

# ANOVA

Hossein Vatani

March 29, 2016

## بنام خدا

### تحلیل پراکنش (تحلیل واریانس) یکطرفه

اغلب اوقات ما در بررسی معنی داری تفاوت بین داده ها نیاز به بررسی میانگین بیش از دو دسته داده بایکدیگر داریم. جهت انجام آن باید از تحلیل پراکنش که در R با دستور `anova` نمایش داده می شود استفاده نماییم.

اگر تحلیل پراکنش (یا همان آزمون فیشر) مشخص کرد که بین میانگین داده ها تفاوت معنی داری وجود دارد، باید با بررسی هایی دیگر مشخص شود که کدامیک از داده ها باهم تفاوت معنی دار دارند.

### الف- تحلیل پراکنش

اولین گام در تحلیل پراکنش، بررسی تصویری میانگین متغیرها مورد نظر در قبال سایر داده هاست. برای ایم امر لازم است که نمودار جعبه ای آنها را پیاده سازی کنیم و این موضوع بسادگی با دستور زیر امکانپذیر است.

```
plot(response ~ factor, data=data_name)
```

که

`response`=متغیر پاسخ یا همان مورد نظر جهت بررسی است  
`factor`=متغیری که داده ها با توجه به آن دسته بندی می شوند  
`data_name`=مجموعه داده ای مورد نظر

مثال-یک شرکت تولید دارو، سه داروی مسکن برای بیماران مبتلا به سردرد میگرد را بر روی ۲۷ داوطلب که بطور تصادفی به سه گروه نه نفره تقسیم شده بودند، آزمایش کرد. به هر گروه داروی مربوطه داده شد و قرار شد در سردرد بعدی آنها مصرف کرده و سپس شدت درد را با عددی بین ۱۰۰۰ گزارش نمایند که نتیجه بشرح زیر شد.

DrugA	4	5	4	3	2	4	3	4	4
DrugB	6	8	4	5	4	6	5	8	6
DrugC	6	7	6	6	7	5	6	5	5

آزمون فرض را اینگونه تعریف می کنیم:  
$$H_0: \mu_1 = \mu_2 = \mu_3 \quad H_1: \exists i, j \text{ such that } \mu_i \neq \mu_j$$

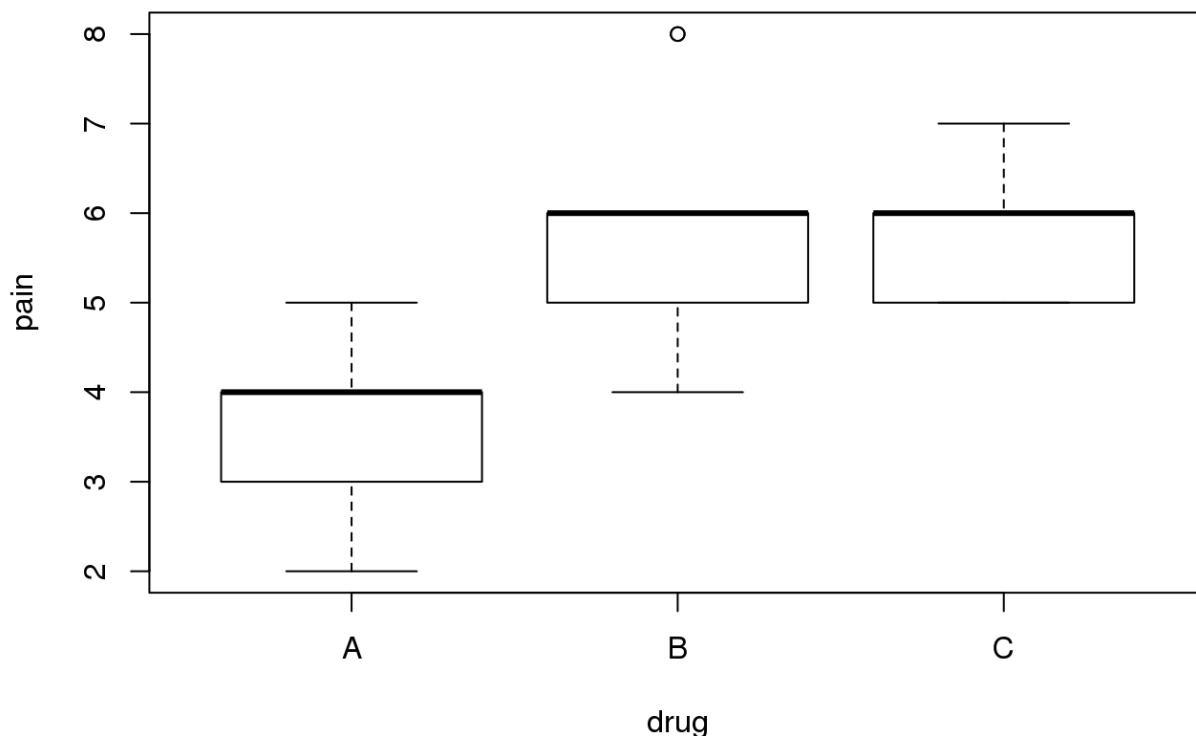
جهت نمایش تصویری، ابتدا لازم است که داده ها را بصورت قاب-داده در نرم افزار وارد کنیم.

```
pain = c(4, 5, 4, 3, 2, 4, 3, 4, 4, 6, 8, 4, 5, 4, 6, 5, 8, 6, 6, 7, 6, 6, 7,
5, 6, 5, 5)
drug = c(rep("A",9), rep("B",9), rep("C",9))
migraine = data.frame(pain,drug)
migraine
```

```
##      pain drug
## 1      4     A
## 2      5     A
## 3      4     A
## 4      3     A
## 5      2     A
## 6      4     A
## 7      3     A
## 8      4     A
## 9      4     A
## 10     6     B
## 11     8     B
## 12     4     B
## 13     5     B
## 14     4     B
## 15     6     B
## 16     5     B
## 17     8     B
## 18     6     B
## 19     6     C
## 20     7     C
## 21     6     C
## 22     6     C
## 23     7     C
## 24     5     C
## 25     6     C
## 26     5     C
## 27     5     C
```

جهت نمایش جعبه ای میزان درد در هر گروه

```
plot(pain~drug, data = migraine)
```



همانطور که از تصویر نمایان است میانگین تاثیر داروی A از دو داروی دیگر کمتر و همچنین میانگین دو داروی B و C با توجه به تصویر تقریباً برابر است.

در مرحله بعد با استفاده از تابع `aov()` به تحلیل موضوع (در واقع همان بررسی شرط آزمون) خواهیم پرداخت. رابطه کلی آن بدین صورت است

```
aov(response ~ factor, data=data_name)
```

که توضیحات عاملهای آن مانند عاملهای نمایش-جعبه ایست. نتیجه این آزمون یک شیء از نوع لیست (List) می باشد که با دستور `summary()` می توان تعداد قابل توجهی از اطلاعات مفید آن را مشاهده کرد.

```
fit = aov(pain ~ drug, data=migraine)
summary(fit)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## drug         2   28.22   14.111    11.91 0.000256 ***
## Residuals    24   28.44    1.185
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

مقدار p-value این آزمون فیشر به ما می گوید که بطور قطع فرض صفر رد می شود.

## ب-مقایسه چندگانه

تحلیل پراکنش آزمون فیشر، برای تحلیلگر مشخص می کند که آیا بین عاملها تفاوت معنی داری وجود دارد یا خیر اما همانطور که گفته شد، ابزار تشخیص که کدام عاملها با یکدیگر تفاوت دارند را در اختیار نمی گذارد. همانطور که مستحضر هستید آزمون هایی دیگر بانام پس-از-آزمون (add-hoc) وجود دارد.

برای مثال قبل که مقایسه برای گروه های یک مجموعه از داده می باشد آزمون `pairwise.t.test()` بکار می رود. رابطه کلی آن

```
pairwise.t.test(reponse, factor, p.adjust = method, alternative = c("two.sided", "less", "greater"))
```

می باشد که عامل های اول و دوم و چهارم توضیح داده شده است

`p.adjust` = نوع های متفاوتی وجود دارد که می توان به `benferroni`, `holm`, `hochbe` `rg`, `hommel`, `BH`, `BY` اشاره کرد

برای مثال قبل داریم:

```
pairwise.t.test(pain, drug, p.adjust="bonferroni")
```

```
##
## Pairwise comparisons using t tests with pooled SD
##
## data:  pain and drug
##
##      A      B
## B 0.00119 -
## C 0.00068 1.00000
##
## P value adjustment method: bonferroni
```

که همانطور که از نتیجه مشخص است، بین مقادیر داروی A با هر دو نوع داروی B و C اختلاف معنی دار وجود دارد که این مورد بین B و C وجود ندارد. لذا فرض  $H_1: \mu_1 \neq \mu_2$  بدین صورت قابل اعلام است:  $H_1: \mu_1 \neq \mu_3$

آزمون دیگری که قابل استفاده است آزمون Tukey می باشد که رابطه آن بشرح زیر است

```
TukeyHSD(fit, conf.level = 0.95)
```

که

`fit` = نتیجه تحلیل پراکنش (که همان شیئی لیست حاصل از دستور `aov` می باشد).  
`conf.level` = سطح اطمینان مورد نظر

مجددا مثال قبل را بررسی می کنیم.

```
TukeyHSD(fit, conf.level = 0.95)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = pain ~ drug, data = migraine)
##
## $drug
##      diff      lwr      upr      p adj
## B-A 2.111111 0.8295028 3.392719 0.0011107
## C-A 2.222222 0.9406139 3.503831 0.0006453
## C-B 0.111111 -1.1704972 1.392719 0.9745173
```