

Datorövning 3

Statistikens Grunder 2

Syfte

1. Lära sig utföra hypotestest för populationsproportionen
2. Lära sig utföra test för populationsmedelvärdet
3. Lära sig utföra test för skillnaden mellan två populationsparametrar då det är parvisa observationer
4. Lära sig utföra test för skillnaden mellan två populationsparametrar för två oberoende stickprov

Exempel

Utföra hypotestest för populationsproportionen

Utgå från övning 17.2 i kompendiet, kapitel 17. Vi läser in datat

```
data work.ovn172;  
input quality$ count;  
datalines;  
defect 10  
good 150  
;  
run;
```

Vi testar om datat verkar komma från en population med proportionen defekta $\pi = 0.08$

```
proc freq data=work.ovn172 order=data;  
weight count;  
tables quality / binomial(p=0.08);  
run;
```

Kommandot `"/binomial(p = 0.08)"` skriver vi för att utföra testet

$$H_0 : \pi = 0.08$$

Kommandot "*order=data*" använder vi för att datat ska komma i den ordning som vi läste in det. Detta ger utskriften:

The FREQ Procedure				
quality	Frequency	Percent	Cumulative Frequency	Cumulative Percent
defect	10	6.25	10	6.25
good	150	93.75	160	100.00

Binomial Proportion for quality = defect	
Proportion	0.0625
ASE	0.0191
95% Lower Conf Limit	0.0250
95% Upper Conf Limit	0.1000
Exact Conf Limits	
95% Lower Conf Limit	0.0304
95% Upper Conf Limit	0.1119
Test of H0: Proportion = 0.08	
ASE under H0	0.0214
Z	-0.8159
One-sided Pr < Z	0.2073
Two-sided Pr > Z	0.4145

Sample Size = 160

Vi ser i utskriften att proportionen defekta i stickprovet blev 0.0625. De två första konfidensintervallgränserna är framräknade med formeln

$$p \pm 1.96 \sqrt{\frac{0.0625(1 - 0.0625)}{160}}$$

De mått på teststatistikan vi är intresserade av är de tre sista; "*Z*", "*one sided*" och "*two sided*". "*Z*" anger det observerade z-värdet från testet. Det vill säga, SAS använder formeln

$$Z = \frac{p - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/n}}$$

(Jämför värdet på utskriften med era egna uträkningar.) "*One sided*" och "*two sided*" anger *p-värdet* för ett enkelsidigt test

$$H_0 : \pi = 0.08$$

$$H_A : \pi < 0.08$$

respektive dubbelsidigt test

$$H_0 : \pi = 0.08$$

$$H_A : \pi \neq 0.08$$

Vi ska i uppgiften testa om $\pi < 0.08$ och därför bör vårt beslut tas genom att titta på *p-värdet* 0.2073. Ska vi förkasta H_0 ?

Utföra test för populationsmedelvärdet

Utgå från övning 17.4 i kompendiet, kapitel 17. Vi läser in datat

```
data work.ovn174;  
input livslangd;  
datalines;  
2970  
3020  
3005  
2900  
2940  
2925  
;  
run;
```

Nu ska vi använda proceduren "*proc ttest*" för att testa

$$H_0 : \mu = 3000$$

$$H_A : \mu < 3000$$

Vi använder koden

```
proc ttest data=work.ovn174 H0=3000;  
run;
```

Vi skriver " $H_0=3000$ " för att ange vår nollhypotes. Vi får utskriften

```

The TTEST Procedure

Variable:  livslangd

N          Mean      Std Dev    Std Err    Minimum    Maximum

6          2960.0     46.7974    19.1050     2900.0     3020.0

      Mean      95% CL Mean      Std Dev    95% CL Std Dev

2960.0      2910.9   3009.1      46.7974     29.2113    114.8

      DF      t Value    Pr > |t|

      5      -2.09      0.0905

```

Utskriften ger oss olika mått. Intressant för hypotestestning av μ är det som står i raden längst ner. Vi ser att teststatistikan är t -fördelad med 5 frihetsgrader. Värdet på teststatistikan är -2.09 och p -värdet för ett dubbelsidigt test är 0.0905 . P -värdet för det enkelsidiga testet är således $\frac{0.0905}{2} = 0.04525$. Ska vi förkasta H_0 ? Vad innebär p -värdet?

Utföra test för skillnader mellan två populationsparametrar då det är parvisa observationer

Utgå från övning 17.9 i kompendiet. Vi läser in datat

```

data work.ovn179;
input pulsfore pulsefter;
datalines;
81 105
81 91
68 87
61 86
67 82
74 78
75 87
64 94
70 93
60 90
;
run;

```

Vi vill nu testa

$$\begin{aligned}
 H_0 &: \mu_{före} - \mu_{efter} = \mu_D = 0 \\
 H_A &: \mu_{före} - \mu_{efter} = \mu_D < 0,
 \end{aligned}$$

det vill säga om pulsen är lägre innan man snusar.

För att testa detta skriver vi koden

```
proc ttest data=work.ovn179 H0=0;  
paired pulsfore*pulsefter;  
run;
```

Vi anger kommandot "*paired*" för att tala om för SAS att vi har beroende mellan observationerna, det vill säga att vi vill använda parvisa observationer. (Kommandot " $H_0=0$ " kan bortses från, då det är default när vi använder kommandot "*paired*".)

Vi får utskriften

The TTEST Procedure					
Difference: pulsfore - pulsefter					
N	Mean	Std Dev	Std Err	Minimum	Maximum
10	-19.2000	8.7534	2.7681	-30.0000	-4.0000
Mean	95% CL Mean		Std Dev	95% CL Std Dev	
-19.2000	-25.4618	-12.9382	8.7534	6.0209	15.9803
DF		t Value	Pr > t		
9		-6.94	<.0001		

Värdet på teststatistikan är -6.94 och p-värdet för ett dubbelsidigt test är mindre än 0.0001. Bör vi förkasta H_0 ?

Utföra ett test för skillnader mellan två populationsparametrar från två oberoende stickprov

Har vi inte parvisa observationer, utan *oberoende* observationer i två stickprov och vill testa om det finns skillnader i medelvärden mellan populationerna, använder vi "*proc ttest*" även här.

Datafilen "*bromssträcka*" innehåller observationer av bromssträcka för två biltyper, A och B. Antag att variabeln "*bromssträcka*" är normalfördelad. Ladda ner filen från hemsidan och läs in datat i SAS. Vi vill testa

$$\begin{aligned}H_0 &: \mu_A - \mu_B = 0 \\H_A &: \mu_A - \mu_B \neq 0\end{aligned}$$

det vill säga om bromssträckan för biltyp A är lika som för biltyp B.

Vi skriver koden

```
proc ttest data=work.broms;  
class bilty;   
var bromsstracka;  
run;
```

Kommandot "*class bilty*" anges för att tala om att variabeln "*bilty*" delar upp hela datamaterialet i två stickprov. Att generera och tolka utskriften finns som en övning nedan.

Uppgifter

Basuppgifter

1. Ladda ner filen "rökdata" från hemsidan. Andelen rökare är $p = 0.6$. Använd koden i exempel "Utföra hypotestest för populationsproportionen" för att testa

$$\begin{aligned} H_0 &: \pi = 0.5 \\ H_A &: \pi \neq 0.5 \end{aligned}$$

Fundera över om kommandot "weight..." i exemplet behöver vara med. Tolka utskriften. Ska vi förkasta H_0 ?

2. Utgå från övning 17.6 i kompendiet. Läs in datat och testa sedan

$$\begin{aligned} H_0 &: \mu = 800 \\ H_A &: \mu \neq 800 \end{aligned}$$

Tolka utskriften. Ska vi förkasta H_0 ?

3. Utgå från övning 17.10 i kompendiet. Läs in datat och testa sedan

$$\begin{aligned} H_0 &: \mu_D = 0 \\ H_A &: \mu_D \neq 0 \end{aligned}$$

Tolka utskriften. Ska vi förkasta H_0 ?

4. Tolka utskriften i exempel "Utföra ett test för skillnader mellan två populationsparametrar från två oberoende stickprov". Gör nödvändiga antaganden. Att fundera på: Kan varianserna i de båda populationerna antas vara lika?