دعای مطالعه

اَللهُمَ اَخرِجنی مِن ظُلُماتِ الوَهمِ وَاَکرِمنی بِنُورِ الفَهمِ

اَللهُمَ افتِح عَلَینا اَبوابَ رَحمَتِکَ وَ انشُر عَلَینا خَزائِنِ عُلُومِکَ

بِرَحمَتِکَ یا اَرحَمَ الرّاحِمینَ

پروردگارا، خارج کن مرا از تاریکی‌های فکر و گرامی بدار به نور فهم

پروردگارا، بگشای بر ما درهای رحمتت را و بگستران گنج‌های دانشت را به امید رحمت

تو ای مهربان‌ترین مهربانان

آناهید گستر خلیلی

AGK

Anahid Gostar Khalili

بانک سوالات ده‌سالانه

آمـــارزیستـــی

«دکتری تخصصی»

(همراه با پاسخنامه تشریحی)

مولفین و گردآورندگان:

دکتـــر رســـول علیـــمی

«استادیار آمارزیستی، دانشگاه علوم‌پزشکی تربت‌حیدریه»

دکتـــر فاطـــمه خراشادیـــزاده

«استادیار آمارزیستی، دانشگاه علوم‌پزشکی نیشابور»

دکتـــر الهام شعرباف عیدگاهی

«دکتری تخصصی آمار زیستی، معاونت پژوهشی و فناوری دانشگاه علوم‌پزشکی مشهد»

سمیـــه برزنـــونی

«کارشناسی ارشد آمار زیستی، معاونت آموزش، تحقیقات و فناوری دانشگاه علوم‌پزشکی تربت‌حیدریه»

عنوان و نام پدیدآور	: AGK بانک سوالات ده‌سالانه آمارزیستی «دکتری تخصصی» (همراه با پاسخنامه تشریحی)/ مولفین و گردآورندگان رسول علیمی ... [و دیگران].
مشخصات نشر	: تهران: آناهید گستر خلیلی، ۱۴۰۳.
مشخصات ظاهری	: ۶۸۰ص. : جدول، نمودار.
شابک	: 978-622-4907-01-1
وضعیت فهرست نویسی	: فیپا
یادداشت	: مولفین و گردآورندگان رسول علیمی، فاطمه خراشادیزاده، الهام شعرباف عیدگاهی، سمیه برزنونی.
موضوع	: زیست‌سنجی -- آزمون‌ها و تمرین‌ها (عالی) Biometry -- Examinations, questions, etc (Higher) دانشگاه‌ها و مدارس عالی -- ایران -- آزمون دوره‌های تحصیلات تکمیلی Universities and colleges -- Iran -- Graduate Record Examination
شناسه افزوده	: علیمی، رسول، ۱۳۶۵- ، گردآورنده
رده‌بندی کنگره	: QH۳۲۳/۵
رده‌بندی دیویی	: ۵۷۴/۱۵۱۹۵
شماره کتابشناسی ملی	: ۹۷۴۶۱۲۲

انتشارات آناهید گستر خلیلی

نام کتاب: AGK بانک سوالات ده‌سالانه آمارزیستی «دکتری تخصصی» (همراه با پاسخنامه تشریحی)

مولفین و گردآورندگان: دکتر رسول علیمی

دکتر فاطمه خراشادیزاده . دکتر الهام شعرباف عیدگاهی . سمیه برزنونی

ناشر: آناهید گستر خلیلی

نوبت و سال چاپ: اول . ۱۴۰۳

شمارگان: ۱۰۰۰

چاپ و صحافی: گیلان

مدیر تولید: اقبال شرقی

مدیر فنی و هنری: مریم آرده

تایپ و صفحه‌آرایی: بیتا اندوژفر

بهاء: ۵۵۰۰۰۰ تومان

آموزشگاه/ فروشگاه: ۰۲۱۶۶۵۶۸۶۲۱

مرکز پخش: ضلع جنوب‌غربی میدان انقلاب . جنب بانک پارسیان . مجتمع تجاری پارس . طبقه اول

مرکز فروش: ۰۲۱۶۶۵۶۹۲۱۶ مدیر فروش: ۰۹۹۰۲۱۰۵۲۹۲

www.agkins.com agkpublisher@gmail.com

طلیعه سخن مؤلفین:

موفقیت در آزمون‌های کنکور و رقابت فشرده برای ورود به مقاطع تحصیلی بالاتر، نیازمند تلاش مستمر، مطالعه بیش‌تر و استفاده از منابع مناسب جهت دستابی به این هدف می‌باشد. مرور و یادگیری سوالات و آزمون‌های سال‌های گذشته به‌عنوان منبعی مناسب برای آمادگی جهت آزمون‌های آینده می‌تواند بسیار مفید و راهگشا باشد.

این مجموعه به‌عنوان یکی از اولین کتاب‌ها با هدف کمک به متقاضیان شرکت‌کننده در کنکور دکتری رشته آمار زیستی تهیه و تدوین شده است.

در نگارش این مجموعه سعی شده است تا سوالات مربوط به آزمون‌های سال‌های ۱۳۸۷ الی ۱۴۰۳ در مباحث آمار زیستی، روش‌های آماری، تحلیل بقاء و کارآزمایی بالینی با راه حل تقریباً کامل و همراه با نکات ضروری ارائه شوند تا دانشجویان عزیز بتوانند با بهره‌گیری از آن درصد بالایی از نمرات این دروس را در کنکور کسب نمایند. امید که این مجموعه توانسته باشد سهم کوچکی در ارتقای علمی خوانندگان محترم داشته باشد.

بی‌شک در این‌گونه مجموعه‌ها که در برگیرنده حجم بالایی از مطالب هستند بروز برخی از نواقص و کاستی‌ها اجتناب‌ناپذیر است. ارائه هر گونه راهکار و پیشنهاد ارزشمندی از سوی خوانندگان محترم این مجموعه در برطرف کردن نقایص احتمالی و بهبود آن می‌تواند ما را در ارتقای این مجموعه یاری دهد.

در این راستا شما می‌توانید از طریق پست الکترونیکی rasulalimi@yahoo.com ما را از نظرات سازنده خود بهره‌مند سازید.

با آرزوی توفیقات روزافزون

طلیعه سخن مؤلفین:

در سال‌های اخیر، کتاب‌های متفاوتی در زمینه‌ی حل سوالات کنکور در رشته‌های مختلف به چاپ رسیده است. اما کتابی با پاسخ کاملا تشریحی و به‌روز در رشته آمار زیستی و در مقطع دکتری موجود نمی‌باشد. از آن‌جا که هر سال تعداد کثیری از دانشجویان متقاضی ادامه تحصیل در رشته آمار زیستی و به طبع آن؛ شرکت در کنکور وزارت بهداشت هستند، ما بر آن شدیم که در کتاب حاضر با پاسخ تشریحی سوالات کنکور دکتری آمار زیستی وزارت بهداشت از سال ۱۳۸۷-۱۴۰۳ راه‌گشایی برای این عزیزان باشیم.

این کتاب شامل حل سوالات دروس استنباط آماری، روش‌های آمار زیستی، تحلیل چند متغیره، تحلیل بقا و کارآزمایی بالینی است. کتب مرجع حل سوالات به شرح زیر است:

Applied Linear Statistical Models- Micheal H. Kutner, Christopher J. Nachtsheim, John Neter, William Li-Fifth Edition.

Applied Multivariate Techniques. Subhash Sharma- Wiley.

تحلیل آماری چند متغیری کاربردی، ریچارد آ جانسون، ترجمه دکتر حسینعلی نیرومند، انتشارات: دانشگاه فردوسی مشهد.

Survival Analysis- David G. Kleinbaum, MitchelKlein- Third Edition.

کارآزمایی‌های بالینی، استوارت پوکاک، ترجمه دکتر آیت‌الهی.

آمار نظری پردازش، تالیف عبدالرضا محمدی (جلد اول و جلد دوم).

امید است که کتاب حاضر گامی مثبت در جهت پیشرفت شما عزیزان برداشته باشد. در جهت هر آن‌چه بهتر کردن این کتاب خواهشمندیم نظرات و پیشنهادات خود را با در میان بگذارید.

با تشکر

الهام شعرباف عیدگاهی
elham.Shaarbaf@gmail.com

سمیه برزنونی
s.barzanouni@gmail.com

فهرست مطالب

عنوان **صفحه**

فهرست مطالب

عنوان **صفحه**

سال

۸۸-۸۷

سوالات

دکتری تخصصی (Ph.D) آمارزیستی

رسول علیمی- فاطمه خراشادیزاده

آمار ریاضی و احتمالات

۱. نمونه تصادفی $x_1, x_2, \ldots, x_n$ را از توزیعی با واریانس σ^2 در نظر می‌گیریم. فرض کنیم $\bar{x} = \sum_{i=1}^{n} X_i / n$. امید ریاضی $T = \sum_{i=1}^{n} (x_i - \bar{x})^2$ برابر است با:

۱) σ^2 ۲) $n\sigma^2$

۳) $(n-1)\sigma^2$ ۴) $\sigma^2 + n$

۲. x_1, x_2, x_3 سه متغیر تصادفی مثبت و هم توزیع می‌باشند. فرض کنید، $\bar{X} = \frac{x_1 + x_2 + x_3}{3}$.

$E(\frac{x_1}{\bar{x}})$

۱) وجود ندارد. ۲) برابر $\frac{1}{3}$ است.

۳) برابر ۳ است. ۴) برابر یک است.

۳. جعبه‌ای شامل ۶ مهره به شماره‌های ۱ تا ۶ می‌باشد. یک نمونه تصادفی چهارتایی بدون جایگذاری از این جعبه بیرون می‌آوریم. احتمال این‌که ۳ مهره در این نمونه دارای شماره‌های زوج باشند، برابر است با:

۱) $\frac{1}{10}$ ۲) $\frac{8}{15}$ ۳) $\frac{2}{10}$ ۴) $\frac{2}{5}$

۴. متغیر تصادفی X دارای چگالی $f(x)=\frac{۱}{\pi(۱+x^۲)}$، $-\infty<x<+\infty$ می‌باشد. تابع توزیع این متغیر تصادفی برابر است با:

۱) $\frac{۱}{۲}+\frac{۱}{\pi}\arctan x$ ۲) $\frac{۱}{\pi}\arctan x$

۳) $\frac{۱}{۲}-\arctan x$ ۴) $\frac{۱}{۲}+\arctan x$

۵. X و Y دو متغیر تصادفی مستقل و هر یک دارای توزیع پواسن با میانگین ۲ و ۳ می‌باشد. $P(x+y=۰)$ برابر است با:

۱) $e^{-۲}+e^{-۳}$ ۲) $e^{+۶}$

۳) $e^{۵}$ ۴) $e^{-۵}$

۶. فرض کنید X دارای میانگین $\mu=۱$ و واریانس $\sigma^۲=\frac{۱}{۴}$ باشد. کران بالا، طبق نامساوی چبیچف، برای $P(|x-۱|\geq ۲)$ کدام عدد است؟

۱) $\frac{۱}{۴}$ ۲) $\frac{۱}{۲}$ ۳) $\frac{۱}{۱۶}$ ۴) $\frac{۱}{۸}$

۷. متغیر تصادفی X دارای چگالی یکنواخت در فاصله (۱ و ۱-) می‌باشد. $P(|x|>\sqrt{۳Var(x)})$ برابر است با:

۱) صفر ۲) یک ۳) $\frac{۱}{۲}$ ۴) $\frac{۱}{۳}$

۸. جفت تصادفی (X , Y) دارای چگالی توأم $f(x,y)=(x+y+۲)/۸$ است، به‌طوری‌که $|x|<۱$، $P(X<Y)$ برابر است با:

۱) ۱ ۲) $\frac{۱}{۲}$ ۳) $\frac{۱}{۲}$ ۴) $\frac{۱}{۴}$

۹. نمونه تصادفی $x_۱, x_۲, ..., x_n$ را از توزیعی با میانگین ۲ و واریانس ۴ در نظر می‌گیریم. $\frac{۱}{n}\sum_{i=۱}^{n}X_i^۲$ هنگامی‌که $n\to\infty$، به کدام در توزیع میل می‌کند؟

۱) به یک متغیر نرمال ۲) به عدد ۸

۳) به عدد ۲ ۴) به عدد ۴

۱۰. $x_۱, x_۲, ..., x_n$ یک نمونه تصادفی از توزیعی با واریانس متناهی $\sigma^۲$ می‌باشد. $S^۲=\sum_{i=۱}^{n}(x_i-\bar{x})^۲/n$ را به‌عنوان برآوردکننده $\sigma^۲$ به‌کار می‌بریم. با کدام n قدر مطلق اریبی $S^۲$ برابر یک دهم $\sigma^۲$ می‌شود؟

۱) ۱۰۰ ۲) ۱۱ ۳) ۹ ۴) ۱۰

۱۱. دو متغیر تصادفی $x_۱$ و $x_۲$، به‌ترتیب با واریانس‌های ۱ و ۴، مستقل می‌باشند. $U=x_۱+x_۲$ و $W=x_۱+cx_۲$ برای چه مقدار c ناهمبسته می‌باشند؟

۱) $\frac{۱}{۲}$ ۲) $\frac{۱}{۴}$ ۳) ۴ ۴) $-\frac{۱}{۴}$

۱۲. میان دو متغیر تصادفی X و Y، با گشتاور دوم متناهی، رابطه $Y - bX = aX^{۲} + c$ برقرار است. ضریب همبستگی X و Y در چه صورت همواره برابر یک می‌شود؟

۱) $a = ۰$ ۲) $b > ۰, a = ۰$

۳) $b < ۰, a = ۰$ ۴) $b = ۰, a = ۰$

۱۳. $x_۱, x_۲, x_۳$ یک نمونه تصادفی از توزیع برنولی با پارامتر P می‌باشد. $E(x_۱ \mid x_۱ + x_۲ + x_۳)$ برابر است با:

۱) $\overline{x}$ ۲) $۳\overline{x}$ ۳) $\frac{۱}{۳}\overline{x}$ ۴) $۲\overline{x}$

۱۴. متغیر تصادفی X دارای چگالی $۰ < x < ۱$ و $f(x) = ۲x$ می‌باشد. چارک اول کدام است؟

۱) ۱ ۲) ۰/۲۵ ۳) ۰/۵ ۴) ۰/۷۵

۱۵. برای یک نمونه تصادفی ده‌تایی از توزیع $N(۰, \sigma^{۲})$، $\sum_{i=۱}^{۱۰} X_i^{۲} = ۱/۶$ را مشاهده کرده‌ایم. MLE انحراف معیار کدام عدد است؟

۱) ۰/۱۶ ۲) ۰/۴ ۳) ۰/۲ ۴) ۰/۰۴

۱۶. برای پارامتر $\sigma^{۲}$ (واریانس)، با استفاده از داده‌ها، یک فاصله اطمینان (۰/۴۹، ۰/۲۵) با ضریب اطمینان ۰/۸۱ به‌دست آمده است. فاصله اطمینان و ضریب اطمینان با همین داده‌ها برای σ کدام است؟

۱) (۰/۴۹ و ۰/۲۵) با ضریب ۰/۹ ۲) (۰/۷ و ۰/۵) با ضریب ۰/۹

۳) (۰/۷ و ۰/۵) با ضریب ۰/۸۱ ۴) (۰/۷ و ۰/۵) با ضریب ۰/۸۲

۱۷. یک نمونه تصادفی از توزیعی با پارامتر θ داریم، U و W و T به‌ترتیب آماره بسنده، برآوردگر گشتاوری و MLE برای θ می‌باشد. همواره داریم؟

۱) $U = W$ ۲) $U = T$

۳) $W = T$ ۴) T تابعی از U

۱۸. $x_۱, x_۲, \ldots, x_n$ یک نمونه تصادفی از توزیع $N(\mu, \sigma^{۲})$ می‌باشد. فرض کنید $\overline{x}$ میانگین این نمونه و $S^{۲}$ واریانس آن باشد. $P((\overline{x} - \mu)(S^{۲} - \sigma^{۲}) > ۰)$ برابر است با:

۱) $\frac{۱}{۲}$ ۲) $\frac{۱}{۴}$ ۳) $\frac{۱}{۸}$ ۴) ۱

۱۹. یک نمونه تصادفی $x_۱, x_۲, \ldots, x_n$ از توزیع پواسن با میانگین λ در نظر می‌گیریم. فرض کنید $S^{۲}$ واریانس نااریب این نمونه باشد. کران پایین برای واریانس $S^{۲}$ برابر است با:

۱) λ ۲) $\frac{۱}{\lambda}$ ۳) $\frac{n}{\lambda}$ ۴) $\frac{\lambda}{n}$

۲۰. فرض کنید $z_۱, z_۲, z_۳, z_۴$ یک نمونه تصادفی از توزیع $N(۰,۱)$ باشد. کدام‌یک از متغیرهای زیر توزیع F دارد؟

۱) $(\frac{z_۱ + z_۲}{z_۲ + z_۲})^{۲}$ ۲) $(\frac{z_۱ + z_۳}{z_۳ + z_۴})^{۲}$

۳) $(\frac{z_۱ - z_۳}{z_۳ + z_۴})^{۲}$ ۴) $(\frac{z_۱ - z_۲}{z_۱ + z_۲})^{۲}$

۲۱. توان آزمون، برای آزمون فرض ساده H_0 در برابر فرض ساده H_1 چیست؟

۱) احتمال رد H_0 هنگامی‌که H_1 نادرست باشد.

۲) احتمال قبول H_0 هنگامی‌که H_1 نادرست باشد.

۳) احتمال رد H_0 هنگامی‌که H_1 درست باشد.

۴) احتمال قبول H_0 هنگامی‌که H_1 درست باشد.

۲۲. می‌خواهیم با مشاهده یک X، فرض این‌که X دارای چگالی $4X^3$ می‌باشد، در برابر این‌که X دارای چگالی $3X^2$ می‌باشد را بیازماییم $0 < x < 1$، اگر خطای نوع اول $\frac{1}{16}$ باشد ناحیه بحرانی بر حسب X به چه صورت است؟

۱) $X < 2$ ۲) $X < \frac{1}{2}$ ۳) $X > 2$ ۴) $X > \frac{1}{2}$

۲۳. نمونه تصادفی $x_1, x_2, ..., x_n$ از توزیع $N(0, \sigma^2)$ را در نظر می‌گیریم. کدام آماره برای σ^2 بسنده است؟

۱) $\bar{x}$ ۲) $\bar{x}^2$

۳) $(\bar{x})^3$ ۴) $e^{\overline{x^2}}$

۲۴. یک نمونه تصادفی از چگالی $x > 0$ و $f(x_i\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}$ در نظر می‌گیریم. برآورد کارا (efficient) برای θ کدام است؟

۱) $\bar{x}$ ۲) $\frac{1}{\bar{x}}$ ۳) $e^{\bar{x}}$ ۴) $\log \bar{x}$

۲۵. یک نمونه تصادفی از توزیع $N(\mu, 1)$ در نظر می‌گیریم. کدام‌یک از آزمون‌های زیر UMP می‌باشد؟

۱) آزمون $\mu = 0$ در برابر $\mu \neq 0$

۲) آزمون $\mu = 0$ در برابر $0 < \mu < \frac{1}{2}$

۳) آزمون $\mu = 0$ در برابر $-\frac{1}{2} < \mu < 0$

۴) آزمون $\mu = 0$ در برابر $\mu > 0$

۲۶. متغیر تصادفی گسسته X دارای تابع مولد گشتاور $M(t) = \frac{1}{4}e^{-t} + \frac{1}{2} + \frac{1}{4}e^{t}$ می‌باشد. Var(X) برابر است با:

۱) $\frac{1}{2}$ ۲) $\frac{1}{4}$ ۳) ۱ ۴) $\frac{1}{8}$

۲۷. $x_1, x_2, ..., x_n$ یک نمونه تصادفی از توزیع نرمال با میانگین μ می‌باشد. $P(\bar{x} < \mu, \bar{x} > x_1)$ برابر است با:

۱) $\frac{1}{2}$ ۲) $\frac{1}{4}$ ۳) $\frac{1}{8}$ ۴) ۱

۲۸. x_1, x_2, x_3 به‌ترتیب دارای انحراف معیارهای σ_3 و σ_4 و σ_1 می‌باشد و دو به دو دارای ضریب همبستگی $\frac{1}{2}$ هستند. $var(\frac{x_1}{\sigma_2} + \frac{x_2}{\sigma_2} + \frac{x_3}{\sigma_3})$ برابر است با:

۱) ۴ ۲) 66 ۳) ۸ ۴) ۱۲

روش‌های آماری

۲۹. اگر در رگرسیون خطی ساده Y روی X که تمام مفروضات آن برقرار است ضریب تعیین برابر یک باشد، کدام گزینه حتماً صحیح است؟

۱) شیب خط رگرسیونی برابر یک است.

۲) شیب خط رگرسیونی منفی است.

۳) مجموع مربعات رگرسیونی با مجموع مربعات خطا برابر است.

۴) مجموع مربعات خطا صفر است.

۳۰. در تحلیل رگرسیونی، تغییر واحد اندازه‌گیری متغیر پاسخ کدام‌یک از موارد زیر را تغییر نمی‌دهد؟

۱) برآورد ثابت رگرسیونی $(\hat{\beta}_0)$

۲) ضریب تعیین (R^2)

۳) برآورد واریانس شسب رگرسیونی $(\mathrm{Var}(\hat{\beta}_1))$

۴) برآورد شیب رگرسیونی $(\hat{\beta}_1)$

۳۱. اگر احتمال بهبودی از یک بیماری P در نظر گرفته شود و در یک تحقیق فقط یک نفر از بین شش بیمار بهبودی پیدا کرده باشد، P-Value آزمون $H_0 : p = 0/5$ در مقابل $H_1 : P < 0/5$ برابر است با:

۱) $\frac{6}{64}$ ۲) $\frac{7}{64}$ ۳) $\frac{12}{64}$ ۴) $\frac{14}{64}$

۳۲. کدام‌یک از مدل‌های زیر قابل تبدیل به حالت خطی نیست؟

۱) $y_i = \frac{1}{1+e^{\beta_0+\beta_1 x_i+e_i}}$ ۲) $y_i = \beta_0+\beta_1\frac{1}{x_i^2-1}$

۳) $y_i = \beta_0+\beta_1 e^{\beta_2 x_i+e_i}$ ۴) $y_i = \beta_0 e^{\beta_1 x_i+e_i}$

۳۳. برای آزمون فرضیه $H_0 : \mu_1 = \mu_2$ در مقابل $H_1 : \mu_1 \neq \mu_2$ با توان ۹۰ درصد و خطای نوع اول ۵ درصد اگر اندازه اثر (effect size) را نصف انحراف معیار مشترک در نظر بگیریم، حداقل نمونه مورد نیاز در هر گروه برابر است با: $(Z_{0/975}+Z_{0/90})^2 = 10/5$

۱) ۴۲ ۲) ۸۴ ۳) ۱۶۸ ۴) ۳۳۶

۳۴. از ویژگی‌های طرح بلوک تصادفی ناقص متعادل کدام است؟

۱) تعداد بلوک‌ها و درمان‌ها (Treatments) برابر نمی‌باشد.

۲) تعداد بلوک‌ها و درمان‌ها (Treatments) برابر می‌باشد.

۳) کلیه درمان‌ها (Treatments) در تمام بلوک‌ها برابر و به دفعات یکسان ظاهر می‌شود.

۴) تکرار هر درمان (Treatment) در کلیه بلوک‌ها برابر و کم‌تر از تعداد کل بلوک‌ها است.

۳۵. کدام‌یک از شاخص‌های زیر مربوط به ارزیابی برازش مدل رگرسیون چندگانه نیست؟

۱) AIC ۲) R_a^2 ۳) PRESS ۴) VIF

۳۶. اگر مدل کامل (Full model) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ و مدل کاهش یافته (reduced model) به‌صورت $y_i = \beta_0 + \varepsilon_i$ باشد، کدام‌یک از گزینه‌های زیر آماره آزمون برازش مدل کاهش یافته نسبت به مدل کامل است؟

۱) $F^* = \frac{SSE(F) - SSE(R)}{df_F - df_R} \div \frac{SSE(F)}{df_F}$

۲) $F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F}$

۳) $F^* = \frac{SSE(F) - SSE(R)}{df_F - df_R} \div \frac{SSE(R)}{df_R}$

۴) $F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(R)}{df_R}$

۳۷. در یک طرح مربع لاتین ۳×۳ با ۲ تکرار درجه آزادی خطا چقدر است؟

۱) ۱۰ ۲) ۸ ۳) ۲ ۴) ۱

۳۸. اگر MSE میانگین مربعات خطا در رگرسیون باشد و توزیع خطاها نرمال نباشد، در این‌صورت:

۱) $\sqrt{MSE}$ برآوردکننده نیرومند (Robust) σ (انحراف معیار خطا) نیست.

۲) $\sqrt{MSE}$ برآوردکننده نیرومند (Robust) σ (انحراف معیار خطا) است.

۳) $\frac{MSE}{n}$ برآوردکننده نیرومند (Robust) σ (انحراف معیار خطا) است.

۴) MSE برآوردکننده نیرومند (Robust) σ (انحراف معیار خطا) است.

۳۹. اندازه نگرش و عملکرد ۶ فرد در مورد مصرف سیگار در مقیاس صفر تا ۱۰۰ نمره‌گذاری شده است. اگر در بین داده‌ها فقط سه زوج ناهماهنگ (Discordant) داشته باشیم و هیچ زوج هم رتبه‌ای نداشته باشیم ضریب همبستگی کندال برابر است با:

۱) ۰/۲ ۲) ۰/۵ ۳) ۰/۶ ۴) ۰/۸

۴۰. در مدل ANOVA با اندازه‌گیری‌های تکراری در r زمان متوالی، منظور از شرط کرویت (Sphericity) چیست؟

۱) کوواریانس مشاهدات در دو زمان متوالی یکسان است.

۲) واریانس مشاهدات در همه تکرارهای زمانی یکسان است.

۳) همبستگی مشاهدات در همه تکرارهای زمانی یکسان است.

۴) همبستگی مشاهدات در دو زمان متوالی یکسان است.

۴۱. در مدل ANOVA با اثرات تصادفی $y_{ij} = \mu_i + \varepsilon_{ij}$ که در آن μ_iها و ε_{ij}ها به‌ترتیب متغیرهای تصادفی مستقل به‌صورت $N(\mu_0, \sigma_\mu^2)$ و $N(0, \sigma^2)$ و $j = 1, \ldots, n, i = 1, \ldots, r$ هستند. کدام گزینه سهم تغییرپذیری μ_iها را آزمون می‌کند؟

۱) $\frac{\sigma_\mu^2}{\sigma^2}$ ۲) $\frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2}$

۳) σ_μ^2 ۴) $\frac{\sigma^2}{\sigma_\mu^2}$

۴۲. در رگرسیون چندگانه X_1، X_2 و X_3 روی y، مقدار $SSR(X_3 \mid x_1, x_2)$ برابر با کدام‌یک از گزینه‌های زیر است؟

۱) $SSR(x_1 x_2, x_3) - SSE(x_1, x_2)$

۲) $SSR(x_1, x_2) - SSE(x_1, x_2, x_3)$

۳) $SSR(x_1, x_2, x_3) - SSE(x_1, x_2)$

۴) $SSR(x_1, x_2) - SSR(x_1, x_2, x_3)$

۴۳. اگر در یک مدل رگرسیون چندگانه خطی با p پارامتر، شاخص مالو (Mallow) را با C_p نشان دهیم. در کدام‌یک از حالات زیر اعتبار (Validation) مدل بالاتر است؟

۱) $C_p \approx p$ ۲) $C_p \leq p$ ۳) $C_p > p$ ۴) $C_p \geq p$

۴۴. کدام‌یک از گزینه‌های زیر، نتیجه قضیه گوس- مارکف در رگرسیون خطی ساده است؟

۱) برآوردهای حداقل مربعات ضرایب مدل، ناریب، سازگار و در بین همه برآوردهای ناریب دارای کم‌ترین واریانس هستند.

۲) برآوردهای حداکثر درست‌نمایی ضرایب مدل، ناریب، سازگار و در بین همه برآوردهای ناریب دارای کم‌ترین واریانس هستند.

۳) برآوردهای حداقل مربعات ضرایب مدل، ناریب و در بین همه برآوردهای ناریب دارای کم‌ترین واریانس هستند.

۴) برآوردهای حداکثر درست‌نمایی ضرایب مدل، ناریب و در بین همه برآوردهای ناریب دارای کم‌ترین واریانس هستند.

۴۵. در تبدیل باکس- کاکس (Box-Cox) برای مدل $y_i^{\lambda} = \beta_0 + \beta_1 x_i + \varepsilon_i$ جهت برآورد λ از کدام روش استفاده می‌شود؟

۱) روش حداکثر درست‌نمایی (Maximum likelihood)

۲) روش کم‌ترین مربعات معمولی (Ordinary least square)

۳) روش کم‌ترین مربعات وزنی (Weighted least square)

۴) روش ناپارامتری Lowess

۴۶. اگر در برازش مدل رگرسیون چندگانه Y روی X_1 و X_2 در یک نمونه تصادفی $n = 19$ و $\sum (y_i - \hat{y})^2 = 64$ باشد، درجه آزادی واریانس خطا برابر است با:

۱) ۲ ۲) ۱۶ ۳) ۱۷ ۴) ۱۸

۴۷. برآورد حجم نمونه با استفاده از روش نمونه‌گیری تصادفی ساده، جهت برآورد شیوع یک بیماری در شهری بزرگ برابر ۱۰۰۰ نفر محاسبه گردیده است. اگر به‌جای نمونه‌گیری تصادفی ساده، نمونه‌گیری خوشه‌ای از خانوارها به‌عمل آوریم، مشخص نمایید. در صورتی‌که متوسط بعد خانوار برابر ۴ و اثر طرح (Design effect) برابر ۳/۲ باشد، حدود چند خانوار بایستی مورد مطالعه قرار گیرند؟

۱) ۲۵۰ ۲) ۴۰۰ ۳) ۶۴۰ ۴) ۸۰۰

۴۸. در رگرسیون چندگانه روش حداقل مربعات وزنی (Weighted least squares) معمولاً در کدام‌یک از موارد زیر استفاده می‌شود؟

۱) در صورتی‌که هم خطی چندگانه وجود داشته باشد.

۲) مقادیر پاسخ، پرت و دور افتاده باشد.

۳) توزیع خطا نرمال نباشد.

۴) واریانس خطا ثابت نباشد.

۴۹. فرض کنید تعداد تصادفات منجر به مرگ در طول هفته دارای توزیع پواسن باشد. اگر تعداد تصادفات در ۱۰ هفته پیاپی برابر با ۳، ۱، ۲، ۴، ۳، ۱، ۰، ۳، ۲، ۱ باشد، احتمال مشاهده‌ی کدام مقدار در طول یک هفته بیش‌تر است؟

۱) ۱ ۲) ۲ ۳) ۳ ۴) ۴

۵۰. در یک طرح آزمایشی کاملاً تصادفی، در صورتی‌که تعداد گروه‌ها ۴ و تعداد مشاهدات در هر گروه مساوی و برابر ۶ باشد، درجه‌ی آزادی خطا برابر است با:

۱) ۵ ۲) ۱۵ ۳) ۱۸ ۴) ۲۰

۵۱. در یک جدول توافقی با $r > 2$ سطر و $c > 2$ ستون، تعداد سطرها یا ستون‌ها را کاهش می‌دهیم. در این‌صورت پراکندگی توزیع با استفاده از ملاک کای- دو:

۱) افزایش می‌یابد. ۲) کاهش می‌یابد.

۳) تغییر نمی‌کند. ۴) برابر $(r-1)(c-1)$ می‌شود.

۵۲. اگر SSE(F) خطای مدل رگرسیونی کامل (Full model) و SSE(R) خطای مدل کاهش یافته (Reduced model) باشد، کدام‌یک از روابط زیر همواره برقرار است؟

۱) $SSE(F) < SSE(R)$ ۲) $SSE(F) \le SSE(R)$

۳) $SSE(F) > SSE(R)$ ۴) $SSE(F) \ge SSE(R)$

۵۳. جهت بررسی وضعیت سلامت و بیماری در خانوارها در یک شهر بزرگ که لیست کامل خانوارها در دست نمی‌باشد، کدام‌یک از روش‌های نمونه‌گیری مناسب‌تر است؟

۱) تصادفی ساده ۲) منظم (سیستماتیک)

۳) طبقه‌بندی شده ۴) خوشه‌ای

۵۴. متغیر تصادفی x دارای میانگین نامعلوم θ و واریانس معلوم σ^2 می‌باشد. اگر $\hat{\theta}$ برآوردکننده‌ای اریب از θ باشد، مقدار اریبی چقدر است؟

۱) $\hat{\theta}-\theta$ ۲) $E(\hat{\theta})-\theta$

۳) $\frac{E(\hat{\theta})-\theta}{\sigma}$ ۴) $\frac{\hat{\theta}-\theta}{\sigma}$

۵۵. در مدل رگرسیون خطی Y روی X_1 و X_2، وقتی‌که X_1 متغیری کمی و X_2 متغیری کیفی در سه رده‌ی مختلف باشد، مشخص نمایید در ازاء مقادیر مختلف X_2، این مدل دارای چند شیب و چند عرض از مبدأ می‌باشد؟

۱) یک شیب و سه عرض از مبدأ ۲) یک شیب و دو عرض از مبدأ

۳) سه شیب و یک عرض از مبدأ ۴) دو شب و دو عرض از مبدأ

۵۶. در مدل رگرسیون لجستیک $Log\frac{\pi}{1-\pi} = \beta_0 + \beta_1 X$، در ازای دو واحد افزایش در متغیر پیشگوی X، نسبت بخت e برابر شده است. مقدار ضریب β_1 مدل چقدر است؟

۱) $\frac{1}{2}$ ۲) ۲ ۳) $\frac{e}{2}$ ۴) $2e$

۵۷. واریانس برآورد میانگین در نمونه‌گیری تصادفی ساده با جایگذاری در مقایسه با نمونه‌گیری تصادفی ساده بدون جایگذاری:

۱) کوچک‌تر یا مساوی است. ۲) مساوی است.

۳) کوچک‌تر است. ۴) بزرگ‌تر است.

۵۸. در مدل رگرسیون خطی که از مبدأ عبور می‌کند، اگر $i = 1$ و...۲ و n ، $e_i = \hat{y}_i$ باشد، برآورد ناریب واریانس خطای مدل کدام است؟

۱) $\frac{\sum_{i=1}^{n} e_i^2}{n-2}$ ۲) $\frac{\sum_{i=1}^{n} e_i^2}{n-1}$ ۳) $(n-1)\frac{\sum_{i=1}^{n} e_i^2}{n-20}$ ۴) $n\frac{\sum_{i=1}^{n} e_i^2}{n-1}$

۵۹. در مدل رگرسیون $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i, i = 1,2,...,n$، برای انجام آزمون $H_0 : \beta_1 = \beta_2$ در مقابل $H_1 : \beta_1 \neq \beta_2$، آماره‌ی آزمون F دارای چه درجه‌ی آزادی می‌باشد؟

۱) $F(1, n-3)$ ۲) $F(2, n-3)$

۳) $F(2, n-4)$ ۴) $F(1, n-4)$

۶۰. در یک مدل تحلیل واریانس سه طرفه اگر عامل‌ها به‌ترتیب دارای ۴، ۳ و ۲ رده (Category) و حجم نمونه برای هر یک از حالات ترکیبی رده‌های این سه عامل که به‌طور تصادفی انتخاب شده‌اند ۱۱ نمونه باشد، درجه آزادی برای مجموع مجذورات باقی‌مانده برابر است با:

۱) ۶۰ ۲) ۲۶۳ ۳) ۱۲۰ ۴) ۲۴۰

۶۱. در ارزیابی تأثیر یک روش گروه درمانی بر کیفیت زندگی سالمندان (y) از طرح قبل و بعد روی نمونه‌ای تصادفی از ۲۰ سالمند استفاده شده است. کدام مدل آماری زیر می‌تواند مناسب باشد؟ $(i = 1,2, j = 1,...,20)$

۱) طرح کاملاً تصادفی ۲) طرح دو عاملی

۳) طرح بلوک‌های تصادفی ۴) طرح آشیانی (nested)

۶۲. محققی می‌خواهد تأثیر نوعی بیماری (D) با ۴ رده (Category) و جنس (S) را بر کیفیت زندگی سالمندان بررسی نماید. اگر حجم نمونه برای هر یک از حالاتی که با ترکیب رده‌های ۲ متغیر فوق تشکیل می‌شود ۸ سالمند باشد، آن‌گاه امید ریاضی اثر متقابل بیماری و جنس عبارت است از:

(σ_e^2 و $\sigma^2 DS$ به ترتیب واریانس خطا و اثرات متقابل‌اند).

۱) $\sigma_e^2 + 8\sigma^2 DS$ ۲) σ_e^2

۳) $\sigma_e^2 + 32\sigma^2 DS$ ۴) $\sigma_e^2 + 16\sigma^2 DS$

۶۳. در یک تحلیل کوواریانس مدل برازش شده به‌صورت $y_{ij} = \alpha_i + \beta_i x + \varepsilon_{ij}$ می‌باشد. در این‌صورت برآورد $E(Y \mid X)$ به‌صورت زیر خواهد بود:

($\bar{Y}_{..}, \bar{X}_{..}$ به‌ترتیب میانگین کل X و Y و $\bar{X}_i$ و $\bar{Y}_i$ میانگین X و Yها در گروه i ام و b_i برآورد β_i است).

۱) $\bar{Y}_{..} - b_i \bar{X}_{..} + b_i X$ ۲) $\bar{Y}_i - b_i \bar{X}_i + b_i X$

۳) $\bar{Y}_i - b\bar{X}_i + bX$ ۴) $\bar{Y}_{..} - \bar{Y}_i - bX$

۶۴. در یک تحلیل رگرسیون چندگانه Y بر حسب X_2 و X_1، آزمون فرض ضریب همبستگی چندگانه $H_0 : \rho^2_{y.12} = 0$ هم ارز کدام آزمون فرض است؟

۱) $H_0 : \beta_0 = \beta_1 = 0$ ۲) $H_0 : \beta_1 = \beta_2 = 0$

۳) $H_0 : \beta_1 = \beta_2 = 1$ ۴) $H_0 : \beta_0 = \beta_2 = 0$

۶۵. اگر سرعت جذب در بدن برای دو نوع داروی مسکن دارای توزیع نرمال با واریانس‌های یکسان و معلوم باشد. آماره لازم برای آزمون فرض $H_0 : \mu_1 = 2\mu_2$ کدام است؟

(μ_1 و μ_2 میانگین سرعت جذب دو دارو است).

۱) $t = \frac{\sqrt{n}(\bar{x}_1 - 2\bar{x}_2)}{S\sqrt{5}}$ ۲) $t = \frac{n(\bar{x}_1 - 2\bar{x}_2)}{10S}$

۳) $z = \frac{\sqrt{n}(\bar{x}_1 - 2\bar{x}_2)}{\sigma\sqrt{10}}$ ۴) $Z = \frac{n(\bar{x}_1 - 2\bar{x}_2)}{5\sigma}$

۶۶. در مطالعه اثر ۴ روش ماساژدرمانی بر فشار خون افراد سه عامل مخدوش وجود دارد و از مدل $y_{ijkm} = \mu + T_i + \beta_j + \gamma_k + w_m + \varepsilon_{ijkm}, i,j,k,m = ۱,...,۴$ استفاده شده است. طرح مطالعه کدام مورد زیر بوده است؟

۱) طرح مربع یونانی- لاتین
۲) تحلیل دو عاملی با دو اثر متقابل
۳) تحلیل سه عاملی با یک اثر متقابل
۴) تحلیل چهار عاملی با یک اثر متقابل

۶۷. در معادله رگرسیون $E(y \mid x) = \alpha + ۲x_۱ - ۳x_۲ + x_۳$، میانگین نمونه‌ای متغیرهای X_i برای $i = ۱,۲,۳$ در رابطه $\bar{x}_i = ۳(۱+i)$ صدق می‌کنند. اگر $\bar{y} = \sum_{i=۱}^{۲} \bar{x}_i$ باشد، برآورد α برابر است با:

۱) -۳۰ ۲) ۲۴ ۳) -۲۴ ۴) ۳۰

۶۸. کدام‌یک از گزینه‌های زیر فرض معنی‌دار بودن ارتباط متغیر $X_۱$ را با متغیر پاسخ در مدل زیر آزمون می‌کند؟

$$E(Y \mid X) = \beta_۰ + \beta_۱X_۱ + \beta_۲X_۲ + \beta_۳X_۳ + \beta_۴X_۱X_۲$$

۱) $H_۰ : \beta_۱ = \beta_۴$ ۲) $H_۰ : \beta_۴ = ۰$
۳) $H_۰ : \beta_۱ = \beta_۴ = ۰$ ۴) $H_۰ : \beta_۱ = ۰$

۶۹. در مقایسه میانگین نمرات کنکور چهار رشته دوره دکتری مدل کوواریانس $y_{ij} = \alpha_i + \beta$ (سن) $+ \gamma$ (بهره هوش) $+ \varepsilon_{ij}$ با فرض $\varepsilon_{ij} \sim N(۰, \sigma_e^۲), i = ۱,...,۴, j = ۱,...,n_i$ برازش شده است. در این‌صورت تفاضل میانگین گروه اول و سوم برابر است با:

۱) $\alpha_۱ + \alpha_۲ + \gamma$ ۲) $\alpha_۱ - \alpha_۲ + \beta$
۳) $\alpha_۱ - \alpha_۲ + \gamma$ ۴) $\alpha_۱ - \alpha_۲$

سال

۸۷-۸۸

پاسخنامه تشریحی

دکتری تخصصی (Ph.D) آمارزیستی

رسول علیمی- فاطمه خراشادیزاده

آمار ریاضی و احتمالات

۱. گزینه (۳)

روش اول:

$$E(T)=E(\sum_{i=1}^{n}(X_i-\bar{X})^2)=E(\sum_{i=1}^{n}(X_i^2+\bar{X}^2-2X_i\bar{X}))$$

$$=E(\sum_{i=1}^{n}X_i^2+n\bar{X}^2-2\bar{X}\sum_{i=1}^{n}X_i)=\underbrace{E(\sum_{i=1}^{n}X_i^2)}_{(1)}+\underbrace{E(n\bar{X}^2)-E(2\bar{X}\sum_{i=1}^{n}X_i)}_{(2)}$$

$$(1)\begin{cases}E(\sum_{i=1}^{n}X_i^2)=\sum_{i=1}^{n}E(X_i^2)=\sum_{i=1}^{n}(Var(X_i)+E(X_i)^2)=n\sigma^2+n\bar{x}^2\\ Var(X_i)=E(X_i^2)-E(X_i)^2=\sigma^2\\ E(X_i)=\bar{X}\end{cases}$$

$$(2)\begin{cases}E(n\bar{X}^2)-E(2\bar{X}\sum_{i=1}^{n}X_i)=nE(\bar{X}^2)-2nE(\bar{X}^2)=-nE(\bar{X}^2)\\ =-n(Var(\bar{X})+E(\bar{X})^2)=-n\left(\frac{\sigma^2}{n}+\bar{X}^2\right)\\ =-\sigma^2-n\bar{X}^2\end{cases}$$

$$(1)\text{ و }(2)\Rightarrow E(T)=n\sigma^2+n\bar{X}^2-\sigma^2-n\bar{X}^2=n\sigma^2-\sigma^2=(n-1)\sigma^2$$

روش دوم:

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

$$\sigma^2=\frac{1}{N}\sum_{i=1}^{N}(X_i-\mu)^2$$

$$E(T)=E(\sum_{i=1}^{n}(X_i-\bar{X})^2)=E((n-1)S^2)=(n-1)E(S^2)=(n-1)\sigma^2$$

نکات:

نکته ۱: نمونه تصادفی $X_1,X_2,...,X_n$ را از جامعه‌ای در نظر می‌گیریم و فرض می‌کنیم $g(x)$ تابعی باشد که برای آن $E(g(x))$ و $Var(g(x))$ وجود دارند. آن‌گاه:

$$E\left(\sum_{i=1}^{n}g(X_i)\right)=n\,E(g(X_1))\ ،\ Var\left(\sum_{i=1}^{n}g(X_i)\right)=n\,Var(g(X_1))$$

نکته ۲: فرض می‌کنیم $X_1,X_2,...,X_n$، نمونه‌ای تصادفی از جامعه‌ای با میانگین μ و واریانس متناهی σ^2 باشد، آن‌گاه داریم:

$$۱)\ E\bar{X}=\mu\ ،\ ۲)\ Var\bar{X}=\frac{\sigma^2}{n}\ ،\ ۳)\ ES^2=\sigma^2$$

۲. گزینه (۴)

$$E(X_i)=E\left(\frac{\overbrace{X_i}^{\text{آماره فرعی}}}{\sum_{i=1}^{n}X_i}\times\underbrace{\sum_{i=1}^{n}X_i}_{\text{آماره کامل}}\right)\ \text{باسو بنابر قضیه}\ E\left(\frac{X_i}{\sum_{i=1}^{n}X_i}\right)\times E\left(\sum_{i=1}^{n}X_i\right)$$

$$\Rightarrow E\left(\frac{X_i}{\sum_{i=1}^{n}X_i}\right)=\frac{E(X_i)}{E\left(\sum_{i=1}^{n}X_i\right)}=\frac{\mu}{n\mu}=\frac{1}{n}\Rightarrow E\left(\frac{X_1}{\bar{X}}\right)=\frac{E(X_1)}{E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)}=\frac{\mu}{\frac{1}{n}(n\mu)}=1$$

نکات: قضیه باسو: فرض کنید X دارای تابع چگالی احتمالی $f_\theta(X)$ ، $\theta\in\Theta$ باشد. اگر T آماره کامل کراندار و بسنده برای پارامتر θ باشد و U هر آماره‌ای باشد که توزیع آن بستگی به θ ندارد، آن‌گاه T و U از هم مستقل هستند.

آماره کامل در خانواده نمایی: فرض کنید مشاهدات $X_1,X_2,...,X_n$، مستقل و هم توزیع از یک خانواده نمایی با تابع چگالی احتمال (pdf) یا تابع جرم احتمال (pmf) به فرم زیر هستند:

$$f(x\mid\theta)=h(x)c(\theta)\exp\left(\sum_{j=1}^{K}w(\theta_j)t_j(x)\right)$$

که $\theta=(\theta_1,...,\theta_k)$

آن‌گاه آماره $T(x)=\left(\sum_{i=1}^{n} t_1(X_i),...,\sum_{i=1}^{n} t_k(X_i)\right)$، یک آماره کامل در فضای پارامتر Θ است.

آماره فرعی خانواده پارامترهای مقیاس: فرض کنید $X_1,...,X_n$ مشاهدات مستقل و هم توزیع از یک خانواده پارامتر مقیاس با تابع توزیع تجمعی $F\left(\frac{x}{\sigma}\right)$، $\sigma>0$ باشند. آن‌گاه هر آماره‌ای که تنها از طریق $n-1$ مقدار $\frac{X_1}{X_n},...,\frac{X_{n-1}}{X_n}$ به نمونه وابسته باشد، یک آماره فرعی است.

برای مثال $\frac{X_1+...+X_n}{X_n}=\frac{X_1}{X_n}+...+\frac{X_{n-1}}{X_n}+1$، یک آماره فرعی است. برای نشان دادن این موضوع داریم:

$$X_i=\sigma Z_i \Rightarrow F(y_1,...,y_{n-1}\mid\sigma)=P_\sigma\left(\frac{X_1}{X_n}\le y_1,...,\frac{X_{n-1}}{X_n}\le y_{n-1}\right)$$

$$=P_\sigma\left(\frac{\sigma Z_1}{\sigma Z_n}\le y_1,...,\frac{\sigma Z_{n-1}}{\sigma Z_n}\le y_{n-1}\right)=P_\sigma\left(\frac{Z_1}{Z_n}\le y_1,...,\frac{Z_{n-1}}{Z_n}\le y_{n-1}\right)$$

۳. گزینه (۳)

$$P\text{ (از ۴ مهره انتخابی، ۳ تا زوج باشند)}=\frac{\binom{3}{3}\binom{3}{1}}{\binom{6}{4}}=\frac{1\times3}{15}=\frac{1}{5}=\frac{2}{10}$$

نکته: فرض کنید $S=\{s_1,...,s_N\}$ یک فضای نمونه متناهی باشد و $P(\{S_i\})=\frac{1}{N}$، آن‌گاه:

$$P\text{ (رخداد پیشامد A)}=\sum_{S_i\in A}P(\{S_i\})=\sum_{S_i\in A}\frac{1}{N}=\frac{\text{تعداد عناصر موجود در A}}{\text{تعداد عناصر موجود در S}}$$

حال برای محاسبه تعداد عناصر موجود در A و S با استفاده از روش‌های شمارش داریم:

تعداد جایگشت‌های ممکن به اندازه ۲ از n شیء		
	بدون جایگذاری	با جایگذاری
دارای اهمیت ترتیب	$\frac{n!}{(n-2)!}$	n^2
بدون اهمیت ترتیب	$\binom{n}{2}$	$\binom{n+2-1}{2}$

با توجه به این‌که در صورت سوال، نمونه‌ها بدون جایگذاری انتخاب می‌شوند و ترتیب مهم نمی‌باشد، هم برای صورت کسر و هم برای مخرج کسر از رابطه $\binom{n}{2}$ استفاده می‌کنیم.

۴. گزینه (۱)

$$F_X(x)=P(X\le x)=\int_{-\infty}^{x} f_x(t)\,dt=\int_{-\infty}^{x}\frac{1}{\pi(1+t^2)}dt=\frac{1}{\pi}(\arctan t)\Big|_{-\infty}^{x}$$
$$=\frac{1}{\pi}\left(\arctan x+\frac{\pi}{2}\right)=\frac{1}{\pi}\arctan x+\frac{1}{2}$$

۵. گزینه (۴)

چون دامنه توزیع پوآسن شامل اعداد غیرمنفی است $\rightarrow P(X+Y=0)=P(X=0,Y=0)$

با توجه به استقلال بین $X,Y \rightarrow =P(X=0)P(Y=0)$

$$=\frac{e^{-\lambda_1}\lambda_1^{0}}{0!}\times\frac{e^{-\lambda_2}\lambda_2^{0}}{0!}=e^{-\lambda_1}e^{-\lambda_2}=e^{-(\lambda_1+\lambda_2)}=e^{-(2+3)}=e^{-5}$$

نکته: متغیر تصادفی X دارای توزیع پوآسن با پارامتر λ $(X\sim P(\lambda))$ است، اگر دارای تابع احتمال زیر باشد:

$$P(X=x\mid\lambda)=f_\lambda(x)=\frac{e^{-\lambda}\lambda^{x}}{x!}\ ,x=0,1,...;\lambda>0$$

در توزیع پوآسن با پارامتر λ داریم:

$$E_\lambda(x)=\lambda$$
$$Var_\lambda(x)=\lambda$$
$$M_X(t)=\exp\left[\lambda(e^{t}-1)\right],t\in R$$

۶. گزینه (۳)

حالت خاصی از نامساوی چبیچف $P\left(|X-\mu|\ge K\sigma\right)\le\frac{1}{K^2}$

$$P(|X-1|\ge\underset{\downarrow}{2})\le\frac{1}{4^2}=\frac{1}{16}$$
$$K\sigma=2\Rightarrow K\left(\frac{1}{2}\right)=2\Rightarrow K=4$$

نکات: نامساوی‌های از نوع چبیچف:

نامساوی مارکف: اگر X یک متغیر تصادفی و $h(0)$ تابعی غیرمنفی باشد که $E\left[h(x)\right]<\infty$، آن‌گاه برای هر $a>0$:

$$P(h(x)\ge a)\le\frac{E(h(x))}{a}$$

نامساوی چبیچف: با انتخاب $h(x)=|x|$، در نامساوی مارکف و با توجه به: $P(|X|\ge a)=P(X^2\ge a^2)$ داریم:

$$P(|X|\ge a)\le\frac{E(X^2)}{a^2}$$

نکته: اگر Z دارای توزیع نرمال استاندارد باشد، آن‌گاه:

$$P(|Z|\ge t)\le\sqrt{\frac{2}{\pi}}\,\frac{e^{-\frac{t^2}{2}}}{t}\quad,\quad t>0$$

نکته: فرض کنید X یک متغیر تصادفی با تابع مولد گشتاور $M_X(t)$ ،$-h<t<h$ باشد. آن‌گاه:

$$P(X\ge a)\le e^{-at}M_X(t)\ ,\ 0<t<h$$
$$P(X\le a)\le e^{-at}M_X(t)\ ,\ -h<t<0$$

۷. گزینه (۱)

$$X \sim U(-1,1) \Rightarrow Var(x) = \frac{(1-(-1))^2}{12} = \frac{4}{12} = \frac{1}{3}$$

$$\Rightarrow P(|X| > \sqrt{3\,var(x)}) = P(|X| > \sqrt{3 \times \frac{1}{3}}) = P(|X| > 1) = P(X > 1) + P(X < -1)$$

با توجه به این‌که مقادیر X برای $X < -1$ و $X > 1$ تعریف نشده است لذا داریم:

$$P(|X| \sqrt{3\,Var(x)}) = 0$$

متغیر تصادفی X دارای توزیع یکنواخت در فاصله (θ_1, θ_2) است $(X \sim U(\theta_1, \theta_2))$، اگر دارای تابع چگالی احتمال زیر باشد:

$$f(x \mid \theta_1, \theta_2) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, \theta_1 \le x \le \theta_2, -\infty < \theta_1 < \theta_2 < \infty \\ 0 \quad \text{سایر نقاط} \end{cases}$$

اگر $X \sim U(\theta_1, \theta_2)$ آن‌گاه:

$$E_{\theta_1,\theta_2}(x) = \frac{\theta_1 + \theta_2}{2}$$

$$Var_{\theta_1,\theta_2}(x) = \frac{(\theta_2 - \theta_1)^2}{12}$$

$$M_X(t) = \begin{cases} \frac{e^{\theta_2 t} - e^{\theta_1 t}}{\theta_2 - \theta_1} \times t & , \quad t \neq 0 \\ 1 & , \quad t = 0 \end{cases}$$

۸. گزینه (۲)

تابع $f(x,y)$ تابع چگالی احتمال توأم بردار تصادفی دو متغیره پیوسته (X,Y) می‌باشد، اگر برای هر $A \subset R^2$ داشته باشیم:

$$P((X,Y) \in A) = \int_A \int f(x,y)\,dx\,dy$$

$$\begin{cases} |X| < 1 \Rightarrow -1 < X < 1 \\ |Y| < 1 \Rightarrow -1 < Y < 1 \end{cases}$$

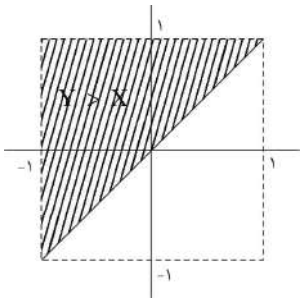
$$P(x < y) = P(-1 < x < y < 1) = \int_{-1}^{1} \int_{-1}^{y} \frac{x + y + 2}{8} dx\,dy$$

$$= \frac{1}{8} \int_{-1}^{1} \left(\frac{1}{2} x^2 + xy + 2x \Big|_{-1}^{y} \right) dy = \frac{1}{8} \int_{-1}^{1} \frac{3}{2} y^2 + 3y + \frac{3}{2}$$

$$= \frac{1}{8} \left(\frac{1}{2} y^3 + 3y^2 + 3y \Big|_{-1}^{1} \right) = \frac{4}{8} = \frac{1}{2}$$

روش دیگر حل مسئله به‌صورت زیر می‌باشد:

با توجه به این‌که دامنه تغییرات X و Y بین ۱- و ۱ می‌باشد، بنابراین مساحت کل برابر با یک است، لذا داریم:



$$P(X<Y)=\frac{۱}{۲}$$

نکات:

اگر $f(x,y)$ تابع چگالی احتمال توأم بردار تصادفی دو متغیره پیوسته (X,Y) باشد، داریم:

$$E(g(x,y))=\int\int g(x,y)f(x,y)dx\,dy$$

$$f_X(x)=\int f(x,y)dy$$

$$f_Y(y)=\int f(x,y)dx$$

۹. گزینه (۲)

با استفاده از قضیه ارگودیک که تعمیمی از قانون اعداد بزرگ می‌باشد، وقتی $n\to\infty$ داریم:

$$\frac{۱}{n}\sum_{i=۱}^{n}h(X_i)=E(h(x))$$

بنابراین:

$$\lim_{n\to\infty}\frac{۱}{n}\sum_{i=۱}^{n}X_i^۲=E(X^۲)=Var(X)+E(X)^۲=۴+۲^۲=۸$$

نکات: مفاهیم همگرایی

همگرایی در احتمال: دنباله متغیرهای تصادفی $X_۱,...,X_n$ در احتمال به متغیر تصادفی X همگراست، اگر برای هر $\varepsilon>۰$:

$$\lim_{n\to\infty}P\left(|X_n-X|\ge\varepsilon\right)=۰ \quad \text{یا} \quad \lim_{n\to\infty}P\left(|X_n-X|<\varepsilon\right)=۱$$

$P\left(|X_n-X|\right)$، احتمال دور بودن X_n از X را نشان می‌دهد. اگر $\{X_n\}$ به X همگرا شود، $P\left(|X_n-X|>\varepsilon\right)$ باید کوچک باشد و با افزایش n بسیار کم‌تر شود. به عبارت دیگر، احتمال این‌که X_n از X فاصله داشته باشد با افزایش n باید به صفر میل نماید. یعنی داریم:

$$\lim_{n\to\infty}P\left(|X_n-X|>\varepsilon\right)=۰$$

قانون ضعیف اعداد بزرگ: فرض کنید $X_۱,X_۲,...$ متغیرهای تصادفی مستقل و هم توزیع با $EX_i=\mu$ و $Var\,X_i=\sigma^۲<\infty$ باشند. میانگین دنباله را به‌صورت $\bar{X}_n=\frac{۱}{n}\sum_{i=۱}^{n}X_i$ تعریف می‌کنیم. آن‌گاه برای هر $\varepsilon>۰$:

$$\lim_{n\to\infty}P\left(|\bar{X}_n-\mu|<\varepsilon\right)=۱$$

یعنی $\bar{X}_n$ در احتمال به μ همگراست.

نکته: فرض کنید $X_1, X_2, ...$ در احتمال به متغیر تصادفی X همگرا باشد و h را یک تابع پیوسته در نظر بگیریم. آن‌گاه $h(X_1), h(X_2), ...$ در احتمال به $h(X)$ همگرا است.

همگرایی قریب به یقین: نوعی از همگرایی که نسبت به همگرایی در احتمال قوی‌تر است. دنباله متغیرهای تصادفی $X_1, X_2, ...$، همگرایی قریب به یقین به متغیر تصادفی X دارند، اگر برای هر $\varepsilon > 0$:

$$P\left(\lim_{n\to\infty} |X_n - X| < \varepsilon\right) = 1$$

اگر وقوع پیشامدی قریب به یقین باشد، در این‌صورت همواره رخ می‌دهد، اما پیشامدهای دیگری هم از لحاظ نظری در فضای احتمال مورد بررسی امکان وقوع دارند. اگرچه با افزایش فضای نمونه، احتمال رخ دادن هر پیشامد دیگری به‌طور مجانبی به صفر می‌گراید. به‌طور مثال، فضای نمونه S با بازه بسته $[0,1]$ با توزیع احتمال یکنواخت را در نظر می‌گیریم. متغیر تصادفی $X(S) = S, X_n(S) = S + S^n$ را تعریف می‌کنیم. برای هر $S \in [0,1]$، اگر $n \to \infty$ آن‌گاه $S^n \to 0$. بنابراین $X_n(S) \to S = X(S)$. با این حال $X_n(1) = 2$ (برای هر n)، بنابراین $X_n(1)$ به عدد $X(1) = 1$ همگرا نمی‌شود. اما از آن‌جایی‌که همگرایی در بازه $[0,1)$ اتفاق می‌افتد و $P([0,1)) = 1$، X_n تقریباً همه‌جا (قریب به یقین) به عدد X همگرا است.

قانون قوی اعداد بزرگ: فرض کنید $X_1, X_2, ...$ متغیرهای تصادفی مستقل و هم توزیع با $EX_i = \mu$، $\mathrm{Var}\, X_i = \sigma^2 < \infty$ باشند و میانگین دنباله به‌صورت $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ تعریف شود. آن‌گاه برای هر $\varepsilon > 0$:

$$P\left(\lim_{n\to\infty} |\bar{X}_n - \mu| < \varepsilon\right) = 1$$

یعنی $\bar{X}_n$ تقریباً همه‌جا (قریب به یقین) به μ همگراست.

همگرایی در توزیع: دنباله متغیرهای تصادفی $X_1, X_2, ...$ در توزیع به متغیر تصادفی X همگرا است اگر $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ (در همه نقاط x که $F_X(x)$ پیوسته است)

نکته: همگرایی در احتمال، همگرایی در توزیع را نتیجه می‌دهد.

قضیه حد مرکزی: فرض کنید $X_1, X_2, ...$ دنباله‌ای از متغیرهای تصادفی مستقل و هم توزیع باشند که تابع مولد گشتاور آن‌ها در همسایگی صفر وجود دارد. هم‌چنین فرض کنید $EX_i = \mu$ و $\mathrm{Var}\, X_i = \sigma^2 < \infty$، $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ و $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ باشد. آن‌گاه برای هر X، $-\infty < x < \infty$ داریم:

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy;$$

یعنی $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$، به‌طور تقریبی دارای توزیع نرمال استاندارد است. به‌عبارت دیگر، قضیه حد مرکزی بیان می‌کند که با فرض شرایط خاص، میانگین تعدادی متغیر تصادفی مستقل، به‌طور تقریبی دارای توزیع نرمال خواهد بود.

قضیه اسلاتسکی: اگر X_n در توزیع به X و Y_n در احتمال به عدد ثابت a همگرا باشند، آن‌گاه:

الف- $Y_n X_n$ در توزیع به aX همگرا است، ب- $X_n + Y_n$ در توزیع به $X + a$ همگرا است.

۱۰. گزینه (۴)

$$\text{Bias}_\theta\hat{\theta} = E(\hat{\theta}) - \theta$$

$$\left|E(S)^2 - \sigma^2\right| = \frac{1}{10}\sigma^2 \Rightarrow \left|E\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}\right) - \sigma^2\right| = \frac{1}{10}\sigma^2 \quad *$$

می‌دانیم: $E(S^2) = E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = \sigma^2 \Rightarrow E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)$

$$= E\left(\frac{n-1}{n} \times \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = E\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}\sigma^2$$

$$* \Rightarrow \left|E\left(\frac{n-1}{n}S^2\right) - \sigma^2\right| = \frac{1}{10}\sigma^2 \Rightarrow \left|\frac{n-1}{n}\sigma^2 - \sigma^2\right| = \frac{1}{10}\sigma^2$$

$$\Rightarrow \left|\frac{\sigma^2}{n}\right| = \frac{\sigma^2}{10} \Rightarrow n = 10$$

نکات: میانگین مربعات خطای (MSE) برآوردگر ω از پارامتر θ، تابعی از θ است که به‌صورت $E_\theta(\omega-\theta)^2$ تعریف می‌شود. به‌عبارت دیگر:

$$MSE = E_\theta(\omega-\theta)^2 = \text{Var}_\theta\omega + (E_\theta\omega - \theta)^2 = \text{Var}_\theta\omega + (\text{Bias}_\theta\omega)^2$$

میزان اریبی برآوردگر ω از پارامتر θ به‌صورت $E_\theta\omega - \theta$ تعریف می‌شود. برآوردگری که میزان اریبی آن برابر صفر شود، یک برآوردگر نااریب نامیده می‌شود، یعنی $E_\theta\omega = \theta$

بنابراین شاخص MSE شامل دو مؤلفه می‌باشد که یکی میزان تغییرپذیری برآوردگر (دقت) و دیگری میزان اریبی آن (صحت) را می‌سنجد.

۱۱. گزینه (۴)

$$\text{Cov}(U,\omega) = \text{Cov}(X_1 + X_2, X_1 + CX_2)$$

$$= \text{Cov}(X_1,X_1) + C\,\text{Cov}(X_1,X_2) + \text{Cov}(X_1,X_2) + C\,\text{Cov}(X_2,X_2)$$

$$= \text{Var}(X_1) + (C+1)\text{Cov}(X_1,X_2) + C\,\text{Var}(X_2)$$

برای این‌که U و ω ناهمبسته باید $\text{Cov}(U,\omega) = 0$ باشد. هم‌چنین با توجه به‌صورت سوال، X_1 و X_2 مستقل هستند پس $\text{Cov}(X_1,X_2) = 0$ است. بنابراین:

$$\text{Var}(X_1) + C\,\text{Var}(X_2) = 0 \Rightarrow 1 + 4C = 0 \Rightarrow C = \frac{1}{4}$$

۱۲. گزینه (۲)

با استفاده از تعریف ضریب همبستگی بین دو متغیر X و Y که به‌صورت $\rho_{XY}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$ می‌باشد داریم:

$$\rho_{xy}=\frac{Cov(X,aX^2+bX+C)}{\sqrt{Var(X)}\sqrt{Var(aX^2+bX+C)}}$$

$$=\frac{Cov(X,aX^2)+Cov(X,bX)+Cov(X,C)}{\sqrt{Var\,X}\sqrt{Var(aX^2)+Var(bX)+2Cov(aX^2,bX)}}$$

$$=\frac{a\,Cov(X,X^2)+b\,Var(X)}{\sqrt{Var\,X}\sqrt{a^2\,var(X^2)+b^2\,Var(X)+2ab\,Cov(X^2,X)}}$$

تنها گزینه‌ای که به جواب منتهی می‌شود، گزینه ۲ است.

۱۳. گزینه (۱)

فرض کنید $S_n=X_1+X_2+...+X_n$، داریم:

$$E(X_1|S_n)=E(X_2|S_n)=...=E(X_n|S_n)$$

$$\Rightarrow E(X_1|S_n)+...+E(X_n|S_n)=E(S_n|S_n)=S_n$$

$$\Rightarrow E(X_i|S_n)=\frac{S_n}{n}\ ,\ i=1,...,n$$

$$\Rightarrow E(X_1|S_3)=\frac{S_3}{3}=\frac{X_1+X_2+X_3}{3}=\bar{X}$$

۱۴. گزینه (۳)

با استفاده از تعریف تابع توزیع که به‌صورت $F_X(x)=P(X\le x)=\int_{-\infty}^{x}f_x(t)dt$ می‌باشد، میانه توزیع برابر با عددی است که در آن مقدار تابع توزیع برابر با $\frac{1}{2}$ باشد یعنی: $P(X\le m)=F_X(m)=\int_{-\infty}^{m}f_x(t)dt=\frac{1}{2}$

با توجه به این‌که میانه برابر با چارک دوم است، برای به‌دست آوردن چارک اول کافیست به‌جای عدد $\frac{1}{2}$، عدد $\frac{1}{4}$ را قرار دهیم. یعنی جایی از توزیع که مقدار توزیع تجمعی تا آن نقطه برابر با $\frac{1}{4}$ شده است:

$$F_X(Q_1)=\int_{-\infty}^{Q_1}f_x(t)dt=\frac{1}{4}\xrightarrow{\text{با توجه به دامنه X}}\int_{0}^{Q_1}2t\,dt=\frac{1}{4}$$

$$\Rightarrow t^2\Big|_0^{Q_1}=\frac{1}{4}\Rightarrow Q_1^2=\frac{1}{4}\Rightarrow Q_1=\pm\frac{1}{2}$$

با توجه به دامنه X، $Q_1=\frac{1}{2}$ درست است.

۱۵. گزینه (۲)

$$X_i \sim N(۰,\sigma^۲) \Rightarrow f(x\mid\sigma^۲)=\frac{۱}{\sqrt{۲\pi}\sigma}e^{\frac{-۱}{۲\sigma^۲}(X_i^۲)}$$

$$\Rightarrow L(\sigma^۲\mid X_۱,...,X_{۱۰})=(۲\pi\sigma^۲)^{-\frac{۱۰}{۲}}\exp\left(-\frac{۱}{۲\sigma^۲}\sum_{i=۱}^{۱۰}X_i^۲\right)$$

$$\Rightarrow \text{Log}\,L(\sigma^۲\mid \underline{X})=-n\ln\sigma-\frac{۱}{۲\sigma^۲}\sum_{i=۱}^{۱۰}X_i^۲$$

$$\Rightarrow \frac{\partial}{\partial\sigma^۲}\text{Log}\,L=\frac{-n}{\sigma}+\frac{۱}{\sigma^۳}\sum_{i=۱}^{۱۰}X_i^۲=۰$$

$$\Rightarrow \sigma^۲=\sum_{i=۱}^{۱۰}X_i^۲/n=\frac{۱/۶}{۱۰}=۰/۱۶\Rightarrow\sigma=۰/۰۴$$

نکات:

روش ML، روشی برای به‌دست آوردن برآوردگر ماکسیمم درست‌نمایی می‌باشد که به اختصار MLE نامیده می‌شود که براساس تابع درست‌نمایی است.

فرض کنیم $X=(X_۱,...,X_n)$، بردار n متغیر تصادفی با تابع چگالی احتمال توأم $f_{\underline{\theta}}(x)$ که $\underline{\theta}\in\Theta\subseteq R^K$ باشد.

برای هر مقدار داده شده‌ی $X=x$، تابع درست‌نمایی X را تابع چگالی احتمال توأم X یعنی $f_{\underline{\theta}}(x)$ تعریف می‌کنیم که به‌صورت تابعی از θ در نظر گرفته می‌شود و آن را با نماد $L(\theta)$ نشان می‌دهیم:

$$L(\underline{\theta})=f_{\underline{\theta}}(x)$$

اگر $X_۱,...,X_n$، نمونه تصادفی n تایی از توزیعی از خانواده‌ی چگالی‌های $\{f_\theta(x):\theta\in\Theta\}$ باشد، آن‌گاه:

$$L(\underline{\theta}\mid\underline{X})=L(\theta_۱,...,\theta_K\mid X_۱,...,X_n)=\prod_{i=۱}^{n}f(X_i\mid\theta_۱,...,\theta_K)$$

MLE پارامتر θ با کمک حل معادله $\frac{\partial}{\partial\theta}L(\theta)=۰$ (یا معادل آن $\frac{\partial}{\partial\theta}\ln L(\theta)=۰$) به‌دست می‌آید. بنابراین اگر $\delta(x)$ برآوردگری برای θ باشد به‌طوری‌که:

i) $P_\theta(\delta(x)\in\theta)=۱\ ,\ \forall\theta\in\Theta$

ii) $L(\delta(X))\geq L(\theta)\ ,\ \forall\theta\in\Theta$

آن‌گاه $\delta(X)$ برآوردگر ماکسیمم درست‌نمایی θ در نظر گرفته می‌شود.

توزیع نرمال: متغیر تصادفی X دارای توزیع نرمال با پارامترهای $\mu,\sigma^۲$ $(X\sim N(\mu,\sigma^۲))$ است، اگر دارای تابع چگالی احتمال زیر باشد:

$$f_{\mu,\sigma^۲}(x)=\frac{۱}{\sqrt{۲\pi\sigma^۲}}\exp\left\{\frac{-۱}{۲\sigma^۲}(x-\mu)^۲\right\},\ x\in R\ ,\ \mu\in R\ ,\ \sigma>۰$$

۱۶. گزینه (۳)

با توجه به خاصیت پایایی برآوردگر ماکسیمم درست‌نمایی (اگر $\hat{\theta}$ برآوردگر ماکسیمم درست‌نمایی θ باشد، آن‌گاه برای هر تابعی از θ مثل $\tau(\theta)$، برآوردگر ماکسیمم درست‌نمایی $\tau(\theta)$ برابر با $\tau(\hat{\theta})$ است)، برای به‌دست آوردن فاصله اطمینان برای σ کافیست از کران‌ها جذر گرفته شود.

ذکر این نکته ضروری است که سطح اطمینان تغییری نخواهد کرد.

۱۷. گزینه (۴)

اگر T یک آماره بسنده برای پارامتر θ باشد و MLE منحصر به فرد θ، یعنی $\hat{\theta}$ وجود داشته باشد، آن‌گاه $\hat{\theta}$ باید تابعی از T باشد و اگر MLE منحصر به فرد پارامتر θ وجود نداشته باشد، آن‌گاه یک MLE پارامتر θ را می‌توان انتخاب کرد که تابعی از T باشد.

۱۸. گزینه (۱)

$(\bar{X},S^2)$ یک آماره بسنده توأم کامل برای (μ,σ^2) است. از طرفی برای هر σ^2 ثابتی، $\bar{X}$ یک آماره بسنده کامل برای μ است و S^2 دارای توزیع کای دو با $n-1$ درجه آزادی است که توزیع آن به μ بستگی ندارد. بنابراین، بنابر قضیه باسو، برای هر μ و σ^2 ثابت، $\bar{X}$ و S^2 مستقل از هم هستند، لذا هر ترکیب خطی از آن‌ها نیز از هم مستقل هستند.

$$P\left((\bar{X}-\mu)(S^2-\sigma^2)>0\right)=P\left(\bar{X}-\mu\right)P(S^2-\sigma^2)>0$$

این رابطه زمانی برقرار است که هر دو احتمال مثبت و یا هر دو احتمال منفی باشند، پس داریم:

$$\begin{aligned}P\left((\bar{X}-\mu)(S^2-\sigma^2)>0\right)&=\left(P(\bar{X}-\mu)>0\right)\left(P(S^2-\sigma^2)>0\right)+\left(P(\bar{X}-\mu)<0\right)\left(P(S^2-\sigma^2)<0\right)\\&=P\left(\bar{X}>\mu\right)P(S^2>\sigma^2)+P(\bar{X}<\mu)P\left(S^2<\sigma^2\right)\\&=\frac{1}{2}\times\frac{1}{2}+\frac{1}{2}\times\frac{1}{2}=\frac{1}{2}\end{aligned}$$

نکات:

خانواده توزیع‌های تولید شده توسط آماره T کامل است، اگر هیچ برآوردگر ناریب صفر غیر از برآوردگر صفر وجود نداشته باشد. یعنی تنها برآوردگر ناریب صفر در این خانواده خود صفر است. به‌عبارت دیگر:

$$E_\theta(g(T))=0\ ,\ \forall\theta\in\Theta$$

$$\Rightarrow P_\theta(g(T)=0)=1\ ,\ \forall\theta\in\Theta$$

باید توجه نمود که کامل بودن یک ویژگی از خانواده توزیع‌های احتمالی است نه یک ویژگی از یک توزیع خاص. به‌طور مثال اگر $X\sim N(0,1)$، با تعریف $g(x)=x$ داریم:

$$Eg(X)=0\Rightarrow P(g(X)=0)=0$$

که مشاهده می‌شود x یک آماره کامل نیست. با این وجود این حالتی از یک توزیع خاص می‌باشد نه یک خانواده از توزیع‌ها. اما اگر $X\sim N(\theta,1)$ آن‌گاه:

$$Eg(X)=0\Rightarrow P(g(X)=0)=1\ ,\ \forall\theta\in\Theta$$

بنابراین خانواده توزیع‌های $N(\theta,1)$ کامل می‌باشد.

۱۹. گزینه (۴)

$$f(x\mid\lambda)=\frac{e^{-\lambda}\lambda^{x}}{x!}\Rightarrow \text{Log}\, f(x\mid\lambda)=-\lambda+x\ln\lambda-\ln(x!)$$

$$\Rightarrow\frac{\partial}{\partial\lambda}\ln f(x\mid\lambda)=-1+\frac{x}{\lambda}\Rightarrow\left(\frac{\partial}{\partial\lambda}\ln f(x\mid\lambda)\right)^{2}=1+\frac{x^{2}}{\lambda^{2}}-\frac{2x}{\lambda}$$

$$E\left[\left(\frac{\partial}{\partial\lambda}\ln f(x\mid\lambda)\right)^{2}\right]=1+\frac{E(x^{2})}{\lambda^{2}}-\frac{2E(x)}{\lambda}=1+\frac{\lambda+\lambda^{2}}{\lambda^{2}}-\frac{2\lambda}{\lambda}=\frac{1}{\lambda}\quad(1)$$

(در توزیع پوآسن مقدار میانگین برابر با واریانس است یعنی $E(X)=\text{Var}(X)=\lambda$)

$$\left[\omega'(\theta)+b'(\theta)\right]^{2}=\left[(\lambda)'+0\right]^{2}=1\qquad(2)$$

$$(1),(2)\Rightarrow \text{Var}(T)\geq\frac{1}{n\times\frac{1}{\lambda}}=\frac{\lambda}{n}$$

نکات:

نامساوی کرامر رائو، کران پایین برای واریانس برآوردگرهای انتخابی ارائه می‌دهد.

فرض کنید $X_1,\ldots,X_n$ دارای توزیع توأم با تابع چگالی احتمال توأم $P_\theta(X_1,\ldots,X_n)$ باشند. اگر T برآوردگر پارامتر $\omega(\theta)$ باشد، آن‌گاه نامساوی کرامر- رائو براساس یک نمونه تصادفی n تایی برای هر $\theta\in\Theta$ به یکی از دو روش زیر خواهد بود:

$$(1)\quad \text{Var}_\theta(T)\geq\frac{\left[\omega'(\theta)+b'(\theta)\right]^{2}}{E_\theta\left[\frac{\partial}{\partial\theta}\ln P_\theta(X_1,\ldots,X_n)\right]^{2}}$$

$$(2)\quad \text{Var}_\theta(T)\geq\frac{(\omega'(\theta)+b'(\theta))^{2}}{n\times E_\theta\left(\left(\frac{\partial}{\partial\theta}\ln f_\theta(x)\right)^{2}\right)}$$

به‌طوری‌که: $b(\theta)=E_\theta(T)-\omega(\theta)$ بیانگر اریبی برآوردگر T در برآورد پارامتر $\omega(\theta)$ است.

۲۰. گزینه (۴)

می‌دانیم:

$$1)\left(Z_1+Z_2\right)\sim N(0,2)\Rightarrow\frac{Z_1+Z_2}{\sqrt{2}}\sim N(0,1)\Rightarrow\left(\frac{Z_1+Z_2}{\sqrt{2}}\right)^{2}\sim X^{2}_{(1)}$$

$$2)\left(Z_1-Z_2\right)\sim N(0,2)\Rightarrow\frac{Z_1-Z_2}{\sqrt{2}}\sim N(0,1)\Rightarrow\left(\frac{Z_1-Z_2}{\sqrt{2}}\right)^{2}\sim X^{2}_{(1)}$$

$$(1),(2)\Rightarrow\frac{\left(Z_1+Z_2\right)^{2}/2}{\left(Z_1-Z_2\right)^{2}/2}=\frac{\left(Z_1+Z_2\right)^{2}}{\left(Z_1-Z_2\right)^{2}}\sim\frac{X^{2}_{(1)}}{X^{2}_{(1)}}\sim F_{(1,1)}$$

۲۱. گزینه (۳)

اگر $\varnothing$ یک تابع آزمون با ناحیه بحرانی C و آماره آزمون $T(X)$ باشد:

$\alpha = P$ (خطای نوع اول) $= P$ (فرض H_0 را رد کنیم وقتی‌که H_0 درست است)

$= P\,T(X)\in C \mid H_0$ درست است $) = P_{H_0}\left(T(X)\in C\right) = E_{\theta_0}\{\varnothing(X)\}$

$\beta = P$ (خطای نوع دوم) $= P$ (فرض H_0 را قبول کنیم وقتی‌که H_1 درست است)

$= P(T(X)\notin C \mid H_1$ درست $) = P_{H_1}(T(X)\notin C) = E_{\theta_1}\left(1-\varnothing(X)\right)$

$\beta^* = \varnothing$ توان آزمون $= 1-\beta = 1-E_{\theta_1}\left(1-\varnothing(X)\right) = E_{\theta_1}\left(\varnothing(X)\right) = P_{\theta_1}\left(T(X)\in C\right)$

$=$ (فرض H_0 را رد کنیم هنگامی‌که H_1 درست است) P

۲۲. گزینه (۲)

با کمک لِم نیمن- پیرسون مسئله را حل می‌کنیم:

$$\begin{cases} H_0 : f_0(x) = 4x^3 \\ H_1 : f_1(x) = 3x^2 \end{cases}$$

$$\frac{f_1(x)}{f_0(x)} > K \Rightarrow \frac{3x^2}{4x^3} > K \Rightarrow x < \frac{3}{4K} \Rightarrow x < C$$

$$\varnothing^*(x) = \begin{cases} 1 & X < C \\ 0 & \text{در غیراین‌صورت} \end{cases}$$

$$\alpha = E_{H_0}(\varnothing^*(X)) = \int_0^C fx^3 dx = x^4 \Big|_0^C = C^4$$

$$\alpha = \frac{1}{16} \Rightarrow C^4 = \frac{1}{16} \Rightarrow C = \frac{1}{2}$$

بنابراین ناحیه بحرانی برابر با $X < \frac{1}{2}$ است.

نکات:

لِم نیمن- پیرسون: اگر X یک متغیر تصادفی با تابع چگالی احتمال $f(x\mid\theta)$ باشد و علاقه‌مند به آزمون $H_1 : \theta = \theta_1$ در مقابل $H_0 : \theta = \theta_0$ باشیم.

اگر $\varnothing^*$ را برای هر $0 \le K \le \infty$ به‌صورت زیر تعریف کنیم:

$$\varnothing^*(X) = \begin{cases} 1 & (x\in C) \quad f(x\mid\theta_1) > K f(x\mid\theta_0) \\ 0 & (x\in C^c) \quad f(x\mid\theta_1) < K f(x\mid\theta_0) \end{cases}$$

(C ناحیه بحرانی و C^c ناحیه پذیرش)

به‌طوری که $\alpha = E_{\theta_0}(\varnothing^*(X))$ باشد، آن‌گاه $\varnothing^*(X)$ پرتوان‌ترین آزمون در سطح α برای آزمون H_0 در مقابل H_1 است.

۲۳. گزینه (۴)

با استفاده از قضیه دسته‌بندی فیشر- نیمن داریم:

$$f(\underline{x}\,|\,\sigma^{۲}) = \underbrace{\left(۲\pi\sigma^{۲}\right)^{-\frac{n}{۲}} e^{-\frac{۱}{۲\sigma^{۲}}\sum_{i=۱}^{n} X_i^{۲}}}_{g(T(\underline{X})|\sigma^{۲})} \times \underbrace{۱}_{h(\underline{x})}$$

در نتیجه $\sum_{i=۱}^{n} X_i^{۲}$ یک آماره بسنده برای پارامتر $\sigma^{۲}$ است.

نکات:

قضیه دسته‌بندی فیشر- نیمن: آماره $T(X)$ برای پارامتر θ بسنده است اگر و تنها اگر برای هر $\theta \in \Theta$ توابع غیرمنفی g و h وجود داشته باشند که: $f(\underline{x}\,|\,\underline{\theta}) = g(T(\underline{X}\,|\,\underline{\theta})h(\underline{X})$

به‌طوری‌که $g(T(X)\,|\,\underline{\theta})$ فقط از طریق $T(X)$ به x بستگی دارد و $h(\underline{x})$ به θ بستگی ندارد. با توجه به این‌که هر تابع یک به یک از آماره بسنده نیز بسنده می‌باشد، بنابراین $e^{\overline{X^{۲}}}$ نیز یک آماره بسنده است.

۲۴. گزینه (۱)

اگر واریانس T_n (برآوردگر ناریب $\omega(\theta)$) با کران پایین کرامر- رائو برابر باشد، T_n را یک برآوردگر θ' را می‌نامیم. کران پایین کرامر- رائو در این حالت به‌صورت $\frac{\omega'(\theta)^{۲}}{I(\theta)}$ ($I(\theta)$ ماتریس اطلاع) است. لذا کارایی برآوردگر ناریب T_n به‌صورت زیر است:

$$e_n(T_n) = \frac{\left(\omega'(\theta)\right)^{۲} / I_{(\theta)}}{Var_{\theta}(T_n)}$$

$e_n(T_n)$ عددی بین صفر و یک است و برای یک برآوردگر کارا، کارایی یک است.

از طرفی اگر T یک برآوردگر UMVU پارامتر $\omega(\theta)$ باشد و T^* نیز یک برآوردگر ناریب $\omega(\theta)$ باشد، آن‌گاه کارایی T^* عبارتست از:

$$e_n^*(T^*) = \frac{V_{\theta}(T)}{V_{\theta}(T^*)}$$

بنابراین $۰ < e_n^*(T^*) \le ۱$ و کلیه برآوردگرهای UMVU با این تعریف برآوردگر کارا هستند. لذا برای به‌دست آوردن برآورد کارا برای پارامتر θ کافیست برآوردگر UMVU پارامتر θ را به‌دست آوریم، لذا داریم:

$$f(x\,|\,\theta) = \frac{۱}{\theta} e^{-\frac{x}{\theta}} \Rightarrow L(\underline{x}\,|\,\theta) = \theta^{-n} e^{-\frac{۱}{\theta}\sum_{i=۱}^{n} X_i}$$

با توجه به فرم نمایی توزیع، $T = \sum_{i=۱}^{n} X_i$، آماره بسنده کامل برای پارامتر θ است. حال تابعی از آماره بسنده کامل را به دست می‌آوریم که برای پارامتر θ ناریب باشد، یعنی $E(g(T)) = \theta$.

$$T = \sum_{i=۱}^{n} X_i \sim \exp(n\theta) \Rightarrow E(T) = n\theta \Rightarrow E\left(\frac{T}{n}\right) = \theta$$

لذا $\frac{T}{n} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$، UMVUE پارامتر θ است.

نکات: قضیه رائو- بلکول- لهمن- شفه: فرض کنید X یک متغیر تصادفی با توزیعی از خانواده توزیع‌های $\{P_\theta : \theta \in \Theta\}$ باشد. اگر $T(X)$ یک آماره بسنده کامل برای θ باشد، آن‌گاه برای هر پارامتر برآوردپذیر $\omega(\theta)$، برآوردگر نااریب یکتا با کم‌ترین واریانس (UMVUE) بر پایه $T(X)$ وجود دارد.

آماره $T(X)$ یک برآوردگر نااریب با کم‌ترین واریانس برای $\omega(\theta)$ می‌باشد اگر:

$$E_\theta(T(X)) = \omega(\theta) \ , \ \forall \theta \in \Theta$$

و برای هر برآوردگر نااریب $h(X)$ پارامتر $\omega(\theta)$ داشته باشیم:

$$Var_\theta(T(X)) \le Var_\theta(h(X)) \ , \ \forall \theta \in \Theta$$

۲۵. گزینه (۴)

پرتوان‌ترین آزمون یکنواخت برای هر آزمون فرضی وجود ندارد و فقط برای حالات خاصی با شرایط معین وجود دارد. حتی در موارد بسیار ساده که علاقه‌مند به آزمون فرض ساده در مقابل فرض مرکب با دو عضو هستیم، ممکن است پرتوان‌ترین آزمون یکنواخت وجود نداشته باشد. در حالت کلی، برای آزمون فرض‌های دو طرفه، پرتوان‌ترین آزمون یکنواخت وجود ندارد. اما به‌طور کلی، در صورتی‌که خانواده توابع توزیع دارای خاصیت MLR باشند، آن‌گاه آزمون UMP برای آزمون فرض‌های یک طرفه وجود دارد.

نکته: خانواده $\{f_\theta : \theta \in \Theta \subset R\}$، دارای خاصیت نسبت درست‌نمایی یکنوا (MLR) در آماره $T(X)$ می‌باشد اگر برای هر $\theta_1 > \theta_0$، نسبت $\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}$ تابعی غیرنزولی از $T(x)$ باشد.

۲۶. گزینه (۱)

اگر X دارای تابع مولد گشتاور $M_X(t)$ باشد، آن‌گاه:

$$E X^n = M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}$$

به‌عبارت دیگر گشتاور n ام برابر با مشتق n ام تابع مولد گشتاور در نقطه $t = 0$ است. از طرفی داریم:

$$Var(X) = E(X^2) - [E(X)]^2$$

لذا برای به‌دست آوردن واریانس کافیست گشتاور اول و دوم متغیر X را محاسبه کنیم:

$$E(X) = \frac{d}{dt} M_X(t)\Big|_{t=0} = \left(\frac{-1}{4} e^{-t} + \frac{1}{4} e^{t}\Big|_{t=0}\right) = 0$$

$$E(X^2) = \frac{d^2}{dt^2} M_X(t)\Big|_{t=0} = \left(\frac{1}{4} e^{t} + \frac{1}{4} e^{t}\Big|_{t=0}\right) = \frac{1}{2}$$

$$\Rightarrow Var(X) = E(X^2) - E(X)^2 = \frac{1}{2}$$

۲۷. گزینه (۲)

در توزیع نرمال داریم: $P(\bar{X} > x_i) = P(\bar{X} < x_i) = \frac{۱}{۲}$

$$P(\bar{X} > \mu) = P(\bar{X} < \mu) = \frac{۱}{۲}$$

در نتیجه: $P(\bar{X} < \mu, \bar{X} > x_۱) = P(\bar{X} < \mu)P(\bar{X} > x_۱) = \frac{۱}{۲} \times \frac{۱}{۲} = \frac{۱}{۴}$

۲۸. با توجه به نادرست بودن صورت سوال حذف شده است.

با فرض $\mathrm{Var}\left[\frac{x_۱}{\sigma_۱} + \frac{x_۲}{\sigma_۲} + \frac{x_۳}{\sigma_۳}\right]$ داریم:

$$\mathrm{Var}\left[\frac{X_۱}{\sigma_۱} + \frac{X_۲}{\sigma_۲} + \frac{X_۳}{\sigma_۳}\right] = \frac{۱}{\sigma_۱^۲}\mathrm{Var}(X_۱) + \frac{۱}{\sigma_۲^۲}\mathrm{Var}(X_۲) + \frac{۱}{\sigma_۳^۲}\mathrm{Var}(X_۳)$$

$$+ \frac{۲}{\sigma_۱\sigma_۲}\mathrm{Cov}(X_۱, X_۲) + \frac{۲}{\sigma_۱\sigma_۳}\mathrm{Cov}(X_۱, X_۳) + \frac{۲}{\sigma_۲\sigma_۳}\mathrm{Cov}(X_۲, X_۳)$$

$$= \frac{۱}{\sigma_۱^۲} \times \sigma_۱^۲ + \frac{۱}{\sigma_۲^۲} \times \sigma_۲^۲ + \frac{۱}{\sigma_۳^۲} \times \sigma_۳^۲ + ۲\rho_{۱۲} + ۲\rho_{۱۳} + ۲\rho_{۲۳} = ۱+۱+۱+۱+۱+۱ = ۶$$

بنابراین با فرض درست بودن سوال، جواب گزینه ۲ است.

نکات:

نکته ۱: اگر X و Y دو متغیر تصادفی باشند و a و b ضرایب ثابت، آن‌گاه:

$$\mathrm{Var}(ax + by) = a^۲\mathrm{Var}(x) + b^۲\mathrm{Var}(y) + ۲ab\,\mathrm{Cov}(x, y)$$

اگر X و Y مستقل از یکدیگر باشند داریم:

$$\mathrm{Var}(ax + by) = a^۲\mathrm{Var}(x) + b^۲\,\mathrm{Var}(y)$$

نکته ۲: ضریب همبستگی بین دو متغیر تصادفی X و Y به‌صورت زیر می‌باشد:

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

روش‌های آماری

۲۹. گزینه (۴)

مدل رگرسیون خطی ساده y بر روی X به‌صورت زیر می‌باشد: $y_i = \beta_۰ + \beta_۱ X_i + \varepsilon_i \qquad , i = ۱, ..., n$

که در آن ε_iها دارای توزیع $N(۰, \sigma^۲)$ می‌باشند و ε_iها دو به دو ناهمبسته هستند یعنی $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = ۰$. این مدل رگرسیونی یک مدل ساده، خطی در پارامترها و خطی در متغیر پیش‌بین می‌باشد. ساده است به این دلیل که تنها یک متغیر پیش‌بین وجود دارد، خطی در پارامترها است چرا که هیچ پارامتری به‌صورت نمایی یا ضرب و تقسیم پارامتری دیگری بیان نشده است و همچنین خطی در متغیر پیش‌بین است چرا که این متغیر با توان یک ظاهر شده است. به این مدل خطی در پارامترها و متغیر پیش‌بین، مدل مرتبه اول می‌گویند.

برای تحلیل مدل رگرسیونی از رویکرد آنالیز واریانس که برمبنای تقسیم مجموع مربعات و درجه آزادی متناظر با متغیر پاسخ y می‌باشد استفاده می‌شود.

$$SSTO = SSR + SSE$$

مجموع مربعات کل (SSTO)، میزان تغییرپذیری مشاهدات پاسخ (y_i) را اندازه می‌گیرد که هرچه تغییرپذیری بین مشاهدات پاسخ بیش‌تر باشد SSTO بزرگ‌تر می‌شود که برابر است با $SSTO = \sum_{i=1}^{n} (y_i - \bar{y})^2$.

مجموع مربعات رگرسیونی (SSR)، معیاری است که آن بخش از تغییرات متغیر پاسخ که توسط خط رگرسیون بیان می‌شود را نشان می‌دهد. به‌عبارت دیگر، هرچه مقدار SSR نسبت به SSTO بیش‌تر شود، اثر متغیر X در محاسبه تغییرات کلی مشاهدات پاسخ بیش‌تر است. مقدار SSR به‌صورت زیر تعریف می‌شود:

$$SSR = \sum(\hat{y}_i - \bar{y})^2$$

مجموع مربعات خطا (SSE)، میزانی از تغییرپذیری متغیر پاسخ است که توسط خط رگرسیون قابل بیان نمی‌باشد و به‌صورت زیر تعریف می‌شود:

$$SSE = \sum(y_i - \hat{y}_i)^2$$

ضریب تعیین (R^2)، معیاری است که با استفاده از آن می‌توان اثر پیش‌گوی X را در کاهش تغییرپذیری متغیر پاسخ (کاهش عدم قطعیت در پیش‌بینی y) محاسبه نمود و برابر است با:

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$

باتوجه به این‌که $0 \le SSE \le SSTO$ لذا $0 \le R^2 \le 1$.

بنابراین ضریب تعیین به‌عنوان میزان کاهش در تغییرپذیری متغیر پاسخ هنگام استفاده از پیش‌گوی X می‌باشد. یعنی هرچه R^2 بزرگ‌تر باشد، استفاده از متغیر X کاهش بیش‌تری در مجموع تغییرات کلی متغیر پاسخ ایجاد می‌کند (اثر متغیر X در محاسبه تغییرات کلی مشاهدات پاسخ بیش‌تر است)

اگر همه مشاهدات بر روی خط رگرسیون برازش شده قرار بگیرند آن‌گاه $SSE = 0$ و $R^2 = 1$. به‌عبارت دیگر تمام تغییرات متغیر پاسخ y_i توسط پیشگوی X قابل بیان است.

هنگامی‌که خط رگرسیون برازش داه شده به‌صورت افقی باشد آن‌گاه $b_1 = 0$ و $\hat{y}_i = \bar{y}$. بنابراین $SSE = SSTO$ و $R^2 = 0$. به‌عبارت دیگر رابطه خطی بین X و y وجود ندارد و پیشگوی x هیچ‌گونه کمکی در کاهش تغییرپذیری مشاهدات y_i با استفاده از مدل رگرسیونی نمی‌کند.

۳۰. گزینه (۲)

تغییر مقیاس متغیر منجر به تغییرات متناظر در مقیاس ضرایب و خطاهای استاندارد خواهد شد، اما هیچ تغییری در معنی‌داری و تفسیر ضرایب ایجاد نمی‌کنند.

اگر متغیر X_j در عدد c ضرب شود آن‌گاه ضریب متناظر با آن بر عدد c تقسیم می‌شود، اما سایر ضرایب مدل ثابت می‌مانند.

اگر متغیر پاسخ در عدد c ضرب شود آن‌گاه تمام ضرایب رگرسیونی در عدد c ضرب می‌شوند. هیچ‌یک از آماره‌های آزمون t و F تحت تأثیر تغییر واحدهای اندازه‌گیری متغیرها قرار نمی‌گیرند.

Dependent / Independent	y	cy	Y
X_1	$\beta_1(se_1)$	$c\beta_1(c\times se_1)$	—
cX_1	—	—	$\beta_1/c(se_1/c)$
X_2	$\beta_2(se_2)$	$c\beta_2(c\times se_2)$	$\beta_2(se_2)$
Intencept	$\beta_0(se_0)$	$c\beta_0(c\times se_0)$	$\beta_0(se_0)$
SSR	SSR	$c^2\times SSR$	SSR
R-squarred	R^2	R^2	R^2

۳۱. گزینه (۲)

X =احتمال بهبودی $\sim Bin\left(n, p=\frac{1}{2}\right)$

$$p-value = p\left(RH_0 \middle| H_0\right) = p(X\le 1) = p(X=0)+p(X=1) = \binom{6}{0}\left(\frac{1}{2}\right)^6 + \binom{6}{1}\left(\frac{1}{2}\right)^6 = \frac{1}{64}+\frac{6}{64} = \frac{7}{64}$$

۳۲. گزینه (۳)

فرم کلی مدل رگرسیون خطی با $p-1$ متغیر پیشگو به‌صورت زیر می‌باشد:

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \qquad ,i = 1,...,n$$

که در آن $\beta_1,\ldots,\beta_{p-1}$ پارامترهای متناظر با متغیرهای $X_1,\ldots,X_{p-1}$ و β_0 عرض از مبداء مدل می‌باشد. مانده‌ها دارای توزیع نرمال $\varepsilon_i \sim N(0,\sigma^2)$ می‌باشند.

خطی بودن در این مدل اشاره به خطی بودن برحسب پارامترها (β_i) و نه خطی بودن سطوح پاسخ است. بنابراین گوئیم رگرسیون برحسب پارامترها خطی می‌باشد اگر بتوان آن را به فرم زیر نوشت:

$$y_i = c_{i0}\beta_0 + c_{i1}\beta_1 + c_{i2}\beta_2 + \ldots + c_{i,p-1}\beta_{p-1} + \varepsilon_i$$

که در آن $c_{i0},\ldots,c_{i,p-1}$ ضرایب β_iها هستند که همان پیشگوهای مدل می‌باشند. برای مثال مدل رگرسیونی $y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$ برحسب ضرایب $c_{i0}=1$ و $c_{i1}=X_{i1}$ و $c_{i2}=X_{i2}$ برحسب پارامترهای خطی می‌باشد.

اما مدل رگرسیونی $y_i = \beta_0 \exp\left(\beta_1 X_i\right) + \varepsilon_i$ یک مدل رگرسیون غیرخطی می‌باشد.

همچنین مدل $y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1}^2 + \beta_3 X_{i2} + \beta_4 X_{i1} X_{i2} + \varepsilon_i$ نیز یک مدل رگرسیون خطی برحسب پارامترها می‌باشد چرا که پارامترها به‌صورت خطی وارد مدل شده‌اند. حال به بررسی خطی بودن مدل‌های رگرسیونی مطرح شده می‌پردازیم:

$$1)\ y_i = \frac{1}{1+e^{\beta_0+\beta_1 X_i + e_i}} \Rightarrow 1+e^{\beta_0+\beta_1 X_i+e_i} = \frac{1}{y_i}$$

$$\Rightarrow \ln\left(\frac{1}{y_i}-1\right) = \beta_0+\beta_1 X_i + e_i$$

لذا با انجام تبدیل لگاریتمی مشاهده می‌شود که مدل تبدیل‌یافته برحسب پارامترهای خطی می‌باشد.

$$2)\ y_i = \beta_0 + \beta_1 \frac{1}{X_i^2-1}$$

با در نظرگرفتن $c_1 = \frac{1}{X_i^2 - 1}$ داریم: $y_i = \beta_1 + c_1\beta_1$ که به فرم خطی تبدیل می‌شود.

$$۳)\ y_i = \beta_0 + \beta_1 e^{\beta_2 X_i + e_i} \Rightarrow \frac{\ln y_i}{\ln \beta_0} = \ln\beta_1 + \beta_2 X_i + e_i$$

$$\Rightarrow \ln y_i = \left(\ln\beta_0\right)\left(\ln\beta_1\right) + \left(\ln\beta_0\right)\beta_2 X_i + \left(\ln\beta_0\right)e_i$$

پس از انجام تبدیل ملاحظه می‌گردد که مدل تبدیل‌یافته برحسب پارامترهای غیرخطی است.

$$۴)\ y_i = \beta_0 e^{\beta_1 X_i + e_i} \Rightarrow \ln y_i = \ln\beta_0 + \beta_1 X_i + \varepsilon_i$$

لذا با تبدیل لگاریتمی، مدل تبدیل‌یافته بر حسب پارامترها خطی می‌باشد.

۳۳. گزینه (۲)

برای به‌دست آوردن حجم نمونه موردنیاز در طرح دو نمونه‌ای در صورتی‌که حجم موردنیاز در یک گروه k برابر حجم گروه دیگر باشد داریم.

$$n_2 = \frac{\left(Z_{\frac{\alpha}{2}} + Z_\beta\right)^2 \sigma^2 (1+1/k)}{\varepsilon^2} \qquad , n_1 = kn_2$$

که در آن σ انحراف معیار مشترک و ε اندازه اثر نامیده می‌شوند. در صورتی‌که حجم نمونه در دو گروه برابر درنظر گرفته شود داریم:

$$n_1 = n_2 = \frac{2\left(Z_{\frac{\alpha}{2}} + Z_\beta\right)^2 \sigma^2}{\varepsilon^2}$$

لذا باتوجه به اطلاعات سؤال داریم:

$$n_1 = n_2 = \frac{2(10/5)\times\sigma^2}{\left(\frac{1}{2}\sigma\right)^2} = \frac{2\times 10/5\times\sigma^2}{\frac{1}{4}\sigma^2} = 8\times 10/5 = 84$$

۳۴. گزینه (۴)

در طرح‌های بلوکی، واحدهای آزمایشی ناهمگن به بلوک‌های همگن تقسیم می‌شوند و سپس تصادفی‌سازی از واحدهای آزمایشی درون هر بلوک انجام می‌شود.

به‌طور مثال فرض کنید در دو کارخانه A و B محصول نان در ۴ درجه حرارت مختلف تولید می‌شود. از آن‌جایی‌که ممکن است حجم نان تولید شده تحت تأثیر فرایندهای مختلف و مواد خام مورد استفاده در این دو کارخانه قرار بگیرد، و باتوجه به این‌که نان‌های تولید شده در هر کارخانه تقریباً مشابه به هم هستند، بنابراین استفاده از طرح بلوکی با اندازه ۴ در هر بلوک توصیه می‌شود.

طرح بلوکی کامل تصادفی

کارخانه (بلوک)	واحدهای آزمایش (درجه حرارت)			
	۱	۲	۳	۴
A	بالا	پایین	خیلی بالا	متوسط
B	متوسط	بالا	خیلی بالا	پایین

لذا مدل آماری خطی باید منعکس‌کننده اثر مداخله (درجه حرارت) و اثر بلوک (کارخانه تولید) باشد:

$$y = [\text{overall constant}] + [\text{Treatment Effect}] + [\text{Block Effect}] + [\text{Experimental Error}]$$

برای مثال فوق اثر درجه حرارت با استفاده از سه ایندیکیتور x_1 و x_2 و x_3 و اثر بلوک با ایندیکیتور x_4 وارد مدل می‌شوند. یعنی: $Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij3} + \beta_4 X_{ij4} + \varepsilon_{ij}$

گاهی مواقع باتوجه به شرایط قادر به انجام بررسی همه‌ی مداخلات (Treatment) در یک بلوک نیستیم و لذا اندازه هر بلوک کم‌تر از اندازه ملاحظات می‌شود.

به‌طور مثال فرض کنید که در یک کارخانه برای بررسی کیفیت یک محصول لبنی از ۵ فرمول‌بندی مختلف استفاده شده است. افراد مختلفی جهت بررسی کیفیت این محصول انتخاب می‌شوند. اما از آن‌جایی‌که فرمول‌بندی این محصولات خیلی مشابه یکدیگر می‌باشند، برای این‌که افراد تشخیص‌های درست‌تری داشته باشند تنها هر فرد ۳ محصول را بررسی خواهد کرد. باتوجه به این‌که افراد مختلف دارای سلایق متفاوتی می‌باشند بنابراین هر فرد به‌عنوان یک بلوک درنظر گرفته می‌شود. باتوجه به این محدودیت، مشاهده می‌شود که هر فرد نمی‌تواند هر پنج فرمول محصول را در یک زمان ارزیابی کند و تنها سه مورد از پنج مورد محصول را می‌توان مورد ارزیابی قرار دهد. لذا هر فرد بیانگر یک بلوک ناقص می‌باشد.

افراد ارزیاب (بلوک)	فرمول محصول				
	۱	۲	۳	۴	۵
۱	×	×	×		
۲	×	×		×	
۳	×	×			×
۴	×		×	×	
۵	×		×		×
۶	×			×	×
۷		×	×	×	
۸		×	×		×
۹		×		×	×
۱۰			×	×	×

در طرح‌های بلوکی تصادفی ناقص متعادل، تکرار هر مداخله (درمان) در کلیه بلوک‌ها برابر و کم‌تر از تعداد کل بلوک‌ها است. در مثال فوق هر مداخله دقیقاً سه بار با مداخله دیگر تکرار شده است. به‌طور مثال، فرمول‌بندی ۱و۲ باهم در بلوک‌های ۱، ۲ و ۳ تکرار شده‌اند.

۳۵. گزینه (۴)

از هر مجموعه‌ای با $p-1$ متغیر پیشگو، 2^{p-1} مدل‌های جایگزین می‌توانند ساخته شوند. این محاسبه براساس این واقعیت است که هر پیشگو را می‌توان از مدل خارج یا وارد مدل نمود. به‌عنوان مثال $2^4 = 16$ مدل مختلف از مجموع ۴ متغیر پیشگو می‌توان تشکیل داد. به این صورت که در ابتدا می‌توان مدل بدون متغیرهای X را به داده‌ها برازش داد یعنی $y_i = \beta_0 + \varepsilon_i$. سپس مدل‌های با تنها یک متغیر پیشگو (X_1, X_2, X_3, X_4)، مدل‌های با دو متغیر پیشگو $(X_1$ و X_2، X_1 و X_3، X_1 و X_4، ...) و غیره را به داده‌ها برازش داد.

با افزایش تعداد متغیرهای پیشگو، بررسی دقیق همه مدل‌های رگرسیونی احتمالی امکان‌پذیر نمی‌باشد. لذا روش‌های انتخاب مدل برای شناسایی گروه کوچکی از مدل‌های رگرسیون که براساس معیارهای مشخصی «مدل‌های مناسب‌تری» هستند توسعه یافته است. معیارهای زیادی برای مقایسه مدل‌های رگرسیونی وجود دارند که مهم‌ترین آن‌ها عبارتند از:

R_p^2، $R_{a,p}^2$، C_p، AIC_p، SBC_p و $PRESS_p$.

معیار R^2_p یا SSE_p: R^2_p که به‌عنوان ضریب تعیین چندگانه تعریف می‌شود، معیاری است که با استفاده از آن می‌توان اثر متغیرهای توضیحی $X_1,\ldots,X_p$ را در کاهش تغییرپذیری متغیر پاسخ محاسبه نمود که عبارتست از:

$$R^2_p = 1 - \frac{SSE_p}{SSTO}$$

دلیل استفاده از معیار R^2_p این نیست که مدلی را انتخاب کنیم که دارای بیش‌ترین مقدار R^2_p باشد. چرا که مقدار ضریب تعیین با اضافه کردن متغیرهای توضیحی به مدل کاهش پیدا نمی‌کند و بنابراین مدلی که شامل همه متغیرهای توضیحی باشد ($p-1$ متغیر) دارای بزرگ‌ترین مقدار R^2_p است.

بنابراین هدف استفاده از معیار R^2_p در انتخاب زیر مدل رگرسیونی مناسب از تعداد $p-1$ متغیر توضیحی اینست که دنبال نقطه‌ای باشیم که بعد از آن اضافه کردن متغیرهای توضیحی تغییر معنی‌داری در مقدار ضریب تعیین ایجاد نکند.

معیار $R^2_{a,p}$ یا MSE_p:

به‌دلیل محدودیت‌های معیار R^2_p، معیار $R^2_{a,p}$ (ضریب تعیین تعدیل شده) به‌عنوان معیاری جایگزین پیشنهاد می‌شود:

$$R^2_{a,p} = 1 - \left(\frac{n-1}{n-p}\right) \times \frac{SSE_p}{SSTO} = 1 - \frac{MSE_p}{\frac{SSTO}{n-1}}$$

این معیار، تعداد پارامترهای مدل رگرسیونی را از طریق درجه آزادی در نظر می‌گیرد.

با توجه به فرمول مقدار $R^2_{a,p}$ افزایش می‌یابد اگر و تنها اگر مقدار MSE_p کاهش پیدا کند چرا که مقدار $\frac{SSTO}{n-1}$ ثابت است.

برخلاف معیار R^2_p، با افزایش تعداد پارامترها مقدار $R^2_{a,p}$ ممکن است کاهش پیدا کند، بدین‌صورت که مقدار افزایش در معیار R^2_p به‌حدی کم است که برای جبران از دست دادن درجه آزادی اضافی کافی نیست.

معیار C_p (آماره مالوس):

این معیار با مجموع میانگین مربع خطای n مقدار برازش شده از هر زیر مدل رگرسیون مرتبط است.

$$\Gamma_p = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}\left(E(\hat{y}_i)-\mu_i\right)^2 + \sum_{i=1}^{n}\sigma^2(\hat{y}_i)\right]$$

یک برآوردگر مناسب Γ_p، C_p می‌باشد:

$$C_p = \frac{SSE_p}{MSE\left(X_1,\ldots,X_{p-1}\right)} - (n-2p)$$

اگر مدل برازش داده شده رگرسیونی با $p-1$ متغیر توضیحی دارای اریبی نباشد آن‌گاه $E\{\hat{y}_i\} \equiv \mu_i$ و داریم:

$$E\{C_p\} \approx p$$

بنابراین هنگامی که مقادیر C_p برای همه مدل‌های رگرسیونی احتمالی در مقابل p ترسیم شوند، مدل‌های با اریبی کم‌تر در نزدیکی خط $C_p = p$ قرار دارند. مدل‌های با اریبی قابل توجه بالای این خط و مدل‌های بدون اریبی پایین این خط قرار می‌گیرند.

در استفاده از معیار C_p به‌دنبال شناسایی زیرمجموعه‌هایی از متغیرهای $X_1,...,X_p$ هستیم که برای آن‌ها (۱) مقدار C_p کوچک است و (۲) مقدار C_p در نزدیکی p است.

مقدار کوچک C_p بیانگر کوچک بودن مجموع میانگین مربعات خطا است و نزدیکی مقدار C_p به p نشان‌دهنده کم بودن اریبی مدل رگرسیونی است.

معیارهای AIC_p و SBC_p:

علاوه‌بر معیارهای $R^2_{a,p}$ و C_p، معیارهای اطلاعاتی آکائیکه (AIC_p) و بیزی‌شوارتز (SBC_p) نیز برای اضافه کردن هر متغیر پیشگو در مدل جریمه‌ای را در نظر می‌گیرند که با درنظر گرفتن این جریمه نتیجه‌گیری می‌شود که آیا اضافه کردن پیشگوی اضافه به مدل باعث به‌دست آورن برآوردهای بهتری می‌شود یا خیر. این معیارها عبارتنداز:

$$AIC_p = n \ln SSE_p - n \ln n + 2p$$

$$SBC_p = n \ln SSE_p - n \ln n + [\ln n]p$$

معیار $RRESS_p$ (مجموع پیش‌بینی مربعات):

معیاری است که با استفاده از آن می‌توان بررسی نمود که مقادیر برازش شده یک زیر مدل رگرسیونی تا چه حد می‌توانند پاسخ‌های مشاهده شده y_i را به‌خوبی پیش‌یبنی نمایند.

مجموع مربعات خطا $SSE = \sum(y_i - \hat{y}_i)^2$ نیز دارای چنین ویژگی است. اما تفاوتی که معیار RRESS با آن دارد در اینست که برای به‌دست آوردن مقدار برازش شده $\hat{y}_i$، فرد iام از مطالعه حذف شده و برآورد تابع رگرسیونی برای زیر مدل مفروض با استفاده از $n-1$ مشاهده دیگر صورت می‌گیرد و با استفاده از آن مقدار پیش‌بینی شده $\hat{y}_{i(i)}$ برای فرد iام به‌دست می‌آید.

لذا مقدار خطای پیش‌بینی آماره PRESS برای فرد iام عبارتست‌از: $y_i - \hat{y}_{i(i)}$ و مقدار $PRESS_p$ عبارتست از مجموع مربعات خطای پیش‌بینی برای همه n فرد، یعنی:

$$PRESS_p = \sum_{i=1}^{n} (y_i - \hat{y}_{i(i)})^2$$

مدل‌های با مقادیر کم‌تر $PRESS_p$ به‌عنوان کاندیدای انتخاب مدل بهینه در نظر گرفته می‌شوند. عامل تورم واریانس (VIF) شاخصی است که برای بررسی وجود هم‌خطی چندگانه در مدل رگرسیونی می‌پردازد.

۳۶. گزینه (۲)

روش آزمون خطی عمومی شامل سه مرحله اصلی است:

۱) مدل کامل

ابتدا با مدلی شروع می‌کنیم که به‌نظر مناسب‌ترین مدل برای داده‌ها باشد که به مدل کامل یا نامحدود معروف است. برای مدل رگرسیونی خطی ساده با خطاهای نرمال، مدل کامل عبارتست از:

$$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

این مدل را با استفاده از روش کم‌ترین مربعات یا روش درست‌نمایی ماکزیمم به داده‌ها برازش داده و مجموع مربعات خطای آن را به دست می‌آوریم.

$$SSE(F) = \sum\left[y_i - (b_0 + b_1 X_i)\right]^2 = \sum(y_i - \hat{y}_i)^2 = SSE$$

۲) مدل کاهش‌یافته

مدل کاهش‌یافته شامل مدلی است که در آن متغیرهایی را که مایل به بررسی اثرات آن‌ها هستیم از مدل خارج می‌نماییم. در مورد مدل رگرسیون خطی ساده با خطاهای نرمال داریم:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

لذا مدل کاهش‌یافته آن عبارتست از:

$$y_i = \beta_0 + \varepsilon_i$$

مشابه حالت قبل با استفاده از روش کم‌ترین مربعات یا روش درست‌نمایی ماکزیمم این مدل را به داده‌ها برازش داده و مجموع مربعات خطای آن را به‌دست می‌آوریم.

$$SSE(R) = \sum (y_i - b_0)^2 = \sum (y_i - \bar{y})^2 = SSTO$$

۳) آماره آزمون

حال به بررسی مجموع مربعات خطای به‌دست آمده از برازش دو مدل کامل و کاهش‌یافته می‌پردازیم یعنی $SSE(F)$ و $SSE(R)$.

همواره داریم $SSE(R) \leq SSE(F)$. دلیل این امر اینست که پارامترهای بیش‌تری در مدل کامل هستند، بنابراین برازش بهتری بر روی داده‌ها ایجاد می‌کنند که دارای انحرافات کم‌تری در اطراف تابع رگرسیون هستند. هنگامی‌که مقدار $SSE(F)$ خیلی کم‌تر از مقدار $SSE(R)$ نباشد، آن‌گاه استفاده از مدل کامل در مقایسه با مدل کاهش‌یافته تأثیر چندانی در کاهش تغییرپذیری مشاهدات y_i در اطراف خط برازش داده شده نخواهد داشت، لذا استفاده از مدل کاهش‌یافته به‌جای مدل کامل توصیه می‌شود.

بنابراین تفاوت کم $SSE(R) - SSE(F)$ بیانگر برقراری فرض H_0 و تفاوت زیاد $SSE(R) - SSE(F)$ بیانگر برقراری فرض H_a می‌باشد.

آماره آزمون مناسب برای بررسی فرضیه H_0 در مقابل H_a عبارتست از:

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F}$$

که تحت فرض H_0 دارای توزیع $F(1-\alpha; df_R - df_F, df_F)$ می‌باشد. لذا قاعده تصمیم‌گیری براساس آن عبارتست از:

$$\begin{cases} \text{If } F^* \leq F(1-\alpha; df_R - df_F, df_F) \text{ , Conclude } H_0 \\ \text{If } F^* > F(1-\alpha; df_R - df_F, df_F) \text{ , Conclude } H_a \end{cases}$$

۳۷. گزینه (۱)

گاهی مواقع در طرح‌های بلوکی به‌جای استفاده از یک متغیر بلوک‌بندی شده، بیش‌تر از یک متغیر بلوک‌بندی شده مورد استفاده قرار می‌گیرد تا از خطاهای آزمایشی، تغییرات مربوط به هریک از متغیرهای بلوک‌بندی شده حذف شوند.

به‌عنوان مثال متغیرهای بلوک‌بندی ممکن است شامل سن و درآمد افراد باشد که هر بلوک خاص شامل افرادی با گروه سنی مشخص و درآمد معلومی می‌باشند.

با این‌حال، استفاده از طرح‌های بلوکی با بیش از یک متغیر بلوک‌بندی، اغلب نیازمند تعداد زیادی واحدهای آزمایشی است. به‌طور مثال اگر سن و درآمد هرکدام شامل ۶ کلاس باشند آن‌گاه ۳۶ بلوک تشکیل می‌شود، حال اگر تعداد درمان‌هایی که قرار است مورد مطالعه قرار بگیرند ۶ درمان باشد، آن‌گاه مطالعه نیازمند ۲۱۶ فرد برای بررسی است.

با درنظر گرفتن ملاحظات مربوط به هزینه ممکن است مجاز به استفاده از ۲۱۶ واحد آزمایشی نباشیم، لذا باید دنبال طرح جایگزینی بود که با استفاده همزمان این دو متغیر بلوکی و تعداد کم‌تری از واحدهای آزمایشی، واریانس خطای آزمایشی را به‌اندازه کافی کاهش دهد. طرح مربع لاتین، طرح مفیدی برای چنین وضعیتی می‌باشد.

فرض کنید آزمایش شامل سه درمان A و B و C باشد. همچنین فرض کنید که روزهای هفته (دوشنبه، سه‌شنبه و چهارشنبه) و اپراتور (شماره ۱ و شماره ۲ و شماره ۳) به‌عنوان متغیرهای بلوک‌بندی استفاده می‌شوند. یک طرح مربع لاتین به‌صورت زیر است:

روز	اپراتور		
	۱	۲	۳
دوشنبه	B	A	C
سه‌شنبه	A	C	B
چهارشنبه	C	B	A

توجه داشته باشید که هر اپراتور هر درمان را اجرا می‌کند و تمام درمان‌ها در هر روز اجرا می‌شوند.

طرح مربع لاتین دارای خصوصیات زیر می‌باشد:

۱) دارای r درمان (Treatment) می‌باشد.

۲) دارای دو متغیر بلوک‌بندی هریک با r کلاس می‌باشد.

۳) هر سطر و هر ستون شامل تمام درمان‌ها می‌باشد.

مزایای این طرح عبارتست از:

۱) استفاده همزمان از دو متغیر بلوک‌بندی به‌جای یک متغیر بلوک‌بندی باعث کاهش بیش‌تری در تغییرات خطاهای آزمایشی می‌شود.

۲) در طرح‌های اندازه‌های تکراری (repeated measure) با استفاده از طرح مربع لاتین قادر به بررسی ترتیب مکانی اثرات درمان‌ها هستیم.

معایب این طرح عبارتنداز:

۱) تعداد کلاس‌های هر متغیر بلوک‌بندی باید برابر با تعداد درمان‌ها باشد که این محدودیت در عمل سخت است.

۲) فرضیه‌های مدل محدود می‌باشند (به‌عنوان مثال هیچ اثر متقابلی بین هریک از متغیرهای بلوکی با درمان و همچنین بین دو متغیر بلوکی وجود ندارد).

۳) استفاده از طرح مربع لاتین زمانی‌که تعداد کمی درمان مورد مطالعه قرار می‌گیرند، به تعداد کمی از درجه آزادی برای خطای آزمایشی منجر خواهد شد. از طرف دیگر، هنگامی‌که تعداد درمان‌های مورد مطالعه زیر باشند، درجه آزادی خطای آزمایشی ممکن است بزرگ‌تر از ضرورت باشد.

باتوجه به این محدودیت هنگامی‌که بیش از ۸ درمان مورد بررسی قرار بگیرند، طرح مربع لاتین به‌ندرت مورد استفاده قرار می‌گیرد. اگر تعداد درمان‌ها کم (کم‌تر از ۴) باشد آن‌گاه نیازمند تکرارهای اضافه درمان‌ها هستیم.

مدل طرح مربع لاتین با درمان‌های ثابت و اثرات بلوکی به‌صورت زیر است:

$$y_{ijk} = \mu_{\circ\circ\circ} + p_i + k_j + \tau_k + \varepsilon_{ijk}$$

که:

$\mu_{\circ\circ\circ}$ is constant

$p_i, k_j \ \tau_k$ are constants subject to the restrictions $\sum p_i = \sum k_j = \sum \tau_k = \circ$

ε_{ijk} are independent $N(\circ, \sigma^2)$

$i = 1,\ldots,r \ ; \ j = 1,\ldots,r \ ; \ k = 1,\ldots,r$

p_i اثر متغیر بلوکی سطری، k_j اثر متغیر بلوک ستونی و τ_k اثر اصلی درمان می‌باشد. جدول زیر، جدول آنالیز واریانس طرح مربع لاتین را با اثرات ثابت نشان می‌دهد:

Source of variation	ss	df	MS	E{MS}
Row blocking variable	SSROW	$r-1$	$MSROW=\frac{SSROW}{r-1}$	$\sigma^2+r\frac{\sum p_i^2}{r-1}$
Column blocking variable	SSCOL	$r-1$	$MSCOL=\frac{SSCOL}{r-1}$	$\sigma^2+r\frac{\sum k_j^2}{r-1}$
Treatments	SSTR	$r-1$	$MSTR=\frac{SSTR}{r-1}$	$\sigma^2+r\frac{\sum \tau_k^2}{r-1}$
Error	SSRem	$(r-1)(r-2)$	$MSRem=\frac{SSRem}{(r-1)(r-2)}$	σ^2
Total	SSTO	r^2-1		

در حالتی‌که n تکرار در هر سلول داشته باشیم مدل طرح به‌صورت زیر می‌شود:

$$y_{ijkm}=\mu_{...}+p_i+k_j+\tau_k+\varepsilon_{ijkm}$$
$$i=1,...,r\ ;\ j=1,...,r\ ;\ k=1,...,r\ ;\ m=1,...,n$$

جدول زیر، جدول آنالیز واریانس طرح مربع لاتین با تعداد n تکرار در هر سلول را نشان می‌دهد:

Source of variation	ss	df	MS	MS
Row blocking variable	SSROW	$r-1$		MSROW
Column blocking variable	SSCOL	$r-1$		MSCOL
Treatments	SSTR	$r-1$		MSTR
Error	SSRem	nr^2-3r+2		MSRem
Total	SSTO	nr^2-1		

باتوجه به اطلاعات مسئله داریم: $\begin{matrix} r=3 \\ n=2 \end{matrix} \Rightarrow df(Error)=nr^2-3r+2=2(9)-(3\times3)+2=11$

۳۸. گزینه (۱)

واریانس (σ^2) عبارت خطا (ε_i) در مدل رگرسیون $y_i=\beta_0+\beta_i X_i+\varepsilon_i$ با خطاهای نرمال برای به‌دست آوردن استنباط‌های مختلف در مورد تابع رگرسیونی و پیش‌بینی مقادیر y باید برآورد شود. یک برآوردگر ناریب برای σ^2، MSE می‌باشد یعنی $E(MSE)=\sigma^2$.

$$S^2=MSE=\frac{SSE}{n-p-1}=\frac{\sum(y_i-\hat{y}_i)^2}{n-p-1}$$

که در آن SSE دارای درجه آزادی $n-p-1$ متناظر با p متغیر توضیحی وارد شده در مدل می‌باشد. اگر توزیع خطاها در مدل رگرسیونی انحراف زیادی از توزیع نرمال داشته باشد آن‌گاه MSE برآوردگر نیرومند σ^2 نخواهد بود. برآوردگر نیرومند، برآوردگری است که درصورت انحراف از توزیع نرمال، $E|MSE|\simeq\sigma^2$. در صورت عدم برخورداری خطاها از توزیع نرمال، برآوردگر انحراف مطلق میانه (MAD) برآوردگری نیرومند برای σ خواهد بود.

$$MAD=\frac{1}{0/6745}\text{median}\left\{\left|e_i-\text{median}\{e_i\}\right|\right\}$$

۳۹. گزینه (۳)

ضریب همبستگی رتبه‌ای کندال (τ) یک آماره ناپارامتری است که برای سنجش همبستگی آماری میان دو متغیر تصادفی رتبه‌ای به‌کار می‌رود.

فرض کنید $(X_1,y_1),(X_2,Y_2),...,(X_n,Y_n)$ مقادیری از دو متغیر با توزیع توأم X و y باشند. دو زوج (X_i,y_i) و (X_j,y_j) را سازگار (concordant) می‌نامیم اگر رتبه‌های دو مشاهده باهم برابر باشند یعنی هرگاه $X_i > X_j$ آن‌گاه $y_i > y_j$ یا اگر $X_i < X_j$ آن‌گاه $y_i < y_j$ برقرار باشند.

دو زوج نازسازگار (discordant) نامیده می‌شوند هرگاه $X_i > X_j$ آن‌گاه $y_i < y_j$ یا اگر $X_i < X_j$ آن‌گاه $y_i > y_j$ باشد. اگر $X_i = X_j$ یا $y_i = y_j$، زوج نه سازگار و نه ناسازگار نامیده می‌شوند. آن‌گاه ضریب همبستگی کندال به‌صورت زیر محاسبه می‌شود:

$$\tau = \frac{\text{تعداد زوج‌های ناسازگار - تعداد زوج‌های سازگار}}{\frac{1}{2}n(n-1)}$$

مقدار این ضریب بین ۱+ و ۱- که این دو مقدار متناظر با تطابق و عدم تطابق کامل است. در صورت سؤال $n = 6$ فرد به‌لحاظ متغیرهای نگرش (X) و عملکرد (y) مورد بررسی قرار گرفته‌اند. تعداد کل زوج‌های مورد بررسی عبارتست از $\binom{6}{2} = 15$ زوج. باتوجه به این‌که ۳ زوج ناسازگار بوده و هیچ زوج هم رتبه‌ای نیز نداشته‌ایم داریم.

$$\tau = \frac{12-3}{\frac{1}{2}\times 6\times(5)} = \frac{9}{15} = \frac{3}{5} = 0/6$$

۴۰. گزینه (۳)

طرح اندازه‌های تکراری افراد یکسانی (شخص، فروشگاه، کارخانه و ...) برای هریک از روش‌های مورد مطالعه مورد استفاده قرار می‌دهد. این کار معمولا به دو صورت رخ می‌دهد:

۱) زمانی‌که افراد چندین بار در طول زمان اندازه‌گیری می‌شوند تا تغییرات روش‌های مورد مطالعه را مشاهده نمایند.

۲) هنگامی‌که افراد تحت وضعیت‌ها یا آزمایش‌های مختلف قرار می‌گیرند و پاسخ به هریک از این وضعیت‌ها با هم مقایسه می‌شوند.

برای مثال بااستفاده از یک طرح اندازه‌های تکراری می‌توان بررسی نمود که آیا مصرف سیگار در بین افراد با مصرف بالای سیگار بعد از یک برنامه هیپنوتراپی (به‌عنوان مثال با سه نقطه زمانی: بلافاصله قبل، یک ماه بعد و ۶ ماه بعد از برنامه هیپنوتیزم) تغییر کرده است.

به‌عنوان مثال دیگر، برای بررسی این‌که آیا رنگ شیشه جلوی ماشین بر میزان سرعت راننده تأثیر می‌گذارد یا خیر می‌توان با استفاده از طرح اندازه‌های تکراری میزان سرعت ماشین را درچهار وضعیت بدون رنگ، رنگ ضخیم، رنگ متوسط و رنگ تیره شیشه جلوی ماشین بررسی نمود.

مزیت اصلی طرح‌های اندازه‌گیری تکراری این است که دقت خوبی برای مقایسه مداخلات ارائه می‌دهند چرا که همه منابع تغییرپذیری بین افراد از خطای آزمایش حذف می‌شوند و فقط تغییرات درون فردی وارد خطاهای آزمایش می‌شوند.

هنگامی‌که قصد استفاده از طرح اندازه‌های تکراری را برای تجزیه و تحلیل داده‌ها داریم، ابتدا باید اطمینان حاصل نمود که این داده‌ها با استفاده از طرح اندازه‌های تکراری قابل تحلیل هستند. بنابراین تنها زمانی استفاده از طرح اندازه‌های تکراری مناسب است که فرضیه‌های موردنیاز این طرح در مورد داده‌ها برقرار باشد. که مهم‌ترین آن‌ها عبارتند از:

۱) عدم وجود داده‌های پرت معنی‌دار در مشاهدات

۲) داده‌ها در گروه‌های مورد مطالعه به‌صورت تقریبی دارای توزیع نرمال باشند.

۳) برقراری شرط کرویت: یعنی همبستگی بین مشاهدات در همه تکرارهای زمانی یکسان است.

۴۱. گزینه (۲)

در طرح آنالیز واریانس و برخی روش‌های دیگر دو نوع از عوامل وجود دارند: عوامل با اثرات ثابت و عوامل با اثرات تصادفی. مناسب بودن هرکدام از این دو نوع بستگی به زمینه مشکل، سؤالات مورد علاقه و نحوه جمع‌آوری داده‌ها دارد.

عامل اثرات ثابت: داده‌ها برای تمام سطوح عامل موردنظر جمع‌آوری شده‌اند. به‌طور مثال: در یک آزمایش هدف مقایسه اثرات سه دوز مختلف یک دارو بر روی پاسخ بیماران می‌باشد. در این‌جا، "مقدار مصرف دوز" عامل، و سه سطح مختلف دوز، سطوح عامل مورد مطالعه می‌باشند و هیچ مطلبی در مورد دوزهای دیگر گفته نشده است.

عامل اثرات تصادفی: عامل مورد نظر دارای سطوح زیادی است که علاقه‌مند به بررسی تمام سطوح عامل می‌باشیم اما تنها یک نمونه تصادفی از سطوح در داده‌ها وجود دارد.

به‌طور مثال یک تولیدکننده بزرگ محصولات لوازم خانگی، علاقه‌مند به مطالعه اثر اپراتور ماشین بر کیفیت نهایی محصول است. محقق یک نمونه تصادفی از اپراتورها را از تعداد زیادی از اپراتورها برای بررسی کیفیت نهایی محصول انتخاب می‌نماید.

تجزیه و تحلیل داده‌ها بسته به این‌که عامل موردنظر ثابت یا تصادفی می‌باشد متفاوت است. بنابراین اگر عامل به‌صورت نادرست طبقه‌بندی شود ممکن است منجر به نتیجه‌گیری نادرست شود. مدل Anova با اثرات تصادفی با حجم نمونه برابر در هریک از سطوح عامل عبارتست از:

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

Where:

$$\mu_i \text{ are independent } N(\mu, \sigma_\mu^2)$$

$$\varepsilon_{ij} \text{ are independent } N(0, \sigma^2)$$

$$\mu_i \text{ and } \varepsilon_{ij} \text{ are independent random variables}$$

$$i = 1,\dots,r \;\; ; \;\; j = 1,\dots,n$$

در مدل Anova با اثرات تصادفی، استنباط‌ها معمولا به‌طور مجزا در مورد هر سطح عامل انجام نمی‌شود و علاقه‌مند به استنباط در مورد کل سطوح عامل هستیم. بنابراین میزان کل تغییرپذیری متغیر پاسخ را که به‌وسیله تغییرپذیری سطوح عامل قابل محاسبه است به‌دست می‌آوریم که عبارتست از:

$$\frac{\sigma_\mu^2}{\sigma^2 y} = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2}$$

که مقداری بین صفر و یک را شامل می‌شود. هرچه درصد تغییرات بیان شده متغیر پاسخ توسط سطوح عامل موردنظر بیش‌تر باشد یعنی σ_μ^2 نسبت به σ^2 بزرگ‌تر باشد، این مقدار به‌سمت یک میل می‌نماید.

لذا برای بررسی این‌که آیا میانگین μ_iها در سطوح مختلف برابر هستند یا خیر آزمون زیر را انجام می‌دهیم:

$$H_0 : \sigma_\mu^2 = 0$$

$$H_a : \sigma_\mu^2 > 0$$

فرضیه H_0 بیانگر اینست که همه μ_iها برابر هستند یعنی $\mu_i \equiv \mu$. اما فرضیه H_a بیانگر این است که μ_iها متفاوت هستند.

تحلیل واریانس Anova با اثرات تصادفی مشابه با Anova با اثرات ثابت می‌باشد.

$$F^* = \frac{MSTR}{MSE} = \frac{\sigma^2}{\sigma^2 + n\sigma_\mu^2}$$

اگر $F^* > F[1-\alpha; r-1, r(n-1)]$، آن‌گاه فرضیه صفر رد می‌شود.

۴۲. گزینه (۲) این سوال حذف شده است.

مجموع مربعات اضافی (Extra sum of squares)، میزان کاهش در مجموع مربعات خطا را هنگامی‌که یک یا چند متغیر پیشگوی جدید وارد مدل رگسیونی می‌شوند (با توجه به این‌که متغیرهای توضیحی دیگر در مدل هستند) را اندازه‌گیری می‌کند.

به‌طور معادل می‌توان مجموع مربعات اضافی را به‌عنوان میزان افزایش در مجموع مربعات رگرسیونی هنگامی‌که یک یا چند متغیر توضیحی وارد مدل می‌شوند تعریف کرد.

به‌طور مثال $SSR(X_2 \mid X_1)$ مجموع مربعات اضافی مربوط به متغیر X_2 وقتی‌که متغیر X_1 در مدل وجود دارد براساس دو تعریف بالا به‌صورت زیر است:

$$\begin{cases} SSR(X_2 \mid X_1) = SSE(X_1) - SSE(X_1, X_2) \\ SSR(X_2 \mid X_1) = SSR(X_1, X_2) - SSR(X_1) \end{cases}$$

همچنین داریم:

$$SSTO = SSR(X_1, X_2) + SSE(X_1, X_2) = SSR(X_1) + SSR(X_2 \mid X_1) + SSE(X_1, X_2)$$

باتوجه به صورت سؤال متغیرهای X_1 و X_2 در مدل وجود دارند، بنابراین مجموع مربعات اضافی مربوط به متغیر x_3 هنگامی‌که دو متغیر X_1 و X_2 در مدل هستند برابر است با:

$$SSR(X_3 \mid X_1, X_2) = SSR(X_1, X_2, X_3) - SSR(X_1, X_2)$$

یا

$$SSR(X_3 \mid X_1, X_2) = SSE(X_1, X_2) - SSE(X_1, X_2, X_3)$$

جواب در بین هیچ‌کدام از گزینه‌ها نمی‌باشد.

۴۳. گزینه (۱)

به سؤال ۳۵ رجوع شود.

۴۴. گزینه (۳)

مدل رگرسیون خطی ساده با یک متغیر پیشگو به‌صورت زیر می‌باشد:

$$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

که در آن ε_iها خطاهای تصادفی با $E(\varepsilon_i) = 0$ و $\sigma^2(\varepsilon_i) = \sigma^2$ می‌باشند و دو به دو ناهمبسته هستند یعنی $\sigma(\varepsilon_i, \varepsilon_j) = 0$ برای همه $i \neq j$.

برای برآورد پارامترهای این مدل از روش حداقل مربعات استفاده می‌شود که این روش میزان انحراف مشاهدات را از مقادیر موردانتظار در نظر می‌گیرد یعنی:

$$Q = \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 X_i)\right)^2$$

باحل این معادله نسبت به پارامترهای β_0 و β_1 برآوردهای کم‌ترین مربعات پارامترها به‌صورت زیر به‌دست می‌آیند:

$$b_0 = \bar{y} - b_1 \bar{X}$$

$$b_1 = \frac{\sum (X_i - \bar{X})(y_i - \bar{y})}{\sum (X_i - \bar{X})^2}$$

قضیه گاوس- مارکف بیان می‌کند که تحت شرایط مدل رگرسیونی، برآوردهای حداقل مربعات b_0 و b_1، برآوردگرهایی نااریب و در بین همه برآوردهای نااریب خطی دارای کم‌ترین واریانس هستند.

نااریب هستند یعنی $E(b_0)=\beta_0$ و $E(b_1)=\beta_1$.

دارای کم‌ترین واریانس در بین همه برآوردهای نااریب خطی هستند یعنی این برآوردگرها نسبت به هر برآوردگر دیگری متعلق به کلاس برآوردگرهای نااریب که یک ترکیب خطی از مشاهدات پاسخ هستند تغییرپذیری کم‌تری دارند.

۴۵. گزینه (۱)

هنگامی‌که یک مدل رگرسیونی برای برازش به داده‌ها مورد استفاده قرار می‌گیرد، معمولا نمی‌توان مطمئن بود که مدل مناسب داده‌ها می‌باشد. یک یا چند ویژگی مدل مانند خطی بودن تابع رگرسیونی یا نرمال بودن مانده‌های مدل، ممکن است برای داده‌های موجود برقرار نباشد. از این‌رو مهم است تا بررسی مناسب بودن مدل برای داده‌ها قبل از انجام استنباط براساس آن مدل صورت پذیرد. اگر مدل رگرسیون خطی برای برازش بر روی داده‌ها مناسب نباشد دو انتخاب اساسی وجود دارد:

۱) رها کردن مدل رگرسیونی و توسعه و استفاده از یک مدل مناسب‌تر

۲) استفاده از تبدیلات مناسب تا مدل رگرسیون برای داده‌های تبدیل شده مناسب باشد.

عدم نرمال بودن خطاها و نابرابری واریانس خطاها معمولا باهم رخ می‌دهند. لذا برای اصلاح این انحراف‌ها در مدل رگرسیون خطی ساده باید تبدیل مناسب بر روی متغیر پاسخ y صورت پذیرد. روش باکس- کاکس به‌طور خودکار یک تبدیل مناسب را از یک خانواده از تبدیلات توانی بر روی متغیر y شناسایی می‌کند.

خانواده تبدیلات توانی به فرم $y'=y^{\lambda}$ می‌باشد که پارامتر λ با استفاده از داده‌ها مشخص می‌شود. لذا مدل رگرسیونی به فرم زیر در می‌آید:

$$y_i^{\lambda}=\beta_0+\beta_1 X_i+\varepsilon_i$$

مدل رگرسیونی شامل یک پارامتر اضافی λ علاوه‌بر سایر پارامترهای مدل می‌باشد که باید برآورد شود. روش باکس- کاکس با استفاده از روش درست‌نمایی ماکزیمم پارامتر λ را مشابه برآورد سایر پارامترها برآورد می‌نماید.

۴۶. گزینه (۲)

فرم کلی مدل رگرسیون خطی با $p-1$ متغیر پیشگو به‌صورت زیر می‌باشد:

$$y_i=\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+\ldots+\beta_{p-1}X_{i,p-1}+\varepsilon_i$$

Where:

$\beta_0,\beta_1,\ldots,\beta_{p-1}$ are parameters

$X_{i1},\ldots,X_{i,p-1}$ are known constants

ε_i are independent $N(0,\sigma^2)$

$i=1,\ldots,n$

فرم ماتریسی مدل فوق به‌صورت زیر می‌باشد:

$$y_{n\times 1}=X_{n\times p}\beta_{p\times 1}+\varepsilon_{n\times 1}$$

که:

$$y_{n\times 1}=\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad , \quad X_{n\times p}=\begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix}$$

$$\beta_{p\times 1}=\begin{bmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_{p-1}\end{bmatrix} \quad , \quad \varepsilon_{n\times 1}=\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{bmatrix}$$

جدول Anova برای مدل رگرسیون خطی به‌صورت زیر می‌باشد:

Source of variation	ss	df	MS
Regression	$SSR = b'x'y - \frac{1}{n}x'Jy$	$p-1$	$MSR = \frac{SSR}{p-1}$
Error	$SSE = y'y - b'x'y$	$n-p$	$MSE = \frac{SSE}{n-p}$
Total	$SSTO = y'y - \frac{1}{n}y'Jy$	$n-1$	

که در آن J یک ماتریس $n\times n$ با مقادیر یک می‌باشد و b برآورد حداقل مربعات ضرایب رگرسیونی مدل می‌باشد یعنی $b=(x'x)^{-1}x'y$.

در مدل رگرسیون خطی، MSE یک برآوردگر نااریب σ^2 می‌باشد یعنی $E(MSE)=\sigma^2$. لذا درجه آزادی واریانس خطا برابر است با $n-p$.

باتوجه به این‌که تعداد مشاهدات برابر $n=19$ و تعداد پارامترهای مدل برابر با $p=3$ می‌باشد، بنابراین درجه آزادی واریانس خطا برابر است با $n-p=16$.

۴۷. گزینه (۴)

حجم نمونه روش نمونه‌گیری تصادفی ساده× اثر طرح = حجم نمونه موردنیاز در نمونه‌گیری خوشه‌ای

$N_{cluster}=3/2\times 1000=3200$ نفر

باتوجه به این‌که متوسط تعداد افراد هر خانوار ۴ نفر می‌باشد لذا $\frac{3200}{4}=800$ خانوار برای برآورد شیوع بایستی مورد مطالعه قرار بگیرند.

۴۸. گزینه (۴)

برای کاهش و یا حذف نابرابری واریانس خطا یک روش استفاده از تبدیلات مناسب بر روی متغیر پاسخ y می‌باشد. مشکلی که با تبدیل متغیر y ممکن است رخ دهد اینست که منجر به ایجاد یک رابطه رگرسیونی نامناسب شود.

زمانی‌که به‌وسیله یک مدل رگرسیونی مناسب تشخیص داده شد که واریانس خطاها ثابت نیستند، روش حداقل مربعات وزنی جایگزین مناسبی برای تبدیلات متغیر پاسخ می‌باشد که روشی براساس تعمیم مدل رگرسیون چندگانه است.

$$y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\ldots+\beta_{p-1}X_{i,p-1}+\varepsilon_i$$

Where:

$\beta_0,\beta_1,\ldots,\beta_{p-1}$ are paramenters

$X_{i1},\ldots,X_{i,p-1}$ are known constants

ε_i are independent $N(0,\sigma_i^2)$

$i=1,\ldots,n$

لذا در این مدل خطاها دارای واریانس ثابت σ^2 نمی‌باشند.

برای برآورد ضرایب رگرسیونی مدل تعمیم‌یافته می‌توان از برآوردگرهای مدل رگرسیون معمولی با واریانس ثابت $\sigma^{۲}$ استفاده نمود. این برآوردگرها نااریب می‌باشند اما دیگر دارای واریانس مینیمم نیستند. برای به‌دست آوردن برآوردگرهای نااریب با کم‌ترین واریانس، باید میزان اطلاعاتی را که هرکدام از مشاهدات در مورد تابع رگرسیون فراهم می‌کنند به‌صورت وزنی وارد مدل نمائیم به این صورت که مشاهدات با واریانس کم‌تر اطلاعات دقیق‌تری را در مورد تابع رگرسیون به‌دست می‌دهند.

روش‌های مختلفی برای وزن‌دهی مشاهدات وجود دارد که در فرم ساده آن اگر واریانس خطاها معلوم باشد $w_i = \frac{۱}{\sigma_i^{۲}}$

در نظر گرفته می‌شود لذا فرم تابع درست‌نمایی مشاهدات به‌صورت زیر می‌شود:

$$L(\beta) = \left[\prod_{i=۱}^{n}\left(\frac{w_i}{۲\pi}\right)^{\frac{۱}{۲}}\right]\exp\left[-\frac{۱}{۲}\sum_{i=۱}^{n} w_i\left(y_i - \beta_۰ - \beta_۱X_{i۱} - \ldots - \beta_{p-۱}X_{i,p-۱}\right)\right]$$

۴۹. گزینه (۲)

اگر متغیر تصادفی x دارای توزیع پوآسن با پارامتر λ باشد آن‌گاه دارای تابع احتمال زیر است:

$$f_\lambda(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad , x = ۰,۱,۲,\ldots, \lambda > ۰$$

اگر $x \sim p(\lambda)$ آن‌گاه $E(x) = \lambda$.

لذا احتمال مشاهده متوسط تعداد تصادفات در طول یک هفته بیش‌تر است یعنی:

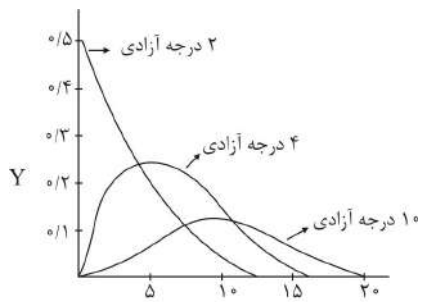
$$E(X) = \frac{۳+۱+\ldots+۲+۱}{۱۰} = \frac{۲۰}{۱۰} = ۲$$

۵۰. گزینه (۴)

به سؤال ۴۱ رجوع شود.

۵۱. گزینه (۱)

اگر متغیر تصادفی X دارای توزیع کای- دو با n درجه آزادی باشد یعنی $X \sim \chi^{۲}_{(n)}$، آن‌گاه میانگین و واریانس آن برابر است با:



chi-square statistics

یکی از ویژگی‌های توزیع کای- دو اینست که با افزایش درجه آزادی شکل توزیع شبیه به توزیع نرمال می‌شود. بنابراین زمانی‌که تعداد سطرها یا ستون‌ها کاهش پیدا کند، درجه آزادی نیز کاهش می‌یابد که این امر باعث می‌شود شکل منحنی از توزیع نرمال فاصله داشته باشد به این معنی که پراکندگی توزیع افزایش می‌یابد.

شماره تماس با نمایندگی‌های فعال و رسمی

آناهیـد گستـر خـلیـلی

نمایندگی	شماره تماس	نمایندگی	شماره تماس
اراک (آقای حسینی)	۰۹۹۱۳۷۱۳۷۵۰	سبزوار (آقای شریفان)	۰۹۱۰۱۷۱۱۸۷۲
اردبیل (خانم عاصمی‌زاده)	۰۹۱۹۹۱۰۱۲۴۸	شاهرود (آقای سلمانی)	۰۹۹۱۳۷۱۳۷۴۱
ارومیه (آقای آرمیون)	۰۹۱۹۹۱۰۱۲۴۱	شهرکرد (خانم تقی‌پور)	۰۹۱۹۹۱۰۱۲۴۹
اسفراین (خانم اسماعیل‌زاده)	۰۹۱۹۶۳۲۱۸۵۲	شیراز (آقای فروردین/ خانم هوشمندی)	۰۹۱۹۵۷۳۰۱۵۲
اصفهان (آقای کیانی)	۰۹۱۹۵۷۳۰۱۵۰	فسا (خانم شعبانی)	۰۹۱۹۳۷۱۳۷۴۲
اهواز (آقای دکتر رضازاده)	۰۹۱۹۵۷۳۰۱۵۴	قزوین (خانم چگینی)	۰۹۱۹۵۷۳۰۱۴۹
ایلام (آقای دکتر بوچانی)	۰۹۱۹۰۹۸۰۳۱۸	کرج (آقای دکتر علیرضاپور)	۰۹۱۹۷۷۸۱۹۴۷
بروجرد (آقای احمدی‌صدر)	۰۹۱۹۶۸۵۳۱۱۶	کرمان (خانم ضمانتی‌یار)	۰۹۱۳۱۹۸۱۲۷۰
بم (خانم سرحدی‌نژاد)	۰۹۹۱۳۷۱۳۷۴۹	کرمانشاه (آقای ابراهیمی)	۰۹۱۹۵۷۳۰۱۴۸
بیرجند (آقای دکتر بهروان)	۰۹۱۹۵۹۰۷۲۰۳	کوهدشت (خانم یاریان)	۰۹۱۹۵۳۷۱۸۹۰
تبریز (خانم عاصمی‌زاده)	۰۹۱۹۵۷۳۰۱۴۷	گرگان (آقای شاه‌علی)	۰۹۱۹۹۱۰۱۲۴۷
جهرم (آقای یاعلی جهرمی)	۰۹۰۱۳۷۳۷۸۹۶	مراغه (آقای صمدی)	۰۹۱۹۵۷۳۳۱۷۹
جیرفت (خانم دکتر محمدی)	۰۹۱۹۹۱۰۱۲۴۰	مشهد (خانم دکتر شجاعی)	۰۹۹۱۳۷۱۳۷۴۴
خرم‌آباد (خانم میر)	۰۹۱۹۲۷۰۵۸۷۸	میاندوآب (آقای صمدی)	۰۹۱۹۵۷۳۳۱۷۹
رشت (خانم تنها)	۰۹۱۹۵۷۳۰۱۵۳	نیشابور (آقای موسوی)	۰۹۱۹۶۳۵۰۷۶۸
رفسنجان (آقای یوسفی)	۰۹۱۹۶۸۲۹۲۸۰	هرمزگان (آقای زارعی)	۰۹۱۹۵۷۳۳۲۸۴
زاهدان (آقای مهمان‌دوست)	۰۹۱۹۹۱۰۱۲۴۵	همدان (آقای سوری)	۰۹۱۹۵۷۳۰۱۵۵
زنجان (خانم دکتر هوشیار)	۰۹۱۹۲۷۰۵۸۷۱	یاسوج (آقای مرتضوی)	۰۹۱۹۵۹۰۷۲۰۴
ساری (آقای دکتر اکبری)	۰۹۱۹۷۷۸۱۹۴۴	یزد (خانم آزاد)	۰۹۱۹۹۱۰۱۲۴۳