

آمار توصیفی:

آرد بفاهیم برای یک مجموعه داده مانند mydata که شامل چند متغیر است، شاخصی را محاسبه کنیم، از تابع `Summary (mydata)` استفاده می‌کنیم. در بیان مثال آرد بفاهیم میانگین متغیرهای داخل mydata با در فرمی داشته باشیم از دستور:

`Summary (mydata , mean , na.rm = T)`

استفاده می‌کنیم.

* عبارت `na.rm = T` در واقع مشاهده‌ها را حذف می‌کند. آرد از عبارت

`na.rm = F` استفاده کنیم و متغیری دارای مشاهده گمشده باشد، خروجی شاخص

برای آن متغیر NA یا گمشده خواهد بود.

* شاخص‌های دیگری که می‌توان با این تابع محاسبه کرد عبارتند از:

`var` (واریانس نمونه‌ای هر متغیر) - `sd` (انحراف معیار) - `min - max` (حداقل و بیشینه) - `quantile` (شامل چارک اول - میانه - چارک سوم) - `range - median`

آرد بفاهیم اطلاعات میانه، میانگین، چارک اول و چارک سوم و مینیمم و ماکزیمم داده‌ها در فرمی زیر نمایش داده شود، از تابع `Summary ()` استفاده می‌کنیم.

`Summary (mydata)`

روش دیگری که برای بدست آوردن خلاصه‌ای از اطلاعات وجود دارد، استفاده از `Package` است.

از جمله پکیج‌هایی که برای آمار توصیفی به کار گرفته می‌شود، می‌توان به پکیج‌های `psych` یا `pastecs` اشاره کرد.

pastecs

نحوه استفاده از پکیج

پس از بارگذاری این پکیج، ابتدا نصب کنید و سپس از تابع (Stat.desc) استفاده کنید:

```
> library (pastecs)
```

```
> Stat.desc (mydata)
```

psych

فنون سیاتا و ویژگی های پکیج

تایمی که پس از نصب پکیج psych برای بدست آوردن آمار توصیفی نه کار برده می شود، تابع (describe) است:

```
> library (psych)
```

```
> describe (mydata)
```

از دیدن نتایج موجود در این پکیج، می توان به تابع (describe.by) یا

(describeBy) اشاره کرد.

ویژگی این تابع این است که آمار به تفصیل خلاصه های آماری را برای دسته های مختلف یک متغیر (همچون مثال متغیر جنسیت) بدست آوردیم از این تابع استفاده می کنیم:

```
> library (psych)
```

```
> describe.by (mydata, mydata $ sexuality)
```

```
> describeBy (mydata, mydata $ sexuality)
```

فرونی های این تابع، میانگین، انحراف معیار، میانه، دامنه تغییرات، چولگی و کشیدگی هستند.

بدست آوردن فراوانی داده‌های گسسته

برای بدست آوردن جدول فراوانی و مشاهده تعداد فراوانی‌های هر دسته از یک متغیر گسسته، از تابع `table()` استفاده می‌کنیم.

بعنوان مثال اگر متغیر جنسیت (`Sexuality`) در مجموعه داده `mydata` قرار داشته باشد و بخواهیم تعداد فراوانی‌های هر جنسیت را مشاهده کنیم از دستور

```
> table(mydata $ sexuality)
```

استفاده می‌کنیم.

همچنین فرق کنید دو متغیر `A` و `B` هر کدام دارای دو سطح باشند و n مشاهده برای هر متغیر ثبت شده باشد و بخواهیم فراوانی هر سطح از متغیر `A` را در هر سطح از متغیر `B` بیجا کنیم، در این صورت از دستور

```
> table(A, B)
```

استفاده می‌کنیم.

خروجی این برنامه به صورت زیر است:

$$n_{11} + n_{12} + n_{21} + n_{22} = n$$

	B_1	B_2
A_1	n_{11}	n_{12}
A_2	n_{21}	n_{22}

همچنین اگر بخواهیم فراوانی نسبی هر دسته را محاسبه کنیم از تابع `prop.table()` استفاده می‌کنیم. البته قبل از محاسبه فراوانی نسبی باید فراوانی‌ها متغیر مورد نظر در یک `data frame` ذخیره شود.

```
> Ma = table(mydata $ sexuality)
```

```
> prop.table(Ma)
```

```
> MM = table(A, B)
```

که در این صورت خروجی آن به صورت زیر است:

	B_1	B_2
A_1	$\frac{n_{11}}{n}$	$\frac{n_{12}}{n}$
A_2	$\frac{n_{21}}{n}$	$\frac{n_{22}}{n}$

همچنین اگر سه متغیر دسته‌ای مانند A ، B و C داشته باشیم که هر کدام از آنها دارای دو سطح باشند و بخواهیم فواصل اینها را محاسبه کنیم داریم:

> table(A, B, C)

	C_1	
	B_1	B_2
A_1	n_{111}	n_{121}
A_2	n_{211}	n_{221}

	C_2	
	B_1	B_2
A_1	n_{112}	n_{122}
A_2	n_{212}	n_{222}

حال اگر بخواهیم هم‌شدید جدول کوادرنی به ما نمایش داده شود از دستور (Ftable) استفاده می‌کنیم. البته متغیری که داخل این تابع قرار می‌گیرد dataframe یا

> M2 = table(A, B, C)

> Ftable(M2)

		C_1	C_2
A_1	B_1	n_{111}	n_{112}
	B_2	n_{121}	n_{122}
A_2	B_1	n_{211}	n_{212}

آرایه است.

حال آذربخفاهیم خلاصه‌های آماری با در هر یک از سطوح C بدست آوریم به صورت زیر عمل می‌کنیم:

> A1 = Ftable(M2)

> Summary(A1)

خروجی این قسمت، میان، میانگین، مینیمم، ماکزیمم و چارک اول و سوم فراوانی‌ها جدول فراوانی مربوط به هر یک از سطوح‌های C می‌باشد.

آذربخفاهیم این خلاصه‌ها را برای هر یک از سطوح مختلف A در جدول متشکل از سه ستون A، B و C بدست آوریم از دستر زیر استفاده می‌کنیم:

> T1 = table(B, C, A)

> T2 = Ftable(T1)

> Summary(T2)

آذربخفاهیم برای این جدول تعاقبی، آزمون استقلال بین مقی‌ها رسته‌ای را بررسی کنیم از دستر زیر استفاده می‌کنیم:

> T1 = table(A, B, C)

> Summary(T1)

* بنا بر این آذربخفاهیم آوردن خلاصه‌های آماری بر اساس یک دسته خاص بود، اینجا خروجی‌های مربوط به تابع (Ftable) با در یک dataframe ذخیره می‌کنیم و سپس از تابع Summary استفاده می‌کنیم. اینجا آذربخفاهیم آزمون استقلال بین مقی‌ها رسته‌ای انجام دهیم، کافیت اطلاعات بدست آمده از (table) با در یک dataframe ذخیره و سپس از تابع (Summary) استفاده

کنیم. همچنین می‌توان از تابع (chisq.test) و (fisher.test) نیز

> chisq.test(T1)

برای انجام آزمون استقلال استفاده کرد.

تابع ژئومیتریک در R

ساختار کلی تابع در R به صورت زیر است:

نام تابع \uparrow
 آرگومان‌ها تابع \uparrow
`myfun = function (arguments) {`

`expressions` \rightarrow دستورات مورد نیاز برای تابع
`return (output)` \rightarrow خروجی که از تابع می‌فهمیم
`}`

توجه است قبل از اینکه اسم تابع را انتخاب کنیم `enter` بزیرم ببینیم تابعی با این نام قبلاً در R وجود نداشته است یا نه.

مثال: فرض کنیم هدف ما درست آوردن میانگین هندسی یک بردار از اعداد باشد.
 فرمول میانگین هندسی برای بردار مشاهده‌ای (x_1, \dots, x_n) برابر است با:

$$\bar{x}_G = \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}}$$

پس آرگومان‌های مورد نیاز برای درست آوردن این تابع یکی بردار مشاهده‌ای و دیگری تعداد آن‌ها است که البته تعداد مشاهده‌ای همان در خود تابع با `length(x)` مشخص کرد.

> `gmean` نام تابع (دلفیاه)
 Error: پس چنین تابعی در R نداریم

> `gmean = function(u) {`

$$g = (\text{prod}(u))^{(1 / \text{length}(u))}$$

`return (g)`

`}`

> `x = c(1:100)`

> `gmean(x)`

نمایش در مورد تابع نویسی :

۱- درج تابع نوشته شده کوتاه تر باشد، بهر وقت قابل فهم تر است.

۲- می توان به جای (g) return از خود و استفاده کنیم.

```
> gmean = function (u) {
+   g = (prod(u)) ^ (1/length(u))
+   g }
```

۳- اگر به جای یک درجی، چند درجی با ارجاع بعضی به بعضی از list استفاده کنیم.

```
> gmean = function (u) {
+   list (g = prod(u) ^ (1/length(u)), n = length(u))
+ }
```

نوع اول :

تابعی بنویسید که بتوان از طریق آن، میانگین هارمونیک یک بردار از مشاهده ها را بدست آورد.

$$x = (x_1, \dots, x_n)$$

$$\bar{x}_H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

$$x_i \neq 0 \quad \forall x_i$$

بدست آورد.

* نوع دوم :

تابعی بنویسید که از طریق آن بتوان یک نمونه تصادفی n تایی از مقادیر تصادفی X با تابع

$$X \sim f_X(x)$$

$$f_X(x) = \lambda e^{-\lambda(x-\theta)}$$

$$x \geq \theta \quad \lambda > 0 \quad \theta \in \mathbb{R}$$

چگالی زیر تولید کرد :

← راه های : ابتدا n عدد تصادفی از توزیع یکگانه بین (۱،۱) تولید کنید و سپس

چندک های این مقادیر را در توزیع داده شده محاسبه کنید. چند ک های بدست آمده،

اعداد تصادفی از این توزیع می باشند !!