

# HİPOTEZ TESTLERİ

Bilimsel Araştırma Teknikleri-Hafta 8

# İstatistik nedir?

- Herhangi bir konuyu bilimsel açıdan incelemek amacıyla çalışmanın planlanması, verilerin toplanması, değerlendirilmesi (düzenlenmesi, özetlenmesi ve analiz edilmesi) bir karara varılması ve çıkarımlar yapılmasına imkan tanıyan bilim dalıdır

# Hipotez testi

- Hipotez basitçe bir iddia, bir savdır. Çözüm olarak ortaya atılan bu iddia, doğrulanmaya ya da doğrulamamanın daha kolay bir yolu olarak yanlışlanmaya çalışılır.
- Bir hipotez kurup bu hipotezi doğrulamaya da yanlışlama ve sonunda bir karara varma işlemlerinin bütününe hipotez testi denir.

# Hipotez testinin aşamaları

1. Varsayımların kontrolü
2. Hipotezlerin oluşturulması
3. Anlam düzeyinin ( $\alpha$ ) belirlenmesi
4. Seçilen hipotez testinin uygulanması
5. sonuç ve yorum

# 1. Varsayımların Kontrolü

- Hipotez testlerinin bazı temel varsayımları bulunmaktadır.
- Bu varsayımların sağlanması ya da sağlanmaması durumlarına göre test süreci ve bu süreçte kullanılacak test istatistikleri farklılaşmaktadır

# Kontrolu gereken varsayımlar

1. Dağılımın Belirlenmesi- Hipotez testini seçerken önemli
2. Ölçek türü
3. Örneklem büyüklüğü
4. Değişken türü

# 1.1 Dağılımın Belirlenmesi

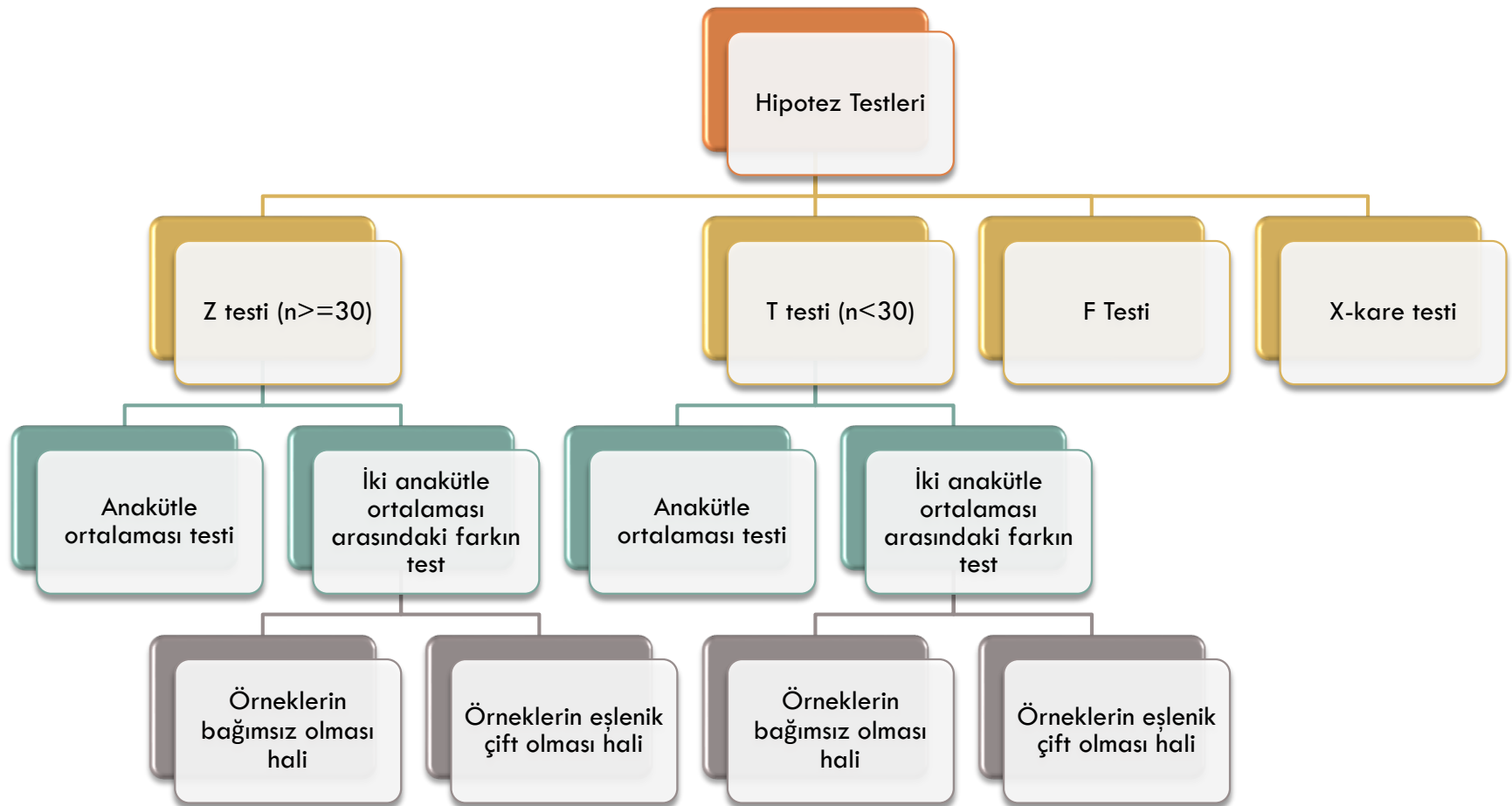
- İstatistiksel testlerde veri dağılımı, kullanılacak test istatistiğinin belirlenmesinde önemli bir ölçüttür.
- Verilerin normal dağılım göstermesi durumunda 'parametrik yöntemler', aksi durumda 'parametrik olmayan yöntemler' kullanılır.
- Parametrik yöntemler, genellikle **ortalama ve varyansa dayalı** algoritmalar içermektedir.
- Parametrik olmayan yöntemler ise **genellikle sıra ve sıra farklarına dayalı** algoritmalarla hesaplamalar içermektedir

# Dağılıma göre kullanılacak testler





# Parametrik Hipotez Testleri



# Non-parametrik testler

## *Non-Parametrik Testlerin Özellikleri*

- 1) Ölçü yerine sıralama, sayma, isaretleme gibi işlemler yapılıyorsa bu test parametrik olmayan (non-parametrik) bir testtir
- 2) Niteliksel (kategorik) veriler için parametrik olmayan testler kullanılır.
- 3) Bunun yanında ölçümle belirtildiği halde veri parametrik test varsayımlarını yerine getiremiyorsa, (varyansların homojenliği ve normal dağılım)
- 4) Denek sayısı az ise ya da
- 5) Değerler yerine sıraları verilmiş ise yine parametrik olmayan testler uygulanır

### **Parametrik Olmayan Hipotez Testlerinin ;**

#### **AVANTAJLARI**

- Uygulanması için birçok varsayıma gerek yoktur.
- Anlaşılması ve uygulanması kolaydır.
- Küçük hacimli bir örnek üzerinden yapılması mümkündür.

#### **DEZAVANTAJLARI**

- Örnek hacminin büyük olması halinde uygulanması güçleşir.
- Bu tür testlerin uygulanmasıyla elde edilen sonuçlar, parametrik testlerin uygulanmasıyla elde edilen sonuçlardan daha az güvenilirirdir.

# 1.2 Ölçek türü



- 4'e ayrılır.
- Sınıflama ölçeği ve sıralama ölçeği düzeyindeki veriler, sınıf ve sıra bilgisi dışında bilgi sağlamamaktadır.
  - Bu ölçek düzeylerinde frekans, yüzde, mod ve bazı durumlarda medyan, bilgi sağlayan istatistiklerdir. Bu ölçek düzeylerinde dört işlem anlamlı değildir. Dolayısıyla bu ölçek düzeylerinde çoğunlukla sıra farklarına dayalı hesaplamalar içeren 'parametrik olmayan yöntemler' kullanılmalıdır.
  - Eşit aralıklı ölçek ve eşit oranlı ölçek düzeyindeki verilerde ise toplama ve çıkarma anlamlıdır. Dolayısıyla bu tür verilerde ortalama, varyans, standart sapma gibi betimsel istatistiklerle bilgi sağlamak mümkündür. Dolayısıyla çoğunlukla ortalama ve varyansa dayalı hesaplamalar içeren 'parametrik yöntemler' kullanılabilir.

# 1.3 Örneklem Büyüklüğü

- Örneklem büyüklüğü, ilgilenilen özellik hakkında anlamlı betimleme ve çıkarımlar yapılabilmesinde önemlidir. Örneklem kuramına göre örneklem büyüklüğü için üç kesme noktasına bağlı olarak üç tanımlama yapılabilmektedir:
- 1.Zengin Örneklem: Gözlem birimi sayısı 30'un (bazı kaynaklarda 20'nin) üzerinde olan örneklemeleri ifade eder.
- 2.Küçük Örneklem: Gözlem birimi sayısı 20 ve altında olan örneklemeleri ifade eder.
- 3.Büyük Örneklem: Gözlem birimi sayısı 1 20'nin üzerinde olan örneklemeleri ifade eder.
- Örneklem büyüklüğüne bağlı olarak elde edilen veri sayısının 20'nin altında olması durumunda parametrik olmayan yöntemlerin kullanılması önerilir.

# 1.4 Değişken türü

- Değişkenin
  - ▣ nitel ya da nicel olmasına,
  - ▣ sürekli ya da kesikli olmasına,
  - ▣ bağımlı-bağımsız değişken ilişkisinin kurulmuş ya da kurulmamış olmasına göre kullanılacak istatistiksel yöntemler de değişmektedir.
- Bu noktada bir genelleme yapmak çok doğru görülmemektedir. Fakat değişken türlerine göre farklı hesaplama biçimleri, algoritma ve formüller içeren bir çok yöntem bulunmaktadır

## 2.H<sub>0</sub> hipotezinin kurulması

- Örneklemden elde edilen ortalama ile anakütleye ait ortalamamanın farkı "sıfır", "0" sayılabilir.
- Bu görüş savunulurken istatistiksel anlamlılık denilen (%99 %97 veya %95) yanılğı payı göz önüne alınır.
- Zaten yapılan işlemlerden sonra farkın çok küçük de olsa sıfırdan farklı olduğu görülür

# 3.Önem düzeyi

## Önem Seviyesi - $\alpha$

- Örneklem dağılımının RED bölgesinin büyüklüğünü gösterir.
- Tipik değerleri: 0.01, 0.05, 0.10
- Araştırmanın başında araştırmacı tarafından seçilir.



## 4.Uygun Hipotez testinin belirlenmesi

- Yukarıda açıklanan kriterler göz önüne alınarak uygun test seçilir

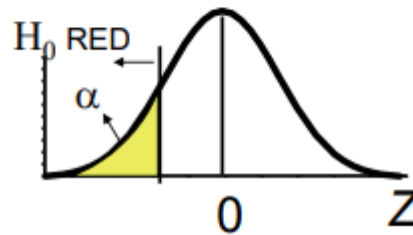
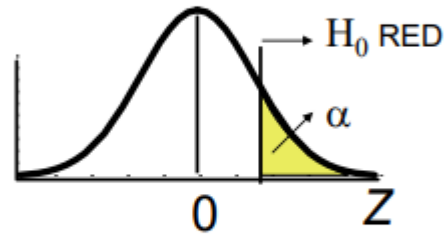
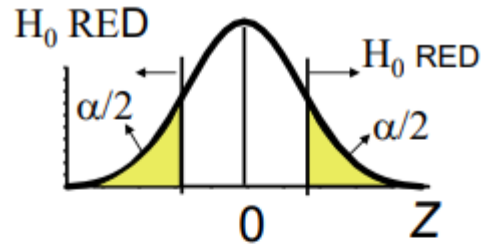
## Parametrik Testler

- Tek örneklemlili t-testi
- İki bağımsız örneklemlili t-testi
- Eşlenik t-testi
- Tek yönlü varyans analizi (ANOVA)
- Korelasyon
- Basit doğrusal Regresyon
- Çoklu regresyon

## Parametrik Olmayan Testler

- Binom testi
- Ki-kare uyum iyiliği testi
- Ki-kare testi
- Wilcoxon-Mann-Whitney testi
- Kruskal Wallis testi
- Wilcoxon işaretli sıra toplamı testi
- Parametrik olmayan korelasyon testi

# 5.Sonuç ve yorumlar



# Z-testi -Parametrik teste örnek (Örneklerin eşlenik çift olması hali)

- A anakütlesindeki her bir bireyin B anakütlesinden bir «ikizi» ve ya «eşi» ile eşlendikten sonra elde edilen çiftlerden seçilen örneğe «eşlenik çift» denir.
- 1. Bir uygulamanın etkinliğini test etmek amacı ile bu eşlerden biri uygulamaya tabi tutulurken diğer eş tabi tutulmaz. Sonra elde edilen gözlem sonuçları arasındaki farklar için test yapılır. Buna «eşlenik-çift örnek testi» denir.
- 2. Bir ana kütleden rastgele seçilen bireylere belli bir uygulama tatbik edildikten önceki ve sonraki gözlem sonuçları da bir eşlenik çift örneği oluşturur.

# Örnek durumlar

- A marka ve B marka ahşap boyaların ahşabı koruma süreleri arasındaki farklılığın testi için;
  - ▣ Bir miktar tahta alınır ve bu tahtalardan herbirinin yarısı A diğer yarısı B ile boyanır. Dayanma süreleri ölçülür. Bu süre **çiftleri arasındaki fark** test edilir.
- İkiz kuzulardan birisi A yemi, diğeri B yemi ile beslendikten sonra,
  - ▣ Bu kuzuların ağırlıkları arasındaki farklar test edilerek A ve B yeminin kilo verimindeki etkisi test edilebilir.

# Örnek durumlar

- Bir grup işçinin belli bir kursa katılmadan önceki ve sonraki verimlilik değerleri arasındaki farkların testiyle, sözkonusu kursun verimliliği artırıp artıtmadığı hakkında karar verilebilir.
- Bir popülasyon üretilir. Bu popülasyona A algoritması ve B algoritması uygulanır. Hangi algoritmanın daha iyi olduğuna karar verilir.

# Önemi

- Eşleme metodunda çiftler aynı olduğundan, bu çiftlerden gözlemlenen sonuçlar arasındaki fark doğrudan doğruya uygulamadan kaynaklanır.
- Bu sebeple eşleme metodunun uygulamada önemli bir yeri vardır.

# 1. Aşama : Hipotezlerin kurulması

- $H_o : \mu_D = 0$

- $H_1 : \mu_D \neq 0$  veya,

- $H_1 : \mu_D > 0$  veya,

- $H_1 : \mu_D < 0$  şeklinde kurulabilir.

- Not: D, Eşlenik çift değerleri arasındaki farkı gösterir.



## 2. Aşama: Test istatistiğinin hesaplanması

- $(X_i, Y_i)$  eşlenik çiftleri göstereceğiz. ( $i=1,2,3,\dots,n \geq 30$ )
  - $D_i = X_i - Y_i$
  - $\bar{D} = \frac{\sum D_i}{n}$
  - $s_D = \sqrt{\frac{\sum (D_i - \bar{D})^2}{n}} = \sqrt{\frac{\sum \bar{D}^2}{n} - \left(\frac{\sum D}{n}\right)^2}$
  - $\bar{Z} = \frac{\bar{D}}{s_d/\sqrt{n}}$

# Örnek-1

**ÖRNEK PROBLEM 3.10** Bir ziraî araştırmacı, çilek yetiştirmede azot ağırlıklı gübrenin, fosfat ağırlıklı gübreye nazaran daha çok mahsul verdireceğini iddia etmektedir. Bu iddiayı test etmek amacıyla 30 parsel çilek dikilmiş ve bu parsellerin her birisinin yarısı azot yarısı da fosfat ağırlıklı gübreyle gübrenmiştir Böylece her bir parsel bir eşlenik çift oluşturmuştur. Daha sonra her bir parselin azotlu kısmından alınan mahsul miktarı ile fosfatlı kısmından alınan mahsul miktarlarının farkları alınarak, ortalama fark  $D=40$  kg. ve standart sapma  $s_D = 110$  kg, bulunmuştur. ( $D=$  Azotlu kısımdan alınan mahsul miktarı - Fosfatlı kısımdan alınan mahsul miktarı)

Bu durumda araştırmacının iddiasını % 1 önem seviyesinde destekleyebilir miyiz?

# Çözüm

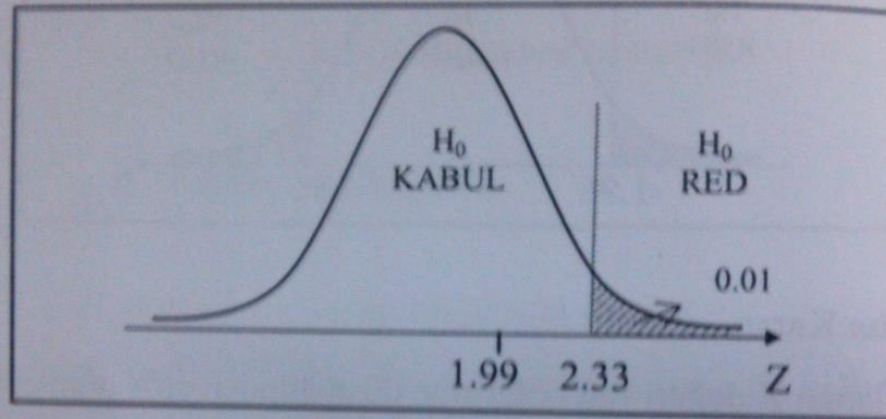
□  $H_0 : \mu_D = 0$

■  $H_1 : \mu_D > 0$  veya,

□  $\bar{Z} = \frac{\bar{D}}{s_d/\sqrt{n}} = \frac{40}{110/\sqrt{30}} = 1.99$

3. Safha: Karar Modeli.

$\alpha = 0.01$  için  $Z_\alpha = 2.33$  tür.



4. Safha: Karar.

Test istatistiğinin değeri kabul bölgesine düştüğünden örnek sonucunu istatistikî açıdan %1 önem seviyesi için, önemli olmadığına, yani, her iki gübreleme sonucunda da aynı miktarda mahsul alındığına %1 önem seviyesinde karar verilir.



Örnek2: 30 işçinin belli bir eğitime tabi tutulmadan önceki ve sonraki verimlilik değerleri aşağıdaki gibidir. Bu eğitimin verimliliği artırıp artırmadığını %5 önem seviyesinde test ediniz?

İşçi No:	Verimlilik Değeri			İşçi No:	Verimlilik Değeri		
	Önce	Sonra	D		Önce	Sonra	D
1	110	121	11	16	107	124	17
2	107	120	13	17	107	118	11
3	112	142	30	18	116	130	14
4	124	138	14	19	95	126	31
5	96	105	9	20	132	135	3
6	103	117	14	21	97	108	11
7	109	110	1	22	104	119	15
8	132	145	13	23	126	131	5
9	112	126	14	24	118	130	12
10	101	118	17	25	104	116	12
11	94	119	25	26	117	120	3
12	99	132	33	27	125	125	0
13	127	143	16	28	109	120	11
14	116	120	4	29	100	116	16
15	108	110	2	30	112	120	8

$$\bar{D} = 12.83$$

$$s_D = 8.27$$

$$Z_{\alpha} = 1.645$$

Z=8.5 bulunur

Ho red edilir

8.5>1.645 dir


# Wilcoxon işaretli-sıra testi-

## Nonparametrik teste örnek

- Wilcoxon işaretli-sıra testi (Wilcoxon signed-rank test), iki örneklem ortalamaları arasında anlamlı olan farklıları tespit etmeyi amaçlar. Parametrik bir istatistik test olan eşlenik z/t-testinin, parametrik olmayan alternatif testidir.

- Eğer iki algoritmanın çıktıların kıyaslanması için kullanılacaksa, test pratik olarak iki algoritmanın karşılıklı davranışlarını değerlendirir. Örneğin  $N_{ds}$  veri kümesinin  $i$ . çıkışı için iki algoritmanın ürettiği çıkış sonuçları arasındaki fark  $d_i$  olsun. Bu farklar mutlak değerlerine göre sıralanır ve her bir farka bir işaret değeri atanır.

- . Mesela, ilk algoritmanın ikinci algoritmayı geçtiği durumlardaki sıra değerlerinin toplamı  $R^+$  iken, ikinci algoritmanın ilk algoritmayı geçtiği durumlardaki sıra değerlerinin toplamı  $R^-$  olduğu kabul edilsin;  $d_i = 0$  olduğu durumlar varsa ve sayısı çift ise, ikiye bölünüp toplamlara eklenir, tekse bir durum ihmal edilirdir ve sayısının çift olması sağlanır.


$$R^+ = \sum_{d_i > 0} \text{sira}(d_i) + \frac{1}{2} \sum_{d_i = 0} \text{sira}(d_i)$$

$$R^- = \sum_{d_i < 0} \text{sira}(d_i) + \frac{1}{2} \sum_{d_i = 0} \text{sira}(d_i)$$

$$T = \min(R^+, R^-)$$



# Örnek

- $H_0$  :SSMA (Garcia ve ark., 2008) ile GYAopt-aiNet algoritmalarının 20 adet veri kümesi üzerinde çalıştırıldıklarında elde edilen sınıflandırma doğruluk oranları arasında **%95 anlam seviyesinde** istatistiksel olarak anlamlı bir fark olup olmadığı Wilcoxon işaretli sıra testi kullanılarak araştırılsın.

- testin hesaplama adımları verilmiştir. Öncelikle, sınıflandırma doğrulukları arasındaki farklar bulunmuş (sütun 4), daha sonra bu farkların mutlak değerleri hesaplanmış (sütun 5), farklar arasındaki işaretler tespit edilmiş (sütun 6) ve son olarak algoritmaların sıra değerleri bulunup, ilgili işaret ile çarpılıp son sütunda gösterilmiştir. Denklem 2.5 ve 2.6'da değerler yerine konulduğunda,  $R^+ = 185$ ,  $R^- = 25$  olarak bulunur. Denklem 2.7'ye göre  $T=25$  olur. Ek tablo-H 'da (Kartal, 2006) 0.05 anlam seviyesi ve  $n=20$  (veri kümesi sayısı) için okunan değer 52'dir. 25 sayısı 52'den küçük olduğu için  $H_0$  hipotezi reddedilir ve algoritmalar arasında anlamsal bir fark vardır denilir.

Veri Tabanı No (N <sub>ds</sub> )	GYAopt-aiNet	SSMA	Fark (d <sub>i</sub> )	Mutlak Fark	İşareti	İşaretli Sıra Değeri
1	0.747171	0.617187	0.129984	0.129984	+	20
2	0.798812	0.814914	-0.0161	0.016102	-	-3
3	0.72619	0.648673	0.077517	0.077517	+	17
4	0.75457	0.707921	0.046649	0.046649	+	7
5	0.993333	0.946667	0.046666	0.046666	+	8
6	0.981313	0.955267	0.026046	0.026046	+	5
7	0.798622	0.743229	0.055393	0.055393	+	11
8	0.715159	0.654514	0.060645	0.060645	+	13
9	0.930909	0.851515	0.079394	0.079394	+	18
10	0.764951	0.88149	-0.11654	0.116539	-	-19
11	0.63	0.56	0.07	0.07	+	15
12	0.939206	0.891693	0.047513	0.047513	+	9
13	0.818072	0.800898	0.017174	0.017174	+	4
14	0.845635	0.803213	0.042422	0.042422	+	6
15	0.929383	0.938013	-0.00863	0.00863	-	-1
16	0.825238	0.772063	0.053175	0.053175	+	10
17	0.79416	0.734141	0.060019	0.060019	+	12
18	0.74803	0.676075	0.071955	0.071955	+	16
19	0.958483	0.968286	-0.0098	0.009803	-	-2
20	0.970303	0.903939	0.066364	0.066364	+	14