آمار در ایمونولوژی

فاطمه السادات حسینی دانشجوی کارشناسی ارشد ورودی ۹۱

بهار۹۲

فهرست مطالب

مقدمه ۲
جنبه‌های خاصی از مطالعات ایمنی در ارتباط با آنالیز آماری ۲
ساختار داده‌های ایمنی ۲
پیچیدگی روابط در پارامترهای ایمنی ۲
تکرارپذیری سنجش پارامترهای ایمنی ۳
آزمایش چندگانه ۳
اهداف پژوهش در مطالعات ایمنی ۳
الگوهای بررسی بدون توجه به روابط علت و معلول ۳
الگوهای بررسی با توجه به روابط علت و معلول ٤
بررسی الگوهایی با هدف‌های چندگانه ٤
سیلیکوایمونولوژی ٤
اصول روش‌های آماری برای آنالیز دادهای ایمونولوژیکی ٤
آنالیز توصیفی داده ها ٥
فرضیات داده ٥
اندازه نمونه ٦
انتخاب روش آماری مناسب برای هر سوال تحقیقی خاص ٦
روش‌های درون وابسته‌ای ۷
آنالیز همبستگی ۷
روش‌های کاهش داده ۷
آنالیز خوشه ای ۷
روش‌های وابستگی آماری ۸
روش وابستگی تک متغیری ۸
روش‌های وابستگی چند متغیری ۸
روش‌های طبقه بندی ۸
روش‌های مقایسه میانگین در گروه‌های چند متغیره ۹
روش‌های رگرسیون چند گانه ۹
روش‌های پیشرفته ۹
آنالیز مسیر و مدل سازی معادله ساختاری(SEM) ۹
مدل‌های تاثیرات ترکیبی ۱۰
نتیجه گیری ۱۲
منابع ۱۲

سمینار دانشجویی

مقدمه

تعداد پارامترهای ایمنی که می‌توانند مورد ارزیابی قرار گیرند به سرعت در حال افزایش است و روز به روز به تعداد آنها افزوده می‌شود. برای تجزیه و تحلیل این داده‌ها نیاز به روش‌های آماری است. این سمینار ارائه دهنده یک نمای کلی از روش‌های مورد استفاده در تجزیه و تحلیل داده‌های آزمایشات ایمونولوژیکی است. در اینجا خلاصه‌ای از روش‌های آماری دو و چند متغیره (تجزیه و تحلیل عاملی، خوشه‌ای و تفکیک تجزیه و تحلیل) و روش‌های پیشرفته‌تر (مسیر تجزیه و تحلیل و مدل سازی معادلات ساختاری) و نشان دادن کاربرد آنها در مطالعات ایمونولوژی است.

اهمیت مکانیسم‌های ایمنی در بیماری‌ها در سال‌های اخیر مورد توجه قرار گرفته است و مطالعات زیادی پیرامون نقش سیتوکاین‌ها ودیگر پارامترهای ایمنی به خصوص در ۱۰ سال اخیر انجام شده است و همچنان نیز رو به افزایش است. تحولات فنی در پردازش نمونه و تکنیک‌های ایمنی پیچیده بستری را برای تجزیه و تحلیل پارامترهای ایمونولوژیک و ارتباط آنها با بیماری و محیط فراهم نموده است.

ایمونولوژیست‌ها تمایل دارند از روش‌های ساده آماری استفاده کنند حتی زمانی که چندین ارتباط میان پارامترها وجود دارد. روش‌های آماری چند متغیره روش هایی هستند که به طور گسترده در دیگر زمینه‌های علمی و مقالات به طور کامل توضیح داده شده‌اند ولی متاسفانه به طور کامل و با جزئیات در اختیار ایمونولوژیست‌ها قرار نگرفته‌اند.

جنبه‌های خاصی از مطالعات ایمنی در ارتباط با آنالیز آماری

ساختار داده‌های ایمنی

قبل از تجزیه و تحلیل داده‌های ایمنی، این موضوع بسیار مهم است که به چگونگی ساختار داده‌ها بپردازیم چرا که روش‌های آماری زمانی به جواب درست می‌رسند که داده‌ها دارای ویژگی مورد نیاز برای استفاده از آن روش آماری را داشته باشند. به طور مثال داده‌ها باید حتی الامکان دارای توزیع نرمال و یا واریانس‌های مشابه در جمعیت‌های مختلف باشند. متاسفانه پارامترهای ایمنی اغلب دارای این پیش فرض‌ها نیستند و محققین مجبورند یا داده‌ها را تغییر دهند (مانند تغییر لگاریتمی به داده‌هایی که دارای توزیع نرمال تقریبا متقارنی هستند) و یا از روش‌های آماری دیگری استفاده کنند که از داده‌هایی با دقت کمتر استفاده می‌کند. (مانند استفاده از روش‌های غیر پارامتریک که به داده‌های دارای توزیع نرمال احتیاج ندارد).

بیشتر پارامترهای ایمنی چندگانه و در عین حال به هم وابسته‌اند. از این رو استفاده از روش‌های آماری که دارای فرضیاتی براساس وابسته نبودن داده‌ها به یکدیگر هستند در مطالعات ایمنی استفاده نمی شوند. و باید روشی استفاده شود که این حقیقت را مورد توجه قرار دهد که مکانیسم‌های بیولوژیکی به هم وابسته‌اند. مکانیسم‌های بیولوژیکی مکانیسم‌هایی هستند که به طور مستقیم قابل مشاهده نیستند ولی دارای اثراتی هستند که قابل اندازه‌گیری‌اند مانند بلوغ سلولی و یا تنظیم سلولی.

پیچیدگی روابط در پارامترهای ایمنی

پارامترهای ایمونولوژیکی معمولا شامل چندین مکانیسم ایمنی به هم مرتبط و متغیراند که ممکن است دارای اثرات متفاوتی در سلول‌های متفاوت باشند (مانند سیتوکاین). هدف ما اغلب توضیح عامل ایجاد کننده یک مکانیسم (ماانند آلرژن) و نتیجه آغاز آن مکانیسم (مانند آتوپی) است. واضح است که تجزیه و تحلیل یک متغیر کار آسانی است ولی این تحلیل آماری قادر نیست که روابط بین مکانیسم‌ها و اثرات آن‌ها را بر یکدیگر مشخص کند. در نتیجه به تکنیک‌های آماری چند متغیره نیاز است تا بتواند به طور همزمان چندین مکانیسم مرتبط با هم را آنالیز کند.

گام اساسی در یک آنالیز آماری ارائه یک فرضیه مشخص در یک چارچوب مفهومی است. چارچوب مفهومی بیان کننده کامل همه ارتباطات درونی میان چندین واکنش و تاثیر متغیرهای مختلف بر واکنش هاست. بدیهی است که چارچوب مفهومی باید کاملا دقیق و واضح باشد تا بتواند ما را به یک نتیجه درست راهنمایی کند. دو جنبه از داده‌های ایمنی عبارتند از:

تکرارپذیری سنجش پارامترهای ایمنی

تکرارپذیری یعنی اینکه با انجام آزمایشات پی در پی به نتایج یکسان برسیم. البته برخی تغییرات در روند آزمایش و آنالیز آماری قابل انتظار است که باید مورد توجه قرار گیرد. با این حال تکرارپذیری تنها برای برخی از پارامترهای ایمنی تعریف شده است مانند اندازه‌گیری آنتی‌بادی علیه Hbs در کسانی که واکسن زده‌اند و یا تعیین نسبت CD4/CD8 در بیماران مبتلا به ایدز. ولی بسیاری از پارامترهای ایمنی قابلیت تکرار پذیری را ندارند. اما مهم قابلیت تکرار بین مراکز مختلف در مطالعات چند مرکزی می‌باشد. یعنی پارامترهای ایمونولوژیکی مشابه، در آزمایشگاه‌های مختلف بررسی شوند.

آزمایش چندگانه

استفاده از آزمایش چندگانه در مطالعات ایمونولوژیکی رو به افزایش است. آزمایش چندگانه به مفهوم این است که بر روی یک سری از داده‌ها آزمون‌های آماری مختلفی انجام دهیم و با مقایسه نتایج به یک نتیجه درست دست یابیم. در آزمون‌های آماری بسیار مهم است که نتایج به دست آمده از روی شانس را از نتایجی که مربوط به واکنش‌های واقعی است تفکیک کنیم. محققان برای هر چه کمتر کردن شانس از سطح معنی داری استفاده می‌کنند (میزان α میزان شانس را نشان می‌دهد). هدف از آزمایش چندگانه کاهش دادن خطای نوع یک است. همان طور که می‌دانیم خطای نوع یک به معنای آن است ما یک واقعیت درست را به غلط اشتباه تلقی کنیم. با انجام آزمایشات چندگانه ما این خطا را کاهش می‌دهیم. در آزمون‌های آماری هرچه قدر که تعداد فرضیات بیشتر باشد احتمال رخ دادن خطای نوع یک بیشتر است و در مقابل هر قدر همبستگی بین داده‌ها بیشتر باشد احتمال رخ دادن خطای نوع یک کمتر می‌شود. کار آزمایشات چندگانه همین ایجاد همبستگی بین داده‌هاست تا اینکه ما به طور جداگانه و مستقل بر روی داده‌ها آزمون‌هایی را انجام دهیم.

اهداف پژوهش در مطالعات ایمنی

در این قسمت ما به بیان سوالات معمولی که در رابطه با مطالعات ایمنی مطرح است می‌پردازیم.

اهداف مشترک مطالعات ایمنی را می‌توان به چهار دسته تقسیم بندی کرد:

الگوهای بررسی بدون توجه به روابط علت و معلول

اهدافی که الگوهای مرتبط با چندین پاامتر ایمونولوژیکی را بدون توجه به روابط علت و معلولی بررسی می‌کنند، در نتیجه غیر وابسته به متغیرها هستند. از جمله مطالعاتی که در این دسته قرار می‌گیرند، می‌توان به نمونه‌های زیر اشاره کرد.

الف: بررسی اهمیت ارتباط میان سایتوکاین‌های مختلف و یا اهمیت تعادل میان سطح سایتوکاین‌های مختلف. مانند بررسی ارتباط میان سایتوکاین‌های پیش التهابی با سایتوکاین‌های ضد التهابی (TNFα, IL10) و یا اندازه‌گیری سطح تعادل میان این دو گروه سایتوکاین (TNFα/IL10).

ب: تعیین سایتوکاین‌های بسیار مرتبط با یکدیگر و قرار دادن آنها در یک گروه. مانند واکنش‌های ایمنی مرتبط با Th1 از جمله IFNγ و TNFα که ممکن است واسطه بیماری‌های التهابی باشند در نتیجه بهتر است از روش‌های آنالیز آماری برای کاهش داده‌ها استفاده کرد و یا اینکه سایتوکاین‌های مرتبط با Th1 را جمع آوری نمود تا تعداد متغیرها کاهش یابد.

ج: شناسایی افراد دارای پروفایل‌های مشابه در پارامترهای ایونولوژیکی و قرار دادن آنها در گروه‌هایی که خوشه نامیده می‌شود. به عنوان مثال دسته‌بندی یک گروه از بیماران بر اساس آزمون‌های پوستی به بیماران آتوپیک و غیر آتوپیک. و یا دسته‌بندی بیماران بر اساس نوع آنتی‌بادی افزایش یافته در بدن به بیماران دارای عفونت حاد و یا مزمن.

الگوهای بررسی با توجه به روابط علت و معلول

دومین گروه از اهداف تحقیقاتی به بررسی روابط علت و معلولی بین یک یا چند پارامتر ایمونولوژیکی می‌پردازد مانند ارتباط میان گروهی از سایتوکاین‌ها و بیماری آسم. در این گروه از تحقیقات باید تعریف صحیحی از متغیرهای وابسته، متغیرهای مستقل و متغیرهای معکوس صورت گیرد و یک روش آماری متناسب انتخاب گردد.

اهداف تحقیقاتی احتمالی برای روابط علت و معلولی به صورت زیر است:

الف: شناسایی و مقایسه بروز سایتوکاین‌ها بین دو یا چند گروه از طریق یک آشکار ساز. مانند اینکه تعیین کنیم که واکسن BCG پس از تزریق چه تاثیری بر سطح IFNγ دارد. باید توجه داشت که معمولا الگوهای ایمونولوژیکی در این مطالعات به عنوان متغیرهای وابسته (پیامد) در نظر گرفته می‌شوند.

ب: شناسایی پیامدهای کلینیکی ناشی از افزایش یا کاهش در برخی پارامترهای ایمونولوژیکی. به عنوان نمونه ایمونولوژیست‌ها علاقه دارند که بررسی کنند که آیا افزایش سطح TNFα یک ریسک فاکتور برای بیماری آرتریت روماتوئید هست یا خیر؟ در این گروه از مطالعات پارامترهای ایمونولوژیکی به عنوان ریسک فاکتور و یا متقیر مستقل بررسی می‌شوند.

بررسی الگوهایی با هدف‌های چندگانه

سومین گروه از مطالعات ایمونولوژیکی شامل سوالات تحقیقاتی پیچیده‌تر است که ممکن است دو یا چند هدف را در برگیرد. این سوالات نقش سایتوکاین را بیش از یک ریسک فاکتور بلکه به عنوان عواملی بررسی می‌کنند که تغییر دهنده یک تاثیر یا پیامد هستند. به عنوان مثال بررسی روابط درون سببی بین عفونت‌های اولیه زندگی و سطح بروز گیرنده‌های شناساگر الگو، فعالیت سایتوکاین‌های التهابی و در نهایت رشد و توسعه بیماری التهابی روده.

سیلیکوایمونولوژی

رشته‌ی سیلیکو ایمونولوژی (آنالیز کامپیوتری با استفاده از انفورماتیک یا ایمونوانفورماتیک) یک رشته در حال توسعه و گسترش است که برای چند نوع سوال تحقیقاتی به کار می‌رود. از جمله

- پیش بینی توالی‌های ایمونولوژیکال میکروبی برای تهیه واکسن مناسب
- پیش بینی توالی‌های پروتئین در آنتی‌بادی‌های درمانی
- شناسایی مولکول‌های تنظیم کننده در سیستم ایمنی ذاتی

در این روش مجموعه بزرگی از داده‌ها (مانند توالی ژنوم) توسط نرم افزار خاص آنالیز می‌شوند و به طور مفصل بررسی می شوند. در این آنالیزها از چندین روش آماری چند متغیره استفاده می‌شود (مانند آنالیز خوشه ای). از آنجاییکه این رشته بسیار تخصصی می‌باشد توضیحی بیشتر در مورد آن داده نمی شود.

اصول روش‌های آماری برای آنالیز دادهای ایمونولوژیکی

با تحقیقی در پایگاه اطلاعاتی MEDLINE روش‌های آماری به کار رفته در ارتباط با داده‌های سایتوکاین‌ها بررسی شد. جدول ۱ نشان دهنده نتایج دقیق این بررسی است. گسترده‌ترین روش به کار رفته، روش‌های آماری ساده است که رابطه بین دو متغیر را بررسی می‌کند. ما همواره در پی روش‌های استانداردی برای مقایسه میانگین پارامترهای ایمونولوژیکی هستیم. مانند مقایسه میانگین بین گروه‌های مستقل (مانند T تست، آنالیز واریانس و یا معادلات غیر پارامتریک)، آنالیز ضریب همبستگی دو متغیره (مانند ضریب همبستگی پیرسون یا اسپیرمن) و رگرسیون خطی تک متغیره. این در حالی است که روش‌های چند متغیره در داده‌های سایتوکاین‌ها کمتر استفاده می‌شود.

برخی از تحقیقات از روش‌های دیگری برای آنالیز آماری استفاده می‌کنند. مانند تحلیل عاملی (برای شناسایی گروه‌هایی از پارامترهای ایمونولوژیکی مرتبط با هم) آنالیز خوشه‌ای (برای شناسایی گروهی از افراد با پروفایل‌های ایمونولوژیکی مشابه) و یا روش‌های متمایز کننده مانند رگرسیون لوجیستیکی (برای شناسایی علل و نتایج پروفایل‌های ایمونولوژیکی). همچنین تحقیقات اندکی از روش‌های مدل‌سازی پیشرفته استفاده کرده اند.

در بخش زیر به طور خلاصه به بررسی روش‌های آماری برای آنالیز داده‌های ایمونولوژیکی می‌پردازیم. تعیین انتخاب بهترین روش آماری وابسته به هدف تحقیق، نوع جمع آوری داده‌ها، تحقق یافتن فرضیات داده و کافی بودن اندازه نمونه است. در ابتدا ما به این موضوعات می‌پردازیم.

Statistical methods	**Number of references**
Univariate techniques	
Analysis of variance	2908
T-test	420
Mann-Whitney U-test	316
Wilcoxon/McNemar test	193
Univariate linear regression	163
Bivariate correlation analysis	157
Kruskal-Wallis H-test	95
Repeated measures analysis of variance	31
Friedman test	7
Non linear regression	5
Multivariate techniques	
Logistic regression	629
Cluster analysis	192
Multivariate analysis of variance	144
Multiple linear regression	91
Factor analysis/Principal components analysis	80
Analysis of covariance	56
Linear discriminant analysis	51
Partial correlation coefficient	24
Multinomial logistic regression	9
Multivariate analysis of covariance	7
Path analysis/Structural equation modelling	4

جدول ۱

آنالیز توصیفی داده ها

اولین مرحله مهم در آنالیز داده‌های ایمونولوژیکی شرح داده هاست. سوال تحقیقاتی هرچه باشد محقق در ابتدا باید با دسته‌بندی داده‌ها، میانگین و انحراف معیار داده را مشخص نماید و از چارت‌ها، هیستوگرام‌ها و طرح‌ها پراکندگی استفاده نماید. معمولا ماتریس‌های اسکتر (Scatter Matris) ابزار قدرتمندی برای بررسی رابطه بین چند پارامتر ایمونولوژیکی هستند.

فرضیات داده

نکته بعدی که محققان هنگام انتخاب روش آماری باید در نظر بگیرند این است که آیا داده‌های آنها با فرضیات داده‌ای سازگار است یا خیر. اولین فرضیه مربوط به مقیاس داده‌هاست. مثلا داده‌ها از نوع گروهی هستند (مانند گروه‌های زن و مرد) از نوع ترتیبی اند (مانند تاریخ تولد) و یا از نوع پیوسته‌اند (مانند مقیاس‌های متریک). این امر روش‌های آماری مورد استفاده را محدود می‌کند.

زمانیکه محقق مجبور است از داده‌های پیوسته استفاده کند دومین فرض آن است که آیا داده‌ها از یک توزیع تئوری (توزیع نرمال) استفاده می‌کنند؟ برای پاسخ به این سوال دو راه وجود دارد

۱ – رسم پلات (هیستوگرام)

۲ – بررسی تبعیت کردن داده‌های ما از ا اطلاعات توزیع نرمال

زمانیکه داده‌های ما فرضیات توزیعی مورد نیاز روش آماری خاص را در بر نگیرند (مانند توزیع نرمال برای کاربرد تست T) روش متداول آن است که داده را به نحوی تغییر داد تا فرضیات را در بر بگیرند. به عنوان نمونه از لگاریتم داده‌ها استفاده کنیم. با این وجود داده‌های ایمونولوژیکی حتی پس از اعمال تغییرات نیز از فرضیات داده پیروی نمی کنند. در چنین مواردی بهتر است از روش آماری دیگری استفاده نمود که نیازمند فرضیات داده‌ای کمتر است. مانند به کاربردن تست‌های غیر پارامتری. روش دیگر استفاده از دسته‌بندی سنجش‌ها با به کار بردن مقادیر معنادار بیولوژیکی است (مانند دسته‌بندی میزان قند خون بالای ۱۲۰ به عنوان مثبت و پایین ۱۲۰ به عنوان منفی).

یک روش استراتژیک غیر مرسوم برای استفاده از روش‌های پارامتری حتی زمانیکه فرضیات داده نقض شوند robust resampling variance است. مفهوم این روش آن است که نمونه‌های تصادفی از داده‌هایمان درست کنیم سپس واریانس هر گروه را به تنهایی مشخص کرده و در نهایت میانگین واریانس‌ها را به عنوان واریانس نهایی استفاده کنیم. مثلا در مطالعه‌ای فرضی میزان IL10 را در ۱۰۰ نفر اندازه‌گیری می‌کنیم. سپس این ۱۰۰ داده را در گروه‌های ۵۰ تایی به طور تصادفی دسته‌بندی می‌کنیم (مثلا ۵۰۰ گروه) و در نهایت از میانگین میانگین‌ها و میانگین واریانس‌ها استفاده می‌کنیم.

اندازه نمونه

دومین موضوع مهم که همواره ایمونولوژیست‌ها با آن مواجه هستند و بر روش آماری تاثیر می‌گذارد برآورد اندازه نمونه در آنالیز آماری است. روش‌های تک واریانسی (مانند تستT و ANOVA) اندازه نمونه را با یک آنالیز تعیین توان محاسبه می‌کنند (هرچه توان یک روش بالاتر باشد، حجم نمونه بیشتری می‌خواهد و در نتیجه مورد اعتماد تر است). آنالیز توان کوئن یک روش مفید برای محاسبه اندازه نمونه در روش‌های تک متغیره و چند متغیره است (مانند رگرسیون چندگانه و کو واریانس). متاسفانه تئوری زیر بنایی برای برآورد کردن اندازه نمونه در مورد سایر روش‌ها چند متغیره کمتر توسعه یافته است.

استفاده از روش تحلیل عاملی (خلاصه کردن موضوعات و متغیرهای کوچک مرتبط با هم در یک متغیر بزرگتر) نسبت به ضریب همبستگی (مقایسه متغیرها به صورت دو به دو با یکدیگر) روش مناسب‌تری است. در آنالیز خوشه‌ای، محاسبه اندازه نمونه به چگونگی تصور محقق در مورد جمعیت مورد مطالعه بستگی دارد. پیشنهاد می‌شود که اگر زیر شاخه‌ها کمتر از ۱۰ خوشه باشند، باید حداقل ۵ الی ۱۰ نفر در هر خوشه قرار گیرند.

یک روش آماری برای محاسبه حجم نمونه استفاده از روش‌های پیشرفته نرم‌افزاری مانند مدل‌سازی معادله ساختاری و آنالیز مسیر است.

انتخاب روش آماری مناسب برای هر سوال تحقیقی خاص

در این بخش به بررسی روش‌های آماری مفید برای داده‌های ایمونولوژیکی می‌پردازیم. مهمترین روش‌های تک متغیره و چند متغیره را به طور خلاصه در جدول بیان می‌کنیم و در نهایت فلوچارتی ارائه می‌کنیم که به ایمونولوژیست‌ها کمک می‌کند تا آنالیز آماری مناسب را طبق هدف تحقیق و تعداد و نوع متغیرهای مطالعه انتخاب کنند اکثر نمونه‌ها به داده‌های سایتوکاین‌ها اشاره دارند، اما کاربرد سایر داده‌ها ایمونولوژیکی رایج نیز درست است.

روش‌های درون وابسته‌ای

اولین گروه به عنوان روش‌های درون وابسته‌ای نامیده می‌شوند. یعنی روش‌های آماری با هدف کشف روابط بین متغیرهای مطالعه بدون آنکه هیچ گونه رابطه‌ی علت و معلولی فرض شود. این روش‌ها زمانی مناسب هستند که محقق نمی‌تواند تعیین کند کدام متغیر به عنوان متغیر وابسته است. به عنوان نمونه این گونه روش‌ها ممکن است در مطالعات ایمونولوژیکی برای بررسی روابط بین سایتوکاین‌های مختلف که در یک فرد اندازه‌گیری می‌شوند مفید باشند.

آنالیز همبستگی

یک روش متداول وابسته به هم شامل آنالیز همبستگی دو واریانسی است که هدف آن ارزیابی روابط خطی بین دو متغیر پیوسته است. به عنوان مثال هارتل و همکارانش یک بررسی مشاهداتی متقابل را در کودکان یک ماهه ، ۸ ساله و بزرگسالان انجام دادند تا تغییرات کلی در تولید سایتوکاین را بررسی کنند. رابطه بین سطوح سایتوکاین و سن با استفاده از ضرایب همبستگی غیر پارامتری مورد آنالیز قرار گرفت.

اگر بیش از دو پارامتر ایمونولوژیکی مورد توجه باشد و چند همبستگی دو واریانسی بدست آید از آنالیز همبستگی چند متغیره استفاده می‌شود. تا درجه چند خطی بودن داده‌ها بررسی شود. برای اینکه بدانیم آیا روابط چندگانه بین سه یا چند متغیر مورد مطالعه وجود دارد، یک روش ساده و مفید آن است که آنالیز همبستگی دو واریانسی در سراسر طبقه‌های تعیین شده با سطح سومین متغیر انجام شود. به عنوان نمونه برای بررسی ارتباط میان سایتوکاین‌های مرتبط با Th1 و Th2 می‌توان سطوح بالا و پایین بروز IL10 را بررسی نمود.

روش‌های کاهش داده

پایگاه‌های داده با چندین پارامتر ایمونولوژیکی مرتبط به هم را می‌توان با روش‌های کاهش داده ساده نمود. این روش‌ها به طور خاص زمانی مناسب هستند که فرض شود اکثر متغیرها بازتاب جنبه‌های فرایند اصلی می‌باشند که به طور مستقیم اندازه‌گیری نمی شوند. متداول‌ترین روش آنالیز مولفه اصلی (PCA) است که نوعی خاص از تحلیل عاملی می‌باشد. ایده اصلی این روش آن است که متغیرهای خلاصه (به عنوان مولفه اصلی) ایجاد کنند که اکثر اطلاعات داده اصلی را در برگیرد. به عنوان نمونه می‌توان گروهی از سایتوکاین‌ها را به عنوان سایتوکاین‌های التهابی (به عنوان مولفه اصلی) دسته‌بندی کرد. روش‌های کاهش داده را باید زمانی استفاده کرد که هدف از تحقیق، بررسی نقش هر پارامتر نباشد بلکه نقش مکانیسم اصلی مورد توجه است.

آنالیز خوشه ای

آنالیز خوشه‌ای (CA) روش آماری مناسبی است که محقق می‌تواند افراد (و نه متغیرها) را طبق ارزش آنها گروه بندی کند. CA افراد را به گونه‌ای گروه‌بندی می‌کند که هر گروه از افرادی که در یک خوشه قرار می‌گیرند دارای خصوصیات مشابه با یکدیگر باشند و همچنین خصوصیات متفاوتی با دیگر گروه‌ها داشته باشند. در آنالیز خوشه‌ای دو روش وجود دارد. اولین روش الگوریتم جمعی است که در آن محقق هر مشاهده را به عنوان یک خوشه تلقی می‌کند و افراد مشابه را در خوشه‌ها طبقه بندی می‌کند. در روش دوم یعنی الگوریتم تفکیکی، آنالیز با یک جمعیت کلی به عنوان یک خوشه شروع می‌شود و بعد افراد مشابه شناسایی شده و در یک گروه قرار می‌گیرند. مناسب‌ترین روش خوشه بندی به نوع داده‌های جمع آوری شده و سوال تحقیقاتی بستگی دارد. خوشه بندی جمعی زمانیکه مقادیر داده‌ها بالاست مناسب‌تر است. CA دارای خواص ریاضی قوی است اما مبنای آماری ندارد. فرضیات داده (مانند خطی بودن و نرمال بودن) در آنالیز خوشه‌ای اهمیت کمی دارند.

در مطالعات ایمونولوژیکی ممکن است آنالیز خوشه‌ای برای شناسایی گروه هایی از افراد با الگوهای ایمونولوژیکی مشابه (مانند سطح سایتوکاین یا آنتی‌بادی) به کار رود. به عنوان نمونه موتاپی و همکارانش افراد مبتلا به شیستوزوما مانسونی را براساس

سطح IgG، IgM، IgA و IgE به خوشه‌هایی گروه‌بندی کردند. محققان دو خوشه را تعیین کردند. یک خوشه با سطح بالای IgM و سطوح پایین سایر آنتی‌بادی‌ها و خوشه‌ای با IgM و IgG1 بالا و IgG4 متوسط و سطوح پایین همه آنتی‌بادی‌های دیگر.

روش مفید دیگر برای مطالعه ارتباط بین پارامترهای ایمونولوژیکی، آنالیز همبستگی کانونی (CCA) است. روشی که هدف آن تعیین همبستگی بین دو مجموعه از پیش تعیین شده از متغیرهاست. در جستجو در MEDLINE به کاربرد CCA در زمینه‌های ایمونولوژیکی دست نیافتیم اما پیشنهاد می‌کنیم این روش آماری برای تعیین بزرگی همبستگی بین دو مجموعه مختلف بسیار مفید خواهد بود. نمونه فرضی آن تعیین همبستگی بین سایتوکاین‌های Th1 و Th2 در افراد مبتلا به عفونت های کرمی روده‌ای و افراد سالم است.

روش‌های وابستگی آماری

دومین گروه از روش‌ها شامل روش‌های وابستگی آماری است. این روش‌ها هنگامی مناسبند که رابطه علت و معلولی بین متغیرها حاکم باشد. به عنوان نمونه بلک و همکارانش واکنش IFNγ (متغیر وابسته) را به مایکوباکتریوم توبرکلوزیس قبل و بعد از دریافتBCG (متغیر مستقل) بررسی نمودند. انتخاب روش وابسته آماری که برای یک تحقیق خاص مناسب‌تر است به طرح تحقیق، تعداد و مقیاس دسته‌بندی متغیرها و سایر فرضیات داده بستگی دارد.

روش وابستگی تک متغیری

این روش‌ها زمانی مناسب هستند که فقط یک متغیر مستقل و یک متغیر وابسته وجود دارد. روش‌های متداول شامل روش‌های مقایسه میانگین بین دو گروه متغیر مستقل و وابسته است. تعداد گروه‌ها و مستقل یا وابسته بودن آنها تعیین می‌کند کدام روش مناسب‌ترین روش برای هر موقعیت است.

در مقابل آنالیز رگرسیون تک متغیری به عنوان یکی از بهترین روش‌هاست که محقق رابطه بین دو متغیر پیوسته را مدل سازی می‌کند و می‌تواند تصمیم بگیرد کدامیک متغیر نتیجه است. متداول‌ترین روش رگرسیون خطی است که روشی با فرضیات دقیق می‌باشد. در هنگامیکه فرضیات داده نقض شوند از روش‌های رگرسیون غیر پارامتری و غیر خطی استفاده می‌گردد. نمونه‌های کاربرد آنالیز رگرسیون در مطالعات ایمونولوژیکی، پیش‌بینی سطوح بروز یک متغیر نتیجه پیوسته (مانند سایتوکاین) در متغیر پیوسته‌ای مانند سن است هنگامیکه یک رابطه علت معلولی فرض باشد.

روش‌های وابستگی چند متغیری

روش‌های وابستگی چند متغیری زمانی مورد نیاز است که سه یا چند متغیر درگیر باشند و حداقل یک متغیر را بتوان به عنوان متغیر وابسته دانست.

روش‌های طبقه بندی

این روش‌ها زمانی مورد نیاز هستند که چند متغیر مستقل (مانند سایتوکاین) و یک نتیجه طبقه بندی شده (مانند آتوپی) وجود داشته باشد. یک روش کلاسیک شامل آنالیز متمایز کننده خطی (LDA) است. LDA نیازمند متغیرهای مستقل با توزیع نرمال است.

روش دیگر که به متغیرهای دارای توزیع نرمال نیاز ندارد و شامل متغیرهای مستقل طبقه‌بندی شده است رگرسیون لوجستیکی نامیده می‌شود.

در مطالعات ایمونولوژیکی معمولا روش‌های طبقه‌بندی می‌توانند در شناسایی پروفایل‌های ایمونولوژیکی بسیار مفید باشند که به بهترین نحو بین دو یا چند گروه از پیش تعیین شده تمایز ایجاد کنند. به عنوان نمونه گاما و همکارانش از LDA بدین منظور استفاده کردند تا شاخص ایمونولوژیکی را بر اساس شش سایتوکاین (IL2,IL4,IL10,IL12,IFNγ,TNF-α) شناسایی

کنند و بین شکل‌های کلینیکی و بدون نشانه لشمانیوز تمایز قائل شوند. این محققان دریافتند که TNFα، IL10، IL4 مرتبط با شکل کلینیکی هستند در حالیکه IL2، IL12، IFNγ مرتبط با شکل غیر نشانه‌ای می‌باشند.

روش‌های مقایسه میانگین در گروه‌های چند متغیره

این روش‌ها همانند آنالیز چند متغیره واریانس (ANOVA) یا آنالیز چند مسیری، برای مقایسه توزیع یک یا چند متغیر پیوسته بین گروه‌ها از طریق یک یا چند فاکتور به کار می‌رود. بر خلاف LDA که فرض می‌شود گروه‌ها یک پیامد مشخص را تعریف می‌کنند (مانند آسم یا غیر آسم)، و متغیر مستقل (سایتوکاین) برای تمایز بین گروه‌ها به کار می‌رود در (M)ANOVA گروه‌ها توسط محقق با توجه به یک یا چند متغیر مستقل (مانند درمان و وضعیت واکسیناسیون) تعیین می‌شوند. یک کاربرد بسیار مفید MANOVA در مطالعات ایمونولوژیک آن است که به طور همزمان سطوح دو یا چند سایتوکاین را بین دو گروه مقایسه نمود. با استفاده از MANOVA به جای کاربرد مکرر ANOVA یک محقق می‌تواند از مسئله افزایش خطای نوع I در کل آزمایش مصون بماند.

روش‌های رگرسیون چند گانه

رگرسیون متعدد زمانی مناسب است که سوال تحقیق درباره پیش‌بینی یک متغیر وابسته و مستمر با مجموعه‌ای از متغیرهای مستقل طبقه‌بندی شده و یا پیوسته است. روش استاندارد متداول معمولا رگرسیون خطی چندگانه است اما هنگامیکه فرضیات داده رعایت نشود راهکارهای دیگری (مانند رگرسیون چندگانه غیر خطی و غیر پارامتری) وجود دارد. آنالیز رگرسیون را می‌توان در مطالعات ایمونولوژیکی برای پیش بینی بروز سایتوکاین‌ها (یا سایر پارامترهای ایمونولوژیکی) به عنوان متغیر نشان‌گر یا پیش بینی مقادیر نتیجه (شدت عفونت انگلی) با بروز یک یا چند سایتوکاین به کار برد. به عنوان نمونه دودو و همکارانش از این روش برای پیش‌بینی نتایج مرتبط با مالاریا در سطح سایتوکاین‌های مختلف (IFNγ, TNFα, IL12, IL10, TGFβ) استفاده نمودند.

رگرسیون PLS یک نوع رگرسیون خطی چندگانه برای ایجاد مدل‌های پیش بینی شده است که عوامل بسیار متعدد دارد. این روش برای آنالیز داده‌ها ایمونولوژیکی بسیار مفید است. مثلا زمانیکه هدف پیش بینی یک نتیجه از طریق مجموعه بزرگی از پارامترهای ایمونولوژیکی مرتبط با هم است.

روش‌های پیشرفته

آنالیز مسیر و مدل سازی معادله ساختاری(SEM)

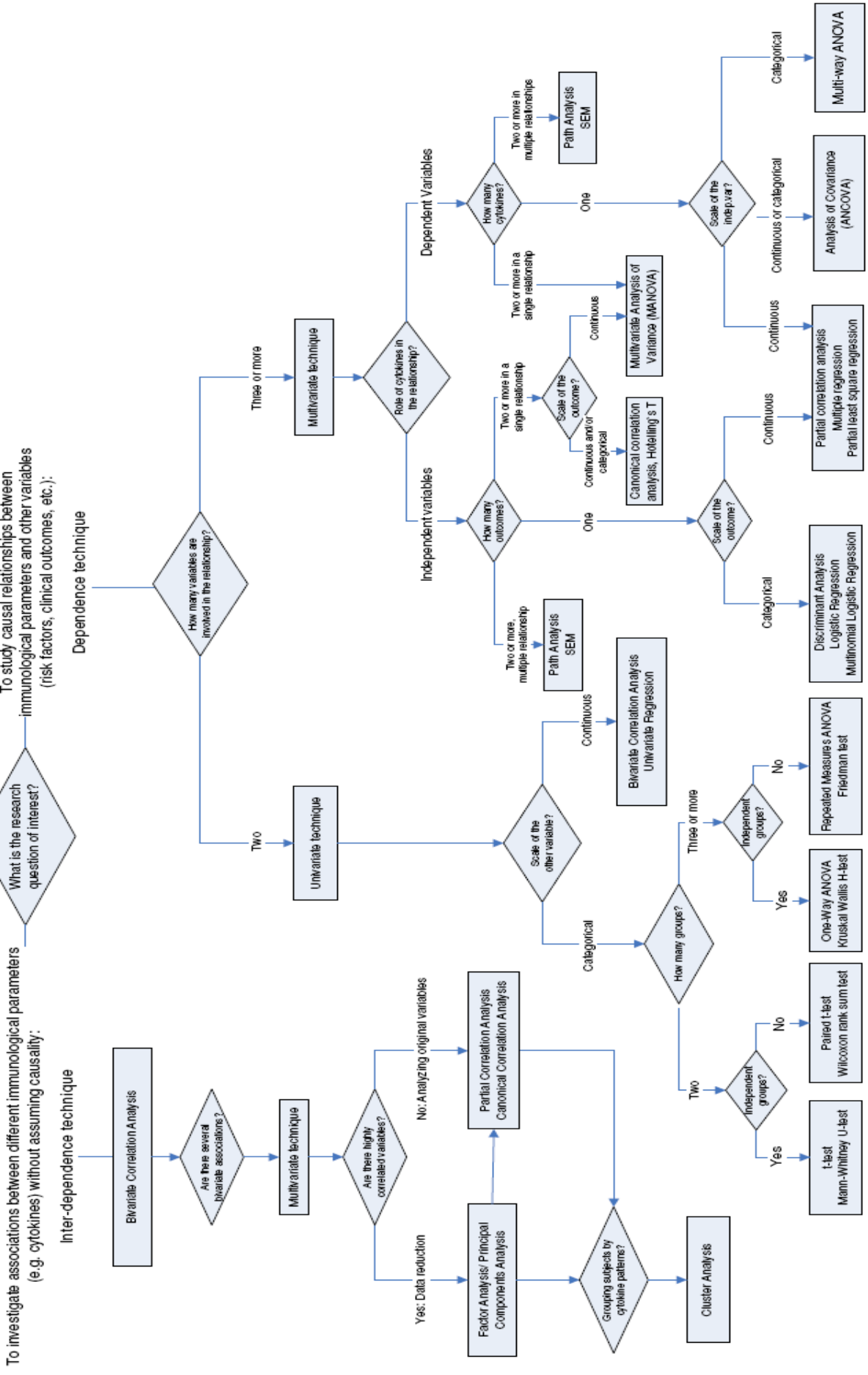
همه روش‌های آماری چند متغیره که گفته شد دارای یک محدودیت مشترک هستند و آن این است که هرچند دارای متغیرهای فراوانی می‌باشند، اما همه آنها وجود یک رابطه را فرض می‌کنند. به هر حال در مطالعات ایمونولوژیکی مدرن محققان اغلب روابط چندگانه بین پارامترهای ایمونولوژیکی و سایر متغیرهای مطالعه را فرض می‌کنند به نحوی که روش چند متغیره ساده نمی تواند به خوبی بازتاب پیچیدگی فرایند ایمونولوژیکی اصلی باشد. به عنوان نمونه در یک مطالعه‌ای که برای بررسی عوامل ریسک آسم و آلرژی انجام شد، محققان یک چارچوب مفهومی را تعریف نمودند که روابط متعدد بین عوامل ریسک (واکسن‌ها، آلرژی‌زاها، عفونت‌های اولیه زندگی)، طرح‌های ایمونولوژیکی (سطوح بروز سایتوکاین) و بروز نتایج و پیامدها (آسم و آتوپی) را مشخص می‌سازد. برای مدل‌سازی این فرایندهای ایمونولوژیکی پیچیده نکته مهم آن است که به طور هم زمان همه این روابط را مدل‌سازی نمود. SEM روش توسعه یافته توسط اقتصاددانان و مهندسان ژنتیک است که چند رابطه بین چند متغیر را به طور همزمان در نظر می‌گیرد. SEM گستره‌ای از آنالیز مسیر است که متغیرهای پنهان نیز نامیده می‌شود. SEM دارای دو مولفه است: مدل اندازه‌گیری که چگونگی ارتباط متغیرهای قابل اندازه‌گیری را با متغیرهای پنهان، تعریف می‌کند و مدل ساختاری که روابط فرضی بین متغیرهای مشاهده شده و یک یا چند متغیر پنهان را تعیین می‌نماید. مفهوم متغیرهای پنهان در مطالعات ایمونولوژیکی مفید است زیرا همواره ایمونولوژیست‌ها فرض می‌کنند که پارامترهای

ایمونولوژیکی اندازه‌گیری شده، نتیجه فرایند اصلی پیچیده‌ای هستند. به عنوان نمونه فرض کنید که سایتوکاین‌های مختلف بدین منظور بررسی می‌شوند تا دو مولفه ایمونولوژیکی تعیین شوند. شکل ۳ نشان دهنده نمودار مسیر مطالعه است که چندین رابطه بین متغیرها تصور می‌گردد. در این نمونه توجه می‌کنیم که سایتوکاین‌های IL12,IFNγ,TNFα نشان‌دهنده متغیر پنهان می‌باشند و واکنش ایمنی مرتبط با Th1 و سایتوکاین های IL4,IL9,IL13 نشان‌دهنده متغیرهای پنهان واکنش ایمنی مرتبط با Th2 هستند. به علاوه در این مدل فرض می‌شود بین دو متغیر پنهان و تاثیرات آنها و متغیر نتیجه (آتوپی یا آسم) نیز رابطه وجود دارد. به دلیل افزایش دانش درباره مکانیزم‌های اصلی ایمونولوژیکی پیچیده باید گفت این روش‌ها نقش مهمی در آنالیز داده‌های ایمونولوژیکی مدرن دارند.

مدل‌های تاثیرات ترکیبی

کاربرد مدل‌های تاثیرات ترکیبی یا تصادفی برای آنالیز آماری داده‌های خوشه‌ای یا طولی یک روش متداول است. ایده اصلی این روش آن است که آنالیز داده‌ها را برای تاثیر خوشه‌بندی از طریق معرفی تاثیر تصادفی به کار برد. در مطالعات ایمونولوژیکی معمولا کاربرد این روش زمانی مفید است که خوشه‌بندی در داده‌ها صورت گیرد. از جمله اندازه‌گیری سایتوکاین های مختلف در بیماران.

سمینار دانشجوئی


To investigate associations between different immunological parameters (e.g. cytokines) without assuming causality:
Inter-dependence technique
To study causal relationships between immunological parameters and other variables (risk factors, clinical outcomes, etc.):
Dependence technique
What is the research question of interest?
Bivariate Correlation Analysis
Are there several bivariate associations?
Multivariate technique
Are there highly correlated variables?
No: Analyzing original variables
Partial Correlation Analysis
Canonical Correlation Analysis
Yes: Data reduction
Factor Analysis/ Principal Components Analysis
Grouping subjects by cytokine patterns?
Cluster Analysis
How many variables are involved in the relationship?
Two
Univariate technique
Scale of the other variable?
Continuous
Bivariate Correlation Analysis
Univariate Regression
Categorical
How many groups?
Two
Independent groups?
No
Paired t-test
Wilcoxon rank sum test
Yes
t-test
Mann-Whitney U-test
Three or more
Independent groups?
No
Repeated Measures ANOVA
Friedman test
Yes
One-Way ANOVA
Kruskal Wallis H-test
Three or more
Multivariate technique
Role of cytokines in the relationship?
Dependent Variables
How many cytokines?
One
Scale of the indep.var?
Continuous or categorical
Analysis of Covariance (ANCOVA)
Categorical
Multi-way ANOVA
Two or more in multiple relationships
Path Analysis SEM
Two or more in a single relationship
Multivariate Analysis of Variance (MANOVA)
Independent variables
How many outcomes?
One
Scale of the outcome?
Continuous
Partial correlation analysis
Multiple regression
Partial least square regression
Categorical
Discriminant Analysis
Logistic Regression
Multinomial Logistic Regression
Two or more in a single relationship
Scale of the outcome?
Continuous
Multivariate Analysis of Variance (MANOVA)
Continuous and/or categorical
Canonical correlation analysis, Hotelling's T
Two or more, multiple relationship
Path Analysis SEM

نتیجه‌گیری

هدف از این سمینار ارائه بررسی فعالیت ایمونولوژیکی مدرن کاربردی در تجزیه و تحلیل داده‌ها بود. این تحقیق به ایمونولوژیست‌هایی که دانش گسترده‌ای درباره آمار ندارند کمک می‌کند تا با مسائل آنالیز آماری داده‌ها آشنا شوند و روش های آماری مناسب را انتخاب کنند که باعث بهترین استنباط از داده‌های گردآوری شده شود. به هر حال چارچوب آماری ارائه شده در این جا به عنوان جایگزینی برای متخصص آمار زیستی با تجربه نیست. در منابع این حقیقت مشخص می‌شود که در اکثر تحقیقات ایمونولوژیکی از روش‌های آماری ساده برای داده‌های ایمونولوژیک استفاده می‌شود حتی زمانیکه چند رابطه بین چند متغیر مطالعه وجود دارد. به نظر ما این موضوع مهم است که از روش‌های آماری پیچیده‌تر برای داده‌های ایمونولوژیکی پیچیده استفاده کرد که باعث درک بهتر مکانیسم شود. تمرکز ما بر روی روش‌های چند متغیری است که باعث آنالیز همزمان چند متغیر می‌شود و امید می‌رود مثال‌های ما از این مطالعات واقعی و فرضی، بتواند ایمونولوژیست‌ها را ترغیب کند که بیشتر از این روش‌ها استفاده کنند. به علاوه به خاطر داشته باشید با توجه به پیچیدگی سوالات تحقیقاتی مطالعات ایمونولوژیکی مدرن یکی از بهترین روش‌ها SEM است که با تهیه یک چارچوب به محقق اجازه می‌دهد چند رابطه را بین چند متغیر به طور همزمان آنالیز کند.

منابع

- Genser B, Cooper PJ, Yazdanbakhsh M, Barreto ML, Rodrigues LC. Guide to modern statistical analysis of immunological data, BMC Immunology, 2007, 26;8:27.
- Donnenberg AD. Statistics of Immunological Testing. In “Handbook of human immunology” edited by M.R.G. O’Gorman, A.D. Donnenberg, 2nd ed., CRC Press, New York, 2008, 29-62.

سمینار دانشجویی