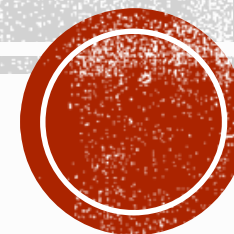


# مقدمه ای بر علم آمار

مقدمه ای بر علم آمار و آزمون های آماری



# آمار چیست؟

- آمار علمی است که به روش گردآوری، خلاصه کردن و طبقه بندی، توصیف و تحلیل داده ها و استفاده از آنها برای پیش بینی، برآورد و تخمین رویدادها می پردازد.
- آمار یکی از مهمترین ابزارهای پردازش داده ها و تصمیم گیری است و در لغت به معنای شمارش و حساب می باشد.
- آمار آمار در اصل علم بررسی داده ها است. داده ها تمامی اطلاعاتی است که ما از محیط مشخصی به صورت علمی و با ابزارهای مشخص بدست می آوریم. آمار علمی است که وظیفه دارد این داده ها را بررسی کند و آن ها را طبقه بندی نماید و گزارش هایی از آن بدست آورد. با علم آمار است، که ما داده ها را بررسی می کنیم و شاخص های مختلفی را محاسبه می کنیم و الگوهایی را از این داده ها و نحوه توزیع آن ها بدست می آوریم. در پایان ما می توانیم از دل این داده ها، پیش بینی هایی را از آینده بدست آوریم.



# نرم افزارهای آماری

## نرم افزارها:

- SPSS
- SAS
- Weka
- Orange
- KNIME
- R
- MATLAB
- Python

■ تجزیه و تحلیل آماری نمی تواند با دست و یا ذهن محقق انجام شود، به خاطر این که عملاً تعداد داده‌ها بسیار زیاد است و زمان زیادی را از آماردان خواهد گرفت، از طرفی آماردان با کوچکترین خطا در محاسباتش ممکن است نتایج کاملاً اشتباهی را بدست آورد. نرم افزارهای آماری برای این طراحی شده اند که آماردانان با استفاده از آن‌ها به راحت ترین شکل ممکن داده‌ها را تحلیل نمایند. نرم افزارها، خیلی سریع و به راحتی می‌توانند جداول آماری و نمودارهایی را طراحی کنند. تعداد زیادی نرم افزار آماری داریم که هر کدام ویژگی‌های خاص به خود را برای انجام تجزیه و تحلیل آماری دارا می‌باشند. برخی از این نرم افزارها شامل موارد فوق است:



آمار استنباطی

آمار توصیفی

## دو شاخه اصلی در آمار:

- متداول‌ترین روش برای جمع‌آوری، سازمان‌دهی و نمایش داده‌هاست. روش‌هایی که با استفاده از آنها، داده‌ها در قالب نمودارها، جداول فراوانی و شاخص‌های عددی خلاصه می‌گردند.
- در این بخش آمار به توصیف و خلاصه‌سازی داده‌ها پرداخته می‌شود، ولی نتیجه‌گیری در مورد داده‌ها و یا تعمیم نتایج به جامعه در این بخش صورت نمی‌گیرد.
- «آمار توصیفی» رشته‌ای است که داده‌های نمونه را با استفاده از میانگین یا انحراف معیار جمع‌بندی می‌کند.



# دو شاخه اصلی در آمار:

آمار توصیفی

آمار استنباطی

- روشهایی که با استفاده از آنها اطلاعات موجود در نمونه با سطح اطمینان مشخصی، به کل جامعه تعمیم داده می شود.
- «آمار استنباطی» بر نتیجه گیری از داده های نمونه تمرکز می کند که ممکن است تغییرات تصادفی را نشان دهد.
- در اصل در این بخش به بررسی فرضیه های پروپوزال پرداخته می شود و نتیجه گیری در مورد داده ها و یا تعمیم نتایج به جامعه صورت می گیرد.



# آمار توصیفی تک متغیره

## Descriptive Statistics

### شاخص های گرایش به مرکز

- میانگین
- میانه
- مد

### شاخص های پراکندگی

- انحراف معیار
- واریانس
- ضریب تغییرات
- دامنه
- چندک ها

- در این بخش به توصیف یک متغیر مشخص در بین متغیرهای تحقیق می پردازیم.
- شاخص های پراکندگی ویژه ی داده های کمی است ولی بیشترین بخش شاخص های گرایش به مرکز ویژه توصیف متغیرهای کیفی است.
- همین طور در بخش آمار توصیفی ما باید به بررسی شکل توزیع داده ها پردازیم. آیا داده های ما دارای توزیع نرمال است یا دارای چوگلی و کشیدگی است؟ این سوال در این بخش بررسی می شود.
- شاخص های کشیدگی در مورد جمع شدن شکل توزیع یا پهن بودن آن می باشد. علاوه بر این موارد ما می توانیم به بررسی توزیع داده ها به صورت ترسیم نمودار هستیوگرام و... نیز پردازیم.



# آمار توصیفی دو متغیره

- در این مدل از تحلیل، توصیفی از رابطه بین دو متغیر مختلف بررسی می گردد. در این مورد داریم:
- جداول توافقی
- نمودارهای پراکنش
- مقادیر کمی وابستگی
- توصیف توزیع های شرطی
- **Descriptions of conditional distributions**



# آمار توصیفی چند متغیره

- در این مدل از تحلیل، توصیفی از رابطه بین چندین متغیر مختلف بررسی می گردد. در این مورد داریم:
- جداول توافقی
- نمودارهای پراکنش





# Inferential Statistics

# آمار استنباطی

■ در آمار استنباطی به بررسی فرضیات پروپوزال پرداخته می شود. آزمون ها و روش های مورد استفاده در تحلیل براساس اهداف مطالعه و نحوه جمع آوری داده ها و ... مشخص می گردد.

■ از جمله ویژگی های اصلی آمار تحلیلی یا استنباطی امکان تعمیم نتایج نمونه به جامعه است.

■ در این بخش دو نوع رویکرد تحلیلی داریم:

■ تحلیل آماری پارامتری **Parametric**

■ تحلیل آماری ناپارامتری **non-parametric**



# تجزیه و تحلیل آماری داده‌ها

■ در حوزه تجزیه و تحلیل داده‌ها توزیع جامعه آماری که نمونه از آن گرفته شده، مهم است زیرا هر چه اطلاعات بیشتر در زمینه رفتار داده‌ها و شکل پراکندگی و توزیع آن‌ها وجود داشته باشد، نتایج قابل اعتمادتر و دقیق‌تر خواهند بود. در مقابل، وجود اطلاعات کم از توزیع جامعه آماری مربوط به نمونه، باعث کاهش اعتماد به نتایج حاصل از روش‌های معمول (پارامتری) آماری می‌شود. بنابراین در این حالت مجبور به استفاده از روش‌های ناپارامتری هستیم که برای اجرای آن‌ها فرضیاتی در مورد توزیع داده‌ها وجود ندارد. به همین علت به روش‌های ناپارامتری گاهی روش‌های توزیع-آزاد نیز می‌گویند.



# آزمون های پارامتریک

نمونه ای از آزمون های پارامتریک:

✓ آزمون تی-مستقل

✓ آزمون تی-وابسته

✓ تحلیل واریانس

✓ تحلیل واریانس چندعاملی

✓ تحلیل کوواریانس

■ زمانی که می خواهیم از یک آزمون مشخص برای تجزیه و تحلیل داده های خود استفاده کنیم، ابتدا باید فرض های آماری آن آزمون را مطالعه نماییم. هر آزمون فرض های خاص خود را دارد و اگر رعایت شود، از آزمون های پارامتری استفاده می شود.

■ این آزمون ها به تجزیه و تحلیل داده هایی که در سطح مقیاس فاصله ای و نسبی سنجیده شده اند، می پردازند.



# آزمون های پارامتریک

- داده های پارامتری به نمونه ای گفته می شود که از توزیع جامعه آماری آن مطلع هستیم. معمولاً این توزیع آماری برای داده های کمی، نرمال یک یا چند متغیره در نظر گرفته می شود. در این حالت از آزمون های آماری پارامتری مثل آزمون تی، آزمون  $F$  و یا آزمون  $Z$  استفاده می کنیم.
- همچنین برای اندازه گیری میزان همبستگی بین متغیرهای دو یا چند بعدی نیز از ضریب همبستگی پیرسون استفاده خواهیم کرد.
- اگر حجم نمونه در روش های تجزیه و تحلیل آمار پارامتری بزرگ انتخاب شود، معمولاً توان آزمون مناسب خواهد بود و به راحتی می توان نتایج حاصل از آزمون فرض را به جامعه نسبت داد.



# آزمون‌های ناپارامتریک

نمونه ای از آزمون های ناپارامتریک:

- ✓ ویلکا کسون
- ✓ یومن-ویتنی
- ✓ کروسکال-والیس
- ✓ فریدمن
- ✓ مک نمار
- ✓ آزمون Q کوکران
- ✓ ضریب همبستگی اسپیرمن

■ اگر فرض های آزمون مشخصی رعایت نشده باشد و ویژگی های لازم را برای اجرای آزمون پارامتریک نداشته باشد، ما باید از جایگزین آن آزمون در بخش آزمون های ناپارامتری، برای تجزیه و تحلیل داده هایمان استفاده کنیم.

■ این آزمون ها بیشتر برای داده هایی به کار می رود که در سطح مقیاس اسمی و رتبه‌ای سنجیده شده باشند.



# آزمون‌های ناپارامتریک

- اگر توزیع جامعه آماری نامشخص باشد و از طرفی حجم نمونه نیز کوچک باشد بطوری که نتوان از قضیه حد مرکزی برای تعیین توزیع حدی یا مجانبی جامعه آماری، استفاده کرد، از تحلیل‌های ناپارامتری استفاده می‌شود، زیرا در این حالت کارآمدتر از روش‌های پارامتری هستند.
- به این ترتیب در زمانی که توزیع جامعه مشخص نباشد و یا حجم نمونه کم باشد، روش‌ها و آزمون‌های ناپارامتری نسبت به روش‌ها و آزمون‌های پارامتری از توان آزمون بیشتری برخوردارند و نسبت به آن‌ها ارجح هستند.



# آزمون‌های ناپارامتریک

- روش‌های ناپارامتری می‌تواند راهگشا باشد و به محقق و تحلیل‌گر داده برای شناخت داده‌ها یاری برساند، اگر:
- برای داده‌ها، نتوان توزیع آماری مناسبی در نظر گرفت.
- وجود داده‌های پرت، وجود چند نما و ... امکان انتخاب توزیع نرمال را برایشان میسر نکند.
- کم بودن حجم نمونه برآورد پارامترهای توزیع نرمال مانند میانگین و بخصوص واریانس را دچار مشکل کند و در عمل امکان بررسی توزیع نرمال به علت حجم کم نمونه برای جامعه وجود نداشته باشد.



# آزمون‌های ناپارامتریک

■ گاهی در پژوهش‌ها، داده‌هایی گردآوری می‌شوند که دارای مقیاس اسمی یا رتبه‌ای می‌باشند. همچنین ممکن است داده‌ها دارای مقیاس فاصله‌ای باشند ولی توزیع داده‌ها نرمال نیست. در چنین مواردی پژوهشگر ملزم به استفاده از آزمون‌های ناپارامتریک است. این آزمون‌ها در کلیه مواردی که پژوهشگر نمی‌تواند از آزمون‌های پارامتریک استفاده کنید ابزار مناسبی برای آزمون فرضیه‌ها هستند. بطور کلی می‌توان گفت که این آزمون‌ها در مورد داده‌هایی بکار می‌روند که:

■ ۱- مقیاس اندازه‌گیری آنها اسمی یا رتبه‌ای باشد (از نوع داده‌های ناپیوسته و یا منفصل و بنابراین حاصل شمارش هستند).

■ ۲- بر نرمال بودن توزیع در جامعه استوار نیستند.





# چرا همیشه از روش‌های ناپارامتریک استفاده نمی‌کنیم؟

■ اگر روش‌های ناپارامتریک همیشه قابل استفاده است و برای داده‌های نرمال نیز می‌توان از آن استفاده کرد پس چرا همیشه از روش‌های ناپارامتریک استفاده نکنیم؟

■ علت اصلی آن است که روش‌های پارامتریک نتایج دقیق‌تر و درست‌تری را ارائه می‌کنند بنابراین اگر شرایط مهیا بود و داده‌ها نرمال بودند بهتر است از روش‌های پارامتریک استفاده شود.

**نکته:** باید توجه داشت که اگر توزیع جامعه آماری قابل تحقیق و تعیین باشد، اجرای روش‌های پارامتری بر روش‌های ناپارامتری ارجح هستند زیرا روش‌های پارامتری نسبت به روش‌های ناپارامتری از دقت بیشتری برخوردارند. بنابراین فقط زمانی که از توزیع جامعه آماری مطلع نیستیم، به اجبار از روش‌های ناپارامتری استفاده می‌کنیم. البته اگر حجم نمونه بزرگ باشد، در اکثر موارد، نتایج حاصل از آزمون‌های پارامتری و ناپارامتری با یکدیگر همخوانی دارند.



# آزمون پارامتریک و ناپارامتریک

الف) آزمون‌های میانگین جامعه

یک میانگین از یک جامعه

مثال: بررسی رضایت دانشجویان از مدرس

روش پارامتریک: آزمون تی- تک نمونه

روش ناپارامتریک: آزمون علامت تک نمونه و

آزمون دوجمله‌ای (نسبت موفقیت)

دو میانگین از یک جامعه

مثال: بررسی اختلاف میانگین رضایت دانشجویان پیش و پس از آزمون نهایی از مدرس

روش پارامتریک: آزمون تی-زوجی

روش ناپارامتریک: آزمون علامت زوجی و آزمون

ویلکاکسون



# آزمون پارامتریک و ناپارامتریک

یک میانگین از دو جامعه

مثال: بررسی اختلاف میانگین رضایت دانشجویان دختر و پسر از مدرس

روش پارامتریک: آزمون تی- مستقل

روش ناپارامتریک: آزمون یو-من-ویتنی

یک میانگین از چند جامعه

مثال: بررسی اختلاف میانگین رضایت دانشجویان رشته‌های مختلف از مدرس

روش پارامتریک: آزمون آنالیز واریانس ANOVA

روش ناپارامتریک: آزمون H کروسکال-والیس



# آزمون پارامتریک و ناپارامتریک

(ب) آزمون‌های همبستگی

انواع تحقیق همبستگی و رگرسیون دومین دسته از روش‌های پرکاربرد آمار می‌باشند.

همبستگی ساده

مثال: بررسی رابطه میان اعتماد، رضایت و وفاداری دانشجویان به مدرس

روش پارامتریک: آزمون همبستگی پیرسون

روش ناپارامتریک: آزمون همبستگی اسپیرمن و آزمون همبستگی تاو کندال



# آزمون پارامتریک و ناپارامتریک

## رگرسیون

مثال: بررسی تاثیر رضایت بر اعتماد و وفاداری دانشجویان به مدرس

روش پارامتریک: رگرسیون خطی یا چندگانه

روش ناپارامتریک: رگرسیون ناپارامتری

## مدل معادلات ساختاری

مثال: بررسی تاثیر رضایت بر اعتماد و وفاداری دانشجویان به مدرس

روش پارامتریک: مدل معادلات ساختاری

روش ناپارامتریک: حداقل مربعات جزئی



# آزمون های آماری

## پارامتریک :

- هنگامی که توزیع جامعه نرمال است.
- برای متغیرهایی از نوع فاصله ای و نسبی.
- حداقل شاخص آماری: میانگین و واریانس

## ناپارامتریک :

- هنگامی که توزیع جامعه نرمال نیست.
- برای متغیرهایی از نوع اسمی و ترتیبی
- حداقل شاخص آماری: میانه و مد.

Managerial

- T تک نمونه ای
- T وابسته
- T دو نمونه مستقل
- T ولج
- T هتلینگ
- تحلیل واریانس (F/ANOVA)
- تحلیل واریانس چند عاملی
- تحلیل کواریانس چند عاملی
- ضریب همبستگی گشتاوری پیرسون
- علامت تک نمونه ای
- علامت زوجی
- ویلکاکسون (علامت)
- مان - ویتنی (آزمون U)
- کروسکال - والیس (آزمون H)
- فریدمن
- نیکوئی برزش (خی دو)
- کولموگروف - اسمیرتف
- تقارن توزیع
- میانه
- مک شمار
- آزمون کوکران
- ضریب همبستگی اسپیرمن

- برای آزمون فرض پیرامون میانگین یک جامعه استفاده می شود.
- برای آزمون فرض پیرامون دو میانگین از یک جامعه استفاده می شود.
- جهت مقایسه میانگین دو جامعه استفاده می شود. با فرض برابر بودن واریانس دو جامعه.
- جهت مقایسه میانگین دو جامعه استفاده می شود. با فرض عدم برابری واریانس دو جامعه.
- برای مقایسه چند میانگین از دو جامعه استفاده می شود. (مقایسه دو جامعه بر اساس میانگین چندین صفت)
- از این آزمون به منظور بررسی اختلاف میانگین چند جامعه آماری استفاده می شود.
- از این آزمون (MANOVA) به منظور بررسی اختلاف چند میانگین از چند جامعه آماری استفاده می شود.
- چنانچه در MANOVA بخواهیم اثر یک یا چند متغیر کمکی را حذف کنیم، استفاده می شود.
- برای محاسبه همبستگی (شدت و جهت ارتباط دو متغیر) دو مجموعه داده استفاده می شود.
- برای آزمون فرض پیرامون میانگین یک جامعه استفاده می شود.
- برای آزمون فرض پیرامون دو میانگین از یک جامعه استفاده می شود.
- همان آزمون علامت زوجی است که در آن اختلاف نسبی تفاوت از میانگین لحاظ می شود.
- جهت مقایسه میانگین دو جامعه استفاده می شود.
- از این آزمون به منظور بررسی اختلاف میانگین چند جامعه آماری استفاده می شود.
- جهت آزمون متغیرهایی با معیاس رتبه ای در k گروه وابسته و بررسی تفاوت آن در گروه ها.
- برای ارزیابی هموارگی متغیرهای اسمی به کار می رود. (مقایسه یک توزیع نظری با توزیع مشاهده شده)
- مقایسه یک توزیع نظری با توزیع مشاهده شده و نشان دادن نرمال بودن توزیع داده ها نوعی آزمون نیکوئی برزش)
- برای بررسی مشاهدات زوجی درباره متغیرهای دو ارزشی استفاده می شود.
- جهت مقایسه میانه دو جامعه استفاده می شود و برای k جامعه نیز قابل تعمیم است.
- برای بررسی مشاهدات زوجی درباره متغیرهای دو ارزشی استفاده می شود.
- تعمیم آزمون مک شمار در k نمونه وابسته است.
- برای محاسبه همبستگی دو مجموعه داده که به صورت ترتیبی قرار دارند استفاده می شود.

شکل ۱- رده بندی آزمون های آماری

# آزمون های پارامتریک

- **آزمون تی-تک نمونه:** برای آزمون فرض پیرامون میانگین یک جامعه استفاده میشود. در بیشتر پژوهش هایی که با مقیاس لیکرت انجام میشوند، جهت بررسی فرضیه های پژوهش و تحلیل سوالات تخصصی مربوط به آنها از این آزمون استفاده میشود.
- **آزمون تی-وابسته:** برای آزمون فرض پیرامون دو میانگین از یک جامعه استفاده میشود. برای مثال اختلاف میانگین رضایت کارکنان یک دانشگاه قبل و بعد از تغییر مدیریت یا زمانی که نمرات یک کلاس با پیش آزمون و پس آزمون سنجش می شود.
- **آزمون تی-دو نمونه مستقل:** جهت مقایسه میانگین دو جامعه استفاده می شود. در این آزمون برای دو نمونه مستقل فرض میشود واریانس دو جامعه برابر است. برای نمونه به منظور بررسی معنی دار بودن تفاوت میانگین نمره نظرات پاسخ دهندگان بر اساس جنسیت در خصوص هر یک از فرضیه های پژوهش استفاده میشود.



# آزمون های پارامتریک

- **آزمون تی-ولچ:** این آزمون نیز مانند آزمون تی-دو نمونه جهت مقایسه میانگین دو جامعه استفاده میشود. در آزمون تی-ولچ فرض می شود واریانس دو جامعه برابر نیست. برای نمونه به منظور بررسی معنی دار بودن تفاوت میانگین نمره نظرات پاسخ دهندگان بر اساس جنسیت در خصوص هر یک از فرضیه های پژوهش استفاده میشود.
- **آزمون تی-هتلینگ:** برای مقایسه چند میانگین از دو جامعه استفاده میشود. یعنی دو جامعه براساس میانگین چندین صفت مقایسه شوند.
- **تحلیل واریانس:** از این آزمون به منظور بررسی اختلاف میانگین چند جامعه آماری استفاده میشود. برای نمونه جهت بررسی معنی دار بودن تفاوت میانگین نمره نظرات پاسخ دهندگان بر اساس سن یا تحصیلات در خصوص هر یک از فرضیه های پژوهش استفاده میشود.
- **تحلیل واریانس چندعاملی:** از این آزمون به منظور بررسی اختلاف چند میانگین از چند جامعه آماری استفاده میشود.
- **تحلیل کوواریانس چندعاملی:** چنانچه در تحلیل واریانس چندعاملی بخواهیم اثر یک یا چند متغیر کمکی را حذف کنیم استفاده میشود.





# آزمون های ناپارامتریک

- **آزمون علامت تک نمونه:** برای آزمون فرض پیرامون میانگین یک جامعه استفاده میشود.
- **آزمون علامت زوجی:** برای آزمون فرض پیرامون دو میانگین از یک جامعه استفاده میشود.
- **ویلکاکسون:** همان آزمون علامت زوجی است که در آن اختلاف نسبی تفاوت از میانگین لحاظ میشود.
- **من-ویتنی:** به آزمون یو نیز موسوم است و جهت مقایسه میانگین دو جامعه استفاده میشود.
- **کروسکال-والیس:** از این آزمون به منظور بررسی اختلاف میانگین چند جامعه آماری استفاده می شود . این آزمون نیز موسوم است و تعمیم آزمون یو من ویتنی می باشد. آزمون کروسکال-والیس معادل روش پارامتریک آنالیز واریانس تک عاملی است.



# آزمون های ناپارامتریک

▪ **فریدمن:** این آزمون معادل روش پارامتریک آنالیز واریانس دو عاملی است که در آن  $k$  تیمار به صورت تصادفی به  $n$  بلوک تخصیص داده شده اند.

▪ **کولموگروف-اسمیرنف:** نوعی آزمون نیکوئی برازش برای مقایسه یک توزیع نظری با توزیع مشاهده شده است.

▪ **آزمون میانه:** جهت مقایسه میانه دو جامعه استفاده می شود و برای  $k$  جامعه نیز قابل تعمیم است.

▪ **مک نمار:** برای بررسی مشاهدات زوجی درباره متغیرهای دو ارزشی استفاده میشود.

▪ **آزمون Q کوکران:** تعمیم آزمون مک نمار در  $k$  نمونه وابسته است.

▪ **ضریب همبستگی اسپیرمن:** برای محاسبه همبستگی دو مجموعه داده که به صورت ترتیبی قرار دارند استفاده میشود.



امیدواریم مطالب ارائه شده مفید واقع گردد.

مدیریت آمار و فناوری اطلاعات دانشگاه علوم پزشکی تبریز

