Labor 11 – 2017

Statistische Teste
Spezifische Octave/Matlab Befehle: ztest, ttest,vartest

1. **Beispiel:** Betrachten wir die Stichprobe :

3150 , 3249 , 3059 , 3361 , 3248 , 3254 , 3259 , 3353 , 3145 , 3051
vom Umfang n=10 für die Lebensdauer in Stunden von 100 Watt-Glühbirnen, die von einer Firma produziert wurden. Hier wurde die Lebensdauer von 10 zufällig aus der Produktion eines bestimmten Zeitraums ausgewählten Glühbirnen bestimmt. Diese Stichprobe dient zur Untersuchung des Merkmals (Zufallsgröße) X der Lebensdauer in der Grundgesamtheit der Glühbirnenproduktion der Firma.

a) Welches ist die durchschnittliche Lebensdauer der Glühbirnen? Welches ist die Stichprobenvarianz?
b) Stimmt die Aussage des Herstellers, dass diese Glührbirnen im Mittel mehr als 3100 Stunden dauern? Man nimmt an α=0.01 und X hat normale Verteilung.

```
X=[3150 , 3249 , 3059 , 3361 , 3248 , 3254 , 3259 , 3353 , 3145 , 3060];
m=mean(X)
v=var(X)

h=ttest(X,3100,'alpha',0.01,'Tail','left')
```

h=0 → H0 wird akzeptiert → die Aussage des Herstellers stimmt

```
hh=ttest(X,3100,'alpha',0.01,'Tail','right')
```

hh=1 → H1 wird akzeptiert

ttest(X,M, 'alpha',..., 'Tail', '...') performs the T-test against the **alternative hypothesis** specified by TAIL:

'both' -- "mean is not zero (or M)" (two-tailed test) H1: mean $\neq$ M; H0: mean = M
'right' -- "mean is greater than zero (or M)" (right-tailed test) H1: mean > M ; H0: mean $\leq$ M
'left' -- "mean is less than zero (or M)" (left-tailed test) H1: mean < M ; H0: mean $\geq$ M

ODER

E(X)= durchschnittliche Lebensdauer ; theoretischer Erwartungswert für die Lebensdauer (der Glühbirnen)
H0: E(X) $\geq$ 3100 , H1: E(X) < 3100

Kurzes Programm:

```
X=[3150 , 3249 , 3059 , 3361 , 3248 , 3254 , 3259 , 3353 , 3145 , 3060]
n=length(X);
t=(mean(X)-3100)/(var(X)/sqrt(n));
q=tinv(0.01,n-1)
if t>q
  disp('man akzeptiert H0')
else
  disp('man lehnt H0 ab')
end
```

2.Beispiel: Ein Medienversender bietet Bücher und DVDs an. Bei einer Überprüfung der Marketingziele dieser Produktpalette stellt er an einer Stichprobe von 100 Kunden fest, dass das durchschnittliche Alter der Besteller 49.92 Jahre beträgt. Ursprünglich war geplant, dass diese Palette eine Kundengruppe mit einem durchschnittlichen Alter von mindestens 54 Jahren erreichen sollte. Um die in der Stichprobe gefundene Abweichung von seinem Ziel zu bewerten, befragt er einen Statistiker. Der Statistiker soll nun prüfen, ob der ermittelte Wert noch zu der Annahme passt, dass das *durchschnittliche* Kundenalter *mindestens* 54 Jahre beträgt. Deshalb führt der Statistiker einen Gauß-Test (Z-Test) durch. Der Statistiker hat diesen Test gewählt, da es bekannt ist, dass die Standardabweichung des Alters der Besteller bei 12 Jahren liegt.

Daten:

Alter=[19 26 27 28 29 30 34 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 69]

Anzahl=[1 2 1 2 1 2 2 2 1 2 1 4 2 2 2 4 1 1 2 2 5 2 5 5 5 5 3 2 4 1 5 5 2 2 3 2 1 1 2 3]

Kunden = Vektor welche alle Daten enthält (aus den Daten folgt *mean*(Kunden)=49.92)

```
>> ztest(Kunden,54,12,'alpha',0.01,'Tail','both')
>> ztest(Kunden,54,12,'alpha',0.01,'Tail','left')
>> ztest(Kunden,54,12,'alpha',0.01,'Tail','right')
```

ztest(X,M,SIGMA, 'alpha',..., 'Tail', '...') performs the Z-test against the **alternative hypothesis** specified by TAIL:
'both' -- "mean is not zero (or M)" (two-tailed test) H1: mean $\neq$ M; H0: mean = M
'right' -- "mean is greater than zero (or M)" (right-tailed test) H1: mean > M ; H0: mean $\leq$ M
'left' -- "mean is less than zero (or M)" (left-tailed test) H1: mean < M ; H0: mean $\geq$ M

Aufgaben (Theorie in Vorlesung 10):

1. Ein Hersteller möchte die beste Kontrolle über den Durchmesser der produzierten Rohre haben. Der durchschnittliche Durchmesser der Rohre ist wichtig, aber wichtiger ist die Varianz der Durchmesser: Wenn die Varianz zu groß ist, können die Rohre nicht richtig zusammengefühgt werden. Daher möchte der Hersteller die Durchmesservariation so gering wie möglich halten. Eine Abweichung von 1 mm wird als akzeptabel angesehen, aber wenn die Abweichung größer als dieser Wert ist, wird die Produktion unterbrochen, um die erforderlichen Anpassungen vorzunehmen. Die Entscheidung wird unter Verwendung der statistischen Hypothesen mit einer Risikowahrscheinlichkeit von $\alpha = 0{,}01$ getroffen. Erstellen Sie einen Test für Varianz mit Hilfe der statistischen Daten $x_1, \ldots, x_n$ (man weiß, dass die Daten einer Normalverteilung entstammen).

2. Laut Werbung eines Unternehmens für einen bestimmten Batterietyp sollten diese Batterien nach dem Laden *mindestens* 350 Stunden halten. Anhand einer Stichprobe wurden folgende Daten erhalten

Stunden (wie lange die Batterie funktioniert)	343	345	347	348	349	350	352	353	355
Absolute Häufigkeit der Batterien	25	15	132	84	40	34	51	8	11

Man teste (mit α=0.01) :
a) ob die Werbung des Herstellers begründet ist;
b) ob die Werbung des Herstellers begründet ist, wenn man zusätzlich weiss, dass die Funktionszeit solcher Batterientypen eine Standardabweichung von 2 Stunden hat.

3. Ein Computerchiphersteller behauptet, dass nicht mehr als 2% der hergestellten Chips, defekt sind. Eine Elektronikfirma ist von dieser Behauptung beeindruckt und hat eine große Menge solcher Chips gekauft. Um zu überprüfen, ob die Aussage des Herstellers wahr ist, entschied die Firma, 300 Chips zu testen. Darunter wurden 10 fehlerhafte Chips gefunden. Kann die Behauptung des Herstellers akzeptiert werden oder sind die Daten nicht ausreichend um die Ausage abzulehnen? (α = 0.01)

4. Eine Abfüllanlage füllt 0.33 Liter Dosen mit Saft. Wenn die Varianz der abgefüllten Menge zu hoch ist, wird es viele Dosen geben, die zu voll sind und viele Dosen, die nicht genügend gefüllt sind. Daher möchte das Unternehmen die Variation der Menge in jeder Dose kontrollieren und die Varianz so niedrig wie möglich halten. Diese Firma möchte feststellen, wann die Varianz der Saftmenge in jeder Dosis außer Kontrolle gerät. Eine Abweichung von 0,04 Liter wird als akzeptabel angesehen, aber das Unternehmen wird die Abfüllmaschine anpassen, wenn die Varianz größer als dieser Wert wird. Erstellen Sie für α = 0.05 einen mit Hilfe der statistischen Daten $x_1, \ldots, x_n$ (man weiß, dass die Daten einer Normalverteilung entstammen).

5. Ein Bankinstitut ist an der durchschnittlichen Zeit interessiert, wie lange die Kunden warten bis sie an einem freien Schalter bedient werden. Anhand einer Umfrage an der 100 Personen teilgenommen haben, wurde festgestellt, dass sie im Durchschnitt 10 Minuten mit einer Standardabweichung von 2 Minuten warten. Mit α = 0,05 sollte getestet werden

a) ob die die mittlere Wartezeit 9 Minuten beträgt oder nicht;
b) ob die Standardabweichung der Wartezeit kleiner als 1 Minute ist oder nicht.

Laborator 11 – 2017

Teste statistice
Comenzi specifice Octave/Matlab: ztest, ttest,vartest

Exemplu: Se dau datele :
3150 , 3249 , 3059 , 3361 , 3248 , 3254 , 3259 , 3353 , 3145 , 3051
pentru un eşantion de n=10 becuri; datele reprezintă timpul de funcţionare în ore a unor becuri de 100 W, produse de o anumită firmă. Acest eşantion se foloseşte pentru cercetarea caracteristicii X, care reprezintă timpul de funcţionare a becului până la defectare.

a) Estimaţi timpul mediu de funcţionare , respectiv varianţa empirică pe baza datelor existente.
b) Producătorul afirmă că astfel de becuri funcţionează cel puţin 3100 ore. Este acceptată afirmaţia producătorului? Se presupune α=0.01, iar X are distribuţie normală.

```
X=[3150 , 3249 , 3059 , 3361 , 3248 , 3254 , 3259 , 3353 , 3145 , 3060];
m=mean(X)
v=var(X)

h=ttest(X,3100,'alpha',0.01,'Tail','left')
```

h=0 → H0 se acceptă → afirmaţia producătorului este acceptată

```
hh=ttest(X,3100,'alpha',0.01,'Tail','right')
```

hh=1 → H1 se acceptă

ttest(X,M, 'alpha',..., 'Tail', '...') performs the T-test against the **alternative hypothesis** specified by TAIL:

'both' -- "mean is not zero (or M)" (two-tailed test) H1: mean $\neq$ M; H0: mean = M
'right' -- "mean is greater than zero (or M)" (right-tailed test) H1: mean > M ; H0: mean $\leq$ M
'left' -- "mean is less than zero (or M)" (left-tailed test) H1: mean < M ; H0: mean $\geq$ M

SAU
H0: E(X) $\geq$ 3100 , H1: E(X) < 3100

Programul:

```
X=[3150 , 3249 , 3059 , 3361 , 3248 , 3254 , 3259 , 3353 , 3145 , 3060]
n=length(X);
t=(mean(X)-3100)/(var(X)/sqrt(n));
q=tinv(0.01,n-1)
if t>q
  disp('se accepta H0')
else
  disp('se respinge H0 ')
end
```

Aplicaţii:

1. Un producător de ţevi metalice doreşte să aibă un control cât mai bun asupra diametrului ţevilor produse. Diametrul mediu al ţevilor este important, dar mai importantă este varianţa diametrului: dacă varianţa este prea mare, atunci ţevile nu se vor mai îmbina corespunzător. Astfel, producătorul doreşte menţinerea varianţei diametrului la un nivel cât mai scăzut posibil, dar şi întreruperi cât mai rare ale producţiei (în cazul necesităţii unor recalibrări). O varianţă de 1 mm este considerată acceptabilă, dar dacă varianța este mai mare decât această valoare, producția se va întrerupe pentru a efectua reglajul necesar. Decizia va fi luată, folosind verificarea ipotezelor statistice cu probabilitatea de risc α=0.01. Să se construiască un test pentru testarea varianţei pe baza datelor statistice x_1, ... , x_n (ştiind că datele provin dintr-o distribuţie normală).

2. O companie face publicitate pentru un anumit tip de baterie despre care susţine că rezistă cel puţin 350 de ore, după ce a fost încărcată. Pe baza unor verificări s-au obţinut următoarele date

Orele de funcţionare	343	345	347	348	349	350	352	353	355
Frecvenţele absolute	25	15	132	84	40	34	51	8	11

Să se testeze (cu α=0.01) :

a) Dacă publicitatea companiei este justificată.

b) Dacă publicitatea companiei este justificată, având în vedere faptul că abaterea standard pentru timpul de funcţionare al bateriilor este de 2 ore.

3. Un producător de cipuri de calculatoare afirmă că nu mai mult de 2% din cipurile, pe care le produce, ar fi defecte. O companie de electronice, este impresionată de această afirmație şi a cumpărat o cantitate mare de astfel de cipuri. Pentru a verifica dacă afirmaţia producătorului este adevărată, compania a decis să testeze un eșantion de 300 de cipuri şi constată că 10 dintre aceste cipuri sunt defecte. Poate fi afirmaţia producătorului acceptată sau sunt suficiente dovezi pentru a fi respinsă? (α=0.01)

4. O companie de produse răcoritoare are o maşină de îmbuteliat, care umple cu răcoritoare doze de 0.33 l. Cantitatea medie din fiecare doză este importantă, dar cantitatea medie corectă nu asigură că maşina funcţionează corect. Dacă varianţa este mare, vor fi multe doze care sunt prea umplute şi multe doze care sunt mai puţin umplute . De aceea, compania doreşte să controleze *variaţia* cantităţii din fiecare doză şi să menţină *varianţa* (*dispersia*) la un nivel cât mai scăzut posibil. Această companie doreşte să detecteze când variabilitatea cantităţii de răcoritoare pusă în fiecare doză scapă de sub control. O varianţă de 0.04 l este considerată acceptabilă, dar compania va regla maşina de îmbuteliat, dacă varianţa devine mai mare decât această valoare. Decizia va fi luată, folosind verificarea ipotezelor statistice.

Pentru α=0.05, să se construiască un test pentru testarea varianţei pe baza datelor statistice x_1, ... , x_n (ştiind că datele provin dintr-o distribuţie normală).

5. O instituţie bancară este interesată de timpul mediu pe care clienţii îl pierd, aşteptând să fie serviţi la unul din ghişee. La sondaj au participat 100 de persoane şi s-a constatat că în medie au aşteptat 10 minute cu o deviaţie standard de 2 minute.

Pentru α=0.05,

a) să se testeze dacă este verificată ipoteza că timpul mediu de aşteptare este de 9 minute sau nu este;

b) să se testeze dacă este verificată ipoteza că abaterea standard a timpului de aşteptare este mai mică decât 1 minut sau nu.