Beispiel Vergleich Mittelwerte

April 26, 2016

1 Übung: Differenz von Mittelwerten

Berechnen Sie zwei Mittelwerte $\bar{x}_1$ und $\bar{x}_2$ über zwei Zeiträume und deren Differenz $\Delta x = \bar{x}_2 - \bar{x}_1$. Wie wahrscheinlich ist es, dass die Differenz nicht rein zufällig ist? Beschreiben Sie die Signifikanz der Änderung. Nutzen Sie eine Monte-Carlo-Methode zur Berechnung der Signifikanz (z.B. durch zufälliges Umsortieren der Zeitreihe).

In [8]:
```
%pylab inline
# X Zeitreihe September-Eisausdehnung von 1972-1992 aus Satellitendaten. Einheit: Millionen Qua
X=array([7.3,7.7,7.5,7.3,7.3,7.2,7.9,7.2,7.9,7.3,7.5,7.5,7.2,6.9,7.5,7.5,7.5,7.0,6.2,6.6,7.6])
print('Durchschnitt der ersten 10 Jahre:',(mean(X[0:10])))
print('Durchschnitt der letzten 10 Jahre:',(mean(X[11:21])))
Differenz=mean(X[0:10])-mean(X[11:21])
print('Differenz:',Differenz)
```

```
Populating the interactive namespace from numpy and matplotlib
Durchschnitt der ersten 10 Jahre: 7.46
Durchschnitt der letzten 10 Jahre: 7.15
Differenz: 0.31
```

In [9]:
```
N=10000
D=[]
for i in range(0,N):
    shuffle(X)
    D.append(mean(X[0:10])- mean(X[11:21]))  #Differenz
D=array(D)
A=sum(D>=Differenz)
P=A*100./N         #Irrtumswahrscheinlichkeit in %
S=100.-P          #Signifikanz in %
print('Anzahl:', A, 'von ',N)
print('Irrtumswahrscheinlichkeit',P, '%')
print('Signifikanz',S, '%')
```

```
Anzahl: 395 von  10000
Irrtumswahrscheinlichkeit 3.95 %
Signifikanz 96.05 %
```

2 Prüfverfahren: Vergleich zweier Mittelwerte

Siehe Schönwiese, Abschnitt 8.2.1 (Seite 124, 4. Auflage)

Gleicher Stichprobenumfang $n = n_a = n_b$:

$$\hat{t} = \frac{|\bar{a} - \bar{b}|\sqrt{n}}{\sqrt{s_a^2 + s_b^2}}$$

Mit Freiheitsgrad $\Phi = 2n - 2$

In [10]:
```
import scipy.stats as st
t_test=st.t.cdf # Student t
z_test=st.norm.cdf # Normalverteilung

n=10
X=array([7.3,7.7,7.5,7.3,7.3,7.2,7.9,7.2,7.9,7.3,7.5,7.5,7.2,6.9,7.5,7.5,7.5,7.0,6.2,6.6,7.6])
a=mean(X[0:10])
b=mean(X[11:21])
sa=std(X[0:10],ddof=1)
sb=std(X[11:21],ddof=1)

phi=2*n-2
t=(a-b)*sqrt(n)/(sqrt(sa**2+sb**2))
print(t,phi)
```

```
1.80014986263 18
```

In [13]:
```
a,sa
```

Out[13]:
```
(7.4600000000000009, 0.27568097504180455)
```

In [14]:
```
b,sb
```

Out[14]:
```
(7.1500000000000004, 0.46963342678684566)
```

In [24]:
```
t_test(t,phi)
```

Out[24]:
```
0.95569007258302929
```

In [25]:
```
z_test(t)
```

Out[25]:
```
0.96408151096942074
```

Signifikanz 95.6% (t-Test) bzw. 96.4% (Normalverteilung) für Ablehnung der Nullhypothese.
Die Untersuchung zeigt: die Mittelwerte sind verschieden mit einem Signifikanzniveau von >95%.

2.1 Octave

In [8]:
```
n=10;
X=[7.3,7.7,7.5,7.3,7.3,7.2,7.9,7.2,7.9,7.3,7.5,7.5,7.2,6.9,7.5,7.5,7.5,7.0,6.2,6.6,7.6];
a=mean(X(1:10));
b=mean(X(12:21));
sa=std(X(1:10));
sb=std(X(12:21));

phi=2*n-2;
t=(a-b)*sqrt(n)/(sqrt(sa**2+sb**2));
disp(num2str(t));
disp(num2str(phi));
```

Out[8]:
```
1.8001
18
```

In [10]:
```
tcdf(t,phi)
```

```
Out[10]: ans =  0.95569
```

```
In [13]: tcdf(t,10000)
```

```
Out[13]: ans =  0.96407
```

Die kumulative Verteilungsfunktion der t-Verteilung wird genutzt, um die Quantile zu berechnen. Im Grenzfall für große Feiheitsgrade/Anzahl von Stichproben entspricht das Ergebnis dem der Normalverteilung.