

Øvelser i epidemiologi og biostatistik, 6. april 2010

Baseline-informationer fra Ebeltoft datasættet

Eksempel på besvarelse

1. Hvor stor en andel af deltagerne var mænd? Var der samme andel i de tre randomiseringsgrupper?

```
. use "ebeltoftdata.dta", clear
. tab2 random sex , row chi
```

-> tabulation of random by sex

```
+-----+
| Key    |
+-----+
| frequency |
| row percentage |
+-----+
```

Randomiseringsgruppe	Køn		Total
	1. Female	2. Male	
1. Spørgeskema	104	136	240
	43.33	56.67	100.00
2. Health-check	113	127	240
	47.08	52.92	100.00
3. Helbredssamtale	119	121	240
	49.58	50.42	100.00
Total	336	384	720
	46.67	53.33	100.00

Pearson chi2(2) = 1.9085 Pr = 0.385

2. Hvor stor en andel af mændene var rygere ved baseline? Hvor stor en andel af kvinderne?

```
. recode tobak_0 (0=0) (1/max=1) , generate(ryger_0)
. label variable ryger_0 "Ryger, baseline"
. label define nej_ja 0 "nej" 1 "ja"
. label values ryger_0 nej_ja
. tab2 ryger_0 sex , col chi
```

-> tabulation of ryger_0 by sex

```
+-----+
| Key |
+-----+
| frequency |
| column percentage |
+-----+
```

Ryger,	Køn		
baseline	1. Female	2. Male	Total
nej	182	213	395
	54.17	55.47	54.86
ja	154	171	325
	45.83	44.53	45.14
Total	336	384	720
	100.00	100.00	100.00

Pearson chi2(1) = 0.1227 Pr = 0.726

3. Hvor stor en andel af mændene var storrygere (≥ 15 gram/dag) ved baseline? Hvor stor en andel af kvinderne? Var der signifikant forskel på andelen blandt mænd og kvinder?

```
. recode tobak_0 (0/14=0) (15/max=1) , generate(storryger_0)
. label variable storryger_0 "Ryger mindst 15 g/dag, baseline"
. label values storryger_0 nej_ja
. tab2 storryger_0 sex , col chi
```

-> tabulation of storryger_0 by sex

```
+-----+
| Key |
+-----+
| frequency |
| column percentage |
+-----+
```

Ryger	Køn		
mindst 15	1. Female	2. Male	Total
g/dag,			
baseline			
nej	245	265	510
	72.92	69.01	70.83
ja	91	119	210
	27.08	30.99	29.17
Total	336	384	720
	100.00	100.00	100.00

Pearson chi2(1) = 1.3235 Pr = 0.250

4. Hvor stor en andel drak mere end Sundhedsstyrelsens genstandsgrænser (21/uge for mænd, 14/uge for kvinder) ved baseline?

```
. gen storalk_0=0
. replace storalk_0=1 if sex==1 & alko_0 > 14 & alko_0 < .
. replace storalk_0=1 if sex==2 & alko_0 > 21 & alko_0 < .
```

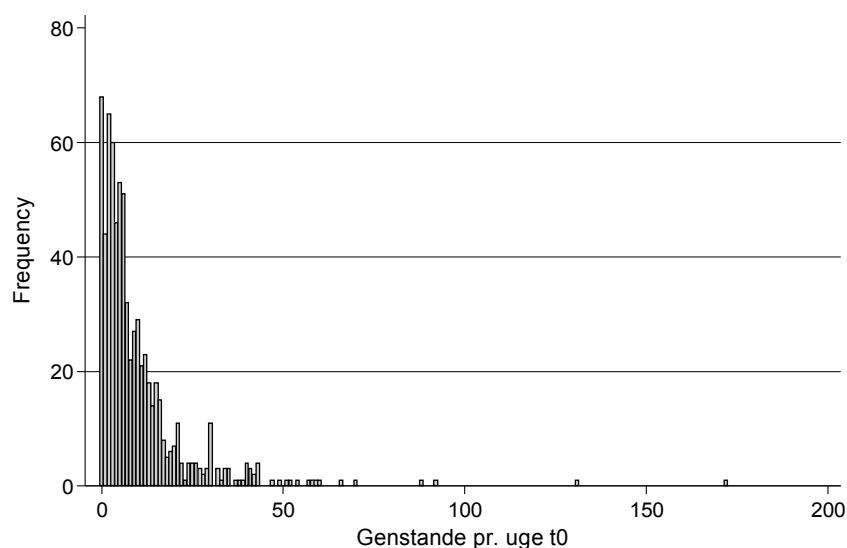
```
. label variable storalk_0 "Alk>14/21, baseline"
. label values storalk_0 nej_ja
. tab1 storalk_0
```

-> tabulation of storalk_0

Alk>14/21, baseline	Freq.	Percent	Cum.
nej	624	86.67	86.67
ja	96	13.33	100.00
Total	720	100.00	

5. Lav et histogram der viser fordelingen af alkoholforbruget ved baseline.

```
. histogram alko_0 , frequency discrete
```



6. Udskriv en liste med baseline-informationer for de deltagere der oplyste at drikke over 50 genstande pr. uge.

```
. list id bdate-tobak_0 if alko_0>50
```

	id	bdate	sex	date_0	alko_0	bmi_0	tobak_0
11.	11	15nov1945	1. Female	24jan1992	59	21.7	25
37.	37	15apr1943	2. Male	20may1992	70	31.6	5
65.	65	15dec1943	2. Male	08jun1992	58	26.8	0
170.	170	15sep1956	2. Male	06may1992	57	33.5	0
221.	221	15jan1960	2. Male	23may1992	88	23.2	15
223.	223	15sep1959	2. Male	24apr1992	172	20.5	27
243.	243	15jul1950	2. Male	08feb1992	60	26	30
393.	393	15may1954	2. Male	08jan1992	131	32.1	23
488.	488	15mar1956	2. Male	12feb1992	66	24	0
509.	509	15jul1942	2. Male	24jan1992	92	28.1	20
642.	642	15jan1957	2. Male	12jan1992	51	25.5	20
696.	696	15nov1949	1. Female	17feb1992	54	23	0
719.	719	15dec1958	2. Male	02apr1992	52	24.6	40

7. Er der forskel på mænds og kvinders BMI?

Uden at tænke nærmere over det, kunne vi prøve med et t-test:

```
. ttest bmi_0 , by(sex)
```

Two-sample t test with equal variances

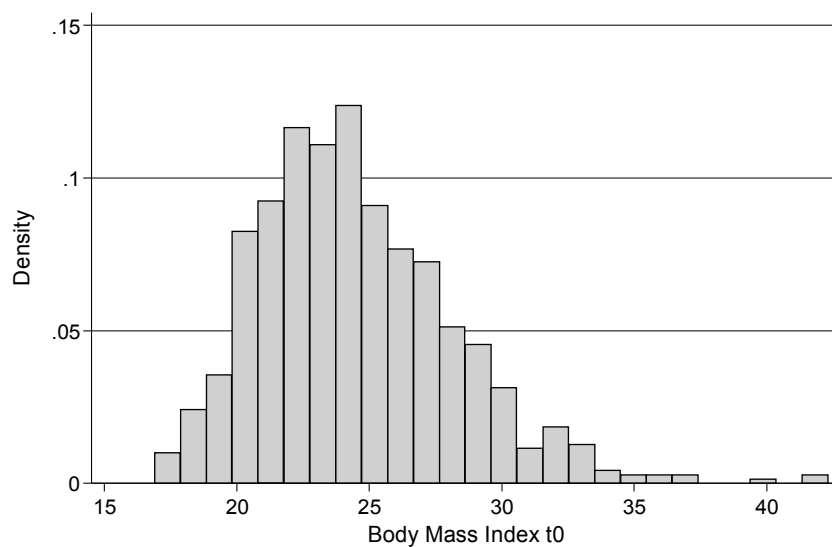
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1. Femal	336	23.20685	.1950856	3.575979	22.8231	23.59059
2. Male	384	25.66771	.1756461	3.441947	25.32236	26.01306
combined	720	24.51931	.1383339	3.711888	24.24772	24.79089
diff		-2.460863	.2618388		-2.974924	-1.946802

diff = mean(1. Femal) - mean(2. Male) t = -9.3984
Ho: diff = 0 degrees of freedom = 718

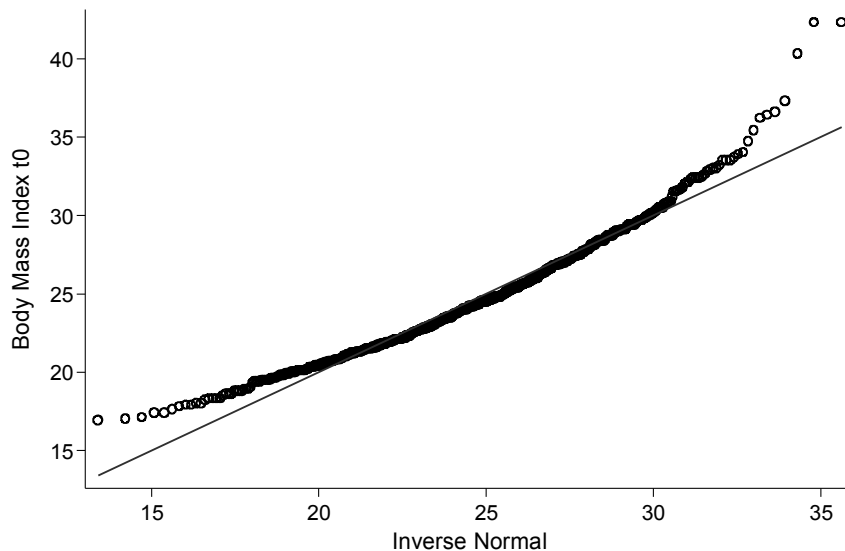
Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 1.0000

Det er den midterste P-værdi (0.0000, dvs. $P < 0.0001$) der gælder. Mændene er tungest med en forskel på 2.5 BMI-enheder. **Imidlertid: Er BMI normalfordelt? Det er jo en betingelse for at resultaterne af t-testet er gyldigt!**

```
. histogram bmi_0
```



```
. qnorm bmi_0
```



BMI er måske ikke helt normalfordelt og t-testet giver ikke nødvendigvis det rette svar; både p-værdier, gennemsnit, standard deviations, standard errors og konfidensintervaller kan være misvisende. Det bedste ville være at undersøge nærmere, hvilken fordeling BMI har, for at kunne give en bedre beskrivelse af forskellen mellem mænd og kvinder. I en snæver vending kan man imidlertid bruge det ikke-parametriske test Kruskal-Wallis, som dog kun giver en p-værdi:

```
. kwallis bmi_0 , by(sex)
```

Test: Equality of populations (Kruskal-Wallis test)

sex	Obs	Rank Sum
1. Female	336	92657.50
2. Male	384	166902.50

```
chi-squared = 104.560 with 1 d.f.
probability = 0.0001
```

```
chi-squared with ties = 104.573 with 1 d.f.
probability = 0.0001
```

Gem til slut dine ændringer til datasættet, så du ikke behøver skrive den samme kode igen:

```
. save "ebeltoftdata2.dta" , replace
```