



İST156 İSTATİSTİĞE GİRİŞ II

UYGULAMA 6

HACETTEPE ÜNİVERSİTESİ
İSTATİSTİK BÖLÜMÜ

Ar. Gör. Dr. Derya Turfan – Ar. Gör. Leyla Bakacak Karabenli

1- Sağlık taramasının yapıldığı bir çalışmada X hastalığına yakalananların oranı tahmin edilmek isteniyor. %95 güven düzeyinde hoşgörü miktarı (kabul edilebilir hata veya d) 0.01, 0.05 ve 0.10 olarak alındığında bu çalışma için örneklem büyüklüğü ne olmalıdır?

- Hastalığa yakalanma oranı üzerinden tahmin yapabilmek amacıyla gerekli örneklem büyüklüğünü elde edebilmek için “*samplingbook*” paketi kullanılacaktır.

```
install.packages("samplingbook", repos='http://cran.us.r-project.org')
##
## Installing package into 'C:/Users/Documents/R/win-library/3.5'
## (as 'lib' is unspecified)
## package 'samplingbook' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:/Users/AppData/Local/Temp/RtmpmHFZV/downloaded_packages
##
library(samplingbook)
##
## Warning: package 'samplingbook' was built under R version 3.5.3
## Loading required package: pps
## Warning: package 'pps' was built under R version 3.5.2
## Loading required package: sampling
## Warning: package 'sampling' was built under R version 3.5.3
## Loading required package: survey
## Warning: package 'survey' was built under R version 3.5.3
## Loading required package: grid
## Loading required package: Matrix
## Loading required package: survival
##
## Attaching package: 'survival'
## The following objects are masked from 'package:sampling':
##
##   cluster, strata
##
## Attaching package: 'survey'
## The following object is masked from 'package:graphics':
##
##   dotchart
```

- Bu pakette “*sample.size.prop*” fonksiyonu kullanılacak ve girdi olarak; hoşgörü miktarı (e), beklenen oran değeri (P) ve güven düzeyi ($level$) tanımlanması gerekmektedir. Hoşgörü miktarı soruda verildiği gibi 3 değer için ayrı olarak tanımlanmıştır. Oran değeri, maksimum örneklem büyüklüğünü elde edebilmek için 0.5 ve güven düzeyi ise soruda verildiği gibi 0.95 olarak tanımlanmıştır.

```
sample.size.prop(e = 0.01, P = 0.5, level = 0.95)
##
## sample.size.prop object: Sample size for proportion estimate
```

```
## Without finite population correction: N=Inf, precision e=0.01 and expected proportion P=0.5
##
## Sample size needed: 9604
sample.size.prop(e = 0.05, P = 0.5, level = 0.95)
##
## sample.size.prop object: Sample size for proportion estimate
## Without finite population correction: N=Inf, precision e=0.05 and expected proportion P=0.5
##
## Sample size needed: 385
sample.size.prop(e = 0.10, P = 0.5, level = 0.95)
##
## sample.size.prop object: Sample size for proportion estimate
## Without finite population correction: N=Inf, precision e=0.1 and expected proportion P=0.5
##
## Sample size needed: 97
```

- Tahmin sonuçlarına göre örneklem büyüklükleri; 0.01 hoşgörü miktarı için 9604, 0.05 için 385 ve 0.10 için 97 olmalıdır. Oran değeri ve güven düzeyi sabit tutulduğunda hoşgörü miktarı arttıkça gerekli örneklem büyüklüğünde azalma görülmüştür.

Ek: Fonksiyon kullanmak yerine tercih edilebilecek kodlar

```
# Kitle büyüklüğü(N) bilinmediğinde
z_table<-1.96
d<-0.01
P<-0.5
Q<-1-P
n<-(z_table*sqrt(P*Q))^2/(d^2); n
## [1] 9604

# Kitle büyüklüğü(N) bilindiğinde
N<-4000000
n<-(N*z_table^2*P*Q)/((N-1)*(d^2)+((z_table^2)*P*Q)); n
## [1] 9580.998
```

2- Bir sigorta şirketi kasko yaptıran müşterilerinin araçları ile ilgili bir araştırma yapmak istiyor. Önceki bilgilerden araçların yaşlarına ilişkin varyansı 5'tir. Kabul edilebilir hata 0.28 kabul edildiğinde, 0.05 anlamlılık düzeyinde örneklem büyüklüğü ne olmalıdır?

- 1.soruda olduğu gibi ortalama tahmini için gerekli örneklem büyüklüğünü hesaplayabilmek amacıyla "samplingbook" paketi kullanılacaktır. Bu yüzden tekrar paket yüklemesi yapılmasına gerek yoktur fakat fonksiyon olarak "sample.size.mean" kullanılacaktır. Bu fonksiyonda girdi olarak; hoşgörü miktarı (e), standart sapma (S) ve güven düzeyi (level) kullanılacaktır.

```
sample.size.mean(e = 0.28, S = sqrt(5), level = 0.95)
##
## sample.size.mean object: Sample size for mean estimate
## Without finite population correction: N=Inf, precision e=0.28 and standard deviation
S=2.2361
##
## Sample size needed: 245
```

- 0.28 hoşgörü miktarı, 5 varyans değeri ve %95 güven düzeyinde araçların yaşlarına ilişkin yapılacak çalışmada gerekli örneklem büyüklüğü 245 olarak elde edilmiştir.

Ek: Fonksiyon kullanmak yerine tercih edilebilecek kodlar

```
# Kitle büyüklüğü(N) bilinmediğinde
z_table<-1.96
d<-0.28
sigma<-sqrt(5)
n<-(z_table*sigma)^2/(d^2); n
## [1] 245

# Kitle büyüklüğü(N) bilindiğinde
N<-40000
n<-(N*(z_table*sigma)^2)/((N-1)*(d^2)+(z_table*sigma)^2); n
## [1] 243.5146
```

3- 14 kadının altı haftalık bir egzersiz programına katıldıktan sonraki ağırlık değişimleri kaydedilmiş. Ortalama ağırlık değişiminin 0'dan farklı olduğuna dair yeterli kanıt olup olmadığını test edebilmek için örneklem büyüklüğü ne olmalıdır?

- İlk iki soruda oran ve ortalama tahmini için gerekli örneklem büyüklükleri hesaplanmıştı. Şimdi bu sorulardan farklı olarak ortalama üzerinden kurulacak olan hipotez testi için gerekli örneklem büyüklüğünün hesaplanması gerekmektedir. Bunun için “pwr” paketinin yüklenmesi gerekmektedir.

```
install.packages("pwr", repos='http://cran.us.r-project.org')
##
## Installing package into 'C:/Users/Documents/R/win-library/3.5'
## (as 'lib' is unspecified)
## package 'pwr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:/Users/AppData/Local/Temp/RtmpmHFZV/downloaded_packages
##
library(pwr)
##
## Warning: package 'pwr' was built under R version 3.5.3
```

- Örneklem büyüklüğünün hesaplanması için “pwr.t.test” fonksiyonu kullanılacaktır. Girdi olarak; etki büyüklüğü (d), anlamlılık değeri (sig.level), güç (power) ve t testi türünün (type) tanımlanması gerekmektedir.
- Etki büyüklüğü, test edilen ortalama değeri ile gerçek ortalama arasındaki farkın standart sapmaya oranlanmasıyla elde edilmektedir. Bunun için ağırlıklardaki değişim miktarları “veri” olarak isimlendirilen vektörle tanımlanmış ve standart sapması hesaplanmıştır. Daha sonra ortalamalar arasındaki farkın (5 olarak varsayılmıştır), standart sapmaya bölümünden elde edilen değerin yuvarlanmış hali olan 0.38 değeri etki büyüklüğü olarak kullanılmıştır.

```
veri<-c(17, 7, -4, -18, 2, 9, 12, 9, -12, -9, -18, -14, -18, -20)
standart_sapma<-sqrt(var(veri))
```

```
etki_buyuklugu<-5/standart_sapma
etki_buyuklugu
## [1] 0.3821938
```

- Hipotez testinde seçenек hipotezinin yönünün belirtilmesi gerekmektedir. “*pwr.t.test*” fonksiyonunda herhangi bir tanımlama yapılmadığı sürece seçenек hipotezinin yönü iki yönlü olarak algılanacaktır. Soruda da 0’a eşitliğin kontrol edilmesi gerekmekte ve bu yüzden aşağıdaki fonksiyonda herhangi bir seçenек hipotezi yönünün belirtilmesine gerek kalmamıştır. Tek yönlü bir hipotez için *alternative = c (“less”, “greater”)* tanımlamalarından birinin kullanılması gerekmektedir.
- Anlamlılık değeri 0.05, güç 0.80 ve t testi türü tek örneklem olarak belirlendikten sonra örneklem büyüklüğü hesaplanmıştır.

```
pwr.t.test(d = 0.38 , sig.level = 0.05 , power = 0.80, type = "one.sample")
##
##      One-sample t test power calculation
##
##              n = 56.30592
##              d = 0.38
##      sig.level = 0.05
##              power = 0.8
##      alternative = two.sided
```

- Kadınların ortalama ağırlık değişiminin 0’a eşit olup olmadığının testinin yapılabilmesi için gerekli örneklem büyüklüğü yaklaşık 56 olarak elde edilmiştir.

Ek: Kullanılan fonksiyonların hangi girdilerle çalıştığının incelenebilmesi için help menüsünün arama tuşundan yararlanılabilir. Ayrıca Console’da her bir fonksiyonun başına tek/çift soru işareti koyulması ile de yardım menüsüne erişim sağlanacaktır. Çift soru işareti, aratılan fonksiyonu içeren tüm sekmelere ulaşılmasını sağlayacaktır.

```
?sample.size.prop
?sample.size.mean
?pwr.t.test
```