Subject:
Year: Month: Date:

جلسه ۸ دیه عمومی جلسه اول

آمار و احتمالات ، استاد خیلی خوشبرخورد و مهربان (نمره کم نمیده)

سرفصل ها

۱- مقدمه ای بر آمار و احتمالات

۲- احتمال

۳- آمار توصیفی

۴- آمار استنباطی

۵- مدل های احتمالی و توزیع های آماری

۶- رگرسیون و همبستگی

۷- آمار چند متغیره

۱- آمار چیست ؛ آمار علمی است که به جمع آوری ، تجزیه ، تحلیل و تفسیر داده ها می پردازد. آمار این امکان را می دهد که از داده های موجود به نتایج قابل اعتماد برسیم و رابطه های عمومی را شناسایی کنیم. آمار به دو دسته توصیفی و استنباطی تقسیم می شود

آمار
- توصیفی
- استنباطی

در آمار توصیفی داده ها به شکلی ساده و مختصر نمایش داده می شود ، شاخص های اصلی در آمار توصیفی شامل میانگین ، میانه ، واریانس و انحراف معیار است

آمار استنباطی از داده های نمونه ای برای استنباط و تصمیم دارن

به یک جمعیت بزرگتر استفاده می شود ، هدف از آمار استنباطی

Subject :
Year : Month : Date :

آزمون فرضیه ها و تخمین پارامترهای جمعیت می باشد

احتمالات به بررسی پدیده های تصادفی و پیش بینی نتایج آنها می پردازد

احتمال : مقدار عددی می باشد که نشان دهنده احتمال وقوع یک رویداد است و مقداری بین
نیز ۰ و ۱ دارد.

احتمال صفر : رویدادی که غیرممکن است

احتمال یک : رویدادی که قطعاً رخ می دهد.

احتمال های بین صفر و یک ، احتمال وقوع رویدادهایی است که بصورت تصادفی یا
احتمالی ممکن است اتفاق بیفتد.

رابطه آمار و احتمال : در حالی که آمار به بررسی داده ها و استنباط اطلاعات
از آنها می پردازد احتمال در واقع به مدل سازی پدیده های تصادفی کمک می کند. احتمال
پایه و اساس بسیاری از روش های استنباطی در آمار می باشد.

مفاهیم پایه در احتمال : فضای نمونه : مجموعه ای از تمام نتایج ممکن یک آزمایش
تصادفی . رویداد : هر زیر مجموعه ای از فضای نمونه که نتیجه ای خاص را بیان
می کند. احتمال شرطی : احتمال وقوع یک رویداد با فرض وقوع رویداد دیگری

Subject :
Year : Month : Date :

قوانین احتمال : قانون جمع : اگر دو رویداد A و B مستقل باشند، احتمال وقوع جمع هر یک از آنها برابر است با

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

قانون ضرب : برای رویداد های مستقل ، احتمال وقوع هر دو رویداد ، برابر است با

$$P(A \cap B) = P(A) \times P(B)$$

آغاز تحلیل داده ها : هنگامی که داده ها جمع آوری می شوند ، اولین قدم در تحلیل آنها استفاده از آمار توصیفی است. تا تصویر کلی از داده ها بدست آید ؛ سپس می توان از آمار استنباطی برای انجام آزمون های فرضیه یا تخمین پارامترهای جمعیت استفاده کرد.

تمرین ۱ : در یک بازی یک تاس شش وجهی انداخته می شود ، احتمال اینکه عدد ۴ ظاهر شود چقدر است ؟

$$A = \{1, 2, 3, 4, 5, 6\} \quad p = \frac{1}{6}$$

تمرین ۲ : در یک کلاس ۶۰ درصد از دانش آموزان دختر هستند و ۴۰ درصد پسر هستند. از میان دخترها ۷۰ درصد نمرات خوب دارند ، از میان پسرها ۵۰ درصد نمرات خوب دارند اگر یک دانش آموز انتخاب شود که نمرات خوب دارد احتمال دختر بودن چقدر است.

D دختر
P پسر
G نمره خوب

$P(G/D) = 0.7$
$P(G/P) = 0.5$ نمره خوب پسر
$P(D/G) = \frac{P(G/D)\,P(D)}{P(G)}$ دختر

$P(D) = 0.6$ دختر
$P(P) = 0.4$ پسر

$P(G) = P(G/D)\,P(D) + P(G/P)\,P(P)$

$P(D/G) = \frac{(0.7)\times(0.6)}{(0.5)\times(0.4)+(0.7)\times(0.6)} = \frac{0.42}{0.62}$

$P(D/G) = \frac{P(G/D)\,P(D)}{P(G)}$

$P(D/G) = \frac{0.7 \times 0.6}{0.62} = \frac{0.42}{0.62}$

0.677

Subject :
Year | Month | Date

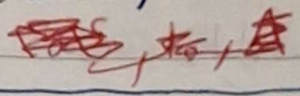
فرض کنید آزمایشی فرض مثبت بالا شد احتمال اینکه آن فرد واقعا بیمار باشد

چقدر است.

B	فرد بیمار	$P(B) \Rightarrow$ بیمار
H	فرد سالم	$P(H) \Rightarrow$ سالم
+	نتیجه آزمایش	$P(+) \rightarrow$ آزمایش

$$\frac{P(+|B)\ P(B)}{P(+|B)P(B)+P(+|H)P(H)}$$

$$\Rightarrow P(B|+)$$



۱۵ آذر ۱۴۰۴ جلسه دوم

متغیر تصادفی: یک تابع است که به هر نتیجه از یک آزمایش تصادفی یک عدد اختصاص می دهد. به عبارت دیگر متغیر تصادفی، مقدار عددی یک رویداد تصادفی را نشان می دهد. متغیر های تصادفی به دو دسته تقسیم می شوند.

۱- متغیر تصادفی گسسته ۲- متغیر تصادفی پیوسته

در متغیر تصادفی گسسته مقادیر این نوع متغیر محدود و شمارش پذیر می باشد مثلاً تعداد شاس هایی که در یک بازی می افتد یا تعداد مشتریانی که وارد یک فروشگاه می شوند، در متغیر تصادفی پیوسته این مقادیر بین متغیر بین دو عدد پیوسته است و می تواند هر مقداری را در یک بازه مشخص بگیرد. این مقادیر معمولاً با اعداد حقیقی نشان داده می شوند، مثلاً حد یک شخص یا

۱۱

Date: Subject:

دمای هوا در یک روز خاص.

توزیع های احتمالی

توزیع های احتمالی گسسته: برای متغیر های تصادفی گسسته توزیع های مختلفی وجود دارند که هرکدام ویژگی های خاص خودرا دارند. مهم ترین توزیع های احتمالی گسسته عبارتند از الف) توزیع برنولی توزیع برنولی برای آزمایش هایی است که تنها دو نتیجه ممکن دارند. مثلاً پرتاب سکه، نتایج شیر یا خط می باشند. این توزیع در واقع یک متغیر تصادفی گسسته است که تنها دو مقدار صفر یا یک را می گیرد. احتمال وقوع نتیجه یک (مثلا شیر) را با p نشان می دهیم و احتمال وقوع نتیجه صفر (مثلا خط) را با 1-p نشان می دهیم

$$P(x=1)=p$$

$$P(x=0)=1-p$$

ب) توزیع دوجمله ای زمانی که آزمایش برنولی چندین بار تکرار شود، توزیع دوجمله ای بدست می آید. در این توزیع تعداد موفقیت ها (مثلاً تعداد شیرها در چندین پرتاب سکه) بصورت تصادفی و براساس تعداد آزمایش ها و احتمال موفقیت در هر آزمایش مشخص می شود.

تعداد آزمایش ها (تکرار)

$$P(x=k)=\binom{n}{k}p^{k}(1-p)^{n-k}$$

احتمال موفقیت

تعداد موفقیت ها (تعداد موفقیت در n آزمایش)

تمرین: فرض کنید یک سکه داریم که احتمال شیر (نتیجه موفقیت) برابر با $p=0,6$ و احتمال خط (نتیجه شکست) برابر با $1-p=0,4$ می‌باشد. حال محاسبه کنید که احتمال وقوع هرکدام از این دو نتیجه چقدر است؟

$$P(x=1)=p=0,6 \longrightarrow 60\%$$

$$P(x=0)=1-p=0,4 \longrightarrow 40\%$$

تمرین: فرض می‌کنیم که یک سکه داریم که احتمال شیر (موفقیت) در هر پرتاب $p=0,6$ و احتمال خط (شکست) در هر پرتاب $1-p=0,4$ حالا ۵ بار این سکه را پرتاب می‌کنیم و می‌خواهیم احتمال وقوع دقیقاً سه شیر (موفقیت) را در پنج پرتاب محاسبه کنیم.

$$P(x=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

$$P(x=3)=\binom{5}{3}(0,6)^3(0,4)^{5-3=2}=0,3456$$

($\binom{5}{3}=10$, $0,6^3=0,216$, $0,4^2=0,16$)

$$\binom{5}{3}=\frac{5!}{3!(5-3)!}=\frac{5\times4\times3\times2\times1}{(3\times2\times1)(2!)}=10$$

ج) توزیع پواسون: این توزیع برای مدل سازی تعداد اتفاقاتی که در یک بازه زمانی یا فضای ثابت اتفاق می‌افتد (مثل تعداد تماس‌های دریافتی در یک مرکز خدماتی یا تصادفات در یک ساعت) این توزیع معمولاً برای اتفاقاتی استفاده می‌شود که احتمال وقوع آن‌ها به طور متوسط در یک بازه زمانی مشخص ثابت باشد.

$$P(x=k)=\frac{d^k e^{-d}}{k!}$$

Date: Subject:

عدد نپر: ۲,۷۱۸

$P(x=k) = \frac{\lambda^k e^{-\lambda}}{k!}$ ← ثابت نپر

k: تعداد اتفاق‌هایی که احتمال وقوع آن‌ها را محاسبه می‌کنیم

میانگین تعداد اتفاقات

تمرین: فرض کنید یک مرکز تماس در یک روز خاص به طور متوسط ۴ تماس در هر ساعت دریافت می‌کند. احتمال اینکه در یک ساعت خاص این مرکز تماس دقیقاً ۳ تماس دریافت کند را محاسبه کنید.

$P(x=k) = \frac{\lambda^k e^{-\lambda}}{k!}$

$\lambda = 4$, $k = 3$

$P(x=3) = \frac{(4)^3 (2,718)^{-4}}{3!} = \frac{4 \quad 2,718}{3!} = (2,718)^{-4} = 0,018$

$\frac{64 \times (0,018)}{3 \times 2 \times 1} = 0,195$

تمرین: در یک کارخانه تولیدی به طور متوسط هر روز دو نقص فنی در خط تولید رخ می‌دهد. محاسبه کنید که احتمال اینکه در یک روز خاص دقیقاً سه نقص فنی در خط تولید رخ می‌دهد را محاسبه کنید

$P(x=3) = \frac{2^3 \quad 2,718}{3} = \frac{8 \times 0,135}{3 \times 2 \times 1} = 0,18$

$k = 3$

$\lambda = 2$

۱۸٪

احتمال وقوع

Date / / Subject

تمرین: یک فروشگاه آنلاین به طور متوسط ۵ سفارش به طور متوسط دریافت می‌کند؛ احتمال اینکه در یک ساعت دقیقه ۷ سفارش دریافت شود را محاسبه کنید.

$p = \alpha = k$

$\alpha = ۷ \quad \frac{۵^{۷}(۲,۷۱۸)^{-۵}}{۷!} = ۰,۱۰۴۶۷ = ۰,۱۰۴۶۷$

۱۰,۴۶۷٪

احتمال وقوع

$۷! = ۷ \times ۶ \times ۵ \times ۴ \times ۳ \times ۲ \times ۱$

قضیه حد مرکزی — جلسه سوم

این قضیه بیان می‌کند که وقتی یک متغیر تصادفی به تعداد زیاد نمونه برداری شود، توزیع میانگین های نمونه ها تقریبا به توزیع نرمال نزدیک خواهد شد حتی اگر توزیع اصلی داده ها نرمال نباشد. فرض کنید شما داده هایی از یک جمعیت با توزیع دلخواه دارید (مثلا ممکن است داده ها به صورت تصادفی و نامنظم توزیع شده باشند) حالا اگر از این داده ها، چندین نمونه تصادفی بگیرید و میانگین هر نمونه را محاسبه کنید، توزیع این میانگین ها تقریبا با شکل یک منحنی نرمال (زنگی) درخواهد آمد به ویژه زمانی که اندازه نمونه ها بزرگتر شود.

فرض کنید یک جمعیت با توزیع دلخواه داریم و می خواهیم نمونه هایی از این جمعیت بگیریم اگر اندازه نمونه ها، بزرگ باشد توزیع میانگین های نمونه ها (که همان میانگین هر نمونه تصادفی است) به شکل زیر رفتار خواهد کرد. میانگین توزیع نمونه ها برابر با میانگین جمعیت اصلی خواهد بود، انحراف معیار توزیع نمونه ها به نام خطای استاندارد شناخته می‌شود و به اندازه نمونه بستگی دارد.

Date: Sat Sun Mon Tue Thu Wed Fri Subject: ----------

هرچه اندازه نمونه بزرگتر باشد، انحراف معیار توزیع میانگین ها کوچکتر می شود، این ویژگی باعث می شود که بتوانیم بدراحتی با استفاده از توزیع نرمال تحلیل های آماری را انجام دهیم، حتی زمانی که داده های اصلی به طور کلی نرمال نیستند.

اگر μ (مو) میانگین جمعیت باشد، میانگین توزیع میانگین ها نیز مو μ خواهد بود
اگر σ (سیگما) انحراف معیار جمعیت باشد، انحراف معیار توزیع میانگین ها به صورت زیر محاسبه می شود.

$$SE = \frac{\sigma}{\sqrt{n}}$$

n ← اندازه نمونه

به کمک قضیه حد مرکزی، توزیع میانگین های نمونه ها با افزایش اندازه نمونه، به یک توزیع نرمال تبدیل می شود، به این معنا که حتی اگر توزیع اصلی جمعیت نرمال نباشد، توزیع میانگین های نمونه ها، برای n بزرگ به نرمال نزدیک خواهد شد.

تمرین: فرض کنید در یک کارخانه تولیدی تعداد قطعات معیوب از میان ۱۰۰۰ قطعه تصادفی اندازه گیری شده و میانگین آن ۵ درصد است، همچنین انحراف معیار تعداد قطعات معیوب در هر ۱۰۰۰ قطعه ۲ درصد است، حال فرض کنید که ما یک نمونه تصادفی از ۱۰۰ قطعه انتخاب می کنیم، حال احتمال اینکه میانگین قطعات معیوب از ۶ درصد بیشتر باشد را محاسبه کنید.

$\mu = 5$

$\sigma = 2$

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{100}} = \%2$$

Date: / / Sat Sun Mon Tue Thu Wed Fri Subject:

تمرین: فرض کنید در یک مرکز تماس، زمان پاسخگویی کارمندان به تماس مشتریان به دقیقه از یک توزیع غیر نرمال پیروی می‌کند، میانگین کل زمان پاسخگویی ۶ دقیقه می‌باشد، انحراف معیار کل ۳ دقیقه می‌باشد حال، نمونه هایی از ۳۶ تماس را تصادفی انتخاب کرده ایم، میانگین زمان پاسخگویی آنها را حساب کنید.

انحراف معیار توزیع میانگین‌ها

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0.5$$

اندازه نمونه‌ها

$\mu = 6$

$\sigma = 3$

تمرین: می دانیم که زمان لازم برای درست کردن یک فنجان قهوه در یک کافی شاپ توزیع نرمال ندارد، اما از داده ها می دانیم که $\mu = 4$ ، $\sigma = 1.2$ فرض کنید از این کافی شاپ نمونه‌هایی با اندازه $n = 36$ می‌گیریم طبق قضیه حد مرکزی انحراف معیار توزیع میانگین ها را برسی کنید.

انحراف معیار توزیع میانگین‌ها

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{1.2}{\sqrt{36}} = 0.2$$

اندازه نمونه

$\mu = 4$

$\sigma = 1.2$

$n = 36$

مفهوم میانگین:

میانگین همان معدل است یعنی مجموع داده ها تقسیم می‌شود.

مثال: فرض کنید نمرات ۵ دانشجو در یک آزمون ریاضی به صورت زیر است میانگین نمرات دانشجویان را بدست آورید.

Subject :
Year : Month : Date :

۱۴
۱۳
۱۲
۱۱
۱۰

$\mu = \frac{۱۴+۱۳+۱۲+۱۱+۱۰}{۵} = ۱۲$ میانگین

$\bar{X}$ ایکس بار

توزیع میانه

میانه: عددی است که در وسط داده های مرتب شده قرار دارد، یعنی نیمی از داده ها کوچکتر، و نیمی بزرگتر از آن هستند، اگر تعداد داده ها زوج بود، دو میانه برابر است با میانگین دو عدد وسط

مثال: برای داده های زیر میانه را حساب کنید.

۲, ۴, ۶, ۸

$\frac{۴+۶}{۲} = ۵$ میانه

مثال: قرعه کشی قد ۵ نفر بر سانتی متر به صورت زیر می باشد.

۱۵۰

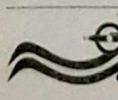
۱۶۰

۱۶۵ میانه

عدد وسطی میانه است

۱۷۰

۱۷۵



Subject :
Year : Month : Date :

تمرین: نمرات ۶ دانشجو به شکل زیر می باشد. میانگین

۱۰ ۱۲ ۱۵ ۱۸ ۱۹ ۲۰

$\frac{۱۵+۱۸}{۲} = ۱۶,۵$ (۳۳)

مد

عددی است که بیشترین تکرار را دارد، یعنی عددی که بیش از بقیه، در داده ها ظاهر شده است.

مثال: نمرات ۸ دانشجو در درس آمار و احتمالات به شکل زیر می باشد. مد آن را بیابید.

۷, ۱۵, ۱۸, ۱۷, ۱۳, ۱۴, ۱۵, ۱۵

مد آن ۱۵ است.

تمرین: به ۷ دانشجو، پرونده ای داده ایم، داده های آنها به شرح زیر می باشد میانگین، میانه و مد این داده ها را محاسبه کنید.

۵۰ ۵۵ ۶۰ ۶۵ ۷۰ ۱۰۰۰ ۷۵

میانه

Date: Subject:

میانگین

$\bar{X} = \frac{50 + 55 + 40 + 45 + 70 + 100 + 75}{7} = \frac{435}{7} = 62.14$

میانه

$X : 50, 55, 40, 70, 45, 75, 100 = 55$

مرتب

میانه

مد

مد ندارد.

مثال: درآمد ماهانه ۷ نفر از زنان روستا کاری به شرح زیر است

$1500 + 2000 + 2200 + 2500 + 3000 + 10000 + 2500 = 23500$

میانه

$\frac{23500}{7} = 3357$

میانگین و میانه و مد

مد = ۲۵۰۰

انحراف معیار

اندازه‌ای است که نشان می‌دهد داده‌ها به طور متوسط چقدر از میانگین فاصله دارند. انحراف معیار کم نشان‌دهنده این است که داده‌ها به طور کلی نزدیک به میانگین قرار دارند و انحراف معیار زیاد نشان‌دهنده این است که داده‌ها از میانگین پراکنده شده‌اند.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - M)^2}{N}}$$

انحراف معیار (σ)

داده

میانگین حسابی

تعداد کل داده‌ها

Date: Subject:

نمرات ۵ دانشجو به شرح زیر است

۹۰ ۷۰ ۸۰ ۹۰ ۱۰۰

توضیح سوال: برای محاسبه انحراف معیار ابتدا میانگین داده ها را محاسبه می‌کنیم سپس کم کردن میانگین از هر داده و سپس مربع کردن آنها بعد از آن جمع کردن مربع های اختلافات، بعد از آن تقسیم بر تعداد داده ها می‌کنیم و در نهایت جذر را از حاصل تقسیم می‌سنجیم می‌کنیم

① $\bar{X} = \frac{۹۰ + ۷۰ + ۸۰ + ۹۰ + ۱۰۰}{۵} = ۸۰$

$(۹۰-۸۰)^۲ = ۱۰۰$

$(۷۰-۸۰)^۲ = ۱۰۰$

$(۸۰-۸۰)^۲ = ۰$

$(۹۰-۸۰)^۲ = ۱۰۰$

$(۱۰۰-۸۰)^۲ = ۴۰۰$

② هر داده را از میانگین کم کرده و تقسیم بر تعداد کل داده‌ها می‌کنیم

③ $۴۰۰ + ۱۰۰ + ۰ + ۱۰۰ + ۴۰۰ = ۱۰۰۰$

④ $\frac{۱۰۰۰}{۵} = ۲۰۰$

⑤ $\sqrt{۲۰۰} = ۱۴/۱۴$

مثال: داده های آزمایشگاهی ۵ دانشجو به شرح زیر می‌باشد: انحراف معیار جمعیت آن را حساب کنید.

۴۰, ۴۵, ۵۰, ۴۵, ۵۰

$\frac{۴۰+۴۵+۵۰+۴۵+۵۰}{۵} = ۴۶$

$(۴۰-۴۶)^۲ = ۳۶$

$(۴۵-۴۶)^۲ = ۱$

$(۵۰-۴۶)^۲ = ۱۶$

$(۴۵-۴۶)^۲ = ۱$

$(۵۰-۴۶)^۲ = ۱۶$

جمع ۷۰

$\frac{۷۰}{۵} = ۱۴$

$\sqrt{۱۴} = ۳/۷۴$

جلسه چهارم

Date: Subject:

تمرین: داده های زیر میزان آلودگی در ۵ نمونه آب متفاوت را نشان می دهد. واریانس و انحراف معیار داده را بدست آورید.

۵۰	$\frac{۵۰+۵۵+۴۰+۴۵+۷۰}{۵} = ۴۰$
۵۵	
۴۰	$(۵۰-۴۰)^۲ = ۱۰۰$
۴۵	$(۵۵-۴۰)^۲ = ۲۵$
۷۰	$(۴۰-۴۰)^۲ = ۰$ $\quad S^۲ = \frac{۲۵۰}{۵} = ۵۰$
	$(۴۵-۴۰)^۲ = ۲۵$
	$(۷۰-۴۵)^۲ = ۱۰۰$ $\quad \sqrt{۵۰} = ۷/۰۷$

$۱۰۰+۲۵+۱۰۰+۲۵ = ۲۵۰$ انحراف معیار

آزمون های فرضی

ابتدا مبحث، ابزارهای آماری است که به ما کمک می کند در مواجه با داده ها تصمیمات علمی و معتبر بگیریم. آزمون فرضی به طور کلی یک روش آماری است که برای ارزیابی صحت یا نادرستی یک فرضیه در مورد جمعیت بکار می رود. این فرایند شامل استفاده از داده های نمونه برای ارزیابی اینکه آیا شواهد کافی برای رد یا پذیرش فرضیه خاص وجود دارد یا خیر که می باشد. آزمون فرضی معمولا در دو مرحله انجام می شود

Date: Subject:

① تعریف فرضیه ها ② ارزیابی داده ها و تصمیم گیری بر اساس شواهد موجود

① در هر آزمون فرض، دو فرضیه اصلی وجود دارد

فرضیه صفر (H_0): این فرضیه معمولاً فرض است که می خواهیم آن را رد کنیم، معمولاً فرضیه صفر بیان می کند که هیچ تفاوت یا اثر خاصی در داده ها وجود ندارد.

فرضیه جایگزین (H_1 یا H_a) این فرضیه، مخالف فرضیه صفر می باشد و معمولاً بیان می کند که یک تفاوت یا اثر خاصی در داده ها وجود دارد. به عنوان مثال اگر بخواهیم آزمون کنیم که آیا میانگین قد افراد در یک جامعه برابر با ۱۷۰ سانتی متر است، فرضیه ها به صورت زیر خواهند بود.

(H_0): میانگین قد افراد برابر با ۱۵۰ سانتی متر است

$(H_a$ یا $H_1)$: میانگین قد افراد متفاوت از ۱۷۰ سانتی متر است

② پس از تعیین فرضیه ها باید یک آزمون آماری انتخاب کنیم که به کمک آن می توانیم شواهد را بررسی کنیم و تصمیم بگیریم که آیا فرضیه صفر را رد کنیم یا نه؟ مراحل کار شامل موارد زیر می باشد.

① محاسبه آماره آزمون

این مقدار معمولاً با استفاده از داده های نمونه و فرمول های خاص هر آزمون آماری محاسبه می شود.

② انتخاب سطح معنی دار (آلفا) (α) این سطح احتمال است که حاکی از احتمال اشتباه در رد فرضیه صفر می باشد، معمولاً ($\alpha = 0.05$) انتخاب می شود که به معنای این است که ۵ درصد احتمال داریم که فرضیه صفر را اشتباه رد کنیم.

Date: Sat Sun Mon Tue Thu Wed Fri Subject:

③ مقایسه آماره ی آزمون با مقدار بحرانی یا استفاده از P-Value

احتمال مشاهده داده های نمونه با p داده های شدیدتر، تحت فرضیه صفر می باشد، اگر کوچکتر از سطح معنادار α آن باشد، فرضیه صفر رد می شود.

④ مقدار بحرانی

در این روش باید آماره آزمون را (مثل توزیع) با مقدار بحرانی که از جدول توزیع مربوطه بدست می آید مقایسه کنیم

انواع آزمون های فرضی

آزمون های فرضی انواع مختلفی دارند که براساس نوع داده ها و نوع آزمون بکار می روند. برخی از مهمترین آزمون ها را توضیح خواهیم داد.

① آزمون t - student

برای مقایسه میانگین های دو نمونه از داده ها استفاده می شود. معمولاً زمانی که اندازه نمونه ها کوچک است، این آزمون می تواند، برای آزمون یک نمونه (برای مقایسه میانگین با یک مقدار ثابت) یا دو نمونه مستقل (برای مقایسه دو میانگین نمونه ای) استفاده می شود.

② آزمون Z

برای مقایسه میانگین ها یا نسبت ها در داده هایی که از توزیع نرمال پیروی می کنند و معمولاً برای نمونه های بزرگ ($n > 30$) استفاده می شود، آزمون Z همچنین برای مقایسه یک میانگین با مقدار مشخص در زمانی که انحراف معیار جامعه را می دانیم بکار می رود.

③ آزمون chi - square

Date: / / Subject:

اینا آزمون برای بررسی ارتباط بین دو متغیر کیفی یا مقایسه توزیع های مشاهده شده با توزیع های مورد انتظار استفاده می شود، اینا آزمون به طور گسترده در مطالعات نظر سنجی و تحلیل جدول های توافقی بکار می رود

۴) آزمون ANOVA (تحلیل واریانس)

اینا آزمون برای مقایسه میانگین ها در سه یا بیشتر گروه های مستقل استفاده می شود و هدف آن بررسی اینا می باشد که آیا اختلاف معناداری بین میانگین گروه ها وجود دارد یا نه

مفاهیم کلیدی در آزمون های فرضی

خطا نوع اول (α): احتمال رد فرضیه صفر، زمانی که فرضیه نوع صفر درست است به عبارتی دیگر، خطا نوع اول یعنی اشتباه در رد فرضیه صحیح

خطا نوع دوم (β): احتمال پذیرش فرضیه صفر، زمانی که فرضیه صفر غلط می باشد، یعنی وقتی که فرضیه صفر باید رد شود ولی اشتباها پذیرفته می شود

مثال فرض کنید می خواهیم بررسی کنیم که آیا میانگین قد دانش آموزان یک کلاس با ۱۷۰ سانتی متر متفاوت است یا خیر، برای اینا کار مراحل زیر را طی می کنیم

۱) فرضیه ها: میانگین قد دانش آموزان برابر با ۱۷۰ سانتی متر می باشد (H_0)

میانگین قد دانش آموزان متفاوت با ۱۷۰ سانتی متر می باشد (H_1)

Date: Subject:

۲) انتخاب سطح معنادار: فرض می کنیم که $\alpha = 0.05$

۳) جمع آوری داده ها: فرض می کنیم اندازه نمونه ۳۵ نفر باشد و میانگین قد نمونه ۱۷۲ سانتی متر و انحراف معیار ۱۰ سانتی متر باشد

۴) آزمون Z: برای انجام آزمون Z از فرمول زیر استفاده می کنیم.

میانگین فرضی — میانگین نمونه

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \text{انحراف معیار} \quad \text{تعداد داده ها}$$

$$Z = \frac{172 - 170}{\frac{10}{\sqrt{35}}} = 1.095$$

۵) مقایسه با مقدار بحرانی یا استفاده از p-Value

با توجه به جدول توزیع Z و سطح معنادار $\alpha = 0.05$ مقدار بحرانی برای آزمون دو طرفه حدود ۱٫۹۶ می باشد چون $1.96 > 1.095$ فرضیه صفر را رد نمی کنیم، در اینجا مثل چون مقدار Z می بینید کمتر از مقدار بحرانی می باشد نمی توانیم فرضیه صفر را رد کنیم و نتیجه می گیریم که شواهد کافی برای ادعا اینکه میانگین قد دانشجویان با ۱۷۰ سانتی متر متفاوت است وجود ندارد.

کاربردهای آزمون های فرضی

در تحقیقات علمی برای تایید یا رد فرضیه های علمی در آمار پزشکی برای بررسی اثربخشی داروها یا درمان ها در آمار اقتصادی برای تحلیل داده های اقتصادی و پیش بینی روند های اقتصادی، در بازاریابی برای بررسی اثر تبلیغات یا تغییرات در قیمت بر رفتار مصرف کننده کاربرد

Date: Subject:

مثال: میزان آفات ۶ درخت به صورت زیر ارائه شده است مطلوب است واریانس؟ میانگین؟ انحراف معیار؟ مد؟ میانه؟

۲۲
۲۴
۲۱
۲۰
۲۵
۲۴

$\bar{x} = \frac{۲۲+۲۴+۲۱+۲۰+۲۵+۲۴}{۶} \Rightarrow (۲۲,۶)$ میانگین

مد = بیشترین تکرار = (۲۴)

$\tilde{x}_{میانه} = \frac{۲۲+۲۴}{۲} = \frac{۴۶}{۲} = (۲۳)$ میانه

$(۲۰ - ۲۲,۶)^۲ =$	۵,۲
$(۲۱ - ۲۲,۶)^۲ =$	۳,۲
$(۲۲ - ۲۲,۶)^۲ =$	۰,۳۶
$(۲۴ - ۲۲,۶)^۲ =$	۲,۷۶
$(۲۴ - ۲۲,۶)^۲ =$	۲,۷۶
$(۲۵ - ۲۲,۶)^۲ =$	۵,۷۶

(۱)

(۲) $۵,۲ + ۳,۲ + ۰,۳۶ + ۲,۷۶ + ۲,۷۶ + ۵,۷۶ = ۲۸,۰۴$

(۳) $s^۲$ واریانس $= \frac{۲۸,۰۴}{۶} = ۴,۶۷$

(۴) $\sigma = \sqrt{۴,۶۷} = ۲,۱۶$

مثال: در یک تحقیق میانگین قد مردان در یک کشور برابر با ۱۷۵ cm است یک نمونه‌ای از ۴۰ مرد به طور تصادفی انتخاب شده است، میانگین قد آن‌ها ۱۷۸ سانتی متر می‌باشد انحراف معیار جامعه برابر با ۸ سانتی متر است مطلوب است محاسبه آماره آزمون Z؟

Date: Subject:

$$Z = \frac{۱۷۸ - ۱۷۵}{\frac{۸}{\sqrt{۶۰}}} = \frac{۳}{۱,۲۴۵} = ۲,۴۷$$

$۲,۴۷ > ۱,۹۹$ فرضیه رد می‌شود

$\alpha = ۰,۰۵$

Date: ۴۰۲ ۹ ۱ Subject:

جلسه ۱۰ ام

میانگین های هارمونیک (همساز)

اگر $x_1, x_2, \dots, x_n$ نمونه‌ای از مشاهدات بدست آمده از یک جامعه باشد میانگین های هارمونیک آنها به صورت زیر تعریف می شود.

$$\overline{x}_h = \left[\frac{1}{n} \left(\sum_{i=1}^{n} \frac{1}{x_i} \right) \right]^{-1} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

(میانگین هارمونیک — اندازه نمونه — تعداد نمونه)

وقتی مشاهدات از نسبت دو کمیت مختلف بوجود آمده باشد (مانند سرعت که نسبت مسافت به زمان است) و مقادیر مخرج کسر برابر مشاهدات مختلف در نمونه متغیر باشند از میانگین های هارمونیک برای محاسبه مکان متوسط داده ها استفاده می شود.

$$\overline{V} = \frac{\Delta x}{\Delta t}$$

تمرین یک اتومبیل به مدت t ثانیه با سرعت ثابت $90m$ بر ثانیه و به مدت t ثانیه دیگر با سرعت ثابت $30m$ بر ثانیه در حرکت است میانگین سرعت این خودرو طی این دو مدت چقدر بوده است؟

حل: اتومبیل در t ثانیه اول، مسافتی برابر با x_1 و در t ثانیه بعد، مسافتی برابر با x_2 را طی کرده است. چون سرعت در t ثانیه اول با سرعت در t ثانیه دوم برابر نیست بنابراین مسافت طی شده در این زمان ها یکسان نخواهد بود $x_1 \neq x_2$

Date: Subject:

$t_1 = t_2 = t$ از مواقعی زمان طی دو زمان مذکور تا هم برابر است در نتیجه می‌توان نوشت:

$$\bar{V} = \frac{x_1 + x_2}{t_1 + t_2}$$

$$\bar{V} = \frac{x_1 + x_2}{2t} = \frac{1}{2}\left(\frac{x_1}{t} + \frac{x_2}{t}\right) = \frac{1}{2}(V_1 + V_2) = \frac{1}{2}$$

$t_1 = t_2 = t$ چون

$$\frac{1}{2}(90 + 30) = 60 \text{ m/s}$$

Date: ۰۴ ۹ ۱۵ Subject: جلسه ششم

چندک

با توجه به تعاریفی که برای چندک جامعه وچندک نمونه‌ای وجود دارد، چندک k ام کمیتی است که در اصطلاح k چندم داده ها کوچکتر یا مساوی آن هستند به عنوان مثال صدک هفتم مقداری است که $\frac{7}{100}$ مشاهدات کوچکتر یا مساوی آن هستند، برخی از چندک ها به دلیل کاربرد فراوان، اسامی خاصی دارند به عنوان مثال چارک i ام، چندکی است که $\frac{i}{4}$ ام مشاهدات کوچکتر یا مساوی آن هستند و با Q_i نمایش داده می‌شود. یا صدک یا صده i ام نقطه‌ای است که $\frac{i}{100}$ و یا i٪ مشاهدات کوچکتر یا مساوی آن می‌باشد مفاهیم دیگری مانند دهک و پنجاهک نیز به طریق مشابه قابل تعریف هستند، چندک‌ها لزوماً بیان کننده مکان تمرکز مشاهدات نیستند، بلکه مکانی از توزیع را نشان می‌دهند که کسری از مشاهدات کوچکتر یا مساوی آن هستند.

نحوه محاسبه چندک‌ها

① مشاهدات را به صورت صعودی مرتب کرده و مشاهده مرتب شده i ام را به صورت $x_{(i)}$ نشان می‌دهیم

② اگر p نسبت چندک یا کمیت متناظر با آن و n اندازه نمونه باشد آنگاه کمیت c را از رابطه زیر محاسبه می‌کنیم:

$$c = (n+1)\,p$$

③ مرحله کلیدی قسمت صحیح کمیت c را با r و قسمت اعشاری آن را با w نشان دهیم در اینصورت چندک به صورت زیر بدست می‌آید.

$$(1-w)\,x_{(r)} + w\,x_{(r+1)} = x_{(r)} + w\left[x_{(r+1)} - x_{(r)}\right]$$

Date: Subject:

تعریف سن یازده نفر از مراجعین به کلینیک اطفال به صورت

۱۱, ۱۰, ۹, ۱۳, ۱۴, ۱۲, ۱۳, ۱۵, ۱۷, ۱۶, ۱۷, ۱۸

گزارش شده است، پنجمین هفتک این داده ها را حساب کنید

① ابتدا مشاهدات را به صورت صعودی مرتب می کنیم

۱۰, ۱۱, ۱۲, ۱۳, ۱۳, ۱۴, ۱۵, ۱۶, ۱۷, ۱۷, ۱۸

② کسر متناظر با پنجمین هفتک را به صورت $p = \frac{5}{7}$ می نویسیم و c را محاسبه می کنیم.

$$C = (n+1)p \longrightarrow C = (11+1)\frac{5}{7} = 8 + \frac{4}{7}$$

$$12 \quad \frac{5}{7} = \frac{60}{7} = 8 + \frac{4}{7}$$

(r) (w)

③ حال با توجه به عدد بدست آمده $\frac{60}{7}$ آن را به صورت $8+\frac{4}{7}$ می نویسیم چون می خواهیم مقدار r و w را بدست آوریم

$r = 8 \quad w = \frac{4}{7}$

$$X + w\left[X_{(9)} - X_{(8)}\right] =$$

$$14 + \frac{4}{7}\left[17 - 14\right] = 14,571$$

Date: Subject:

معیارهای مقایسه

معیارهای مقایسه، معیارهایی هستند که مزایا پراکندگی حول پارامترهای مکانی یک توزیع را اندازه‌گیری می‌کنند. انواع معیارها واریانس، مانند قدر مطلق انحرافات، دامنه بین چارکی و فرست تغییرات، از معیارهای پراکندگی معروف می‌باشند.

دامنه

اگر a_n و ... و a_2 و a_1 یک نمونه n تایی بدست آمده از یک جامعه باشد، آنگاه دامنه نمونه اختلاف مقادیر کوچکترین و بزرگترین مشاهده می‌باشد.

$$R = Max(x_i) - Min(a_i)$$

با جایگزینی اندازه جامعه N بجای جایگزاری اندازه n دامنه جامعه را نیز می‌توان بدست آورد.

دامنه به وجود نقاط دور افتاده بشدت حساس است و در صورت وجود نقطه دورافتاده، مقدار دامنه به‌صورت غیرطبیعی بزرگ می‌شود. همچنین دامنه معیاری است که از اطلاعات موجود در تمام مشاهدات استفاده نمی‌کند و فقط به اطلاعات کوچکترین و بزرگترین مشاهده نیاز دارد. بنابراین ممکن است نمایندگی خوبی از پراکندگی مشاهدات نباشد، اما مزیت آن سادگی استفاده از آن می‌باشد.

نکته: یکی از مهمترین مزایا انواع مقدار، نسبت به معیارهای مانند دامنه را می‌توان با شکل بررسی کرد. همانطورکه از فرم تابعی انواع توزیع‌ها، مشخص است، انواع به شکل پراکندگی متفاوتی دارند. اما دامنه هر دو توزیع تقریبا با هم برابر است. درحالی که انواع معیار آنها متفاوت می‌باشد. دو شکل زیر توزیع‌های

Date: Subject:

با رابطه برابری انواری بسیار متفاوت را نشان می دهند.

جلسه هفتم ۹۲/۹/۲۲

تحلیل واریانس یک طرفه

فرض کنید k جامعه مختلف با توزیع نرمال و میانگین های به ترتیب $\mu_1, \mu_2, \dots, \mu_k$ و واریانس مشترک و مجهول σ^2 که داریم و می خواهیم فرض برابری میانگین های این جوامع را آزمون کنیم. در این حالت، فرض صفر و فرض مقابل آن را می توان به صورت زیر نمایش داد.

$$\begin{cases} H_0: \mu_1 = \mu_2 = \mu_k \\ H_1: \text{حداقل یکی از } \mu_i \text{ ها با بقیه متفاوت است} \end{cases}$$

برای آزمون این فرض k نمونه تصادفی n تایی از جوامع به صورت زیر در نظر می گیریم.

جامعه ۱	X_{11}	X_{12}	...	X_{1n}
جامعه ۲	X_{21}	X_{22}	...	X_{2n}
.	.	.	.	.
.	.	.	.	.
,	.	.	.	.
جامعه k	X_{k1}	X_{k2}		X_{kn}

X_{ij} مشاهده ی ام از نمونه جامعه i ام می باشد که به طور مستقل دارای توزیع نرمال با میانگین μ_i و واریانس σ^2 می باشد به طور عادی X_{ij} را می توان به صورت زیر نیز نمایش داد

$$X_{ij} = \mu_i + \varepsilon_{ij}$$

$$i = 1, 2, \dots, k$$

$$j = 1, 2, \dots, n$$

که ε_{ij} دارای توزیع نرمال با میانگین صفر و واریانس σ^2 می باشد.

Date: Sat Sun Mon Tue Thu Wed Fri Subject:

$$\varepsilon_{ij} = x_{ij} - \mu_i$$

$$i = 1, 2, \dots, k$$

$$j = 1, 2, \dots, n$$

برای اینکه امکان تصمیم‌گیری در مدل اثر به مدل‌های پیچیده‌تر تحلیل واریانس وجود داشته باشد، معمولاً آن را به صورت زیر نمایش می‌دهند:

$$x_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

$$i = 1, 2, \dots k$$

$$j = 1, 2, \dots n$$

که در اینجا μ را میانگین کل، τ_i را اثر جامعه i ام یا اثر تیمار i ام می‌نامند.

$$\sum_{i=1}^{k} \tau_i = 0$$

برای انجام آزمون، از تغییرپذیری موجود در داده‌ها استفاده می‌شود، بدین منظور ابتدا تغییرپذیری کل موجود در نمونه‌های همه جوامع را به صورت زیر نشان می‌دهیم.

$$\sum_{i=1}^{k} \sum_{j=1}^{n} \left(x_{ij} - \bar{\bar{x}} \right)^2$$

$\bar{\bar{x}}$: میانگین نمونه‌ای کل داده‌ها

Date: Sat Sun Mon Tue Thu Wed Fri Subject:

$$\bar{\bar{X}} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}}{kn}$$

ضریب تغییرات

اگر $X_1, X_2, \dots, X_n$ مقادیر بدست آمده از یک نمونه باشند، آنگاه ضریب تغییرات نمونه را می‌توان از رابطه زیر بدست آورد.

$$C.V = \frac{S}{\bar{X}}$$

نکته: چنانچه از رابطه ضریب تغییرات بدست یابی این معیار بدون واحد است در نتیجه برای مقایسه پراکندگی دو مشخصه آماری که واحدهای یکسانی ندارند می‌توان از این معیار استفاده کرد، همچنین برای مقایسه پراکندگی بین دو مشخصه که میانگین‌های متفاوتی دارند نیز از این معیار استفاده می‌شود.

مثال: میانگین و انحراف معیار وزن و قد یک گروه از افراد برحسب جدول زیر می‌باشد، کدام یک از این دو مشخصه (قد و وزن) پراکندگی بیشتری دارد؟

	قد (اینچ)	وزن (پوند)
$\bar{X}$	۶۹	۱۴۵
S	۲/۵	۱۲

$$C.V_1 = \frac{12}{69} = 0/1739$$

$$C.V_2 = \frac{25}{145} = 0/1724$$

Date: Sat Sun Mon Tue Thu Wed Fri Subject:

نکته: چون دو شاخص قدر وزن واحد متفاوتی دارند، بهترین شاخص برای اندازه‌گیری پراکندگی، ضریب تغییرات است، بنابراین پراکندگی نسبی قد و وزن در این نمونه تقریبا یکسان است، اما پراکندگی نسبی قد کمی بیشتر از وزن است.

معیارهای شکل تابع

گاهی اوقات دو توزیع مختلف معیارهای مکانی و پراکندگی یکسان دارند اما شکل آنها کاملا با هم متفاوت است، به عبارت دیگر نشاور اول و دوم این توزیع ها یکسان بوده اما نشاور های بالاتر آنها با هم فرق می‌کند، وجود معیار هایی که بتواند گشتاور های بالاتر را اندازه گیری کند می‌تواند در بسیاری از تحلیل ها مفید باشد.

چولگی (Skewness)

اگر $x_1, x_2, \dots, x_n$ یک نمونه n تایی از یک جامعه باشد آنگاه چولگی نمونه به صورت زیر تعریف می‌شود.

$$b = \frac{m_3}{(m_2)^{\frac{3}{2}}}$$

جلسه هشتم

Date: Subject: ۹۹/۹/۲۹

تمرین: فرض کنید که یک تاس شش وجهی استاندارد را یک بار پرتاب کنیم، احتمال اینکه عدد ۴ یا بیشتر بیاید چقدر است؟

$P(4) = \frac{3}{6} = \frac{1}{2}$

$\frac{1}{2}$ احتمال اینکه عدد ۴ یا بیشتر بیاید

تعداد حالات مطلوب ۴ و ۵ و ۶

تمرین: یک سکه را ۳ بار پرتاب می‌کنیم، احتمال اینکه دقیقاً دو بار پشت سکه بیاید را محاسبه کنید.

$2^3 = 8$ $\quad p(2) = \frac{3}{8}$

تمرین: یک دانشجو نمرات زیر را در ۵ درس پاس کرده است، میانگین، میانه و مد و واریانس نمرات این دانشجو را حساب کنید

۱۲ ۱۴ ۱۶ ۱۸ ۲۰

میانه

واریانس: $\frac{\sum (x_i - \mu)^2}{n}$

$\frac{12+14+16+18+20}{5} = \frac{80}{5} = 16$

$(12-16)^2 = 16$
$(14-16)^2 = 4$
$(16-16)^2 = 0$
$(18-16)^2 = 4$
$(20-16)^2 = 16$

واریانس $\frac{40}{5} = 8$

انحراف معیار $\sqrt{8} = 2.82$

Date: / / Subject:

تمرین: قد افراد در یک جامعه به طور نرمال توزیع شده باشد با میانگین ۱۷۰ سانتی متر و انحراف معیار ۱۰ سانتی متر احتمال اینکه فردی که تصادفی از این جامعه انتخاب می‌شود قدی بین ۱۶۰ و ۱۸۰ سانتی متر داشته باشد چقدر است؟ با توجه به آنکه میدانیم $P(-1 < Z < 1) = 0.682$

$$\frac{170-160}{10} = \frac{10}{10} = 1 \qquad \frac{170-180}{10} = \frac{-10}{10} = -1$$

تمرین: در یک مسابقه شطرنج هر بازیکن دو حرکت می‌تواند انجام دهد اگر دو بازیکن در هر نوبت، دو انتخاب متفاوت داشته باشند تعداد کل حالت‌های برای یک دور بازی چقدر خواهد بود.

$$2 \times 2 = 4$$