



سلام بر تمام دانشجویان عزیز، چند نکته مهم:

۱. حجم گزارش به هیچ عنوان معیار نمره‌دهی نیست، در حد نیاز توضیح دهید.
۲. نکته‌ی مهم در گزارش نویسی روشن‌بودن پاسخ‌هاست، اگر فرضی برای حل سؤال استفاده می‌کنید حتماً آن را ذکر کنید، اگر جواب نهایی عددی است به صورت واضح آن را بیان کنید.
۳. تحویل تنها کد یا نمودار بدون شرح روند فکری و تحلیل‌های مربوطه موجب کسر نیمی از نمره تمرین مربوطه خواهد شد.
۴. مهلت تحویل تمرین‌ها تمدید نخواهد شد.
۵. هرگونه پرسش در ارتباط با این تمرین را با raminnakhli@gmail.com در میان بگذارید.

۱. با رسم QQ-Plot در مورد درستی قضیه زیر بحث کنید. جهت تولید تابع توزیع تجمعی از یک تابع دلخواه، مثلاً گوسی، استفاده کنید.

$$G(x) = F\left(\frac{x - \mu}{\sigma}\right)$$

if $\sigma \neq 0$ and μ constant,

$$\text{then } y_p = \mu + \sigma x_p$$

$$x_p = F^{-1}(p), y_p = G^{-1}(p)$$

۲. فرض کنید برای سنجش نکویی برازش به کمک QQ plot، آماره زیر (فاصله از خط $X=Y$) را برگزیده‌ایم:

$$D_n = \max_{i=1:n} \left| F^{-1}\left(\frac{i}{n}\right) - x_i \right|$$

$$x_1 < x_2 < x_3 < \dots < x_n$$

به کمک شبیه‌سازی بحث کنید که مشکل این آماره چیست و بیان کنید چگونه آماره Kolmogorov – Smirnov این مشکل را حل می‌کند.

۳. برای فرض $H_0: \mu = 0$ و $H_1: \mu_1 \neq \mu_2$ رابطه‌ی $power$ با تعداد (n) و فاصله دو توزیع (Δ) را نشان دهید. (فرض کنید H_0 و H_1 از یک توزیع نرمال با انحراف معیار ۱۰ تبعیت می‌کنند).
 - هنگام کشیدن رابطه با n ، Δ را ثابت بگیرید.
 - هنگام کشیدن رابطه با Δ ، n را ثابت بگیرید.

نموداری بکشید که یک محور آن Δ و محور دیگر آن n باشد و $power$ های مختلف را با یک بازه رنگی روی آن نشان دهید. ($color plot$)

۴. همگرایی $Kolmogorov - Smirnov$ را به کمک شبیه‌سازی نشان دهید. نقش توان را در D_n با عوض کردن مقدار توان، به کمک شبیه‌سازی در آماره $Kolmogorov - Smirnov$ بررسی کنید.

۵. جدول زیر نشان‌دهنده تعداد فرزندان پسر در خانواده‌هایی است که دارای ۱۲ فرزند می‌باشند.

تعداد فرزندان پسر	تعداد خانواده‌ها
۰	۷
۱	۴۵
۲	۱۸۱
۳	۴۷۸
۴	۸۲۹
۵	۱۱۱۲
۶	۱۳۴۳
۷	۱۰۳۳
۸	۶۷۰
۹	۲۸۶
۱۰	۱۰۴
۱۱	۲۴
۱۲	۳

با این فرض که X_i متغیر تصادفی مربوط به تعداد فرزندان پسر در این ۶۱۱۵ خانواده باشد،

الف) آماره‌ای معرفی کنید که به وسیله آن‌ها بتوان فرض زیر را تست کرد:

$$H_0: X_1, X_2, \dots, X_{6115} \sim \text{Binomial}(12, 0.5)$$

ب) با استفاده از شبیه‌سازی، توزیع آماره معرفی‌شده را به دست آورید و هیستوگرام آن را رسم کنید.

راهنمایی: داده‌های مورد نیاز را با استفاده از توزیع *Binomial* ذکر شده به دست آورید.

ج) بر اساس توزیع به دست آمده، برای آماره محاسبه شده در قسمت الف، $p - value$ را به دست آورید. آیا با $H_0, \alpha = 0.05$ رد می‌شود؟

۶. فرض کنید داریم:

$$X_1, X_2, \dots, X_n \sim N(0, \sigma_x^2)$$

$$Y_1, Y_2, \dots, Y_n \sim N(0, \sigma_y^2)$$

الف) برای تست فرضیه $H_0: \sigma_x = \sigma_y$ چه روشی را پیشنهاد می‌کنید؟

ب) با استفاده از نمونه برداری از دو توزیع نرمال با واریانس‌های ۳ و ۱۰ روش فوق را بررسی کنید.

۷. داده *StudentPerformance.csv* را در نظر می‌گیریم. این داده شامل نمرات دانش‌آموزان در دو دوره مختلف است. اما معلوم نیست که کدام نمره متعلق به کدامین دانش‌آموز است.

الف) با استفاده از یک تست پارامتریک و یک تست بدون پارامتر مشخص کنید که آیا میانگین دو داده با یکدیگر متفاوت است؟

ب) چقدر احتمال دارد که جمعیت A میانگین بیشتری نسبت به جمعیت B داشته باشند؟

در مرحله بعد، فرض می‌کنیم به ازای هر دانش‌آموز، نمره در درس A و B ثبت شده است. مراحل الف و ب را دوباره انجام دهید.

چرا داده مرحله دوم بهتر از مرحله اول است؟

۸. برای این سوال از دیتاست *Galton* قرار گرفته در کنار تمرین استفاده کنید. این دیتاست شامل قد دانش‌آموزان و پدر آنها می‌باشد که در این سوال از قد فرزندان استفاده می‌کنیم. (در این سوال فرض کنید انحراف معیار جامعه نامشخص است)

الف) با استفاده از نمونه برداری تعداد ۲۰۰۰۰ نمونه با اندازه ۶۰ تولید کنید. سپس بازه اطمینان ۹۷ درصد را برای هر یک از این نمونه‌ها را حساب کنید. حال بررسی کنید که چند درصد بازه‌ها میانگین واقعی جامعه را در بر می‌گیرند.

ب) آزمایش فوق را برای ۱۰۰۰۰ نمونه با اندازه ۱۰ تکرار کنید. این بار بازه اطمینان ۹۰ درصد را محاسبه کنید. از انجام این دو آزمایش چه نتیجه ای میگیرید.

۹. با استفاده از نمونه برداری از یک توزیع بتا با پارامترهای $a = 2$ و $b = 5$ یک نمونه با اندازه ۵۰ تولید کنید.

الف) میانه نمونه فوق را M در نظر بگیرید. با تست غیر پارامتریک فرض زیر را بررسی کنید.

$$H_0: M = 0.4$$

$$H_1: M > 0.4$$

ب) $power$ تست فوق را محاسبه کنید.

ج) میزان ۰.۴ را در قسمت اول را به ۰.۶ افزایش دهید و دو قسمت را تکرار کنید.

د) جواب به دست آمده در بخش ج را با الف مقایسه کنید و تغییر $Power$ را توجیه کنید.

۱۰. با استفاده از دیتاست $Galton$ و اطلاعات قد پدران به این سوال پاسخ دهید.

الف) از داده ذکر شده یک نمونه با اندازه ۷۰ داده جدا کنید. سپس فرض زیر را بررسی کرده و $power$ را محاسبه کنید.

$$H_0: \mu = 60$$

$$H_1: \mu \neq 60$$

ب) حال یک نمونه با اندازه ۱۰ داده جدا کنید و تست فوق را تکرار کنید و نتایج را گزارش کنید.

ج) از مقایسه نتایج فوق چه نتیجه ای میگیرید؟