

## Tentamen i Statistiska metoder 2008-03-27

Skrivtid: 14-18

Hjälpmedel: Med skrivningen häftad formelsamling. Räknedosa.

Jourhavande lärare: Lotta Hallberg

Redovisa och motivera kort alla dina lösningar

Obs! Skriv namn och personnummer på varje papper du lämnar in.

### 1

Vårdtiden för 10 patienter som vårdats på sjukhus för en viss diagnos var följande: 8, 6, 8, 5, 42, 8, 6, 9, 8, 10 dagar.

- a) Beräkna genomsnittlig vårdtid samt median för de 10 patienterna. Vilket av dessa lägesmått anser du var bäst i detta material? Motivera. 1p
- b) Beskriv materialet i ett lådagram. 2p
- c) Om en vårdplats kostar 2400:-kr per dag, hur stor blir då den totala kostnaden för de 10 patienterna. 1p

### 2

Följande data är hämtat från brottsförebyggande rådets hemsida. Variabeln 'misshandel' är totala antalet anmälda misshandelfall i Sverige för det givna året.

misshandel	år	misshandel	år
7382	1950	37511	1988
8605	1955	39641	1989
8684	1960	40690	1990
11689	1965	40454	1991
18349	1970	45232	1992
21509	1975	50926	1993
21378	1976	53665	1994
23596	1977	54380	1995
22868	1978	53731	1996
23171	1979	55109	1997
24668	1980	56878	1998
24314	1981	59918	1999
28200	1982	58846	2000
29220	1983	59461	2001
30785	1984	61631	2002
31996	1985	65177	2003
32805	1986	67089	2004
34757	1987	72645	2005

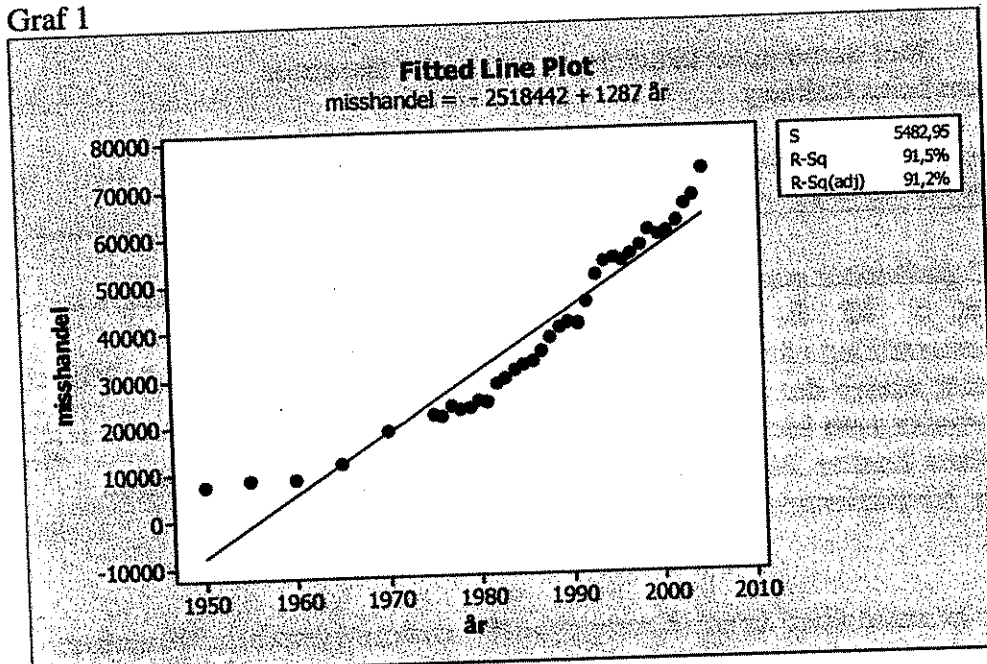
Man vill göra en prognos för totala antalet misshandelfall år 2006. Dessutom vill man skatta hur stor ökning av antalet fall det är i genomsnitt varje år.

Nedan visas två grafer där en regressionslinje har lagts in.

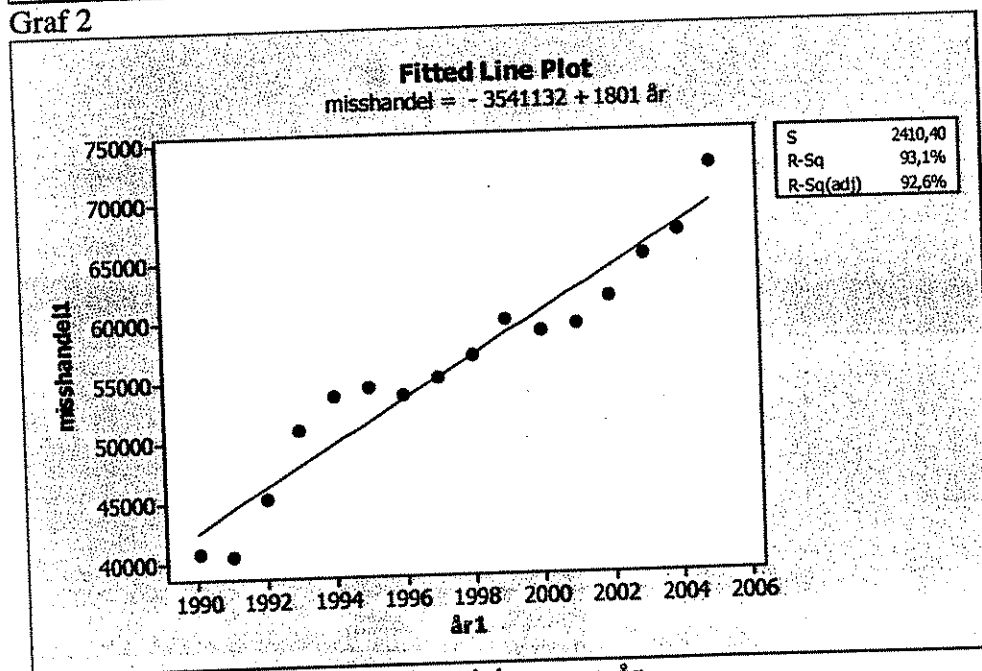
- a) I den första grafen där alla data har använts är inte den anpassade linjen en bra modell. Förklara varför. 1p
- b) Gör en prognos för totala antalet misshandelfall år 2006, dels med hjälp av regressionslinjen i den första grafen och dels med den andra. Vilken prognos anser du vara bäst? Motivera. 2p

- c) Skatta hur stor ökning av antalet fall det är i genomsnitt varje år med hjälp av den andra grafen. 1p
- d) Pröva på 1% nivå om det är signifikant ökning av antalet fall per år. Använd modellen i graf 2 samt utskriften som kommer efter grafen. 1p

Graf 1



Graf 2



**Regression Analysis: misshandel versus år**

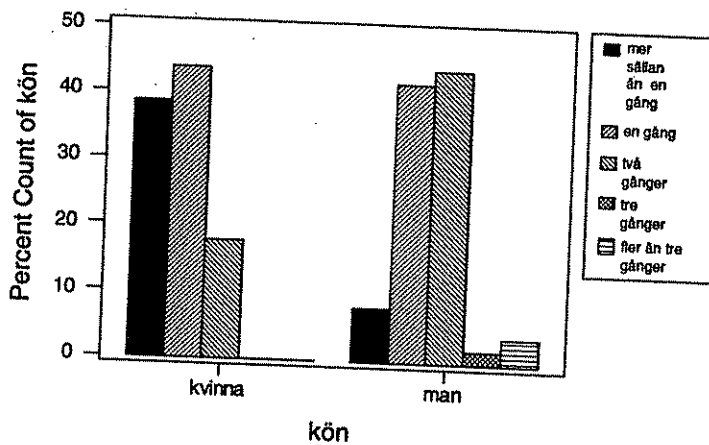
The regression equation is  
 misshandel = - 3541132 + 1801 år

Predictor	Coef	SE Coef
Constant	-3541132	261118
År	1800,8	130,7

S = 2410,40    R-Sq = 93,1%    R-Sq(adj) = 92,6%

**3**

Vid en undersökning av studenternas alkoholvanor vid campus valla hösten 03 valdes ett urval av studenter ut och följande fråga ställdes: 'Hur ofta dricker du alkohol per vecka?'. Uppdelat på kön blev resultatet.



Som vi ser i figuren så kunde man svara med 5 olika alternativ på frågan. Följande omkodning gjordes

1= mer sällan än 1 gång

2= en gång

3= 2 eller fler gånger

Den sista är alltså en sammanslagning av de tre sista alternativen. Detta pga för få observationer här.

Nedan visas resultatet i en frekvenstabell

Antal ggr \ kön	Man	Kvinna
1	4	37
2	21	42
3	25	17

Pröva med ett  $\chi^2$  - test om det finns ett signifikant samband mellan kön och hur ofta studenterna dricker per vecka. Tolka resultatet

3p

4

I ett företag vill man jämföra sjukfrånvaron på två avdelningar A och B. Antalet sjukdagar under ett visst år för ett urval av anställda för de båda avdelningarna ges nedan:

Avd A: 0, 4, 13, 52, 8, 0, 2, 62, 1, 3, 12, 45, 3, 1, 25, 15, 0, 0

Avd B: 7, 9, 2, 12, 8, 4, 10, 6, 9, 7, 10, 8, 1, 0, 0, 5, 12, 5

samt sammanställningen  $\sum A = 256$ ,  $\sum A^2 = 10020$ ,  $\sum B = 115$ ,  $\sum B^2 = 983$

För att utföra analyserna nedan är egentligen stickprovsstorlekarna för små men anta ändå här att normalfördelade (z) testfunktioner är tillåtna. Använd signifikansnivå 10%. Var noga med att ange rätt hypoteser samt testfunktion.

a) Pröva om medelantalet sjukdagar skiljer sig åt mellan de båda avdelningarna.

3p

b) Pröva om andelen anställda som har 3 eller färre sjukdagar per år skiljer sig åt mellan avdelningarna.

2p

c) Kan man med ledning från resultaten ovan dra någon slutsats om att någon avdelning har en 'bättre' statistik på antalet sjukdagar gentemot den andra avdelningen? Ured. (Om du inte löst a och b ovan så kan du försöka utreda med hjälp av lådagram)

1p

d) Beräkna p-värdena i a- och b-uppgiften ovan.

2p

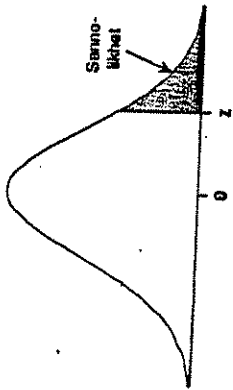


# Tabell 1 Normalfördelningen

Tabellen ger sannolikheten att en standardiserad normalfördelad variabel är större än ett visst värde z.

Exempel: Sannolikheten att få ett värde större än 1,23 är 10,9%.

Den standardiserade normalfördelningen är symmetrisk kring punkten noll. Därför ges sannolikheterna endast för positiva z-värden.

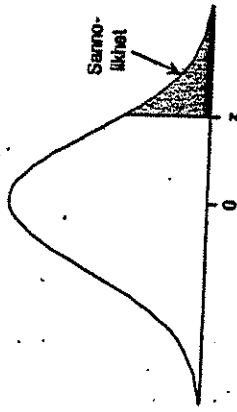


z	0	1	2	3	4	5	6	7	8	9
0,0	50,0	49,6	49,2	48,8	48,4	48,0	47,6	47,2	46,8	46,4
0,1	46,0	45,6	45,2	44,8	44,4	44,0	43,6	43,3	42,9	42,5
0,2	42,1	41,7	41,3	40,9	40,5	40,1	39,7	39,4	39,0	38,6
0,3	38,2	37,8	37,4	37,1	36,7	36,3	35,9	35,6	35,2	34,8
0,4	34,5	34,1	33,7	33,4	33,0	32,6	32,3	31,9	31,6	31,2
0,5	30,9	30,5	30,2	29,8	29,5	29,1	28,8	28,4	28,1	27,8
0,6	27,4	27,1	26,8	26,4	26,1	25,8	25,5	25,1	24,8	24,5
0,7	24,2	23,9	23,6	23,3	23,0	22,7	22,4	22,1	21,8	21,5
0,8	21,2	20,9	20,6	20,3	20,0	19,8	19,5	19,2	18,9	18,7
0,9	18,4	18,1	17,9	17,6	17,4	17,1	16,9	16,6	16,4	16,1
1,0	15,9	15,6	15,4	15,2	14,9	14,7	14,5	14,2	14,0	13,8
1,1	13,6	13,3	13,1	12,9	12,7	12,5	12,3	12,1	11,9	11,7
1,2	11,5	11,3	11,1	10,9	10,7	10,6	10,4	10,2	10,0	9,9
1,3	9,7	9,5	9,3	9,2	9,0	8,9	8,7	8,5	8,4	8,2
1,4	8,1	7,9	7,8	7,6	7,5	7,4	7,2	7,1	6,9	6,8
1,5	6,7	6,6	6,4	6,3	6,2	6,1	5,9	5,8	5,7	5,6
1,6	5,5	5,4	5,3	5,2	5,1	4,9	4,8	4,7	4,6	4,6
1,7	4,5	4,4	4,3	4,2	4,1	4,0	3,9	3,8	3,8	3,7
1,8	3,6	3,5	3,4	3,4	3,3	3,2	3,1	3,1	3,0	2,9
1,9	2,9	2,8	2,7	2,7	2,6	2,6	2,5	2,4	2,4	2,3
2,0	2,3	2,2	2,2	2,1	2,1	2,0	2,0	1,9	1,9	1,8
2,1	1,8	1,7	1,7	1,7	1,6	1,6	1,5	1,5	1,5	1,4
2,2	1,4	1,4	1,3	1,3	1,3	1,2	1,2	1,2	1,1	1,1
2,3	1,1	1,0	1,0	1,0	1,0	0,9	0,9	0,9	0,9	0,8
2,4	0,8	0,8	0,8	0,8	0,7	0,7	0,7	0,7	0,7	0,6

z	0	1	2	3	4	5	6	7	8	9
2,5	0,6	0,6	0,6	0,6	0,6	0,5	0,5	0,5	0,5	0,5
2,6	0,5	0,5	0,4	0,4	0,4	0,4	0,4	0,4	0,4	0,4
2,7	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3
2,8	0,3	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
2,9	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,1	0,1
3,0	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1

# Tabell 2 Normalfördelningen

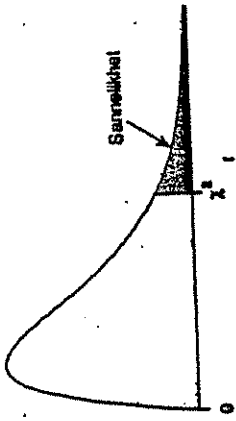
Kritiska värden. z-värden för vissa sannolikheter.



Sannolikhet	z	Sannolikhet	z	Sannolikhet	z
50%	0,00	9%	1,34	2,5%	1,96
40%	0,25	8%	1,41	1,0%	2,33
30%	0,52	7%	1,48	0,5%	2,58
20%	0,84	6%	1,55	0,1%	3,09
10%	1,28	5%	1,64	0,05%	3,29

# Tabell 4 $\chi^2$ -fördelningen

Kritiska värden.  $\chi^2$ -värden för vissa sannolikheter.

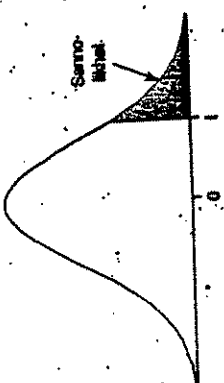


fg.	Sannolikhet		
	5%	1%	0,1%
26	38,89	45,64	54,05
27	40,11	46,96	55,48
28	41,34	48,28	56,89
29	42,56	49,59	58,30
30	43,77	50,89	59,70
31	44,99	52,19	61,10
32	46,19	53,48	62,49
33	47,40	54,78	63,87
34	48,60	56,06	65,25
35	49,80	57,34	66,62
36	51,00	58,62	67,99
37	52,19	59,89	69,35
38	53,38	61,16	70,71
39	54,57	62,43	72,06
40	55,76	63,69	73,41
41	56,94	64,95	74,75
42	58,12	66,21	76,09
43	59,30	67,46	77,41
44	60,48	68,71	78,75
45	61,66	69,96	80,08
46	62,83	71,20	81,39
47	64,00	72,44	82,72
48	65,17	73,68	84,03
49	66,34	74,92	85,35
50	67,50	76,15	86,66

fg	Sannolikhet		
	5%	1%	0,1%
1	3,84	6,63	10,83
2	5,99	9,21	13,82
3	7,81	11,34	16,27
4	9,49	13,28	18,47
5	11,07	15,09	20,52
6	12,59	16,81	22,46
7	14,07	18,48	24,32
8	15,51	20,09	26,12
9	16,92	21,67	27,88
10	18,31	23,21	29,59
11	19,68	24,72	31,26
12	21,03	26,22	32,91
13	22,36	27,69	34,53
14	23,68	29,14	36,12
15	25,00	30,58	37,70
16	26,30	32,00	39,25
17	27,59	33,41	40,79
18	28,87	34,81	42,31
19	30,14	36,19	43,82
20	31,41	37,57	45,31
21	32,67	38,93	46,80
22	33,92	40,29	48,27
23	35,17	41,64	49,73
24	36,42	42,98	51,18
25	37,65	44,31	52,62

# Tabell 3 t-fördelningen

Kritiska värden. t-värden för vissa sannolikheter.



fg	Sannolikhet				
	5%	2,5%	1%	0,5%	0,05%
5	2,02	2,57	3,36	4,03	6,87
6	1,94	2,45	3,14	3,71	5,96
7	1,89	2,36	3,00	3,50	5,41
8	1,86	2,31	2,90	3,36	5,04
9	1,83	2,26	2,82	3,25	4,78
10	1,81	2,23	2,76	3,17	4,59
11	1,80	2,20	2,72	3,11	4,44
12	1,78	2,18	2,68	3,05	4,32
13	1,77	2,16	2,65	3,01	4,22
14	1,76	2,14	2,62	2,98	4,14
15	1,75	2,13	2,60	2,95	4,07
16	1,75	2,12	2,58	2,92	4,01
17	1,74	2,11	2,57	2,90	3,97
18	1,73	2,10	2,55	2,88	3,92
19	1,73	2,09	2,54	2,86	3,88
20	1,72	2,09	2,53	2,85	3,85
21	1,72	2,08	2,52	2,83	3,82
22	1,72	2,07	2,51	2,82	3,79
23	1,71	2,07	2,50	2,81	3,77
24	1,71	2,06	2,49	2,80	3,75
25	1,71	2,06	2,49	2,79	3,73
26	1,71	2,06	2,48	2,78	3,71
27	1,70	2,05	2,47	2,77	3,69
28	1,70	2,05	2,47	2,76	3,67
29	1,70	2,05	2,46	2,76	3,66
30	1,70	2,04	2,46	2,75	3,65
40	1,68	2,02	2,42	2,70	3,55
60	1,67	2,00	2,39	2,66	3,46
120	1,66	1,98	2,36	2,62	3,37
$\infty$	1,64	1,96	2,33	2,58	3,29

# Beskrivande statistik

Medelvärde

$$\bar{x} = \frac{\sum x}{n}$$

Standardavvikelsen

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$$

Regressionslinjen

$$y = a + b \cdot x$$

$$a = \bar{y} - b \cdot \bar{x}$$
$$b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

Korrelationskoefficienten

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

Residualspridningen

$$s_e = \sqrt{\frac{\sum (y - \hat{y})^2}{n-2}} = \sqrt{\frac{(n-1) \cdot s_y^2 \cdot (1-r^2)}{n-2}}$$

# Konfidensintervall

Ett stickprov, populationens andel

$$p \pm z \cdot \sqrt{\frac{p(1-p)}{n}}$$

$$n = z^2 \cdot \frac{p(1-p)}{(d/2)^2}$$

(stickprovets storlek,  $d$  är intervallets bredd)

Ett stickprov, populationens medelvärde

$$\bar{x} \pm z \cdot \frac{s}{\sqrt{n}}$$

(stort stickprov)

$$\bar{x} \pm t \cdot \frac{s}{\sqrt{n}}$$

(normalfördelad variabel,  $n - 1$  frihetsgrader)

$$n = z^2 \cdot \frac{s^2}{(d/2)^2}$$

TVå oberoende stickprov, skillnader mellan andelar

$$(p_1 - p_2) \pm z \cdot \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

TVå oberoende stickprov, skillnader mellan medelvärden

$$(\bar{x}_1 - \bar{x}_2) \pm z \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(\bar{x}_1 - \bar{x}_2) \pm t \cdot \sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

där

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

( $n_1 + n_2 - 2$  frihetsgrader)

# Hypotesprövning, testfunktioner

Ett stickprov, populationens andel

$$z = \frac{p - P_0}{\sqrt{\frac{P_0(1 - P_0)}{n}}}$$

Ett stickprov, populationens medelvärde

$$z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

Two independent samples, differences between proportions

$z = \frac{p_1 - p_2}{\sqrt{p(1-p) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$  ( $p$  is the proportion in the combined sample)

Two independent samples, differences between means

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$\chi^2$ -metoden

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$