آمار و احتمال مهندسی

نیم‌سال اول ۱۴۰۳ ـ ۱۴۰۲

دانشکده‌ی مهندسی کامپیوتر

مدرس: دکتر نجفی

تمرین سری ششم

موعد تحویل: ۱۶ دی

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص‌شده است.
- هم‌کاری و هم‌فکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت هم‌فکری و یا استفاده از هر منابع خارج درسی، نام هم‌فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.

مسئله‌ی ۱. (۱۴ نمره)

مادربزرگ مهدی که چایخور قهاری است، ادعا می‌کند که می‌تواند تشخیص دهد که ابتدا چای در فنجان ریخته شده یا آب جوش! اما مهدی که به تازگی در کلاس آمار و احتمال مهندسی با آزمون فرض آشنا شده تصمیم می‌گیرد که ادعای مادربزرگ را بررسی کند. به همین منظور او ۸ فنجان چای آماده می‌کند و در ۴تای آن‌ها ابتدا چای و سپس آب جوش، و در ۴تای دیگر ابتدا آب جوش و سپس چای را اضافه می‌کند. حال او ۸ فنجان به طور تصادفی روبروی مادربزرگش قرار می‌دهد تا او ۴ فنجانی که ابتدا در آن‌ها چای ریخته شده است را تشخیص دهد.

آ) برای آزمایش طراحی شده فرض صفر و جایگزین را بنویسید.

ب) آزمون فیشر را اجرا کرده و مقادیر p-value را بنویسید.

مسئله‌ی ۲. (۱۵ نمره)

اداره هواشناسی شهر تهران، ۴ دستگاه سنجش آلودگی هوا را در یک منطقه قرار داده است. فرض کنید شاخص آلودگی هوا در این منطقه ثابت است اما این دستگاه‌ها دقیق نیستند و شاخص آلودگی را با کمی نویز گزارش می‌دهند. در یک روز نسبتا آلوده، مقادیر گزارش شده توسط این ۴ دستگاه به شرح زیر است:

$$۱۵۳, ۱۴۸, ۱۵۲, ۱۵۳$$

آ) با کمک داده‌های جمع‌آوری شده، یک بازه اطمینان ۹۹ درصدی برای شاخص آلودگی هوا در این منطقه ارائه دهید.

ب) اگر شاخص آلودگی هوا بیش از ۱۵۰ باشد، هوا در شرایط ناسالم برای تمامی گروه‌ها قرار می‌گیرد و مدارس و دانشگاه‌ها تعطیل می‌شوند. برخی از صاحب نظران معتقد که از آنجایی که میانگین شاخص آلودگی هوا بیش از ۱۵۰ بوده، هوا ناسالم است و مدارس و دانشگاه‌ها می‌بایست تعطیل شوند. از طرفی گروهی دیگر از صاحب نظران بر این باورند که میانگین بدست آمده به طور محسوس و معناداری از ۱۵۰ بیشتر نبوده و مدارس و دانشگاه‌ها نباید تعطیل شوند. با توجه به اینکه p-value قابل قبول برای تعطیلی مدارس حداکثر ۰/۱ است، بررسی کنید که آیا مدارس تعطیل شوند یا خیر.

پ) برای کاهش خطای نوع اول و دوم باید مقدار $p-value$ قابل قبول را افزایش دهیم یا کاهش؟ چرا؟

مسئله‌ی ۳. (۱۴ نمره)

متغیر تصادفی پیوسته X با تابع چگالی احتمالاتی زیر را در نظر بگیرید:

$$f_X(x) = \begin{cases} ۳x^{۲} & x \in [۰, ۱] \\ ۰ & o.w. \end{cases}$$

اگر $Y|X = x \sim \text{Geometric}(x)$ باشد، تخمین MAP متغیر تصادفی X به شرط $Y = ۳$ را بدست آورید.

مسئله‌ی ۴. (۲۰ نمره)

مسئله‌ی رگرسیون خطی ساده را در نظر بگیرید که در آن ورودی‌های $\{(x_i, y_i)\}_{i=1}^{n}$ داده شده است. مقادیر $\{x_i\}_{i=1}^{n}$ یقینی و مشخص هستند. اما به ازای هر x_i مقدار y_i از طریق رابطه‌ی زیر به دست می‌آید:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

که $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. اگر بدانیم که β_0، β_1 و σ نامنفی هستند، به پرسش‌های زیر پاسخ هید.

آ) اثبات کنید که تخمین بیشینه درست‌نمایی دو پارامتر β_0 و β_1 معادل انتخاب مقادیری برای β_0 و β_1 است که میانگین مربعات خطا را کمینه می‌کند.

ب) اثبات کنید که تخمین‌های بدست آمده در بخش پیشین نااُریب بوده و از توزیع‌های زیر پیروی می‌کنند:

$$\hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2}\right), \quad \hat{\beta}_0 \sim \mathcal{N}\left(\beta_0, \frac{\sigma^2 \sum_i x_i^Y}{n \sum_i (x_i - \bar{x})^2}\right)$$

پ) بررسی کنید که آیا تخمینگر بیشینه درست‌نمایی عضوی از خانواده‌ی خطی تخمینگرهای زیر است یا نه؟ اگر هست رابطه‌ی γ_i را برحسب داده‌های ورودی بدست آورید.

$$\tilde{\beta_1} = \frac{\sum \gamma_i y_i}{\sum \gamma_i x_i} \quad \text{such that} \quad \sum_i \gamma_i = 0$$

ت) اثبات کنید هر تخمینگری که عضو خانواده فوق است نااُریب می‌باشد.

ث) اثبات کنید به ازای هر انتخابی از مقادیر γ_i در خانواده فوق داریم $Var(\hat{\beta_1}) \leqslant Var(\tilde{\beta_1})$. نتیجه بدست آمده را توضیح دهید.

مسئله‌ی ۵. (۱۵ نمره)

یک مدل رگرسیون خطی به شکل $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$ در شرایطی که خطای ان از توزیع نرمال $\epsilon \sim N(0, \sigma^2)$ پیروی میکند داریم.

MLE ضرایب $\beta_0, \beta_1, \beta_2$ را بیابید.

مسئله‌ی ۶. (۱۰ نمره)

فرض کنید میانگین معدل همه ی دانشجویان فارغ‌التحصیل‌شده از دانشگاه صنعتی شریف در سال ۲۰۰۵ برابر ۳/۰۵ (بر اساس معیار GPA) بوده است. سازمان ثبت احوال قصد دارد سوابق ۱۰۰ دانشجوی فارغ التحصیل در سال ۲۰۲۲ را بررسی کند تا ببیند میانگین GPA تغییر کرده است یا خیر. در این نمونه‌ی ۱۰۰ تایی، میانگین معدل برابر ۲/۱۵ بوده و انحراف معیار برابر ۱/۵ می‌باشد.

الف. فرصیه های صفر و جایگزین را برای این تحقیق بیان کنید.

ب. قبول/رد فرضیه H_0 را در مقابل H_1 جایگزین را در سطح اهمیت $\alpha = ۰/۰۵$ بررسی کنید. فرض کنید حجم نمونه به اندازه‌ی کافی بزرگ بوده و توزیع میانگین نمونه از توزیع نرمال پیروی می‌کند. (منظور از سطح اهمیت، Significance Level می‌باشد.)

مسئله‌ی ۷. (۱۲ نمره)

داده $\{(w[i], x[i])\}_{i=0}^{N-1}$ مشاهده شده است. می‌دانیم $x[i] = A + w[i]$ که A یک پارامتر نامعلوم با توزیع چگالی احتمال پیشین زیر است.

$$p(A) = \begin{cases} \lambda \exp(-\lambda A) & A \geqslant 0 \\ 0 & A < 0 \end{cases}$$

که $\lambda > 0$ و $W[i]$ها نمونه‌هایی از توزیع نرمال با میانگین ۰ و واریانس σ^2 می‌باشند. همچنین مقادیر $w[i]$ از A مستقل هستند. برآوردگر MAP را روی A بیابید.

موفق باشید :)