

# EKMUD BİYOİSTATİSTİK AKADEMİSİ

**Yrd.Doç.Dr.Gökmen ZARARSIZ**

Erciyes Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Kayseri

Turcosa Analitik Çözümler Ltd Şti, Kayseri

[gokmenzararsiz@hotmail.com](mailto:gokmenzararsiz@hotmail.com)



**17 Şubat 2018 Cumartesi**

Saat	Konu	Anlatıcı
09:00-10:30	Biyoistatistik'te Temel Kavramlar	Dr. Gökmen ZARARSIZ
	Tanımlayıcı İstatistikler	
10:30-11:00	Kahve Arası	
11:00-12:30	Hipotez Testlerine Giriş	Dr. Gökmen ZARARSIZ
	İki Örneklem Testleri (Bağımlı-Bağımsız)	
12:30-13:30	Öğle Arası	
13:30-15:00	İkiden Fazla Örneklem Testleri (Bağımlı-Bağımsız)	Dr. Selçuk KORKMAZ
15:00-15:30	Kahve Arası	
15:30-17:30	Pratik Uygulama	Dr. Nazif Elaldı, Dr. Serap Gencer
17:30-18:00	TURCOSA: Milli İstatistik Yazılımı	Dr. Selçuk KORKMAZ



**18 Şubat 2018 Pazar**

Saat	Konu	Anlatıcı
09:00-11:00	Korelasyon Analizi	Dr. Gökmen ZARARSIZ
	Regresyon Analizleri (Doğrusal, Lojistik)	
11:00-11:30	Kahve Arası	
11:30-13:00	Tanı Testlerine Giriş	Dr. Selçuk KORKMAZ
	ROC Analizi	
13:00-14:00	Öğle Arası	
14:00-15:30	Pratik Uygulama	Dr. Nazif Elaldı, Dr. Serap Gencer
15:30-16:00	Kahve Arası	
16:00-17:30	Sağkalım Analizi	Dr. Selçuk KORKMAZ



# BİYOİSTATİSTİK'TE TEMEL KAVRAMLAR



## İstatistik

Bilimsel bir çalışmada çalışmanın tasarlanması, verilerin toplanması, analiz edilmesi, yorumlanması ile ilgilenen ve doğru karar verilebilmesi için yöntem ve teknik geliştiren bir bilim dalıdır.

TURCOSA



## Materials and methods

### *Study design*

This cross-sectional study was conducted in the elementary and secondary schools of the two central and 10 outlying districts of Kayseri province, Turkey. Data were obtained from the study of the Determination of Anthropometric Measurements of Turkish Children and Adolescents (DAMTCA-I) from February to April 2005. The study protocol was approved by the Erciyes University School of Medicine Ethics Committee and the Kayseri province

### *Data collection*

*Anthropometric indices.* All measurements were performed by well-trained health technicians. Weight, height and WC were measured twice, and the average was recorded for reference charts.

Weight was measured to 0.1 kg using a standard beam balance using a Tefal Ultraslim (France) accuracy, in minimal clothing (bare feet and with light clothing). Height was determined to the nearest 1 mm with portable stadiometers. The portable scales and stadiometers were calibrated daily.

*Questionnaire.* The survey was based on a questionnaire sent home prior to evaluation and collection of anthropometric data from participating children and adolescents. The questionnaire consisted of three sections. A general section concerned self-reported family data such as region of residence (urban/suburban), family income (low/fair/good) and

### *Statistical analysis*

The unpaired *t*-test was used to compare each parameter among gender for each age group. According to AFA classification (underweight, < 5th; healthy nutrition, ≥ 5–84.99th; overweight, ≥ 85th percentiles), comparison of mean BMI, WC and WHtR was performed with one-way ANOVA.

Univariate and multivariate multinomial logistic regression analyses were used to investigate the probability of underweight and overweight versus healthy nutrition according to socio-demographic variables and health-related behaviours. In the first step of multivariate regression modelling, all covariates were simultaneously taken in the regression model. Then from the set of covariates with  $p=0.05$  (or a block of covariates representing



### Data analysis

First, we evaluated the four prediction models previously built in the PPCOS-I population to predict success of ovulation, conception, pregnancy and live birth (Rausch et al., 2009) in the PPCOS-II population. The criteria for defining ovulation and conception have changed. For PPCOS-I, ovulation was defined as having a midluteal serum progesterone level  $>5$  ng/mL and conception was defined as any positive serum human chorionic gonadotrophin level. For PPCOS-II (Legro et al., 2014a,b), ovulation was defined as having a midluteal serum progesterone level  $>3$  ng/mL and conception was defined as having a serum level of human chorionic gonadotrophin  $>10$  mIU per milliliter. For both studies, clinical pregnancy was defined as an intrauterine pregnancy with fetal heart motion as detected by transvaginal ultrasound examination; live birth was defined as the delivery of a viable infant. Logistic regression was used for the analysis of these binary outcomes.

Since our first goal was to replicate the four published models, we considered the same set of baseline clinical and laboratory predictors as in Rausch et al. (2009). These comprised treatment, age, weight, body mass index (BMI), hirsutism score, race, waist measurement, waist/hip ratio, ethnic group, duration of attempting conception, pregnancy history, prior loss of pregnancy, prior parity, history of smoking, baseline total testosterone, baseline free androgen index (FAI), baseline glucose, baseline insulin, baseline pro-insulin, baseline SHBG and baseline white blood cell count. We categorized BMI ( $<30$ ,  $30-34$  and  $\geq 35$  kg/m<sup>2</sup>), age ( $\leq 34$ ,  $>34$  years), proinsulin ( $<23$ ,  $\geq 23$  pmol/l), FAI ( $<10$ ,  $\geq 10$ ), hirsutism ( $<8$ ,  $8-15$ ,  $\geq 16$ ) and duration of attempting conception ( $<1.5$ ,  $\geq 1.5$  years). For history of smoking, as an additional step, we believe it is more appropriate and informative to consider current, quit and never smoking.

We then used backward deletion to rebuild four logistic regression models for ovulation, conception, pregnancy and live birth using the PPCOS-II data. As in Rausch et al. (2009), both age and treatment variables were retained in the final models regardless of their statistical significance. For each covariate selected from the initial backward deletion procedure, interaction effects between treatment (CC or letrozole) and the selected covariates were then assessed, and those interactions that were significant at the 0.10 level or lower were included in the final models. For each outcome, the final models are presented in a table with odds ratios and the corresponding 90% confidence intervals for all covariates following the tables in Rausch

displayed by presenting the odds ratios for the treatment effect given the value of the covariate.

To compare the performance of the prediction models built from PPCOS-I and II data on the PPCOS-II data, we estimated the probability of live birth for each woman in the PPCOS-II study first using the PPCOS-I model and then using the PPCOS-II model. Then, we took the average of the estimated probabilities of live birth within the groups defined by the predictors in the PPCOS-I models. Figure 1 uses different colors to represent the ranges of the estimated probabilities and their differences.

To assess the predictive power of various models and the consistency of the predictors across PPCOS-I and II data, we constructed receiver operating characteristic (ROC) curves and calculated the areas under them (AUC). In this way, we assessed the PPCOS-I-generated models used in Rausch et al. (2009) applied to the PPCOS-II data and the PPCOS-II-generated models applied to the PPCOS-I data.

To assess the predictive power of our models in terms of both the precision and reproducibility, we adopted an intuitive calibration approach used in related studies (Veltman-Verhulst et al., 2012), performed the Hosmer–Lemeshow test (Hosmer and Lemeshow, 1980) for goodness-of-fit of the logistic regression models, and Pearson's chi-square test for independent validation. We stratified our samples into quintiles (i.e. very low, low, intermediate, high and very high success rates) according to the predicted model, obtained the average of the estimated success probabilities from the participants within each strata, and calibrated those averages with the observed percentages of success from the corresponding strata.

All analyses were performed with SAS software, version 9.2 (SAS Institute, Cary, NC, USA) or R ([www.r-project.org](http://www.r-project.org)). Trial Registration: ClinicalTrials.gov number: NCT0068861 (PPCOS-I) and NCT00719186 (PPCOS-II).



**İstatistiksel Olay:** Araştırmaya, incelemeye konu teşkil eden gözlenebilen, deneysel olarak varlığı kanıtlanabilen ve sayılarak, ölçülerek ya da tartılarak sayısal biçimde ifade edilebilen olay







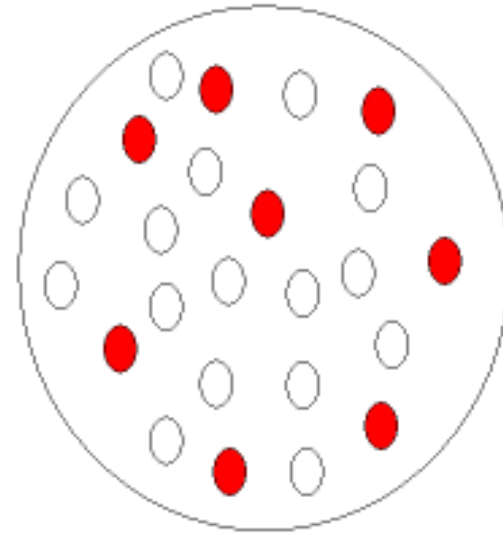
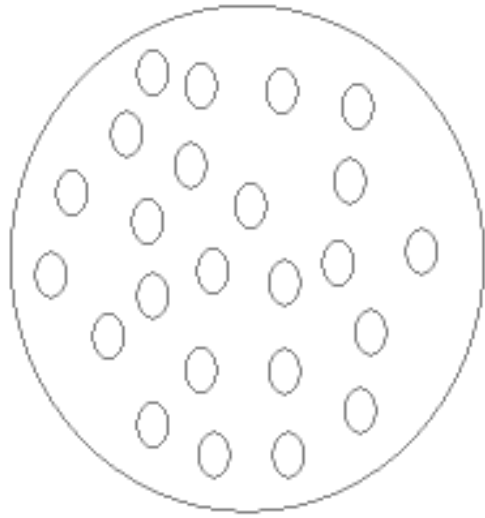
**Evren (Populasyon, Toplum, Kitle):** İstatistiksel olayın gözlenebildiği ve genellenebildiği gözlemler topluluğu

**Parametre :** Evrende ilgilendiğimiz değişkene ait özet değere denir.

TURCOSA



**Örneklem** : Bir evrenin içinde yer alan ve onu temsil edebilen alt küme



**İstatistik** : Örneklemden elde edilen özet değere denir.



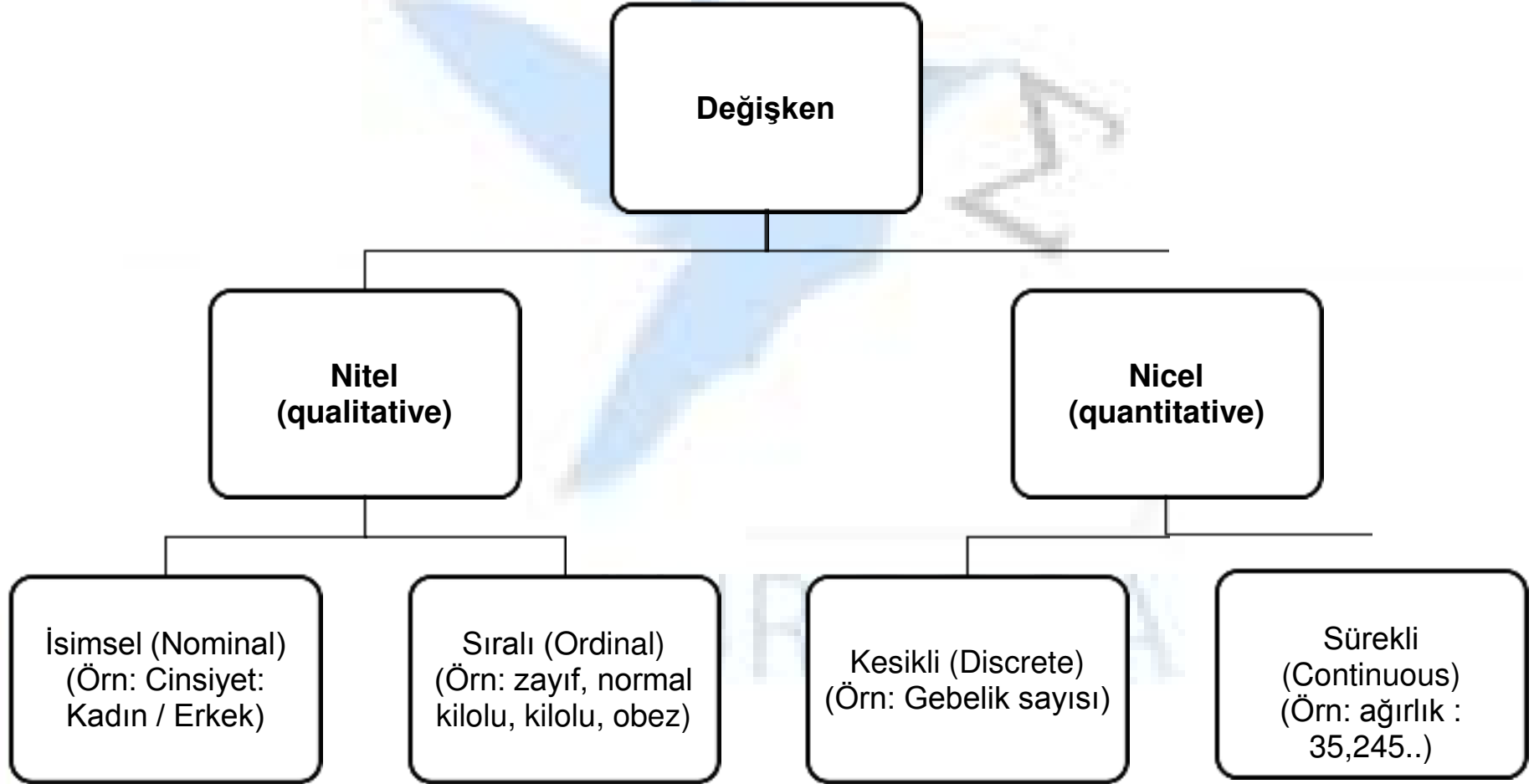
**Gözlem (observation):** İncelenen olayın gözleendiği en küçük toplum parçasına, toplum ögesine birim denir.





**Değişken (variable, feature):** Gözlemden gözleme farklı değerler alabilen ve değerleri sayısal olarak ifade edilebilen özelliklere değişken adı verilir.

TURCOSA





## Nitel Değişkenler

**İsimsel Değişken:** Birimlerin özellikleri isimsel kategorilere göre belirtilir. Herhangi bir sıralama söz konusu değildir. (Örn: Kan grubu: A, B, AB, 0 ; Medeni hal: Evli, bekar)



**Sıralı değişken:** Birimlerin özellikleri sıraya konulmuş biçimde belirtilir. Küçüklük veya büyüklük durumu vardır.

Boy uzunluğu : Kısa<Orta<Uzun  
Gelir durumu: Düşük<Normal<Yüksek)





## Nicel Değişkenler

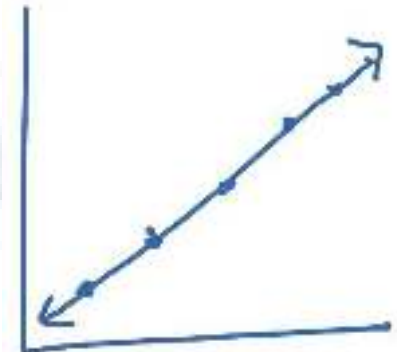
**Kesikli Değişken:** Değerler seti içinde sadece tamsayı değerler alabilen değişkenlerdir. Çocuk sayısı, hastanede kalma gün sayısı vb.

Discrete



**Sürekli değişken:** Değerler seti içinde her türlü değeri alabilen (tamsayı ve kesirli) değişkenlerdir. Boy uzunluğu, ağırlık, yaş, sistolik kan basıncı vb.

Continuous





## Bağımlılık Durumuna Göre Değişkenler

$$P = \frac{1}{1 + e^{-(-.5596 + 1.2528 X)}}$$

**Bağımlı Değişken:** Değeri başka değişkenlerce belirlenen, dışsal faktörlerden etkilenerek değer alan değişkenlerdir.

**Bağımsız Değişken:** Değeri rasgele oluşan, başka değişkenlerin değişimi üzerinde etkili olan değişkenlerdir.





**Veri:** Gözlemlerden elde edilen sayısal ya da sözel değerlerdir.

cinsiyet	gun	gebelik	dogumsayisi	kacincig	anneyas	babayas	anneegit	babaegit	annemeslek	babameslek
erkek	24,00	6,00	3,00	6,00	28,00	33,00	ilkokul	Ortaokul	Ev hanimi	Esnaf
erkek	14,00	3,00	2,00	3,00	25,00	30,00	Ortaokul	lise	Ev hanimi	işçi
erkek	4,00	3,00	3,00	3,00	35,00	43,00	universite v	universite v	Memur	Memur
erkek	22,00	2,00	1,00	2,00	27,00	27,00	lise	lise	Ev hanimi	Diger
kiz	29,00	1,00	1,00	1,00	19,00	22,00	Ortaokul	lise	Ev hanimi	Esnaf
kiz	28,00	2,00	2,00	2,00	28,00	28,00	ilkokul	ilkokul	Ev hanimi	Emekli
kiz	14,00	2,00	2,00	2,00	35,00	34,00	lise	lise	Ev hanimi	işçi
erkek	30,00	5,00	2,00	5,00	26,00	30,00	lise	ilkokul	Ev hanimi	işçi
erkek	27,00	3,00	3,00	3,00	31,00	36,00	ilkokul	universite v	Ev hanimi	işçi
erkek	5,00	4,00	4,00	4,00	34,00	39,00	Okuryazar	ilkokul	Ev hanimi	işçi
erkek	14,00	2,00	2,00	2,00	24,00	28,00	lise	lise	Ev hanimi	işçi
erkek	5,00	1,00	1,00	1,00	21,00	26,00	ilkokul	ilkokul	Ev hanimi	Esnaf
erkek	2,00	5,00	4,00	5,00	41,00	43,00	Ortaokul	ilkokul	Ev hanimi	işçi
erkek	8,00	2,00	2,00	2,00	24,00	29,00	ilkokul	ilkokul	Ev hanimi	Esnaf
kiz	9,00	2,00	2,00	2,00	31,00	30,00	ilkokul	ilkokul	Ev hanimi	Esnaf
kiz	9,00	2,00	2,00	2,00	30,00	41,00	Ortaokul	lise	Ev hanimi	Esnaf
kiz	19,00	1,00	1,00	1,00	20,00	28,00	lise	universite v	Ev hanimi	Esnaf
kiz	16,00	2,00	2,00	2,00	21,00	29,00	Ortaokul	Ortaokul	Ev hanimi	işçi
kiz	3,00	2,00	2,00	2,00	27,00	34,00	Ortaokul	Ortaokul	Ev hanimi	işçi
kiz	31,00	2,00	2,00	2,00	37,00	52,00	lise	lise	Ev hanimi	Emekli
kiz	24,00	2,00	2,00	2,00	27,00	36,00	lise	lise	Memur	Memur
kiz	8,00	3,00	3,00	3,00	28,00	38,00	ilkokul	Ortaokul	Ev hanimi	işçi
erkek	16,00	2,00	2,00	1,00	21,00	25,00	Ortaokul	ilkokul	Ev hanimi	Esnaf



**İstatistiksel testlerin tümünde ortak olan bir varsayım verinin analiz öncesi eksiksiz ve hatasız elde edilmiş olmasıdır.**

TURCOSA



## Biyostatistik Analizlerinde Kullanılan Bazı Yazılımlar





## ***Neden TURCOSA?***

- *TURCOSA, kullanıcı dostu bir yazılımdır.*
- *TURCOSA bulut tabanlı bir istatistik yazılımıdır. Bilgisayar, tablet ya da akıllı telefon üzerinden istatistiksel analiz yapabilirsiniz.*
- *TURCOSA kişiselleştirilmiş proje tabanlı bir istatistik yazılımıdır. Aynı proje içerisinde birden fazla veri seti ile eş zamanlı çalışabilirsiniz. Çalışma arkadaşlarınızla birlikte analiz yapabilirsiniz. Proje ile ilgili makale, iş listesi, analiz sonuçlarını saklayabilirsiniz.*
- *TURCOSA Türkçe ve İngilizce dil seçeneklerini desteklemektedir.*
- *TURCOSA istatistik yazılımı için ömür boyu kullanıcı lisansı almak zorunda değilsiniz, haftalık/aylık/yıllık lisans sahibi olabilirsiniz.*
- *TURCOSA interaktif raporlama seçeneklerine sahiptir.*

