

Tentamen i statistiska databaser och register (Statistical Databases and Registers)

28 april 2009, kl. 13.00-18.00 i Värtasalen.

Skrivningen består av två delar 1) Infologiska aspekter på databaser och 2) Datamining och andra aspekter på analysen av databaser och register. **Svar och lösningar på de två delarna skall skrivas på separata papper.** Examinationen på kursen består av dessa två delar och på delen Statistisk kvalitet och andra aspekter på register (som tenderats muntligt). Alla tre delarna måste vara godkända för att bli godkänd på hela kursen. Slutbetyget baseras på en sammanvägning i proportionerna 40 % (Statistisk kvalitet), 40 % (Infologi) och 20 % (Övrigt). Alla delarna måste vara godkända vid samma kurstillfälle (termin).

Alla antaganden och beteckningar skall vara väl förklarade och definierade (också sådana som använts under kursen). Alla svar, lösningar och motiveringar skall vara lätta att följa. Svar och argument som inte är begripliga ger 0 poäng även om svaret skulle vara rätt.

Del 1: Infologiska aspekter på databaser

Bifogat finns fyra sidor med sammanfattande tabeller ur en rapport från Statens energimyndighet och SCB (*Årliga energibalanser 2006 – 2007, korrigerad version 2009-02-16 EN 20 SM 0804*). De flesta frågor gäller tablå D, slutlig energianvändning.

1. Tablå D, slutlig energianvändning, bygger på sammanställning av data om energianvändning från olika källor. Försök konstruera en konceptuell datamodell för underliggande data, där du pekar ut viktiga objekttyper, objektens variabler och relationerna mellan objekttyperna. Använd den teknik som du har sett många exempel på under kursen. Observera att det finns många riktiga svar på denna fråga. Det kan till exempel bero på vilka antaganden du gör. Men skriv ut dessa antaganden.
2. I en annan tabell (tablå A) finns data om bruttotillförseln av energi. Försök att utvidga den konceptuella modellen till att täcka även den delen. Även på denna fråga finns många riktiga svar.
3. Statistiken som ligger bakom tablå D skulle kunna lagras i en flerdimensionell kub (hyperkub). Beskriv strukturen hos en sådan kub och vilka dimensioner som finns. Diskutera också om man skulle kunna utvidga kuben eller behöver komplettera med flera kuber för att de andra tabellerna på energianvändning skulle kunna tas fram.
4. Beskriv vilka metadata som nämns och tas upp i tablåer eller anslutande text.
5. Ge exempel på metadata som inte finns i tabellen men som olika läsare skulle kunna behöva eller vara intresserade av. Nämn några olika användningar av metadata.

Del 2: Datamining och andra aspekter.

1. Vilken typ av data behövs enligt PW för att studera återköpsbeteende? Hur gör man?
2. Vilka skillnader och likheter finns det mellan datamining och “vanlig” statistik, som den lärs ut på andra kurser?
3. I Datamining används ofta linjär regression för prediktion. Vilka problem med data bör man då tänka på?
4. Ibland används förkortningen/begreppet SEMMA. Beskriv kort vad det innebär! (eller beskriv kortfattat analysprocessen, som ML kallar “The Analytic Workflow” (vilket är ungefär samma sak men med andra ord)).
5. I korganalys används ibland begreppen support och konfidens. Beskriv kortfattat vad de innebär! Exemplifiera gärna!

Statistiken med kommentarer

I tabellerna A – D nedan sammanfattas översiktligt huvuddragen i utvecklingen av energitillförsel, omvandling och slutlig energianvändning under perioden 2002 – 2007. Uppgifterna för 2007 baseras i huvudsak på preliminär statistik.

Tablå A Energitillförsel 2002 - 2007; PJ, TWh, procentuell förändring, procentandelar (Tillförd primär energi och motsvarande)

	Kol, koks	Träd bränsle, avlutar, sopor o.d., torv	Råolja, olja produkter ³	Natur gas	Fjärr värme (via värme pump- par)	Vatten- och vind kraft ¹	Känbränsle /kärnkraft ²		Netto import av el energi	Summa bruttotill- förse ³		Index (alt 1)
							Alt 1	Alt 2		Alt 1	Alt 2	1990 = 100
PJ (Petajoule)												
2002	106,2	357,5	772,1	33,2	27,7	241,1	722,5	245,2	19,3	2 279,6	1 802,3	108,4
2003	108,3	376,5	797,8	37,1	23,8	194,8	718,7	242,7	46,2	2 303,3	1 827,3	109,5
2004	108,7	382,3	793,6	36,6	24,0	221,2	817,4	279,6	-7,6	2 376,2	1 838,4	113,0
2005	100,4	388,0	777,3	34,3	22,2	265,4	775,3	260,6	-26,6	2 336,3	1 821,5	111,1
2006	98,9	388,5	760,7	35,4	21,1	225,7	683,1	241,1	21,8	2 235,1	1 793,2	106,3
2007	101,3	414,4	764,1	38,9	20,1	243,4	661,9	241,1	4,7	2 249,0	1 828,1	107,0
Förändring i % mellan 07/06	2,4%	6,7%	0,5%	9,9%	-4,6%	7,8%	-3,1%	0,0%	..	0,6%	1,9%	
TWh												
2002	29,5	99,3	214,5	9,2	7,7	67,0	200,7	68,1	5,4	633,2	500,6	
2003	30,1	104,6	221,6	10,3	6,6	54,1	199,6	67,4	12,8	639,8	507,6	
2004	30,2	106,2	220,4	10,2	6,7	61,4	227,1	77,7	-2,1	660,1	510,7	
2005	27,9	107,8	215,9	9,5	6,2	73,7	215,4	72,4	-7,4	649,0	506,0	
2006	27,5	107,9	211,3	9,8	5,9	62,7	189,7	67,0	6,1	620,9	498,1	
2007	28,1	115,1	212,3	10,8	5,6	67,6	183,9	67,0	1,3	624,7	507,8	
% fördelning alt 1												
2002	4,7	15,7	33,9	1,5	1,2	10,6	31,7		0,8	100,0		
2003	4,7	16,3	34,6	1,6	1,0	8,5	31,2		2,0	100,0		
2004	4,6	16,1	33,4	1,5	1,0	9,3	34,4		-0,3	100,0		
2005	4,3	16,6	33,3	1,5	0,9	11,4	33,2		-1,1	100,0		
2006	4,4	17,4	34,0	1,6	0,9	10,1	30,6		1,0	100,0		
2007	4,5	18,4	34,0	1,7	0,9	10,8	29,4		0,2	100,0		
% fördelning alt 2												
2002	5,9	19,8	42,8	1,8	1,5	13,4		13,6	1,1		100,0	
2003	5,9	20,6	43,7	2,0	1,3	10,7		13,3	2,5		100,0	
2004	5,9	20,8	43,2	2,0	1,3	12,0		15,2	-0,4		100,0	
2005	5,5	21,3	42,7	1,9	1,2	14,6		14,3	-1,5		100,0	
2006	5,5	21,7	42,4	2,0	1,2	12,6		13,4	1,2		100,0	
2007	5,5	22,7	41,8	2,1	1,1	13,3		13,2	0,3		100,0	

1) Producerad elenergi brutto

2) Alt 1 = Förbrukat kärnbränsle. Alt 2 = Producerad elenergi brutto

3) Bunkring för utrikes sjöfart ingår.

År	2002	2003	2004	2005	2006	2007
PJ	51,3	69,0	81,0	82,8	89,0	88,5

Tablå B Insatt energi för omvandling till andra energislag, PJ

	Kol, koks	Träd bränsle, avlutar, sopor o.d., torv	Råolja, olja pro dukter	Gas (natur-, stads-, mas- & koksugns-)	Fjärr värme (via värme pumpar)	Primär vatten- kraft	Kärn- bränsle- energi	Elenergi	Summa
Insatt för elproduktion									
2002	11,5	19,6	15,9	8,8	-	238,9	722,5	0,2	1017,4
2003	14,8	24,4	21,6	10,8	-	192,5	718,7	0,3	983,1
2004	5,9	36,4	8,7	12,1	-	218,1	817,4	0,2	1098,8
2005	4,2	34,7	7,4	12,3	-	262,0	775,3	0,3	1098,2
2006	4,7	39,7	8,9	9,2	-	222,2	683,1	0,1	967,8
2007	3,2	44,3	5,5	10,7	-	238,3	661,9	0,1	964,0
Insatt för fjärrvärmeproduktion									
2002	4,5	102,9	17,1	13,8	27,7	-	-	4,7	170,7
2003	4,1	107,0	18,1	14,0	23,8	-	-	1,7	168,8
2004	9,5	101,3	14,1	12,5	24,0	-	-	1,3	162,8
2005	8,3	105,9	12,1	11,4	22,2	-	-	1,2	161,1
2006	10,7	110,4	13,3	11,0	21,1	-	-	0,8	167,3
2007	7,9	116,1	9,3	11,8	20,1	-	-	1,2	166,4
Insatt i övriga omvandlingsanl.¹									
2002	66,4	-	823,0	0,3	-	-	-	-	889,7
2003	66,4	-	860,2	0,2	-	-	-	-	926,9
2004	72,7	-	903,9	0,2	-	-	-	-	976,8
2005	71,4	-	879,4	0,2	-	-	-	-	951,0
2006	70,2	-	884,0	0,1	-	-	-	-	954,3
2007	72,1	-	786,8	0,1	-	-	-	-	859,0

1) Oljeraffinaderier, gasverk, koksverk samt för framställning av masugns gas

Tablå C Bruttoproduktion av omvandlad energi, PJ

	Koks	Olje produkter	Gas (stads-, mas-, koksugns-)	Fjärr värme ¹	Elenergi	Summa
2002	30,1	757,0	27,8	183,0	528,2	1 526,2
2003	29,7	799,7	27,7	185,3	486,8	1 529,2
2004	33,1	847,5	29,3	188,9	549,0	1 647,7
2005	33,3	812,4	29,9	186,8	486,8	1 549,1
2006	33,3	838,6	28,8	185,2	512,8	1 598,6
2007	33,5	739,7	28,6	189,5	486,8	1 478,0

1) Inkl mottagen värme, huvudsakligen spillvärme från industrisektorn

Anm. I tablå B ovan redovisas enbart insats av primär energi och motsvarande. Vid beräkning av total energiinsats för el och fjärrvärmeproduktion tillkommer mottagen värme, huvudsakligen spillvärme från industrisektorn enligt följande PJ:

	2002	2003	2004	2005	2006	2007
För elproduktion	0.0	0.0	0.0	0.0	0.0	0.0
För fjärrvärmeproduktion	15.6	19.2	22.9	19.4	19.4	19.3

Tablå D Slutlig användning för energiändamål inom landet 2002 - 2007, PJ

	Kol, koks	Träd- bränsle, avlutar, sopor o.d., torv	Olje- produkter	Gas (natur- stads-, mas- & koksugns-)	Fjärr- värme	Summa bränslen (inkl fjärr- värme)	El- energi	Summa totalt	Index 1990 = 100
Jordbruk, skogsbruk, fiske									
2002	0,0	0,2	26,7	0,8	0,4	28,1	5,8	33,9	112,9
2003	0,0	2,0	27,7	1,0	0,4	31,1	5,0	36,1	120,3
2004	0,0	2,0	23,6	1,0	0,4	27,0	5,5	32,6	108,5
2005	0,0	3,5	23,3	1,0	0,4	28,2	8,3	36,5	121,6
2006	0,0	3,6	23,8	1,0	0,4	28,8	6,3	35,1	117,1
2007	0,0	3,5	21,2	1,0	0,4	26,1	5,6	31,8	105,9
Industri (tillverkning, mineralutv.)¹									
2002	53,3	194,2	68,4	21,4	16,4	353,7	200,4	554,1	110,0
2003	52,1	198,9	76,8	24,7	15,9	368,5	196,2	564,7	112,1
2004	52,9	199,3	70,5	25,3	17,0	365,0	199,3	564,3	112,0
2005	49,4	198,7	62,5	24,9	15,9	351,5	201,3	552,8	109,8
2006	46,0	192,8	62,6	25,5	15,8	342,7	203,6	546,3	108,5
2007	50,8	200,9	57,3	25,9	16,0	350,9	205,7	556,5	110,5
Byggverksamhet									
2002	-	-	10,3	0,2	-	10,5	3,1	13,6	91,2
2003	-	-	6,8	0,6	-	7,4	2,9	10,3	69,3
2004	-	-	6,6	0,5	-	7,1	2,9	10,0	67,2
2005	-	-	6,6	0,5	-	7,1	2,9	10,0	66,8
2006	-	-	7,1	0,5	-	7,6	3,1	10,7	71,9
2007	-	-	7,5	0,5	-	8,0	3,2	11,2	75,1
Offentlig verksamhet									
2002	0,0	0,9	5,6	0,9	23,3	30,6	36,4	67,0	88,9
2003	0,0	1,0	6,0	1,1	23,4	31,5	35,7	67,1	89,0
2004	0,0	1,8	5,3	1,2	23,5	31,8	36,9	68,7	91,1
2005	0,0	1,2	3,2	1,1	23,0	28,6	35,7	64,3	85,3
2006	0,0	1,5	2,7	1,5	20,1	25,8	34,7	60,4	80,1
2007	0,0	1,7	2,9	1,6	20,6	26,8	35,1	61,9	82,1
Transporter									
2002	0,0	-	322,2	0,5	-	322,7	10,3	333,0	110,9
2003	0,0	-	327,9	0,8	-	328,7	10,2	338,9	112,8
2004	0,0	-	340,7	0,8	-	341,5	10,8	352,3	117,3
2005	0,0	-	346,9	0,8	-	347,6	10,1	357,8	119,1
2006	0,0	-	346,1	1,0	-	347,0	10,4	357,4	119,0
2007	0,0	-	356,3	1,0	-	357,3	10,5	367,9	122,5

1) Exkl raffinaderier och koksverk som ingår under "Användning i energisektorn"

Tablå D (forts)

	Kol, koks	Träd- bränsle, aviutar, sopor o.d., torv	Olje- produkter	Gas (natur-, stads-, mas- & koksugns-)	Fjärr- värme	Summa bränslen (inkl fjärr- värme)	El- energi	Summa totalt	Index 1990 = 100
Övriga tjänster									
2002	0,0	0,4	9,8	2,4	29,9	42,5	62,1	104,6	117,1
2003	0,0	0,3	9,0	2,3	31,3	42,9	61,1	104,0	116,4
2004	0,0	0,5	8,4	2,2	32,2	43,2	60,1	103,3	115,6
2005	0,0	0,3	6,7	2,5	33,0	42,4	61,5	103,9	116,4
2006	0,0	0,4	7,2	2,9	33,0	43,4	60,6	104,0	116,5
2007	0,0	0,4	8,7	3,0	33,7	45,8	62,1	107,9	120,9
Hushåll (bostäder och annat)									
2002	0,0	39,3	42,4	3,2	94,3	179,3	153,7	332,9	101,1
2003	0,0	42,9	38,6	3,0	96,5	181,0	154,9	335,8	102,0
2004	0,0	41,0	36,0	3,1	95,2	175,3	153,9	329,2	100,0
2005	0,0	43,7	25,1	2,8	96,5	168,1	152,0	320,1	97,2
2006	0,0	40,2	17,2	2,5	97,7	157,6	149,5	307,0	93,3
2007	0,0	47,5	13,6	2,5	104,3	168,0	149,2	317,2	96,3
Summa kategorifördelad anv.									
2002	53,4	235,0	485,2	29,5	164,3	967,4	471,8	1439,2	107,2
2003	52,1	245,2	492,7	33,6	167,5	991,0	465,9	1457,0	108,5
2004	53,0	244,6	491,0	34,2	168,2	990,9	469,4	1460,3	108,7
2005	49,4	247,4	474,3	33,6	168,8	973,6	471,8	1445,4	107,6
2006	46,0	238,4	466,6	34,9	166,9	952,8	468,2	1421,0	105,8
2007	49,4	247,4	467,5	33,6	168,8	966,8	471,8	1438,6	107,6
Förändring i % mellan 07/06	7,4%	3,8%	0,2%	-3,7%	1,1%	1,5%	0,8%	1,2%	
Differenspost (ej spec anv.)									
2002	-	-	41,4	-	-	41,4	-	41,4	
2003	-	-	49,7	-	-	49,7	-	49,7	
2004	-	-	47,4	-	-	47,4	-	47,4	
2005	-	-	43,4	-	-	43,4	-	43,4	
2006	-	-	24,6	-	-	24,6	-	24,6	
2007	-	-	35,1	-	-	35,1	-	35,1	