

دانشگاه علوم پزشکی کرمانشاه

۱۳۸۰ / ۹ / ۲۰

دانشکده پزشکی

012079

پایان نامه

جهت اخذ درجه دکتراي حرفه ای پزشکی (MD)

موضوع:

بررسی کیفی استفاده از روشهای آماری مقالات چاپ شده

در مجلات علوم پزشکی ایران سال ۱۳۷۷

استاد راهنما:

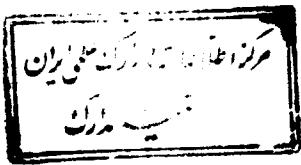
آقای منصور رضایی

نگارش:

بیژن کبودی

اسفند ۱۳۷۸

۳۵۴۷۱



به نام

خداوند بخشنده و مهربان

تقدیم به :

پدر و مادر مهربان، صبور، و دلسوزی که همه
عمر به من درس زندگی آموختند.

و

برادر و خواهران عزیزم بویژه خواهر بزرگم
فریبا راهنمای دلسوز، آگاه و همیشگی من

۱۷۳۴۷۱

تقدیم به اساتید گرامی :

آقایان: دکتر بابک ایزدی، دکتر غلامرضا بهرامی،

دکتر مسعود ناصری پور، و دکتر علی میکائیلی

خانمها: دکتر شهره ملک خسروی و دکتر آذر حدادی

تقدیم به معلمین واقعی من در پژوهش، اخلاق بالینی

و اخلاق انسانی:

دکتر حمید راهی، دکتر محمدرضا عباسی و

دکتر محمدرضا سعیدی

و تقدیم به :

آقای منصور رضایی، استاد راهنمای اینجانب، به

عنوان تشکری کوچک از زحمات ایشان در طول

تحصیلات دانشگاهیم.

رؤیاها

من به دریای خیال خوش خود

سفری خواهم کرد

موجها می آید

قایق کوچک اندیشه من می لرزد

سفرم طوفانی است

لب ساحل اما ، چه هوا آرام است

چه کسی می داند زورق خاطره ام

چه پریشان شده است

در دل یک طوفان ، بیکران تا خود عشق

بیژن کبودی

۱	چکیده فارسی
۲	چکیده انگلیسی
۳	مقدمه
۴	جامعه آماری
۴	اصطلاحات اساسی در آمار توصیفی
۷	اصطلاحات اساسی در نمونه‌گیری
۸	اصطلاحات اساسی در برآورد آماری
۹	اصطلاحات اساسی در آزمون فرض
۱۲	انتخاب آزمون آماری مناسب
۱۵	بیان روشهای آماری در یک مقاله پژوهشی
۱۷	مرور متون
۱۸	اهداف
۱۹	روش کار
۲۱	یافته‌ها
۲۳	بحث
۲۴	نتیجه‌گیری و پیشنهادات
۲۵	جدولها
۲۹	نمودارها
۳۲	تشکرات
۳۳	منابع
۳۷	ضمائم

چکیده

جهت بررسی وضعیت استفاده از روشهای آماری در مجلات تایید شده توسط کمیسیون نشریات علوم پزشکی کشور با روش نمونه گیری تصادفی، نود مقاله مداخله ای یا مشاهده ای تحلیلی که در سال ۱۳۷۷ چاپ شده بودند مورد مطالعه قرار گرفتند. برای هر مقاله نوع تحقیق بیان شده، نوع تحقیق بر اساس مطالعه روش کار، انواع آزمون آماری ذکر شده در قسمت روش کار، انواع آزمون بکار رفته در بخش نتایج، وجود آمارشناس در لیست نویسندگان یا در بخش تشکرات و برخی از سایر ویژگیهای مهم در ارتباط با روشهای آماری، ثبت شد. به علاوه برای هر مقاله آزمونهای آماری پیشنهادی نیز مشخص شد.

در چهل و نه درصد موارد (۴۴ مقاله) اصلاً به نوع تحقیق اشاره نشده بود و بر اساس بررسی بخش روش کار عملاً شصت و دو درصد مقالات از نوع تجربی یا نیمه تجربی و ۲۴/۴ درصد آنها از نوع مقطعی - تحلیلی بودند و تنها ۳/۲ درصد مطالعات به روش مورد - شاهدهی طراحی شده بودند. از نظر نحوه بیان روشهای آماری، در سی و هشت درصد مقالات یا اصلاً روشهای آماری ذکر نشده بود و یا فقط به آوردن نام چند روش آماری در قسمت روش کار (بدون مشخص بودن محل کاربرد آنها) محدود شده بود در عین حال شصت و دو درصد مقالات نیز از نظر نحوه بیان روشهای آماری در حد خوب یا بسیار خوب قرار گرفتند. از نظر استفاده صحیح از روشهای آماری، در هفتاد و شش درصد مقالات مورد بررسی یا روشهای بکار رفته قابل ارزیابی نبودند و یا حداقل یک خطای واضح در استفاده از آزمونهای آماری وجود داشت (به ترتیب ۲۵ و ۵۱ درصد). در ۱۳ درصد مقالات خطای واضحی وجود نداشت اما اطمینان از کاربردهای صحیح نیز ممکن نبود و تنها در یازده درصد مقالات با اطمینان کامل صحت همه آزمونهای بکار رفته تایید شد. فاصله اطمینان ۹۵ درصد برای نسبت مقالاتی که مطمئناً بطور صحیح از روشهای آماری استفاده کرده بودند بین ۵ تا ۱۷ درصد برآورد شد. در بررسی برخی از ویژگیهای مهم بیان نتایج در مقالات (از نظر آماری) تنها در سه درصد موارد برای آماره های محاسبه شده، فاصله اطمینان ارائه شده بود. همچنین ۹۶ درصد مقالات هیچ توجهی به شدت رابطه بدست آمده نداشتند و حداکثر به ذکر PValue اکتفا کرده بودند و در ۴۶ درصد مقالات حتی مقدار دقیق PValue نیز ذکر نشده بود. از ۵۷ مقاله ای که باید به عوامل مخدوش کننده در تفسیر نتایج توجه می کردند تنها در هفت مورد (۱۲ درصد) به این نکته اشاره شده بود. بیشترین آزمونهای بکار رفته نیز به ترتیب آزمون T-Student، آزمون کای دو و روش آنالیز واریانس بود.

در مجموع در مقالات چاپ شده به زبان فارسی کیفیت استفاده از آزمونهای آماری وضعیت مطلوبی ندارد. به محاسبه فاصله اطمینان و همچنین کنترل متغیرهای مخدوش کننده به اندازه کافی توجه نشده است. به نظر می رسد پژوهشگران ما خیلی کم از روش مورد - شاهدهی استفاده می کنند و بطور عمده روشهای مداخله ای یا مقطعی را بکار می برند.

کلمات کلیدی: روشهای آماری، کیفیت، نشریات علوم پزشکی ایران، مقالات

Abstract

The ninety analytic articles of medical journals publishing at 1377 in Iran that proofed by committee of Iranian journals, were selected in random way to assessment quality of statistical presentation and appropriate or inappropriate statistical application . Study design , type of statistical tests, statistician in authorship recorded for each paper. The 62 percent of papers had experimental or quasiexperimental design. The 38 percent of papers void of statistics or had very limit presentation and 62 percent had good quality presentation . The 68 articles could be classified as inappropriate or incomplete usage of statistical tests. The 95 percent confidence interval for appropriate usage of statistics was 5%-17%. The most commonly used statistical methods were t test, contingency tables and analysis of variance. In conclusion the quality of statistical usage in Iranian medical journals need to be improved.

Keywords: Statistical methods, quality, Iranian medical journals, articles.

مقدمه :

بی شک اساس پیشرفت امروزی علم پزشکی ، تحقیقات وسیع و فراوانی است که نتایج آنها بصورت مقالات علمی در مجلات علوم پزشکی چاپ می شود . علم آمار در اغلب پژوهش های علمی جایگاه بسیار مهمی دارد زیرا در مراحل اصلی پژوهش مانند طراحی مطالعه ، انتخاب نمونه ، تعیین حجم نمونه و از همه مهمتر نتیجه گیری و تفسیر مشاهدات نقش مستقیمی دارد (۱). با وجود این ، مطالعات در کشورهای پیشرفته حاکی از آن است که نه تنها اکثر خوانندگان مقالات بلکه درصد قابل توجهی از محققین درک مناسب و کافی از اصول کاربرد روشهای آماری ندارند و در بررسی های متعدد معلوم شده است در سالهای ۱۹۷۰ تا ۱۹۹۰ بیش از ۸۰-۶۰ درصد مقالات بطور واضح روشهای آماری بکار رفته را بیان نکرده اند و یا این روشها بصورت صحیح بکار نرفته است . از رایجترین خطاها در استفاده از روشهای آماری ، فرض غلط درباره توزیع نرمال یا عدم توجه به توزیع صفت ، استفاده اشتباه از انحراف معیار و خطای معیار ، عدم توجه به تصحیحات لازم هنگام انجام آزمونهای چند گانه ، استفاده از آزمونهای پارامتری در شرایطی که پیش فرضهای پارامتری وجود ندارد ، عدم توجه به تقابل بین متغیرها ، عدم توجه به توان ، استفاده اشتباه از روش رگرسیون و عدم درک صحیح از مفهوم همبستگی می باشد (۷۶). عدم توجه سردبیران مجلات نیز عامل مهمی در بوجود آمدن این شرایط محسوب می شود (۵۱). با توجه به مطالب فوق ، در اغلب مجلات معتبر دنیا هر چند وقت یکبار کیفیت مقالات چاپ شده از ابعاد گوناگون بویژه کاربرد روشهای آماری مورد بررسی قرار می گیرد (۲۰۱ و ۳۰۵ و ۶) و بر اساس نتایج حاصل ، دستورالعملهایی برای نویسندگان و حتی سردبیران منتشر می شود (۴ و ۷ و ۸ و ۹).

با بررسی های به عمل آمده در مجلات علوم پزشکی ایران تاکنون مقاله ای با موضوع نحوه استفاده از روشهای آماری در مقالات چاپ شده ، منتشر نشده است. البته صدری و همکاران اخیراً با بررسی ۷۸ گزارش طرح تحقیقاتی مصوب دانشگاه علوم پزشکی همدان طی ۵ سال گذشته ، به بررسی روش نمونه گیری طرحها پرداخته اند در این بررسی در ۴۵ درصد طرحها روش نمونه گیری اشکال داشته است (۱۰). به همین دلیل تصمیم گرفتیم با انتخاب نمونه ای تصادفی از مقالات چاپ شده در مجلات مورد تأیید کمیسیون نشریات علوم پزشکی کشور ، وضعیت کیفی استفاده از روشهای آماری را در مجلات معتبر علوم پزشکی کشور بررسی نمائیم .

جامعه آماری

جامعه آماری مجموعه ای از همه پدیده هایی است که می خواهیم درباره آنها اطلاعاتی کسب کنیم. معمولاً فقط نسبت کوچکی از جوامع آماری قابل بررسی هستند که آنرا نمونه می نامند. کمیت محاسبه شده در یک نمونه آماره نام دارد و به مقدار همین کمیت در کل جامعه پارامتر می گویند. هدف اصلی از محاسبه آماره، استنباط درباره پارامتر جامعه است یعنی ارائه برآوردی از پارامتر یا آزمودن فرضیه ای درباره مقدار صفت در کل جامعه. اگر در جامعه آماری اندازه همه عناصر مشخص شوند نتیجه حاصل توزیع فراوانی صفت در جامعه است. در بسیاری از موارد توزیع فراوانی صفات در جوامع تقریباً از الگوهای ویژه ای پیروی می کند که آنها را توزیع های نظری می گویند. برای آنکه توزیع یک صفت را در جامعه ارائه دهیم باید از روشهای آمار توصیفی استفاده کنیم و اگر بخواهیم بر اساس آماره محاسبه شده در یک نمونه درباره توزیع یا پارامترهای جامعه استنباطهایی داشته باشیم باید روشهای آمار تحلیلی را بکار ببریم.

اصطلاحات اساسی در آمار توصیفی

هدف از کاربرد آمار توصیفی خلاصه نمودن داده های خام است، به صورتی که بتوان برآحتی درباره وضعیت متغیرهای مورد بررسی در یک پژوهش اطلاعاتی کسب نمود. برای نیل به این هدف در مبحث آمار توصیفی سه روش مهم بکار می رود که عبارتند از جدولها، نمودارها و شاخصهای عددی (۱۱).

الف) جدولها:

بسادگی می توان از جدولها برای خلاصه سازی داده های مربوط به متغیرهای کیفی استفاده نمود. اما در مورد متغیرهای کمی معمولاً لازم است ابتدا صفت مورد نظر گروه بندی شود. رایجترین جدولهای مورد استفاده عبارتند از جدولهای فراوانی و جدولهای دو بعدی. جدولهای فراوانی برای بررسی توزیع یک متغیر بکار می روند. در اولین ستون این جدولها سطوح مختلف متغیر مورد بررسی قرار می گیرد و در ستون دوم تعداد یا فراوانی ساده افرادی که در سطوح مختلف متغیر قرار می گیرند نوشته می شود. در اغلب جدولها ستونهای فراوانی نسبی، درصد فراوانی و فراوانی نسبی تجمعی نیز در این جدولها به چشم می خورد. برای محاسبه فراوانی نسبی هر گروه کافی است فراوانی ساده آن گروه را بر تعداد کل افراد مورد بررسی تقسیم شود. اگر فراوانی نسبی را در عدد صد ضرب کنیم، درصد فراوانی بدست می آید. با جمع فراوانی نسبی هر

متغیر کیفی می توان از نمودارهای میله ای خوشه ای استفاده نمود. برای بررسی توزیع همبستگی بین دو متغیر کمی نیز اغلب از نمودار پراکنش استفاده می شود. برای مقایسه میانگین یک صفت در دو یا چند گروه بهتر است از نمودارهای میله - خطا استفاده نمود.

جدول ۱: توزیع فراوانی گروه های خونی

گروه خونی	f	rf	%f	crf
A	۷۰	۰/۳۵	۳۵	۰/۳۵
B	۵۰	۰/۲۵	۲۵	۰/۶
AB	۲۰	۰/۱	۱۰	۰/۷
O	۶۰	۰/۳	۳۰	۱
جمع	۲۰۰	۱	۱۰۰	-

جدول ۲: توزیع توأم گروه خونی و جنس

جنس	مرد	زن	جمع
گروه خون			
A	۴۰	۳۰	۷۰
B	۳۰	۲۰	۵۰
AB	۱۵	۵	۲۰
O	۳۵	۲۵	۶۰
جمع	۱۲۰	۸۰	۲۰۰

گروه با همه فراوانی های نسبی ما قبل خود، فراوانی نسبی تجمعی محاسبه می شود.

در جدول یک توزیع فراوانی داده های فرضی یک پژوهش خلاصه شده است به کمک رسم جدولهای دو بعدی می توان توزیع توأم دو متغیر را بررسی نمود. کافی است سطوح مختلف متغیر اول در ردیف ها و سطوح مختلف متغیر دوم در ستونها ثبت شود اگر متغیر اول m حالت و متغیر دوم n حالت داشته باشد جدول حاصل دارای $m \times n$ خانه اصلی است. در جدول دو توزیع توأم جنس و گروه خونی در یک جامعه فرضی مشاهده می شود.

(ب) نمودارها:

نمودار، روشی تصویری برای نمایش توزیع داده ها است. برای رسم نمودارها اغلب سطوح مختلف صفت روی محور افقی و فراوانی افراد روی محور عمودی مشخص می گردد. نمودارها انواع بسیار زیادی دارند و هر یک در شرایطی بکار می روند. برای نمایش توزیع فراوانی یک متغیر کیفی میتوان از نمودارهای دایره ای یا میله ای استفاده نمود. برای بررسی توزیع متغیرهای کمی نیز نمودارهای چند گوش و هیستوگرام بکار می روند. باید توجه داشت که در نمودار هیستوگرام سطح زیر منحنی متناسب با فراوانی نسبی داده ها است. برای بررسی توزیع توأم دو یا چند

ج) شاخصهای عددی :

برای آنکه وضعیت توزیع یک صفت را بصورت کمی در جامعه معینی مشخص کنیم معمولاً به دو سری از شاخصهای عددی احتیاج داریم که عبارتند از شاخصهای مرکزی و شاخص های پراکندگی. شاخصهای مرکزی اندازه هایی هستند که مرکز یک توزیع یا متوسط داده ها را مشخص می کنند. میانگین، میانه و نما مهمترین این شاخص ها هستند. میانگین برای داده های کمی قابل محاسبه است و انواع متفاوتی دارد که رایجترین آنها میانگین حسابی است برای محاسبه میانگین حسابی باید همه داده ها را با هم جمع و بر تعداد مشاهدات تقسیم نمود این شاخص را در جامعه با حرف یونانی μ نشان می دهند.

از دیگر انواع میانگین می توان به میانگین هندسی و میانگین توافقی اشاره نمود که هر یک در شرایط ویژه ای کاربرد دارند. معمولاً هر جا فقط نام میانگین ذکر می شود منظور میانگین حسابی است. میانه یک توزیع مقداری از صفت است که نیمی از افراد جامعه زیر آن و نیم دیگر بالای آن هستند. (به عبارتی میانه در یک نمودار هیستوگرام نقطه ای از محور طولها است که اگر از آنجا خط عمودی رسم کنیم مساحت زیر منحنی را نصف کند). نما مقدار مشاهده یا مشاهداتی است که بیشترین فراوانی را دارند. یک توزیع می تواند بدون نما یک نمایی یا چند نمایی باشد.

شاخصهای پراکندگی اندازه هایی هستند که چگونگی پراکندگی داده ها را حول شاخص های مرکزی (معمولاً میانگین) توصیف می کنند. دامنه تغییرات بیانگر تفاوت بین بزرگترین و کوچکترین مشاهده موجود است. مهمترین شاخص پراکندگی واریانس است که عبارت است از متوسط مجذور انحرافات از میانگین داده ها. جذر واریانس را انحراف معیار (σ) گویند که دیمانسیون آن مشابه دیمانسیون داده های اصلی است. فرمول محاسبه واریانس در یک جامعه بصورت زیر است :

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{n}$$

برای مقایسه پراکندگی دو صفت با واحدهای متفاوت، نمی توان از واریانسها استفاده نمود. در این موارد از ضریب تغییرات استفاده می شود که واحد ندارد. فرمول محاسبه ضریب تغییرات بصورت زیر است :

$$CV = \frac{\sigma}{\mu} * 100$$

اصطلاحات اساسی در نمونه گیری

اگر نمونه طوری استخراج شده باشد که شانس استخراج کلیه عناصر در تمام استخراجها ثابت و مساوی باشد و همچنین استخراج کلیه نمونه های ممکن نیز ثابت و مشخص باشد، نمونه حاصل را نمونه تصادفی می گویند. این روش حصول احتمالهای ثابت و مستقل را تضمین می کند. اگر همه نمونه های تصادفی ممکن با حجم برابر را از یک جامعه استخراج کنیم و برای همه آنها آماره معینی را محاسبه کنیم، توزیع نمونه ای آماره مورد نظر را بدست آورده ایم.

شکل هر توزیع بر حسب پارامترهای آن توصیف می شود. یکی از رایجترین انواع توزیع در علم آمار توزیع نرمال است که دارای شکلی متقارن و زنگوله مانند است. برای توصیف یک توزیع نرمال دو پارامتر میانگین و انحراف معیار کافی است. توزیع های نمونه ای نیز دارای پارامترهایی هستند که اغلب وابسته به پارامترهای جامعه مادر (اصلی) هستند. میانگین توزیع نمونه ای امید ریاضی آماره نام دارد و انحراف معیار توزیع نمونه ای را خطای معیار آماره می نامند. شکل توزیع نمونه ای و اندازه پارامترهای آن به توزیع صفت در جامعه اصلی، روش نمونه گیری و حجم نمونه وابسته است.

فرض کنید توزیع یک صفت در جامعه دارای توزیع نرمال با میانگین و واریانس μ ، σ^2 باشد. این توزیع را به صورت $N(\mu, \sigma^2)$ نشان می دهیم. اگر تعداد زیادی نمونه تصادفی با حجم n از این جامعه استخراج شود و برای هر کدام از آنها مقدار آماره \bar{X} (میانگین حسابی صفت در نمونه) را محاسبه نماییم، امید ریاضی این آماره ها برابر μ خواهد شد و واریانس آنها به سمت $\frac{\sigma^2}{n}$ میل می کند و شکل توزیع نیز از الگوی نرمال پیروی می کند یعنی $N(\mu, \frac{\sigma^2}{n})$.

قضیه حد مرکزی که از قضایای اساسی در نظریه نمونه گیری محسوب می شود بیان می کند اگر حجم نمونه استخراج شده به اندازه کافی بزرگ باشد و واریانس جامعه نیز متناهی باشد آنگاه توزیع نمونه ای میانگین ها تقریباً نرمال خواهد بود حتی اگر توزیع صفت در جامعه مادر نرمال نباشد. هر چه توزیع صفت در جامعه مادر از توزیع نرمال دورتر باشد، باید حجم نمونه بزرگتری انتخاب شود (۱۲).

اصطلاحات اساسی در برآورد آماری:

مقادیر عددی پارامترها فقط از طریق سرشماری و بررسی همه عناصر جامعه قابل محاسبه است. علم آمار این امکان را می دهد تا بر اساس نتایج حاصل از بررسی یک نمونه درباره مقدار پارامتر برآوردی ارائه دهیم. برای آنکه مشخص کنیم یک آماره، تا چه حد یک پارامتر را به خوبی برآورد می کند باید از ملاکهای معینی استفاده کنیم. نداشتن تورش (اریبی) یکی از این ملاکهاست. یک آماره را نااریب گویند هرگاه امید ریاضی توزیع نمونه ای آن برابر پارامتر مورد برآورد باشد. اگر آماره $\bar{\theta}$ برآورد کننده پارامتر θ باشد اختلاف امید ریاضی $\bar{\theta}$ از θ را مقدار اریبی برآورد کننده می گویند. میانگین نمونه تصادفی از یک جامعه نرمال برآوردی نااریب از میانگین جامعه است زیرا امید ریاضی نمونه ای \bar{X} برابر μ است. انحراف مقدار برآورد از پارامتر جمعیت را دقت واقعی می گویند. چون در عمل پارامتر جامعه معلوم نیست می توان گفت دقت واقعی یک مفهوم نظری فاقد ارزش عملی است. به همین جهت باید دقت احتمالی محاسبه شود. دقت احتمالی یک برآورد کننده را معمولاً بوسیله خطای معیار توزیع نمونه ای همان برآورد کننده مشخص می کنند و هر چه خطای معیار کوچکتر باشد دقت بیشتر می شود. اگر دو برآورد کننده نااریب باشند آنکه خطای معیار کوچکتری دارد برآورد کننده بهتری است.

برآورد می تواند شامل دامنه ای از اعداد باشد که با اطمینان معینی مقدار پارامتر را شامل می شود. این حالت را برآورد فاصله ای می گویند. فیشر به عنوان برآورد کننده مقداری را مورد قبول می داند که درستنمایی نمونه ای را ماکزیمم کند. برآورد کننده هایی که این خاصیت را دارند برآورد کننده حداکثر درستنمایی خوانده می شوند. در واقع اگر X_1 و X_2 و ... X_n به عنوان متغیرهای تصادفی مشاهده شوند و تابع $f(x_1, x_2, \dots, x_n | \theta)$ معرف تابع چگالی احتمال توأم متغیرهای فوق باشد می توانیم $\bar{\theta}$ را طوری تعیین کنیم که $f(x_1, x_2, \dots, x_n | \bar{\theta})$ ماکزیمم باشد در اینصورت $\bar{\theta}$ را برآوردگر حداکثر درستنمایی برای پارامتر θ می گویند (۱۲).

در بعضی از حالات منطقی است θ را که یک پارامتر مجهول است مقدار یک متغیر تصادفی از یک توزیع مفروض در نظر گرفت. این حالت وقتی اتفاق می افتد که قبل از مشاهده نتایج

داده های x_1, x_2, \dots, x_n اطلاعاتی درباره وضعیت θ بصورت یک توزیع احتمال داشته باشیم. این توزیع را توزیع پیشین می گویند و برآورد کننده حاصل را برآورد کننده بیز می نامند.

اگر $\bar{\theta}_i$ برآورد کننده های θ باشند برآورد گر بیز باعث می شود امید ریاضی تابع $(\bar{\theta}_i - \theta)^2$ می نیمم شود. در واقع میانگین توزیع $f(\theta | x_1, x_2, \dots, x_n)$ همان برآورد گر بیز است. در مورد نمونه های بزرگ برآورد کننده های بیز با برآورد کننده های حداکثر درستنمایی تفاوت بسیار اندکی دارند (۱۳، ۱۲).

اصطلاحات اساسی در آزمون فرض :

فرض آماری حکمی درباره پارامتر جامعه است. محقق سعی می کند بر اساس داده های حاصل از نمونه گیری از همان جامعه، حکم مورد نظر را تایید یا رد نماید. آزمون آماری مجموعه قواعدی است که نشان می دهد نتایج حاصل با چه احتمالی بر اساس تصادف توجیه می شوند. بر اساس منطق آزمونهای آماری ابتدا می پذیریم فرض معینی که اغلب بر عدم تفاوت دلالت دارد صحیح است سپس پیامدهای این پذیرش را بر اساس یک توزیع نمونه ای مورد بررسی قرار می دهیم اگر در این شرایط احتمال مشاهده نتایج حاصل بعید به نظر برسد، محقق نتیجه می گیرد فرض اولیه تایید نمی شود.

عبارت "بعید به نظر رسیدن" را بر اساس سطح معنی داری آماری بصورت کمی بیان می کنند و آن سطحی از احتمال است که اگر احتمال مشاهده نتایج از آن کمتر باشد فرض اولیه رد می شود. فرض مورد آزمون را فرض صفر یا فرض خنثی می گویند و در مقابل آن فرض دیگری مطرح می شود که آنرا فرض جایگزین می گویند. عدم تایید فرض صفر باعث می شود فرض جایگزین را بپذیریم.

همانطور که ذکر شد رد فرض صفر بر اساس یک گزاره احتمال اعمال می شود به عبارتی ممکن است فرض صفر واقعاً درست باشد اما بر اساس تصادف نتایج مشاهده شده خلاف آن را نشان دهد در این حالت دچار نوعی خطای آماری شده ایم که آنرا خطای نوع اول می گویند بدیهی است احتمال بروز این خطا حداکثر برابر سطح معنی داری است و اساساً انتخاب سطح معنی داری معین در ابتدای یک طرح پژوهشی برای کنترل این نوع خطاست. از طرفی ممکن است فرض صفر واقعاً اشتباه باشد اما باز هم به دلیل خطای تصادفی، نتایج حاصل از مشاهده نمونه نتوانند فرض صفر را رد کنند و محقق فرض

صفر را تایید کند (در حالی که واقعاً اشتباه است) در اینصورت خطای نوع دوم رخ داده است بدیهی است هر چه خطای نوع دوم کوچکتر باشد توانایی پژوهش انجام شده برای رد یک فرض صفر واقعاً غلط بیشتر می شود به همین دلیل مقدار یک منهای خطای نوع دوم ($1-\beta$) را توان آزمون می گویند. توان یک آزمون را می توان بصورت زیر نیز تعریف نمود (اشتباه بود فرض | H_0 رد فرض صفر) $P = \text{توان}$. در عمل به خطای نوع دوم خیلی کم توجه می شود و در مورد خطای نوع اول نیز گر چه وضعیت بهتر است اما باید توجه داشت استفاده مکرر از سطوح $0/05$ و $0/01$ بنابر قراردادی است که پایه علمی و منطقی ضعیفی دارد (۱۲).

فرایند تصمیم گیری در انجام آزمون فرض امری نسبتاً پیچیده و حرفه ای است به همین دلیل از بیان جزئیات انواع آنها صرف نظر می کنیم و در فصل بعد به اصول و معیارهای انتخاب انواع آزمون آماری می پردازیم. فقط به عنوان نمونه جزئیات فرایند تصمیم گیری در یکی از رایجترین حالتها آزمون فرض یعنی مقایسه دو میانگین مستقل، می پردازیم (۱۴).

هرگاه بخواهیم درباره اختلاف پارامترهای μ_1 و μ_2 بر اساس یافته های حاصل از نمونه گیری یعنی \bar{X}_1 و \bar{X}_2 استنباطی ارائه دهیم با فرض برقراری شرایط پارامتری، قبل از هر چیز باید مشخص کنیم آیا واریانسهای دو جامعه معلوم هستند یا خیر؟ در صورت معلوم بودن واریانسها، وقتی فرض صفر حاکی از عدم تفاوت بین میانگینها باشد، با استفاده از روش Z می توان فرض فوق را آزمون نمود بدیهی است باید بتوان فرض نمود توزیع $(\bar{X}_1 - \bar{X}_2)$ به توزیع نرمال نزدیک است. در عمل اغلب واریانس ها مجهولند و محقق ناچار است از برآورد آنها یعنی S_1^2 و S_2^2 استفاده نماید. در این حالت قبل از هر چیز باید بر اساس آزمون F معلوم شود آیا می توان فرض نمود واریانس های مجهول با هم برابر هستند یا خیر؟ در صورت تایید فرض تساوی واریانس ها باید بر اساس S_1^2 و S_2^2 و حجم نمونه در دو گروه مورد بررسی برآوردی از واریانس مشترک بدست آورد و سپس بر اساس آزمون T -Student فرض صفر مبنی بر عدم تفاوت دو میانگین را آزمون نمود.

اگر فرض تساوی واریانس ها تایید نشود تصمیم گیری بعدی وابسته به حجم نمونه در دو گروه مورد بررسی است. چنانچه حجم نمونه به اندازه کافی بزرگ باشد (اغلب بیشتر از ۳۰ نفر در هر گروه) می توان