



Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2011-04-09
Sal (1) Om tentan går i flera salar ska du bifoga ett försättsblad till varje sal och <u>ringa in</u> vilken sal som avses	TER1
Tid	8-12
Kurskod	732G60
Provkod	TEN1
Kursnamn/benämning Provnamn/benämning	Statistiska metoder Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	4
Jour/Kursansvarig Ange vem som besöker salen	Oleg Sysoev
Telefon under skrivtiden	0735673482
Besöker salen ca kl.	10
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	carita.lilja@liu.se tel 1463
Tillåtna hjälpmedel	Med skrivningen häftad formelsamling. Räknedosa.
Övrigt	
Vilken typ av papper ska användas, rutigt eller linjerat	rutigt
Antal exemplar i påsen	

Tentamen

Linköpings Universitet, Institutionen för datavetenskap, Statistik

- Kurskod och namn: 732G60 Statistiska Metoder
- Datum och tid: 2011-04-09, 8-12
- Jourhavande lärare: Oleg Sysoev
- Tillåtna hjälpmedel: Med skrivningen häftad formelsamling. Räknedosa.
- Betygsgränser: Tentamen omfattar totalt 20p. Godkänt från 12p, väl godkänt från 16p.
Siffrorna i uppgifterna är påhittade.
-

Redovisa och motivera tydligt alla dina lösningar!

Formelsamlingen finns på de sista sidorna

Uppgift 1 (6p)

- a) En partisympatiundersökning som omfattade 1000 personer från Linköping visade att 50.5% var beredda att rösta på parti S medan undersökningen i Norrköping som omfattade 900 personer visade att andelen sympatisörer för S var 53.0%. Beräkna ett 95% konfidensintervall som visar skillnaden mellan andelen sympatisörer i Norrköping och Linköping. Kan man konstatera att andelar sympatisörer är olika i dessa två städer? Motivera. (3p)
- b) Man kan presentera samma information som i A i form av en tabell:

Kategori	Antal
Röstar på S	982
Röstar inte på S	918

Undersök med hjälp av chitvå-test med konfidensgraden 99% om andelar personer som röstar på S och inte röstar på S är samma i populationen. (3p)

Uppgift 2 (3p)

I ett experiment mättes hur lång tid det tar för män som är 20 år gamla att springa uppför en trappa. Tabellen visar tiden för varje person (i sekunder)

Män	9.8	10.2	9.0	11.2	9.5	9.8
-----	-----	------	-----	------	-----	-----

Genomför en hypotesprövning som visar med 99% sannolikhet om männen kan springa uppför denna trappa på mindre än 10 sekunder i genomsnitt (3p)

Uppgift 3 (5p)

Följande tabell visar priser (Ls, lettiska lats) och konsumtion (i 1000-tal Ls) av två olika varor i en butik i Lettland.

	06/2008	01/2009	06/2009	01/2010	06/2010
Pris per kilo, bovete	0.45	0.60	0.53	0.64	0.85
Pris per kilo, ris	0.66	0.74	0.70	0.75	0.70
Konsumtion, bovete	22.5	23.7	22.5	23.2	30.1
Konsumtion, ris	28.2	30.9	27.4	31.2	28.9

- Omräkna priserna till indexserier med tidpunkt 01/2009 som basår. (2p)
- Uttryck konsumtionsvärdena för bovete i priserna av tidpunkt 01/2009. (3p)

Uppgift 4 (6p)

Tabellen nedan redovisar löner av lärare från flera skolor i USA (**Pay**, i 1000-tal dollar) och motsvarande omkostnader per elev (**Spend**, i 1000-tal dollar)

Pay	24,2	27,1	30,1	26,5	27,3	21,6	21,9	20,8	18	20,9
Spend	3,1	3,6	3,7	4,2	3,9	3,5	3,1	3	2,9	3,2

- a) Betrakta en modell med Pay som respons och Spend som förklaringsvariabel och skatta regressionskoefficienter med minsta-kvadratmetoden (redovisa uträkning) och redogör den skattade regressionsekvationen. Tolka båda regressionskoefficienterna. (4p)
- b) Skatta residualerna och rita ett histogram som visar hur residualerna är fördelade. Histogrammets startpunkt ska vara lika med -3 och stolpvidden ska vara lika med 2. Liknar denna fördelning en normal fördelning? (2p)

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

Medianen

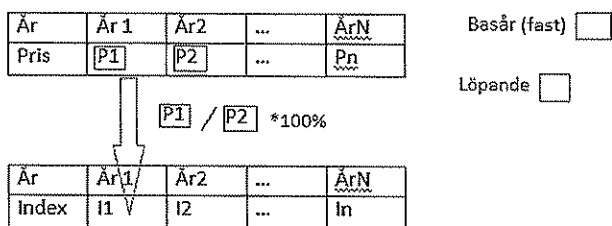
- Udda n $x_m = x_{(n+1)/2}$
- Jämnt n $x_m = \frac{x_{n/2} + x_{n/2+1}}{2}$

Standardavvikelse: $s = \sqrt{\frac{n}{n-1} (\overline{x^2} - \bar{x}^2)}$

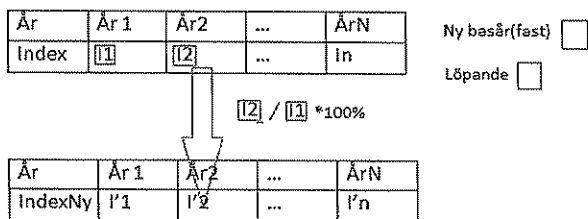
Lådagrammets punkter:

- Uteiggare (punkter $< Q1 - 1.5 * \text{kvartilavstånd}$)
- Max(minvärde, $Q1 - 1.5 * \text{kvartilavstånd}$)
- Q1
- Medianen
- Q3
- Min(maxvärde, $Q3 + 1.5 * \text{kvartilavstånd}$)
- Uteiggare (punkter $> Q3 + 1.5 * \text{kvartilavstånd}$)

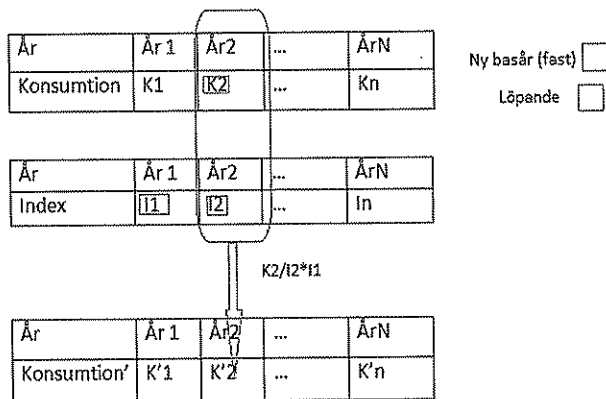
Omräkning till Index



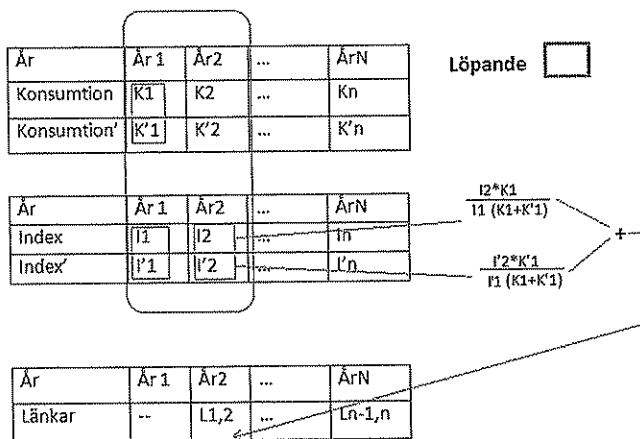
Byte av basår



Deflatering



Kedjeprisindex



$$I_t = L_{1,2} \cdot L_{2,3} \cdot \dots \cdot L_{t-1,t} \cdot 100$$

Regression

$$y = a + bx \quad r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2)(\overline{y^2} - \bar{y}^2)}} \quad b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} \quad a = \bar{y} - b\bar{x}$$

Residualer

$$e = y - \hat{y} \quad \text{där} \quad \hat{y} = a + bx_i$$

Chitvå-metoden

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

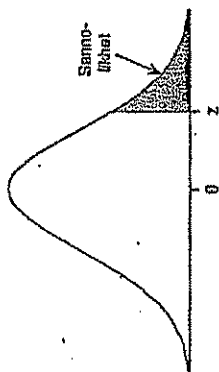
Frihetsgraderna: K-1 för envägsindlad tabell och (R-1)(K-1) för korstabell.

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.

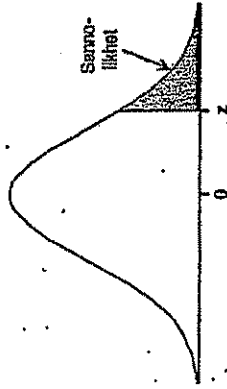


z	0	1	2	3	4	5	6	7	8	9
0,0	50,0	49,6	49,2	48,8	48,4	48,0	47,6	47,2	46,8	46,4
0,1	46,0	45,6	45,2	44,8	44,4	44,0	43,6	43,3	42,9	42,5
0,2	42,1	41,7	41,3	40,9	40,5	40,1	39,7	39,4	39,0	38,6
0,3	38,2	37,8	37,4	37,1	36,7	36,3	35,9	35,6	35,2	34,8
0,4	34,5	34,1	33,7	33,4	33,0	32,6	32,3	31,9	31,6	31,2
0,5	30,9	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	0,1	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

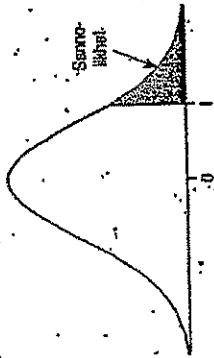
Kritiska värden, z -värden för vissa sannolikheter.



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 3 t-fördelningen

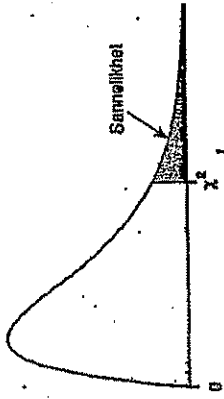
Kritiska värden. t-värden för vissa sannolikheter.



fg	Sannolikheter						
	5%	2,5%	1%	0,5%	0,1%	0,05%	0,01%
5	2,02	2,57	3,36	4,03	5,89	6,87	8,87
6	1,94	2,45	3,14	3,71	5,21	5,96	7,87
7	1,89	2,36	3,00	3,50	4,79	5,41	7,29
8	1,86	2,31	2,90	3,36	4,50	5,04	6,87
9	1,83	2,26	2,82	3,25	4,30	4,78	6,62
10	1,81	2,23	2,76	3,17	4,14	4,59	6,40
11	1,80	2,20	2,72	3,11	4,02	4,44	6,21
12	1,78	2,18	2,68	3,05	3,93	4,32	6,04
13	1,77	2,16	2,65	3,01	3,85	4,22	5,89
14	1,76	2,14	2,62	2,98	3,79	4,14	5,75
15	1,75	2,13	2,60	2,95	3,73	4,07	5,62
16	1,75	2,12	2,58	2,92	3,69	4,01	5,50
17	1,74	2,11	2,57	2,90	3,65	3,97	5,39
18	1,73	2,10	2,55	2,88	3,61	3,92	5,29
19	1,73	2,09	2,54	2,86	3,58	3,88	5,20
20	1,72	2,09	2,53	2,85	3,55	3,85	5,12
21	1,72	2,08	2,52	2,83	3,53	3,82	5,05
22	1,72	2,07	2,51	2,82	3,50	3,79	4,98
23	1,71	2,07	2,50	2,81	3,48	3,77	4,92
24	1,71	2,06	2,49	2,80	3,47	3,75	4,86
25	1,71	2,06	2,49	2,79	3,45	3,73	4,80
26	1,71	2,06	2,48	2,78	3,44	3,71	4,75
27	1,70	2,05	2,47	2,77	3,42	3,69	4,70
28	1,70	2,05	2,47	2,76	3,41	3,67	4,65
29	1,70	2,05	2,46	2,76	3,40	3,66	4,60
30	1,70	2,04	2,46	2,75	3,39	3,65	4,55
40	1,68	2,02	2,42	2,70	3,31	3,55	4,42
60	1,67	2,00	2,39	2,66	3,23	3,46	4,30
120	1,66	1,98	2,36	2,62	3,16	3,37	4,18
∞	1,64	1,96	2,33	2,58	3,09	3,29	4,07

Tabell 4 χ^2 -fördelningen

Kritiska värden. χ^2 -värden för vissa sannolikheter.



fg	Sannolikheter			
	5%	1%	0,1%	0,01%
1	3,84	6,63	10,83	15,49
2	5,99	9,21	13,82	19,00
3	7,81	11,34	16,27	21,78
4	9,49	13,28	18,47	24,00
5	11,07	15,09	20,52	25,99
6	12,59	16,81	22,46	27,73
7	14,07	18,48	24,32	29,33
8	15,51	20,09	26,12	30,83
9	16,92	21,67	27,88	32,19
10	18,31	23,21	29,59	33,41
11	19,68	24,72	31,26	34,53
12	21,03	26,22	32,91	35,57
13	22,36	27,69	34,53	36,53
14	23,68	29,14	36,12	37,41
15	25,00	30,58	37,70	38,29
16	26,30	32,00	39,25	39,16
17	27,59	33,41	40,79	39,99
18	28,87	34,81	42,31	40,81
19	30,14	36,19	43,82	41,62
20	31,41	37,57	45,31	42,43
21	32,67	38,93	46,80	43,23
22	33,92	40,29	48,27	44,02
23	35,17	41,64	49,73	44,81
24	36,42	42,98	51,18	45,60
25	37,65	44,31	52,62	46,39
26	38,89	45,64	54,05	47,18
27	40,11	46,96	55,48	47,97
28	41,34	48,28	56,89	48,76
29	42,56	49,59	58,30	49,55
30	43,77	50,89	59,70	50,34
31	44,99	52,19	61,10	51,13
32	46,19	53,49	62,49	51,92
33	47,40	54,78	63,87	52,71
34	48,60	56,06	65,25	53,50
35	49,80	57,34	66,62	54,29
36	51,00	58,62	67,99	55,08
37	52,19	59,89	69,35	55,87
38	53,38	61,16	70,71	56,66
39	54,57	62,43	72,06	57,45
40	55,76	63,69	73,41	58,24
41	56,94	64,95	74,75	59,03
42	58,12	66,21	76,09	59,82
43	59,30	67,46	77,41	60,61
44	60,48	68,71	78,75	61,40
45	61,66	69,96	80,08	62,19
46	62,83	71,20	81,39	62,98
47	64,00	72,44	82,72	63,77
48	65,17	73,68	84,03	64,56
49	66,34	74,92	85,35	65,35
50	67,50	76,15	86,66	66,14

Beskrivande statistik

Medelvärde

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

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

Standardavvikelsen

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

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

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 intervalls 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)

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

Two oberoende stickprov, skillnader mellan andelar

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

Two 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

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

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

TVå oberoende stickprov, skillnader mellan andelar

$$z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

(p är andelen i det sammanslagna stickprovet)

TVå oberoende stickprov, skillnader mellan medelvärden

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

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

χ^2 -metoden

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