

# داده های مفقود (Missing Data)

مریم معینی

دکترای اقتصاد سلامت





# چرا و به چه علت؟!

- ✓ آزمودنی حاضر به شرکت در مطالعه نیست.
- ✓ آزمودنی حاضر به پاسخگویی به برخی سوالات مطالعه نیست.
- ✓ آزمودنی پاسخ برخی سوالات را نمیداند.
- ✓ آزمودنی سهوا به برخی سوالها پاسخ نمیدهد.
- ✓ آزمودنی واجد شرایط پاسخگویی به برخی سوالات نیست.
- ✓ آزمودنی مطالعه را پیش از پایان رها می کند.
- ✓ .....

# چرا و به چه علت!؟



- ✓ تحلیلگر آزمودنی را سهوا از مطالعه حذف میکند.
- ✓ تحلیلگر فراموش میکند بخشی از اطلاعات آزمودنی را جمع آوری یا ثبت کند.
- ✓ تحلیلگر آگاهانه آزمودنی را از مطالعه حذف میکند.
- ✓ تحلیلگر آگاهانه بخشی از اطلاعات آزمودنی را جمع آوری یا ثبت نمی کند.
- ✓ بخشی از اطلاعات از دست میرود.
- ✓ خطا در طراحی ابزار گردآوری اطلاعات وجود دارد.
- ✓ ....

# خطای ابزار

۱. آیا در سال گذشته چکاب دندانپزشکی انجام داده اید؟

بلی خیر

۲. تعداد دفعات چکاب دندانپزشکی در سال گذشته چند بار بوده است؟

یک بار	دو بار	سه بار	بیش از سه بار
--------	--------	--------	---------------

از کدهای مختلف برای ثبت داده های مفقود استفاده نمایید.



سن	مدت بستری	سابقه سکته قلبی	فشار خون	بیمار
۶۱	۸	۰	۱۳	مینا
-99	۰	۰	۰	رضا
۰	-99	-98	۱۴	سپیده
۵۴	۱۲	-98	-99	منصور
۴۳	۱۱	۱	۱۲	احمد

..فراموشی در ثبت اطلاعات

**99-** پاسخ مبهم

**98-** عدم پاسخگویی

**چرا داده های مفقود مهم هستند؟**

سن ابتلا به بیماری	بیمار
▪	۱
▪	۲
۳۹	۳
▪	۴
۴۲	۵
▪	۶
۳۷	۷
۳۹	۸

سن ابتلا به بیماری	بیمار
۲۱	۱
۲۲	۲
۳۹	۳
۲۰	۴
۴۲	۵
۱۸	۶
۳۷	۷
۳۹	۸



**داده های مفقود به یک اندازه خطرناک نیستند!**

# متهمان ردیف اول!



✓ مفقود شدن داده ها به صورت کاملا تصادفی

✓ مفقود شدن داده ها به صورت تصادفی

✓ مفقود شدن داده ها به صورت غیر تصادفی

## مثال: ترک مطالعه

مطالعه ای برای بررسی اثر یک داروی جدید بر کاهش فشار خون در نمونه ای از بیماران در حال انجام است. تعدادی از بیماران پیش از پایان دوره درمان، مطالعه را ترک کرده اند و در نتیجه بخشی از مقادیر فشار خون آنها مفقود است. ترک مطالعه دلایل متفاوتی دارد و بنابراین تاثیر مقادیر مفقود بر نتایج مطالعه یکسان نیست.





- ✓ تعدادی از بیماران به دلایل شخصی مانند مسافرت، سرماخوردگی، بی توجهی به بیماری و ... درمان را رها کردند.
- ✓ تعدادی از بیماران به دلیل عوارض دوز بالای دارو، درمان را رها کردند.
- ✓ تعدادی از بیماران سالمند به دلیل عدم تطابق با درمان جدید، آن را رها کردند.
- ✓ بیماران دارای فشار خون بالا به دلیل مشکلات فشار خون، درمان جدید را رها کردند.

## مثال: عدم مشارکت در مطالعه

مطالعه ای برای بررسی بیمای های ویروسی منتقله از طریق خون در بیماران بستری در بخش اورژانس یک بیمارستان بر اساس اطلاعات ثبتی بیماران در زمان پذیرش بیماران انجام شده است. اطلاعات مربوط به نتایج آزمایش خون تعدادی از بیماران مفقود است. عدم ثبت اطلاعات دلایل متفاوتی دارد و بنابراین تاثیر مقادیر مفقود بر نتایج مطالعه یکسان نیست.





- ✓ پزشک فراموش کرده است آزمایش خون را درخواست دهد.
- ✓ کارشناس آمار فراموش کرده است نتایج را ثبت کند.
- ✓ به دلیل وخامت حال بیمار، پزشک از تجویز آزمایش منصرف شده است.
- ✓ بیمار مبتلا به دلیل ترس از عدم دریافت خدمات، حاضر به انجام آزمایش نشده است.

## مثال: عدم پاسخگویی



مطالعه ای برای بررسی رفتارهای پرخطر در بیماران مراجعه کننده به بخش روانپزشکی یک بیمارستان انجام شده است. اطلاعات از طریق پرسشنامه جمع آوری شده است. بیماران باسواد شخصا پرسشنامه را پر کرده اند و برای بیماران کم سواد پرسشنامه توسط پرسشگر از طریق مصاحبه حضوری پر شده است. تعدادی از بیماران به برخی سوالات جواب نداده اند. بنابراین تاثیر مقادیر مفقود بر نتایج مطالعه یکسان نیست.



- ✓ چند نفر از بیمارانی که پرسشنامه را پر کرده اند به جای پاسخگویی به تعدادی از سوالات، پاسخ نوشته اند: «سوال را متوجه نشدم»
- ✓ پرسشگر در چند مصاحبه فراموش کرده است برخی سوالات را بپرسد.
- ✓ اکثر بیماران در مصاحبه حضوری به سوالات مربوط به رفتارهای جنسی نامتعارف پاسخ نداده اند.



# چه می توان کرد؟

## پیشگیری ✓

- آیا پیشگیری همیشه بهتر از درمان است؟
- آیا پیشگیری همیشه امکان پذیر است؟
- آیا پیشگیری می تواند غیر اخلاقی باشد؟



✓ سرمایه اجتماعی، داده های مفقود را کاهش میدهد.

✓ تکنولوژی اطلاعات داده های مفقود را کاهش میدهد.

# چه می توان کرد؟

✓ حذف مشاهدات مفقود از محاسبات آماری



■ حذف موردی

■ حذف لیستی

## حذف موردی

وضعیت بیمه	تعداد روز بستری	رضایت بیمار
۱	۱۰	۱
۱	۳	۲
<del>۱</del>	<del>۴</del>	<del>۲</del>
<del>۲</del>	<del>۵</del>	<del>۰</del>
۲	۶	۲
<del>۱</del>	<del>۳</del>	<del>۲</del>
۱	۱۰	۱

وضعیت بیمه	تعداد روز بستری
۱	۱۰
۱	۳
۲	۶
۱	۱۰

## حذف موردی

وضعیت بیمه	تعداد روز بستری	رضایت بیمار
۱	۱۰	۱
۱	۳	۲
.	۴	۲
۲	.	.
۲	۶	۲
.	۳	۲
۱	۱۰	۱

تعداد روز بستری	رضایت بیمار
۱۰	۱
۳	۲
۴	۲
۶	۲
۳	۲
۱۰	۱

# حذف لیستی

وضعیت بیمه	تعداد روز بستری	رضایت بیمار
۱	۱۰	۱
۱	۳	۲
.	۴	۲
۲	.	.
۲	۶	۲
.	۳	۲
۱	۱۰	۱

وضعیت بیمه	تعداد روز بستری	رضایت بیمار
۱	۱۰	۱
۱	۳	۲
۲	۶	۲
۱	۱۰	۱

### List wise deletion

Gender	Manpower	Sales
M	25	343
<del>F</del>	<del>.</del>	<del>280</del>
M	33	332
<del>M</del>	<del>.</del>	<del>272</del>
<del>F</del>	<del>25</del>	<del>.</del>
M	29	326
<del>.</del>	<del>26</del>	<del>259</del>
M	32	297

### Pair wise deletion

Gender	Manpower	Sales
M	25	343
F	<del>.</del>	280
M	33	332
M	<del>.</del>	272
F	25	<del>.</del>
M	29	326
<del>.</del>	26	259
M	32	297

اگر مفقود شدن داده ها به صورت کاملا تصادفی باشد و حجم  
مشاهدات مفقود پایین باشد ( $> 2\%$ )، روش حذف مشاهدات توصیه  
میشود.



# چه می توان کرد؟

## ✓ جانهای مقادیر مفقود

- روش های مختلف جانهای، بر اساس داده های موجود، تخمینی از داده های مفقود ارائه میدهند.





- ✓ روش های جانپي براي داده هاي دموگرافيك و متغيرهاي كيفي کمتر استفاده ميشود.
- ✓ گروه عمده اي از روش هاي جانپي فقط مناسب داده هاي سري زماني است.
- ✓ گروه عمده اي از روش هاي جانپي فقط براي داده هاي مفقود تصادفي مناسب است.
- ✓ داده هاي جايزين شده توسط برخي روش هاي جانپي فقط براي آماره هاي توصيفي مناسب است.
- ✓ اگر نسبت داده هاي مفقود از كل نمونه بالا ( $> 40\%$ ) باشد، جانپي داده ها توصيه نميشود.



## هشدار

✓ هرگز از میانگین مشاهدات موجود به عنوان جایگزین مشاهدات مفقود استفاده نکنید!

# چه می توان کرد؟

✓ کاربرد روش های تحلیل آماری ویژه داده های مفقود



# چه می توان کرد؟

✓ استفاده از منابع دیگر برای دسترسی به مقادیر مفقود شده



✓ افزایش حجم نمونه

## چه نباید کرد؟

~~No!~~ اگر یک نمونه از مطالعه خارج شد، باید نمونه دیگری جایگزین شود.



~~No!~~ تمام پرسشنامه ها باید به طور کامل پر شوند.

~~No!~~ تمام اطلاعات آزمودنی باید به هر نحو ممکن تکمیل شوند.

## دو روی یک سکه

✓ وجود داده های مفقود می تواند یک نشانه خطا در نمونه گیری باشد.

✓ در نمونه گیری صحیح معمولا همیشه داده های مفقود وجود دارد.



## سخن پایانی:

وجود داده های مفقود یک واقعیت اجتناب ناپذیر است  
اما نباید مورد بی توجهی قرار گیرد.





# Spss

- Inspecting for categorical and metric variable:
  - Discreptive , Frequency, Histogram
- Variable view-missing
- Transform-recode
- System missing
- User missing