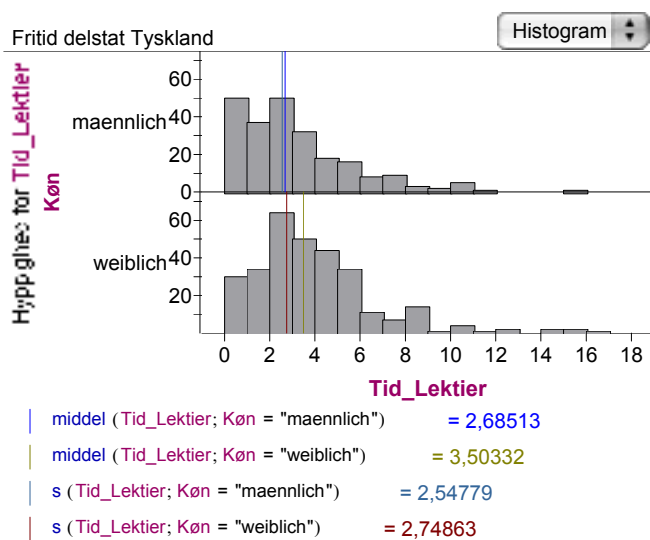


Fritid delstat Tyskland

	Køn	Alder	Højde	Vægt	BMI	Tid_TV	Tid_Musik	Tid_Comp	Tid_Sport	Tid_Lekt...	Tid_Job
=											
1	maennlich	17	1,88	70	19,8053	10	2	4	2	3	
2	weiblich					14	0	35	0	4	
3	weiblich	17	1,7	56	19,3772	7	3	3	1,5	2	
4	weiblich	17	1,7	51	17,6471	10	1	2	4,5	0,5	
5	weiblich	17	1,8	75	23,1481	25	10	8	3	2	



Blandt de 538 unge, der er interviewet, læser pigerne i gennemsnit (middelværdi) 0,8 timer mere lektier end drengene. Vi sammenligner middelværdien for pigerne 3,5 med middelværdien for drengene 2,7.

Hypotese: Gælder dette for alle unge tyskere ?

Vi deler op i to hypoteser H_0 og H_1 .

H_0 : Der er ikke nogen forskel.

H_1 : Der er en forskel.

Det store spørgsmål: Hvordan undersøger vi det ?

Trin et: Vi rører rundt i vores datasæt, så drenge og pigerne fordeles tilfældigt.

Trin to: Vi laver 1000 omrøringer af vores datasæt, så det svarer til 538000 unge mennesker.



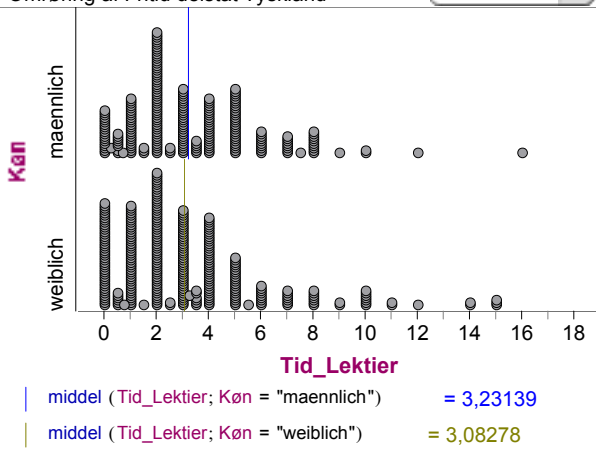
Omrøring af Fritid delstat Tyskland

Omrøring af Fritid delstat Tyskland

	Køn	Alder	Højde	Vægt	BMI	Tid_TV	Tid_Musik	Tid_Comp	Tid_Sport
1	weiblich	17	1,88	70	19,8053	10	2	4	2
2	maennlich					14	0	35	0
3	weiblich	17	1,7	56	19,3772	7	3	3	1,5
4	maennlich	17	1,7	51	17,6471	10	1	2	4,5
5	maennlich	17	1,8	75	23,1481	25	10	8	3
6	maennlich	16	1,9	80	22,1607	15	10	10	4
7	weiblich	17	1,8	85	26,2346	21	0	0	0
8	weiblich	17	1,85	73	21,3294	11	4	3	1
9	weiblich	17	1,6	46	17,9688	15	3	0	3
10	maennlich	17	1,65	53	19,4674	15	1	1	3
11	weiblich	17	1,85	93,7	27,3776	5	5	10	5

Omrøring af Fritid delstat Tyskland

Prikdiagram



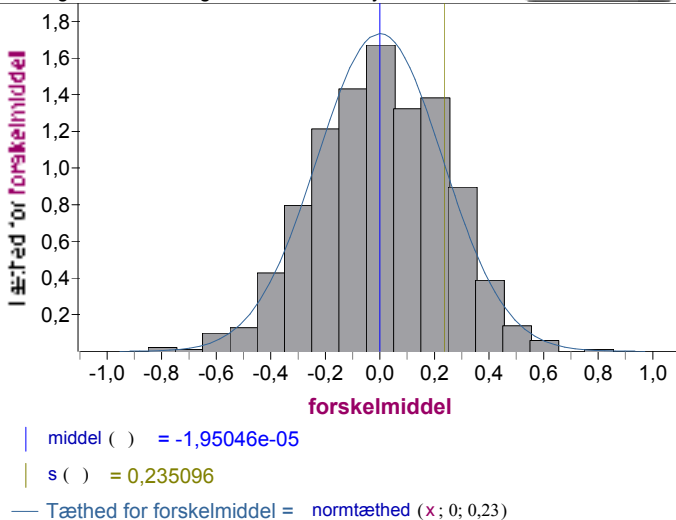
I det omrørte datasæt er forskellen mellem de to middelværdier meget mindre 0,1.



Målinger fra Omrøring af Fritid delstat Tyskland

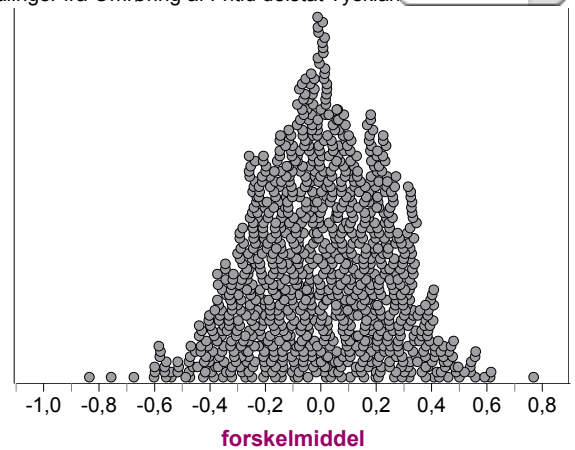
Målinger fra Omrøring af Fritid delstat Tyskland

Histogram



Målinger fra Omrøring af Fritid delstat Tyskland

Prikdiagram



Når drenge og piger fordeles tilfældigt, så er 2 til 3 tilfælde ud af 1005 tilfælde, hvor forskellen i lektielæsningen er lige så stor, som i det oprindelige datasæt. I statistikken siger man, at vi har en eksperimente p-værdi $3/1005=0,003$. Derfor må vi afvise H_0 hypotesen. Derfor er forskellen mellem lektielæsningen ikke kun tilfældig. Omvendt accepterer vi H_1 hypotesen.

Forskellen i middelværdi er normalfordelt med en middelværdi og spredning.

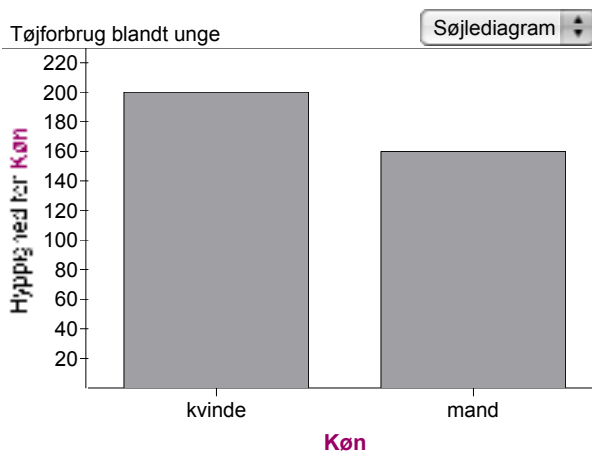
Vi har estimeret middelværdien til 0 og spredningen 0,23.

Grafen for normaldelingen er en smuk klokkeform.



Analyse af data

Tøjforbrug blandt unge



Tøjforbrug blandt unge

	Køn	Tøjforbrug
1	kvinde	Under 15..
2	kvinde	Under 15..
3	kvinde	Under 15..
4	kvinde	Under 15..
5	kvinde	Under 15..
6	kvinde	Under 15..
7	kvinde	Under 15..
8	kvinde	Under 15..
9	kvinde	Under 15..
10	kvinde	Under 15..
11	kvinde	Under 15..
12	kvinde	Under 15..
13	kvinde	Under 15..
14	kvinde	Under 15..
15	kvinde	Under 15..
16	kvinde	Under 15..
17	kvinde	Under 15..
18	kvinde	Under 15..

tæl ()

Tøjforbrug blandt unge

		Tøjforbrug		Række total
		Over 1500	Under 1500	
Køn	kvinde	102	98	200
		112,22222	87,777778	200
	mand	100	60	160
		89,777778	70,222222	160
Søjle total		202	158	360
		202	158	360

R1 = tæl ()

R2 = Forventet

Celler fra oversigtstabel for Tøjforbrug blandt unge

	Køn	Tøjforbrug	R1	R2	chi2bidrag	chi2sum
=					$\frac{(R1 - R2)^2}{R2}$	chi2bidrag + forrige ()
1	kvinde	Over 1500	102	112,222	0,931133	0,931133
2	kvinde	Under 15...	98	87,7778	1,19044	2,12157
3	mand	Over 1500	100	89,7778	1,16392	3,28549
4	mand	Under 15...	60	70,2222	1,48805	4,77353

Vi får en chi2 teststørrelse på 4,77

Et omrørt datasæt - dvs. køn er fordelt tilfældigt



Celler fra oversigtstabel for Tøjforbrug blandt unge



Omrøring af Tøjforbrug blandt unge

Omrøring af Tøjforbrug blandt unge

		Tøjforbrug		Række total
		Over 1500	Under 1500	
Køn	kvinde	111	89	200
		112,22222	87,777778	200
	mand	91	69	160
		89,777778	70,222222	160
Søjle total		202	158	360
		202	158	360

R1 = tæl ()

R2 = Forventet



Celler fra oversigtstabel for Omrøring af Tøjforbrug blandt unge

Celler fra oversigtstabel for Omrøring af Tøjforbrug blandt unge

	Køn	Tøjforbrug	R1	R2	chi2bidrag	chi2sum
=					$(R1 - R2)^2$	chi2bidrag + f
1	kvinde	Over 1500	111	112,222	0,0133113	0,0133113
2	kvinde	Under 15...	89	87,7778	0,0170183	0,0303296
3	mand	Over 1500	91	89,7778	0,0166392	0,0469688
4	mand	Under 15...	69	70,2222	0,0212729	0,0682416

I det omrørte datasæt får vi en chi2 teststørrelse på 0,068.

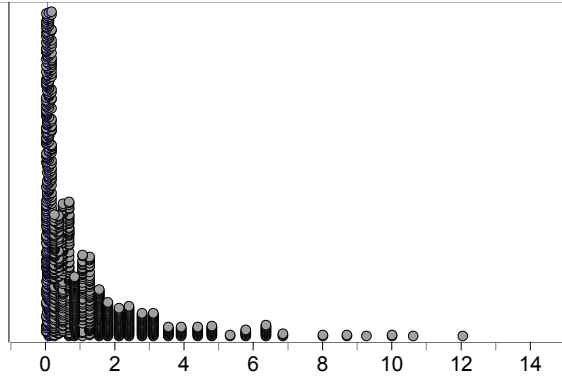
Nu vil vi lave 1000 omrøringer og se, hvordan de er fordelt. Og sammenligne teststørrelserne med den, vi fik i det oprindelige datasæt.

1000 omrørte datasæt



Målinger fra Celler fra oversigtstabel for Omrøring af Tøjforbrug blandt unge

Målinger fra Celler fra oversigtstabel for Omrøring af Tøjforbrug blandt unge Prikdiagram



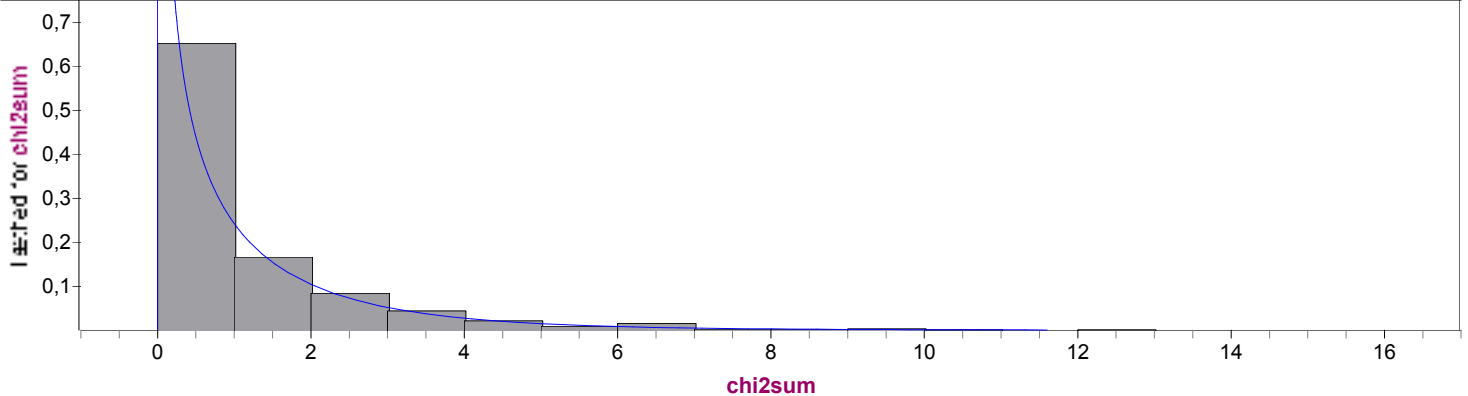
Vi ser, at i de 1000 omrørte datasæt, hvor køn er tilfældigt valgt, da er der 4,4 %, som har en chi2 teststørrelse, der er mindst lige så stor, som i det oprindelige datasæt. De 4,4% er vores eksperimentelle p-værdi. Da den er mindre end vores signifikans niveau på 5%, så er vores sammenhæng mellem køn og tøjforbrug ikke tilfældig.

andel (chi2sum \geq 4,77) = 0,044

Bag de mange chi2teststørrelser ligger der en graf for funktionen ved navn tæthedChi2.

Målinger fra Celler fra oversigtstabel for Omrøring af Tøjforbrug blandt unge

Histogram



Tæthed for chi2sum = $\text{chiTæthed}(x; 1)$

Med en frihedsgrad passer graf og data fint.



OverlevelseOperationstype

OverlevelseOperationstype

	Operationstype	Overlevet
=	$\text{ombyt}(\text{indeks})$	$\text{indeks} \leq 54$ (indeks < 70)
1	Operation A	I live
2	Operation A	I live
3	Operation A	I live
4	Operation A	I live
5	Operation A	I live
6	Operation A	I live
7	Operation A	I live
8	Operation A	I live
9	Operation A	I live

OverlevelseOperationstype

		Overlevet		Række total
		Død	I live	
Operationstype	Operation A	14	40	54
		20,463158	33,536842	54
	Operation B	6	10	16
		6,0631579	9,9368421	16
	Operation C	16	9	25
		9,4736842	15,526316	25
Søjle total		36	59	95
		36	59	95

R1 = tæl ()

R2 = Forventet



Celler fra oversigtstabel for OverlevelseOperationstype

Celler fra oversigtstabel for OverlevelseOperationstype

	Operatio...	Overlevet	R1	R2	Chi2Bidrag	Chi2
=					$(R1 - R2)^2$	sum (Chi2Bid
1	Operatio...	Død	14	20,4632	2,04135	10,5271
2	Operatio...	I live	40	33,5368	1,24557	10,5271
3	Operatio...	Død	6	6,06316	0,000657895	10,5271
4	Operatio...	I live	10	9,93684	0,000401427	10,5271
5	Operatio...	Død	16	9,47368	4,49591	10,5271
6	Operatio...	I live	9	15,5263	2,74326	10,5271



Omrøring af OverlevelseOperationstype

Omrøring af OverlevelseOperationstype

		Overlevet		Række total
		Død	I live	
Operationstype	Operation A	20	34	54
		20,463158	33,536842	54
	Operation B	5	11	16
		6,0631579	9,9368421	16
	Operation C	11	14	25
		9,4736842	15,526316	25
Søjle total		36	59	95
		36	59	95

R1 = tæl ()

R2 = Forventet



Celler fra oversigtstabel for Omrøring af OverlevelseOperationstype

Celler fra oversigtstabel for Omrøring af OverlevelseOperationstype

	Operatio...	Overlevet	R1	R2	chi2bidrag	chi2	<n
=					$(R1 - R2)^2$	sum (chi2bidr	
1	Operatio...	Død	20	20,4632	0,010483	0,713001	
2	Operatio...	I live	34	33,5368	0,006396...	0,713001	
3	Operatio...	Død	5	6,06316	0,186422	0,713001	
4	Operatio...	I live	11	9,93684	0,113749	0,713001	
5	Operatio...	Død	11	9,47368	0,245906	0,713001	
6	Operatio...	I live	14	15,5263	0,150045	0,713001	



Målinger fra Celler fra oversigtstabel for Omrøring af OverlevelseOperationstype

Målinger fra Celler fra oversigtstabel for Omrøring af OverlevelseOperationstype Histogram

