvinna	8	10	12	30
Σ	54	71	65	190

$$E(N_{man, låg}) = \frac{n_{låg} \cdot n_{man\cdot}}{n} = \frac{54 \cdot 160}{190}$$

Övning, forts.

Fördelning: χ^2 med $(2-1)(3-1) = 2$ f.g.

Kritiskt gräns: $\chi_{0,05}^2(2) = 5,99$

Beräkna alla **förväntade antal** under H_0

Ex. $E(N_{man, låg}) = 160 \cdot 54 / 190 = 45,474$

<i>E(N_{x,y})</i>	Y = Låg	Medium	Hög	Σ
X = Man	45,474	59,789	54,737	160
Kvinna	8,526	11,211	10,263	30
Σ	54	71	65	190

$$\chi_{obs}^2 = \frac{(46 - 45,474)^2}{45,474} + \dots + \frac{(12 - 10,263)^2}{10,263}$$

$$\approx 0,543 < 5,99$$

Avsnitt 18.3

Homogenitetstest

Testa om flera fördelningar är lika

Se exempel sidor 7-9, testa om fördelningen är densamma för fem olika länder.

Principen densamma som för oberoendetest och för beräkningen av antal frihetsgrader.

Kan ni läsa på själva!

F10 Lite kort om χ^2 -test

Testvariabel:

$$\chi^2 = \sum_{j=1}^K \frac{(\text{Observerat}_j - \text{Förväntat}_j)^2}{\text{Förväntat}_j}$$

χ^2 fördelat med frihetsgrader bestämt av rader och/eller kolumner

Goodness-of-fit:

kategorier – 1

Goodness-of-fit med skattade parametrar:

kategorier – 1 – # parametrar

Oberoendetest / homogenitetstest:

(# rader – 1) · (# kolumner – 1)

Kap 19 Beslutsteori

Beslutsteori är att definiera

- **Handlingsalternativ**
 - de åtgärder vi kan välja bland
- **Tillstånd**
 - naturtillstånd (state-of-nature), som kommer att äga rum
- Ev. en **sannolikhetsfördelning** för de olika tillstånden
- **Nyttan** av kombinationen alternativ och tillstånd
 - vad vi vinner eller förlorar

Exempel

- **Handlingsalternativ**
 - köpa eller sälja aktier
- **Tillstånd**
 - Hög- eller lågkonjunktur
- **Sannolikhetsfördelning**
 - $P(\text{Hög}) + P(\text{Låg}) = 1$
- **Nyttan**
 - tjäna eller förlora pengar
- **Extra information**
 - förändringar på arbetsmarknaderna i USA och EU

Beslutsmatrix

Beteckna de olika

- handlingsalternativen med A_i (*action*)
- tillstånden med S_j (*state*)
- sannolikheten $P(S_j)$ med p_j
- nyttan med u_{ij} (*utility*)

Nytta u_{ij}	Tillstånd & sannolikhet j		
Handlingsalternativ i	S_1	S_2	S_3
	p_1	p_2	p_3
A_1	u_{11}	u_{12}	u_{13}
A_2	u_{21}	u_{23}	u_{23}
A_3	u_{31}	u_{32}	u_{33}

Beslut under säkerhet

Varje handling har en och endast en konsekvens (tillstånd) och därmed nytta.

Beslutsmatrisen kan se ut som:

Nytta u_{ij}	Tillstånd & sannolikhet j		
Handlingsalternativ i	S_1	S_2	S_3
	-	-	-
A_1	u_{11}		
A_2			u_{23}
A_3			u_{33}

- Endast en nytta per rad i matrisen
- Däremot kan samma tillstånd vara en konsekvens av olika handlingar (kolumn)

Beslut under säkerhet

Sannolikheter p_j för de olika tillstånden har ingen direkt funktion här

Se det som betingade sannolikheter:

$$P(S_1|A_1) = P(S_3|A_2) = P(S_3|A_3) = 1$$

och övriga sannolikheter = 0

Strategi: välj den handling som **maximerar nyttan u**

Kan skippa beskrivningen av oändligt många handlingar och tillstånd sid 6-11, Ex. 19.3 etc.

Beslut under risk

- Man **vet inte med säkerhet** vilket tillstånd som kommer att äga rum för en given handling.
- Men man kan ange en **sannolikhetsfördelning** p_j för tillstånden.

Nytta u_{ij}	Tillstånd & sannolikhet j			
Handlingsalternativ i	S_1	S_2	S_3	Förv. nytta
	p_1	p_2	p_3	
A_1	u_{11}	u_{12}	u_{13}	$E(U_1)$
A_2	u_{21}	u_{23}	u_{23}	$E(U_2)$
A_3	u_{31}	u_{32}	u_{33}	$E(U_3)$

Beslut under risk

För varje handling kan den **förväntade nyttan** beräknas:

$$E(U_i) = u_{i1}p_1 + u_{i2}p_2 + u_{i3}p_3$$

En strategi:

Välj den handling som **maximerar** den **förväntade nyttan** u

Värdet av observationer

- Om man kan skaffa sig **bättre information om framtiden** (vilket tillstånd som kommer att äga rum) kan man fatta bättre beslut.
- Dvs. välja den handling som ger **störst nytta**.
- Perfekt information:
Om man vet att S_j äger rum, välj den handling som maximerar u .
Men man vet ju inte vilket som kommer att äga rum, **men om** ...

Perfekt information

- Beteckna med v_j den största nytta som kan erhållas när S_j äger rum, t.ex.

$$v_1 = \max(u_{11}, u_{21}, u_{31})$$

- Upprepa för varje tillstånd

Nytta u_{ij}	Tillstånd & sannolikhet j			
Handlingsalternativ i	S_1	S_2	S_3	Förv. nytta
	p_1	p_2	p_3	
A_1	u_{11}	u_{12}	u_{13}	$E(U_1)$
A_2	u_{21}	u_{23}	u_{23}	$E(U_2)$
A_3	u_{31}	u_{32}	u_{33}	$E(U_3)$
Nytta perfekt info	v_1	v_2	v_3	$E(V)$

Den största av dessa, osv.

Perfekt information

- Man vet inte vilket av S_1 , S_2 och S_3 som äger rum förrän man har "köpt" den perfekta informationen.
- Men vi har **sannolikheterna**
- **Förväntade** värdet $E(V)$ kan beräknas:

$$E(V) = p_1v_1 + p_2v_2 + p_3v_3$$

Nytta u_{ij}	Tillstånd & sannolikhet j			
Handlingsalternativ i	S_1	S_2	S_3	Förv. nytta
	p_1	p_2	p_3	
A_1	u_{11}	u_{12}	u_{13}	$E(U_1)$
A_2	u_{21}	u_{23}	u_{23}	$E(U_2)$
A_3	u_{31}	u_{32}	u_{33}	$E(U_3)$
Nytta perfekt info	v_1	v_2	v_3	$E(V)$

Perfekt information

- Strategin under risk utan (perfekt) information: välj den största av $E(U_i)$
- Beteckna den största med $E(U^*)$.
- Kan visas att $E(U^*) \leq E(V)$
- Alltså, vi kan **betala $E(V) - E(U^*)$** för att få den perfekta informationen
- Detta är **värdet av perfekt information**
- Det kan bli mer, det kan bli mindre men i **förväntan** är nyttan $E(V)$

Imperfekt information 1

- Ej perfekt information:
Om man vet att S_j äger rum, välj den handling som maximerar u .
Men man vet inte vilket som kommer att äga rum, **men om ...**
Man skaffar sig information men av **sämre kvalitet**.
Trots ny information vet man inte exakt vilket av tillstånden som äger rum.
Men vi får **nya (betingade) sannolikheter** som kan användas isf $\{p_j\}$.

Imperfekt information 2

Man skaffar sig information och får uppdaterade sannolikheter $P(S_j|B_k)$

Nytta u_{ij}	Tillstånd & sannolikhet j			
Handlingsalternativ i	S_1	S_2	S_3	Förv. nytta
	$P(S_1 B_k)$	$P(S_2 B_k)$	$P(S_3 B_k)$	
A_1	u_{11}	u_{12}	u_{13}	$E(U_1 B_k)$
A_2	u_{21}	u_{23}	u_{23}	$E(U_2 B_k)$
A_3	u_{31}	u_{32}	u_{33}	$E(U_3 B_k)$
Nytta perfekt info	V_1	V_2	V_3	$E(V B_k)$

Liknar m.a.o. situationen med perfekt information

Imperfekt information 2

Informationen som inhämtas är behäftad med osäkerhet

- Man kan observera flera utfall $B_1, B_2, \dots$
- Tablå på förra sidan görs för varje utfall
- Man måste ha sannolikheter

$$P(B_1), P(B_2), \dots,$$
$$P(S_1|B_1), P(S_2|B_1), \dots$$
$$P(B_1|S_1), P(B_2|S_1), \dots$$

Blir snabbt svårt att hänga med.

Förslag: på egen hand försök följa Exempel 19.1 om omeletten, kan göra det lättare att förstå.

Beslut under osäkerhet

Man känner endast till

- *handlingsalternativen*
- *tillstånden*
- *nyttan* för varje kombination av handling och tillstånd

Man känner inte till *sannolikheterna* för tillstånden

Kan inte heller skaffa sig *information* om dem.

Beslut under osäkerhet

Fyra olika strategier:

- Maximin
- Maximax
- Minimax regret
- Laplacekriteriet

Allmänt:

Identifiera vad som medför antingen det *värsta* (min-) eller det *bästa* (max-) och maximera eller minimera sedan.

Maximin

För *varje handling* identifiera det värsta som kan hända (*min nytta*)

Välj den handling som har det största värdet av dessa (*max av alla min*)

Garanterar att man åtminstone kommer att få just så mycket

Nytta u_{ij}	Tillstånd & sannolikhet j			
Handlingsalternativ i	S_1	S_2	S_3	Min nytta
	-	-	-	
A_1	2	4	8	2
A_2	5	6	4	4
A_3	10	7	2	2

Välj

Maximax

För *varje handling* identifiera det bästa som kan hända (*max nytta*)

Välj den handling som har det största värdet av dessa (*max av alla max*)

Man maximerar nyttan (om man har tur!)

Nytta u_{ij}	Tillstånd & sannolikhet j			
Handlingsalternativ i	S_1	S_2	S_3	Max nytta
	-	-	-	
A_1	2	4	8	8
A_2	5	6	4	6
A_3	10	7	2	10

Välj

Minimax regret 1

1. För **varje tillstånd** identifiera den handling som medför **max nytta**
2. **Välj** den handling som medför att den största **alternativförlusten minimieras**

Nytta u_{ij}	Tillstånd & sannolikhet j		
Handlings-alternativ i	S_1	S_2	S_3
	-	-	-
A_1	2	4	8
A_2	5	6	4
A_3	10	7	2
Max nytta	10	7	8

Minimax regret 2

Beräkna differenser mot max nyttan för varje tillstånd 10-2, 10-5, 10-10, 7-4 osv.

Nytta u_{ij}	Tillstånd & sannolikhet j			
Handlings-alternativ i	S_1	S_2	S_3	Max regret
	-	-	-	-
A_1	8	3	0	8
A_2	5	1	4	5
A_3	0	0	6	6
Max nytta	10	7	8	

Välj

Minst

Genom att välja A_2 kommer man att "ångra sig" minst.

Laplacekriteriet

Som med beslut under risk men okända sannolikheter p_j

Antag att alla tillstånd lika möjliga

Dvs. $p_j = (\text{antal tillstånd})^{-1}$

Välj den handling med största förväntade nytta.

Nytta u_{ij}	Tillstånd & sannolikhet j			
Handlings-alternativ i	S_1	S_2	S_3	Förv. nytta
	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	
A_1	2	4	8	4,33
A_2	5	6	4	5,00
A_3	10	7	2	6,33

Övningar inför tentan

Tenta 2010-10-27, uppgift 1.

a) Konfidensintervall

- Anta att $\sigma = 5,5$
- Beräkna ett 95 % KI för μ ($\alpha = 0,05$)
- Tillkommer: anta att $X_i \sim N(\mu, \sigma^2)$

- Punktskattningen: $\bar{x} = \frac{302,4}{6} = 50,4$

- Formel: $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

- Insättning ger $50,4 \pm 1,96 \cdot \frac{5,5}{\sqrt{6}}$

el. $50,4 \pm 4,4$ el. $(46,0 ; 54,8)$

Övning 1, forts.

b) Beräkna stickprovsstorlek

- Antaganden givna i a)
- Längden på KI får vara högst 6

dvs. $\bar{x} \pm 3$

$$\text{dvs. } z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq 3 \Leftrightarrow n \geq \frac{z_{\alpha/2}^2 \sigma^2}{3^2}$$

- Insättning ger

$$\frac{1,96^2 \cdot 5,5^2}{3^2} = 12,91204 \leq 13$$

- Svar: minst 13

Övning inför tentan

Tenta 2010-10-27, uppgift 3.

a) Homogenitetstest

- $n = 300$ fördelat över kön och rökvanor
- Avgör på 5 % -nivån om kvinnor och män skiljer sig åt map rökvanor
- Data

	0	1-5	6-15	>15	Σ
Män	121	18	10	3	152
Kvinnor	123	3	17	5	148
Σ	244	21	27	8	300

- H_0 : Kvinnor och mäns fördelning lika
- H_1 : Kvinnor och mäns fördelning ej lika

Övning 3, forts.

- Testvariabel och dess fördelning:

$$\chi^2 = \sum_{\text{kön}} \sum_{\text{vanor}} \frac{(\text{Obs} - \text{Förv})^2}{\text{Förv}} \sim \chi^2(?)$$

- Beräkna förväntat antal under H_0 :

	0	1-5	6-15	>15	Σ
Män	123,63	10,64	13,68	4,05	152
Kvinnor	120,37	10,36	13,32	3,95	148
Σ	244	21	27	8	300

- Cellerna som motsvarar vanan ">15" har för lågt förväntat antal; tumregel ≥ 5
- Kollapsa kategorierna "6-15" och ">15"

Övning 3, forts.

- Ny tabell:

	0	1-5	>6	Σ
Män	121	18	13	152
Kvinnor	123	3	22	148
Σ	244	21	35	300

- Förväntade värden, fortfarande under H_0 :

	0	1-5	>6	Σ
Män	123,63	10,64	17,73	152
Kvinnor	120,37	10,36	17,27	148
Σ	244	21	35	300

Övning 3, forts.

- Testvariabel och dess fördelning:

$$\chi^2 = \sum_{k \text{ön vanor}} \sum \frac{(\text{Obs} - \text{Förv})^2}{\text{Förv}} \sim \chi^2(2)$$

ty (rader – 1)(kolumner – 1) = 2

- Förkasta H_0 om

$$\chi^2_{obs} \geq \chi^2_{0,05}(2) = [\text{enl. tab.}] \approx 5.99$$

- Insättning ger

$$\chi^2_{obs} = 12,9939 > \chi^2_{0,05}(2) = 5.99$$

- Slutsats: Vi förkastar H_0 på 5 % nivån; kvinnors och mäns rökvanor skiljer sig åt

SG2: 2011-03-18 Uppgift 4

- a) Beslutsmatris.

Nytta u_{ij}	Motbud & sannolikhet		
	Inget bud	Bud 50'	Bud 70'
Mitt bud	p_0	p_{50}	p_{70}
Bud 40'	60'	0	0
Bud 60'	40'	40'	0
Bud 80'	20'	20'	20'

Om Mitt bud > Motbud:

$$\text{Nytta} = \text{Avkastning} - \text{Mitt bud} = 100' - \text{Mitt bud}$$

Om Mitt bud < Motbud:

$$\text{Nytta} = 0$$

SG2: 2011-03-18 Uppgift 4, forts.

- b) Maximin:

- Finn sämsta nyttan för varje handling
- Välj max av dessa

Nytta u_{ij}	Tillstånd & sannolikhet j			Min nytta
	Inget bud	Bud 50'	Bud 70'	
Handlingsalternativ i	p_0	p_{50}	p_{70}	
Bud 40'	60'	0	0	0
Bud 60'	40'	40'	0	0
Bud 80'	20'	20'	20'	20

Välj

Max

SG2: 2011-03-18 Uppgift 4, forts.

- c) Beslut under risk:

- Förväntad nytta per handling
- Välj max av dessa

Nytta u_{ij}	Tillstånd & sannolikhet j			Förv. nytta $E(U_i)$
	Inget bud	Bud 50'	Bud 70'	
Handlingsalternativ i	0,25	0,50	0,25	
Bud 40'	60'	0	0	15
Bud 60'	40'	40'	0	30
Bud 80'	20'	20'	20'	20

Välj

Max

SG2: 2011-04-08

Uppgift 1

Goodness-of-fit test, dvs. χ^2 -test med

H_0 : Sannolikheterna är

$$\pi_{rosa} = 0,50, \pi_{vita} = 0,25, \pi_{röda} = 0,25$$

H_A : Sannolikheterna ej enligt

3 klasser/kategorier ger 2 frihetsgrader

Testvariabel är $\chi^2(2)$ -fördelad

Kategori	Rosa	Vita	Röda	Σ
π_j (under H_0)	0,50	0,25	0,25	1,00
Förv. $= n\pi_j$	60	30	30	120
Observ. n_j	75	20	25	120
$(O-F)^2/F$	3,75	3,33	0,83	7,92

SG2: 2011-04-08 Uppgift 2

a) Beslutsmatris.

Nytta u_{ij}	Skada & sannolikhet		
	Ingen skada	Lätt skada	Svår skada
Mitt bud	p_0	p_L	p_S
Teckna försäkring	-490	-1990	-1990
Teckna ej försäkring	0	-6000	-12000

Om "Tecknar försäkring": **(0 el. 1500)**

$$\text{Nytta} = \text{Kostnad} - \text{Självrisk} = -490 - \text{Självrisk}$$

Om "Tecknar ej försäkring":

$$\text{Nytta} = \text{Kostnad} - \text{Självrisk} = 0 - \text{Självrisk}$$

(0, 6000 el. 1200)

SG2: 2011-04-08 Uppgift 2, forts.

b) Beslut under risk:

- Förväntad nytta per handling
- Välj max av dessa

Nytta u_{ij}	Skada & sannolikhet			Förv. nytta $E(U_i)$
	Ingen skada	Lätt skada	Svår skada	
Mitt bud	p_0	p_L	p_S	
Teckna försäkring	-490	-1990	-1990	-514
Teckna ej försäkring	0	-6000	-12000	-102

Välj

Max

SG2: 2011-04-08 Uppgift 2, forts.

Extra: Maximin och Maximax:

- Förväntad nytta per handling
- Välj max av dessa

Nytta u_{ij}	Skada & sannolikhet			Min	Max
	Ingen skada	Lätt skada	Svår skada		
Mitt bud	p_0	p_L	p_S		
Teckna försäkring	-490	-1990	-1990	-1990	-490
Teckna ej försäkring	0	-6000	-12000	-12000	0

Maximin

Maximax

SG2: 2011-04-08 Uppgift 2, forts.

Extra: Minimax regret

Varje tillstånd: handling som medför max nytta

Nytta u_{ij}	Skada & sannolikhet		
	Ingen skada	Lätt skada	Svår skada
Mitt bud	p_0	p_L	p_S
Teckna försäkring	-490	-1990	-1990
Teckna ej försäkring	0	-6000	-12000

SG2: 2011-04-08 Uppgift 2, forts.

Extra: Minimax regret, forts.

Beräkna alternativ förlusten inom varje tillstånd

Välj handling med minst alternativ förlust

Nytta u_{ij}	Skada & sannolikhet			Max
	Ingen skada	Lätt skada	Svår skada	
Mitt bud	p_0	p_L	p_S	
Teckna försäkring	490	0	0	490
Teckna ej försäkring	0	4010	10010	10010

Välj

Minst