



آمار و احتمال مهندسی

بهار ۱۴۰۴

استاد: علی شریفی زارچی

دانشگاه صنعتی شریف

گردآورندگان: ارشیا دادرش، ارشیا یوسف نیا، امیدرضا معصومی، امیراردلان دهقان پور، عسل مسکین

دانشکده مهندسی کامپیوتر

مهلت ارسال: ۲۳ خرداد

آزمون‌های آماری

تمرین ششم

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- در طول ترم امکان ارسال با تاخیر پاسخ همه‌ی تمارین تا سقف ۴ روز و در مجموع ۱۰ روز، وجود دارد. پس از گذشت این مدت، پاسخ‌های ارسال شده پذیرفته نخواهند بود. همچنین، به ازای هر روز تأخیر غیر مجاز ۱۰ درصد از نمره تمرین به صورت ساعتی کسر خواهد شد.
- همکاری و هم‌فکری شما در انجام تمرین مانعی ندارد اما پاسخ‌های ارسال شده 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.

سوالات نظری (۱۰۰ نمره)

۱. (۱۵ نمره) فرض کنید میانگین IQ همه افراد جامعه ۱۰۰ است. قصد داریم IQ ۲۵ دانشجوی دانشگاه صنعتی شریف را بررسی کرده تا ببینیم آیا میانگین IQ دانشجویان دانشگاه صنعتی شریف تفاوت معناداری با IQ جامعه دارد یا خیر. در این نمونه ۲۵ تایی میانگین IQ برابر ۱۰۷ و انحراف معیار ۱۵ گزارش شده است.

(آ) فرض صفر و جایگزین را برای این تحقیق بیان کنید.

(ب) اگر p-value برای تأیید فرض جایگزین حداکثر ۰/۰۱ باشد، آیا می‌توان نتیجه گرفت IQ دانشجویان شریف تفاوت معناداری با جامعه دارد؟

(ج) اگر بخواهیم بررسی کنیم که آیا IQ دانشجویان شریف بیشتر هست یا نه، به قسمت (ب) پاسخ دهید.

۲. (۱۵ نمره) دانشجویی که تصور می‌کند در درس آمار و احتمال تسلط کافی به مطالب دارد تصمیم می‌گیرد تا ادعای یک استاد را مورد بررسی قرار دهد. آن استاد ادعا کرده است که پس از انجام تغییرات مهمی در ساختار درس، توانسته است وضع تحصیلی دانشجویان خود را بهتر کند. شما می‌خواهید با اطلاعاتی که بدست آورده‌اید، ببینید آیا این مورد درست است یا نه. شما اطلاعاتی در مورد ۸۱ نفر دانشجویان استاد قبل از اعمال تغییرات دارید که میانگین نمره آن‌ها در درس ۶۴ بوده است. همچنین می‌دانید ۳۶ یک کران بالا برای واریانس داده‌ها است و این داده‌ها از یک توزیع گاوسی آمده‌اند. علاوه بر این داده‌ها، اطلاعاتی در مورد ۱۶ نفر از دانشجویان استاد پس از اعمال تغییرات دارید که در آن میانگین نمره‌ها ۶۸ شده و می‌دانید ۶۴ یک کران بالا برای واریانس داده‌ها است (این داده‌ها نیز از توزیع گاوسی آمده‌اند). حال هدف شما این است تا متوجه شوید که آیا این تغییر در میانگین نمرات معنادار است یا به صورت تصادفی رخ داده است. فرض کنید تمامی داده‌ها به صورت جمع‌آوری شده‌اند اما مستقل از هم هستند.

(آ) فرض‌های H_0 و H_1 را برای این مورد مشخص کنید و یک آماره برای توصیف این مسئله ارائه دهید. دقت کنید که این آماره باید نسبت به درستی H_0 ، میانگین و واریانس مشخص داشته باشد. آن‌ها را مشخص کنید.

(ب) با فرض اینکه حداکثر p-value قابل قبول برابر با ۰/۰۱ است، صحت ادعای استاد را بررسی کنید.

(ج) فرض کنید مقداری که برای p-value بدست آورده‌اید، به اندازه‌ای که می‌خواهید کم نیست و می‌خواهید این مقدار را بهبود ببخشید. اگر می‌توانستید اندازه‌ی نمونه یکی از گروه‌ها را تغییر دهید، کدام مورد را انتخاب می‌کردید؟ (با فرض ثابت ماندن تعداد نمونه‌ی دیگر گروه.)

۳. (۱۰ نمره) تیم سیگنال شریف در مقابل یک تیم خاص از یک توزیع نرمال پیروی می‌کند، ولی ما اطلاعی در مورد میانگین (θ) و واریانس آن (σ^2) نداریم. فرض کنید طول ۱۰ مسابقه اخیر این دو تیم (اعدادها از تیم دانشگاه شریف به صورت زیر است) (داده‌ها از هم مستقل هستند):

۵۹, ۶۲, ۵۹, ۷۴, ۷۰, ۶۱, ۶۲, ۶۶, ۶۲, ۷۵

(آ) یک بازه با اطمینان ۹۵ درصدی برای θ پیدا کنید.

(ب) حال فرض کنید $\sigma^2 = ۲۵$. با اطمینان ۹۵ درصدی درباره θ پیدا کنید. جواب این قسمت و قسمت قبل را با هم مقایسه کنید. به نظر شما دلیل تفاوت بین بازه‌ها چیست؟

۴. (۱۰ نمره) یک سکه ۶۴ بار پرتاب می‌شود و ۲۲ بار شیر می‌آید. فرضیه‌ی منصفانه بودن سکه را با سطح معناداری برابر با ۰/۰۵ آزمون کنید. سپس یک سکه را ۱۶ بار پرتاب می‌کنیم و k بار شیر ظاهر می‌شود. اگر k به گونه‌ای باشد که $k_1 \leq k \leq k_2$ ، فرض می‌کنیم سکه منصفانه است (با سطح معناداری $\alpha = ۰/۰۵$). k_1 و k_2 را بیابید و مقدار خطای β را نیز محاسبه کنید.

۵. (۱۰ نمره) فرض کنید ۴۹ داده داریم با میانگین نمونه ۶/۲۵ و واریانس نمونه ۳۶. می‌خواهیم فرضیه‌های زیر را آزمون کنیم:

(آ) H_0 : داده‌ها از $N(۴, \sigma^2)$ آمده‌اند که در آن σ نامشخص است.

H_A : داده‌ها از $N(\mu, \sigma^2)$ آمده‌اند که در آن $\mu \neq ۴$.

آزمون را در سطح معناداری $\alpha = ۰/۰۵$ انجام دهید. از t-table برای پیدا کردن مقدار p استفاده کنید.

(ب) یک نمودار رسم کنید که تابع چگالی null pdf، ناحیه رد و ناحیه‌ای که برای محاسبه p-value در قسمت (a) استفاده می‌شود را نشان دهد.

۶. (۱۰ نمره) طبق اطلاعات شرکت «شیرین عسل»، هر بسته اسمارتیز باید شامل ۲۰٪ دانه‌ی آبی، ۲۰٪ دانه‌ی قهوه‌ای، ۲۰٪ دانه‌ی سبز، ۱۵٪ دانه‌ی نارنجی، ۱۵٪ دانه‌ی قرمز و ۱۰٪ دانه‌ی زرد باشد. آرزو تصمیم می‌گیرد این ادعا را بررسی کند. او ۲۰ بسته اسمارتیز تهیه می‌کند. هر بسته شامل ۵۰ عدد دانه است و در مجموع آرزو ۱۰۰۰ دانه اسمارتیز دارد. او تعداد هر رنگ را شمرده و نتایج زیر را به دست می‌آورد:

تعداد مشاهده شده	آبی	قهوه‌ای	سبز	نارنجی	قرمز	زرد	مجموع
۱۸۰	۱۹۰	۱۸۵	۱۶۰	۱۶۵	۱۲۰	۱۰۰۰	

در سطح معناداری ۰/۰۵، فرضیه‌ای را آزمون کنید که آیا نسبت رنگ‌های اعلام شده توسط شرکت شیرین عسل با رنگ‌های مشاهده شده تفاوت دارد یا نه. ابتدا به دقت مشخص کنید فرضیه‌ی صفر و فرضیه‌ی مقابل چیست و سپس نتیجه‌گیری کنید که با توجه به داده‌ها چه تصمیمی می‌گیرید. (عبارت عددی کامل مربوط به آماره آزمون را بنویسید. برای استفاده از جدول‌ها لازم نیست مقدار نهایی را دقیق تا اعشار محاسبه کنید، فقط عبارت عددی را ارائه دهید.)

۷. (۱۵ نمره) فرض کنید X_1, \dots, X_n نمونه تصادفی از توزیع $n(\mu_X, \sigma_X^2)$ و Y_1, \dots, Y_m نمونه تصادفی مستقل از توزیع $n(\mu_Y, \sigma_Y^2)$ باشند. هدف ما آزمون فرضیه‌های زیر است:

$$H_0: \mu_X = \mu_Y \quad \text{در مقابل} \quad H_1: \mu_X \neq \mu_Y$$

با فرض اینکه $\sigma_X^2 = \sigma_Y^2 = \sigma^2$.

(آ) LRT را برای این فرضیه‌ها استخراج کنید. نشان دهید که این آزمون بر اساس آماره زیر قابل انجام است:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{S_p^2 \left(\frac{1}{n} + \frac{1}{m} \right)}}$$

که

$$S_p^2 = \frac{1}{n+m-2} \left(\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2 \right)$$

(مقدار S_p^2 گاهی به عنوان برآورد واریانس تجمعی شناخته می‌شود.)

(ب) نشان دهید که تحت H_0 ، $T \sim t_{n+m-2}$. (این آزمون با نام Two-sample t test نیز شناخته می‌شود.)

(ج) نمونه‌هایی از چوب از بخش مرکزی و حاشیه‌ای یک کلیسای بیزانسی به دست آمده و سال چوب اندازه‌گیری شده است که داده‌ها به صورت زیر است:

مرکزی		حاشیه	
۱۲۵۱	۱۲۹۴	۱۲۷۴	۱۲۸۴
۱۲۴۸	۱۲۷۹	۱۲۶۴	۱۲۷۲
۱۲۴۰	۱۲۷۴	۱۲۵۶	۱۲۵۶
۱۲۳۲	۱۲۶۴	۱۲۵۰	۱۲۵۴
۱۲۲۰	۱۲۶۳		۱۲۴۲
۱۲۱۸	۱۲۵۴		
۱۲۱۰	۱۲۵۱		

با استفاده از آزمون t دو نمونه‌ای بررسی کنید که آیا میانگین سن بخش مرکزی و بخش حاشیه‌ای برابر است یا خیر.

۸. (۱۵ نمره) میانگین ضربان قلب در حالت استراحت به طور معمول ۷۱ ضربه در دقیقه است و انحراف معیار آن ۴ ضربه در دقیقه می‌باشد. گروهی از پژوهشگران معتقدند که ضربان قلب مردان زمانی که در انتظار ورود به مصاحبه شغلی هستند، افزایش می‌یابد. برای بررسی این فرضیه، گروهی از ۹ مرد که برای مصاحبه شغلی آمده‌اند، به دستگاه اندازه‌گیری ضربان قلب بی‌سیم مجهز شدند تا در یک ساعت قبل از مصاحبه ضربان قلب آن‌ها ثبت شود. میانگین ضربان قلب هر یک از آن‌ها در جدول زیر آورده شده است:

شرکت‌کننده	ضربان قلب (تعداد در دقیقه)
۱	۸۰
۲	۷۴
۳	۷۳
۴	۷۲
۵	۷۸
۶	۷۵
۷	۷۰
۸	۷۴
۹	۶۹

^۱ Likelihood-ratio test

- (آ) برای بررسی اینکه آیا شواهد معناداری وجود دارد که نشان دهد ضربان قلب مردان هنگام انتظار برای مصاحبه شغلی افزایش می‌یابد، باید از آزمون z استفاده شود یا آزمون t ؟
- (ب) آزمون را با سطح معناداری ۵٪ انجام دهید و نتیجه را تفسیر کنید.
- (ج) بازه اطمینان ۹۰٪ برای میانگین جامعه محاسبه کنید.