به نام خداوند جان و خرد

دانشگاه تهران

دانشکدگان فنی

دانشکده مهندسی برق و کامپیوتر

استنباط آماری

تمرین شماره ۳

نام و نام خانوادگی: علی خرم فر

شماره دانشجویی: ۸۱۰۱۰۲۱۲۹

علت تاخیر ۱ روزه در ارسال: محاسبات اشتباه گریس از روی سامانه LMS و مشاهده عدد ۱۷ و عدم محاسبه ساعات اضافه - حل تمامی سوالات

بهمن‌ماه ۱۴۰۲

فهرست مطالب

فهرست اشکال

۱_ پاسخ مسئله شماره ۱

در این مسئله پایه حقوق توسعه‌دهندگان نرم‌افزار بر حسب هزار دلار به شرح زیر ارائه شده است:

43, 47, 52, 68, 72, 55, 61, 44, 58, 63, 54, 59, 77, 36, 80, 53, 60

با توجه به شرایط مسئله درخواست شده که بررسی کنیم آیا تفاوت معناداری بین اینکه میانه حقوق برابر ۵۰هزار دلار باشد وجود دارد یا خیر. به این منظور مشخص است که تست از نوع NonParametric است. زیرا در این تست میانه مورد بحث است و فرض صفر بر اساس آن تنظیم شده است.

۱-۱_ تست Wilcoxon signed rank test

در این تست ابتدا اختلاف داده‌ها را از میانه به دست می‌آوریم سپس مقدار این اختلاف را مرتب (Rank) می‌کنیم. بدیهی است اختلافات فارغ از علامت بوده و بدون توجه به علامت مرتب‌سازی انجام می‌شود. سپس مقدار Rank برای مقادیر مثبت با یکدیگر جمع شده که حاصل آن w+ خواهد بود و بر اساس این test statistics و مقدار significance level تصمیم‌گیری درباره رد و یا عدم رد فرض صفر صورت می‌پذیرد.

فرض صفر H0 : میانه حقوق برابر ۵۰ هزار دلار است

فرض مقابل H1 : میانه حقوق بزرگتر از ۵۰ هزار دلار است.

مقدار Significant level : ۵٪

H0 : M= 50

H1 : M>50

Significant level: 5%

در این تست باتوجه به اینکه تعداد نمونه n کم است از تقریب نرمال استفاده نمی‌کنیم و برای بررسی فرض صفر ابتدا اختلاف تمامی اعداد ورودی را با میانه مورد نظر یعنی ۵۰ بدست می‌آوریم.

اختلافات با میانه برابر ۵۰:

d=-7, -3, 2, 18, 22, 5, 11, -6, 8, 13, 4, 9, 27, -14, 30, 3, 10

اختلافات با میانه برابر ۵۰ بدون توجه به علامت:

d= 7, 2, 1, 14, 15, 5, 11, 6, 8, 12, 4, 9, 16, 13, 17, 3, 10

محاسبه W^+ یعنی جمع رنک‌های آن‌هایی که مقدار اختلاف آن‌ها با میانه مثبت است که پس از مرتب‌سازی لیست بالا بدست می‌آید:

W= 1+14+15+5+11+8+12+4+9+16+17+3+10= 125

حال می‌توان تقریب نرمال زد و مقدار z آماره مورد نظر را محاسبه کرد:

$$\mu_t = \frac{n(n+1)}{4} = \frac{17 * 18}{4} = 76.5$$

$$Var = \frac{n(n+1)(2n+1)}{24} = \frac{17 * 18 * 35}{24} = 446.25$$

$$\sigma_t = \sqrt{446.25} = 21.12$$

$$z = \frac{W_x - E(W_x)}{\sqrt{Var_x}} = \frac{125 - 76.5}{21.12} = 2.296 \rightarrow pvalue = (p > 2.296) = 1 - (p < 2.296) = 0.01083$$

$$\alpha = 0.05$$

باتوجه به اینکه مقدار $pvalue$ از $signifitance\ level$ کمتر است پس فرض صفر رد می‌شود و نتیجه گرفته می‌شود که میانه حقوق بیشتر از ۵۰ هزار دلار است.

۲-۱_ پاسخ قسمت ۲

تست Mann-Whitney برای حالتی است که میانه ۲ گروه باهم مقایسه می‌شود. در صورتی که این مسئله درباره ۱ گروه است. می‌توان یک گروه جدید ساخت به صورتی که بر اساس فرض صفر این مسئله باشد. یعنی گروهی که میانه آن برابر ۵۰ باشد. در این صورت می‌توان گروه دوم را همان نمونه مسئله در نظر گرفت و مسئله را به این شیوه حل کرد.

دو گروه x و y در این مسئله در نظر گرفته می‌شود.

گروه x ورودی مسئله:

43, 47, 52, 68, 72, 55, 61, 44, 58, 63, 54, 59, 77, 36, 80, 53, 60

گروه y ساخته شده با میانه ۵۰ به عنوان فرض صفر:

50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50

از دو روش می‌توان U را محاسبه که یکی از آن‌ها انتخاب می‌شود:

W1 و W2 رنک poolشده هستند. یعنی پس از ترکیب دو گروه و n=m

$$U_1 = W_1 - \frac{n(n+1)}{2}$$

$$U_2 = W_1 - \frac{n(n+1)}{2}$$

43,47,52,68,72,55,61,44,58,63,54,59,77,36,80,53,60,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50,50

36,43,44,47,50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50,50, 52,53,54,55,58,59,60,61,63,68,72,77,80

$$W_1 = 1 + 2 + 3 + 4 + 22 + 23 + 24 + 25 + 26 + 27 + 28 + 29 + 30 + 31 + 32 + 33 + 34 = 374$$

$$W_2 = 5 + 6 + 7 + 8 + 9 + 10 + 11 + 12 + 13 + 14 + 15 + 16 + 17 + 18 + 19 + 20 + 21 = 221$$

$$U_1 = W_1 - \frac{n(m+1)}{2} = 374 - \frac{17 * 18}{2} = 374 - 153 = 221$$

$$U_2 = W_1 - \frac{m(m+1)}{2} = 221 - \frac{17 * 18}{2} = 221 - 153 = 68$$

U=MIN(U1,U2)

$$\mu_{U_2} = \frac{n * m}{2} = \frac{17 * 17}{2} = 144.5$$

$$\sigma_{U_2} = \sqrt{\frac{nm(n+m+1)}{12}} = \sqrt{\frac{289(35)}{12}} = 29.03$$

مقدار $pvalue$ برابر با

$$z = \frac{U_2 - \mu_{U_2}}{\sigma_{U_2}} = \frac{68 - 144.5}{29.03} = -2.635 \rightarrow pvalue = 0.0042$$

پس فرض صفر رد می‌شود و نتیجه مجددا از این روش حاصل می‌شود.

*** پس از ددیلان متوجه شدم که باید برای مشاهده‌های یکسان تقسیم شود رنک. روش صحیح است ولی محاسبات ایراد دارد.

۲_پاسخ مسئله شماره ۲

در این مسئله میزان قدرت هر گوش چپ و راست سنجیده می‌شود. آزمون فرض تشکیل می‌شود به نحوی که فرض صفر بیان می‌کند که تفاوتی بین گوش چپ و راست وجود ندارد. داده‌ها دو گروه هستند ولی هر گوش چپ و راست برای هر نفر از یکدیگر مستقل نبوده و یک جفت هستند.

۱-۲_ پاسخ قسمت ۱

در نگاه اول به نظر می‌رسد که با محاسبه di ها بین گوش چپ و راست اختلاف وجود دارد. اما باید توجه داشت که برای اثبات وجود تفاوت باید از تست‌های آماری استفاده کرده و استنباط آماری انجام شود. زیرا که بعضی از خروجی‌های این نمونه‌برداری مانند شخص ۱۲ مشخص است که گوش چپ او دارای مشکل است و این موضوع باعث مشکل در نتایج می‌شود. و یا فرد ۳ که هر دو گوش او مشکل دارد اما هدف مشخص شدن اختلاف است. پس به طور کلی برای بیشتر افراد این تفاوت وجود دارد ولی اینکه دلیل تفاوت چیست در اینجا با توجه به نوع آزمایش مشخص نمی‌شود. فقط باکمک تست‌های آماری می‌توان نتیجه گرفت که آیا تفاوت معناداری بین گوش چپ و راست باتوجه به این نمونه‌برداری وجود دارد یا خیر.

۲-۲_ پاسخ قسمت ۲

ابتدا داده‌های هردو گوش را ترکیب کرده سپس در جدول رنک می‌شوند:

Left Ear		Right Ear	
xi	rank	yi	rank
25	7	32	19.2
29	11.5	30	14.3
10	3	7	2
31	16	36	24
27	9.5	20	4.5
24	6	32	19.2
27	9.5	26	8
29	11.5	33	23
30	14.3	32	19.2
32	19.2	32	19.2
20	4.5	30	14.3
5	1	32	19.2

$$W_1 = 11.5 + 3 + 16 + 9.5 + 6 + 9.5 + 11.5 + 14.3 + 19.2 + 4.5 + 1 = 106$$

$$W_2 = 19.2 + 14.3 + 2 + 24 + 4.5 + 19.2 + 8 + 23 + 19.2 + 19.2 + 14.3 + 19.2 = 186$$

$$U_1 = W_1 - \frac{n(m+1)}{2} = 106 - \frac{12 * 13}{2} = 106 - 78 = 28$$

$$U_2 = W_2 - \frac{m(m+1)}{2} = 186 - \frac{12 * 13}{2} = 186 - 78 = 108$$

U=MIN(U1,U2)

$$\mu_{U_1} = \frac{n * m}{2} = \frac{12 * 12}{2} = 72$$

$$\sigma_{U_1} = \sqrt{\frac{nm(n+m+1)}{12}} = \sqrt{\frac{12 * 12(25)}{12}} = 17.32$$

مقدار $pvalue$ برابر با

$$z = \frac{U_2 - \mu_{U_2}}{\sigma_{U_2}} = \frac{28-72}{17.32} = -2.54 \rightarrow pvalue = 0.0055$$

فرض صفر رد می‌شود. پس با هم متفاوت هستند.

یک راه دیگر هم استفاده از جدول بجای تقریب نرمال است که در آن صورت هم باز فرض صفر رد می‌شود.

۳_ پاسخ مسئله شماره ۳

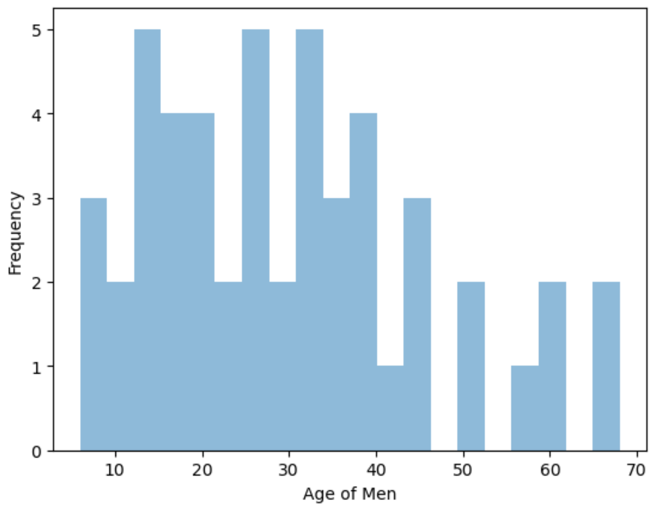
در این سوال سن زن و مرد مقایسه شده‌است. هدف آزمون فرض بررسی این است که آیا طرفداران مرد از طرفداران زن پیرتر هستند یا خیر. پس هدف مقایسه توزیع سن زن و مرد است تا متوجه شویم تفاوتی میان آن‌ها وجود دارد یا خیر. ورودی مسئله به شرح زیر است:

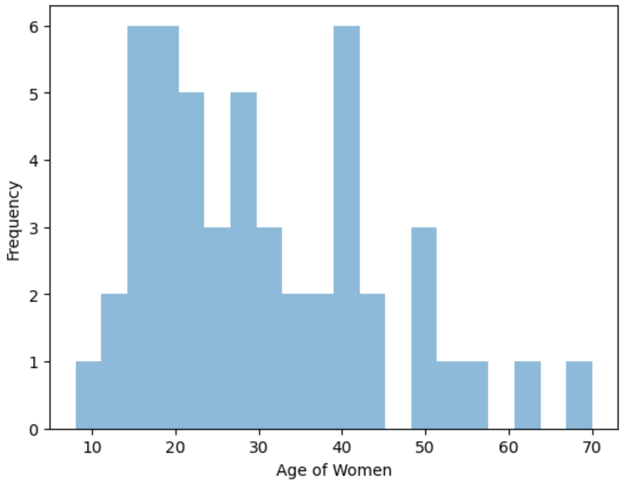
Men’s Age: 52, 18, 27, 12, 24, 17, 68, 25, 12, 9, 51, 44, 42, 34, 44, 15, 21, 66, 61, 32, 31, 20, 6, 13, 34, 38, 45, 17, 16, 15, 36, 21, 29, 21, 29, 9, 33, 15, 37, 27, 31, 15, 57, 37, 27, 31, 38, 27, 60, 23

Women’s Age: 36, 49, 20, 31, 51, 31, 15, 16, 39, 70, 52, 16, 39, 34, 18, 34, 30, 18, 26, 18, 25, 16, 39, 49, 22, 37, 39, 21, 16, 63, 45, 43, 17, 28, 29, 23, 42, 23, 28, 55, 41, 18, 23, 8, 13, 26, 13, 27, 28, 18

۱-۳_ پاسخ قسمت ۱

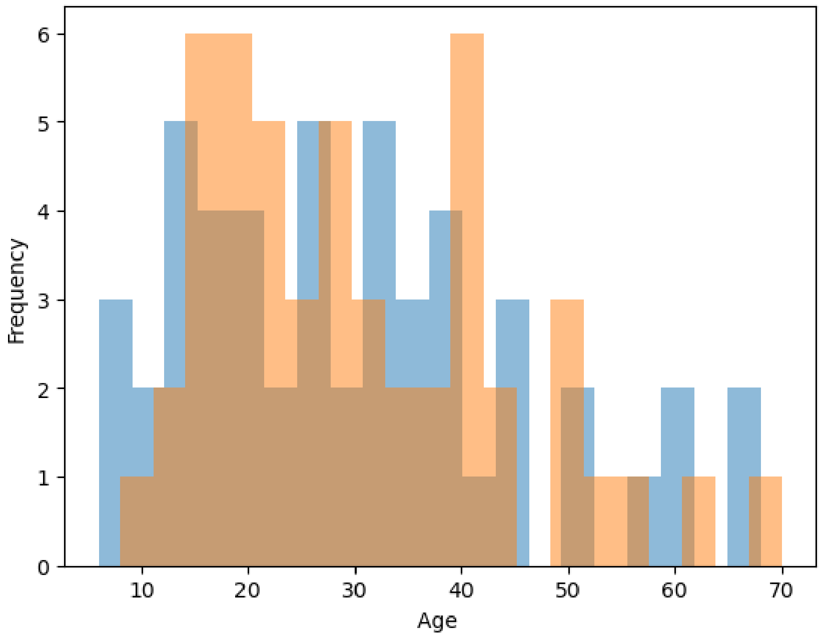
هیستوگرام سن زن و مرد به شکل زیر است(کد نوت‌بوک پیوست شد):





شکل ۱نمودار هیستوگرام سن زن و مرد

هیستوگرام برای مقایسه:



شکل ۲ هیستوگرام مقایسه توزیع سن زن و مرد

۲-۳_ پاسخ قسمت ۲

باتوجه به راهنمایی‌های انجام‌شده در گروه درسی تلگرام اگر از روش MLE برویم فرض صفر رد نمی‌شود ولی اثبات نمی‌شود که فرض صفر برقرار است یا خیر. از نمودارها مشخص است که توزیع نرمال نیست ولی باید از روش‌های آماری برای اثبات این موضوع استفاده شود.

باید از تست Shapiro-Wilk استفاده شود:

فرض صفر به این صورت است که داده‌ها نرمال هستند. برای بدست آوردن test statistics از تابع Shapiro استفاده شد که نتیجه زیر حاصل شد. کد مربوطه در نوت‌بوک پیوست شد:

```
Test Statistics of men= ShapiroResult(statistic=0.9460382461547852, pva
lue=0.023510854691267014)
Test Statistics of women= ShapiroResult(statistic=0.9431541562080383, p
value=0.01799035631120205)
```

باتوجه به اینکه مقدار pvalue کمتر از ۰.۰۵ است پس فرض صفر که نرمال بودن داده‌هاست رد شده و توزیع نرمال نیست.

۳-۳_ پاسخ قسمت ۳

خیر امکان استفاده از تست‌های parametric وجود ندارد. زیرا که برای این دو گروه از داده مستقل از هم، توزیع داده‌ها مشخص نیست. برای نمونه اگر بخواهیم از two sample t test استفاده کنیم باید توزیع داده‌ها نرمال بوده تا بتوان مقدار test statistics بر حسب نرمال استاندارد محاسبه شده و یا بر حسب توزیع t بیان شود. تا بتوان بر اساس آن برای فرض صفر نتیجه گرفت. پس اگرچه که شرط استقلال داده‌ها در این مسئله برقرار است ولی بخاطر اینکه توزیع مشخص نیست نمی‌توان از تست‌های parametric استفاده کرد. این مورد در قسمت قبلی سوال بررسی شد.

همچنین در این مسئله شرط اینکه هردو متغیر که بررسی می‌شود یکسان باشند برقرار است و از این نظر مشکلی نیست در هر دو گروه سن مورد بررسی قرار می‌گیرد.

۳-۴_ پاسخ قسمت ۴

در این قسمت از مسئله باید تبدیلی روی داده‌ها انجام شود، تا توزیع نرمال شود. مثلا تبدیل لگاریتمی. سپس مجددا تابع Shapiro فراخوانی شود تا نتیجه نرمال شدن یا نشدن داده‌ها بررسی شود. به این منظور تبدیل‌های مختلفی تست شد که در نهایت با کمک تبدیل لگاریتمی داده‌ها نرمال شدند:

```
Test Statistics of transformed men= ShapiroResult(statistic=0.977242112
159729, pvalue=0.4424073100090027)
Test Statistics oftransformed men= ShapiroResult(statistic=0.9834237694
740295, pvalue=0.7023096084594727)
```

باتوجه به مقدار pvalue جدید، پس فرض صفر رد نمی‌شود و نتیجه می‌شود داده‌ها نرمال هستند.

۳-۵_ پاسخ قسمت ۵

باتوجه به نتیجه قسمت قبلی ، شرایط استفاده از تست‌های parametric فقط موضوع نرمال‌بودن توزیع نقض شده ببود که در قسمت ۴ از سوال ۳ این موضوع حل شد. حال می‌توان برای این دو گروه از two sample t test استفاده کرد و نتیجه را بررسی کرد.

به این منظور از تابع ttest_ind استفاده شد که نتایج خروجی به شکل زیر است:

```
Test Statistics of two sample t test= -0.38345358326164847
P value= 0.7022140661639309
```

باتوجه به اینکه مقدار pvalue بیشتر از significance level برابر ۰.۰۵ بوده پس فرض صفر رد نمی‌شود. و نتیجه می‌شود که تفاوتی میان میانگین سن زن و مرد در این تست وجود ندارد.

۳-۶_ پاسخ قسمت ۶

برای حل این قسمت از مسئله از یک تست non parametric مانند mannwhitneyu بر روی داده‌ها استفاده می‌کنیم. باتوجه به مطلوب مسئله، این تست بر روی داده‌های اولیه قبل از تبدیل اجرا می‌شود. به این منظور تابع mannwhitneyu بر روی داده‌های اولیه اجرا شد و نتایج زیر حاصل شد:

```
Test Statistics of Mann Whitney test= 1213.0
P value= 0.801221571030353
```

در این تست مقدار pvalue بیشتر شد. یعنی با شدت بیشتری اطمینان حاصل می‌شود که میانگین زنان و مردان نزدیک به هم بوده و تفاوت قابل توجهی بین میانگین آن‌ها وجود ندارد. پس در این تست نیز فرض صفر رد نمی‌شود.

۳-۷_ پاسخ قسمت ۷

به طور کلی تست‌های پارامتریک مثل t test قوی‌تر هستند ولی باید توجه داشت که در این مسئله داده‌های اولیه توزیع نامشخص داشته و t test پس از تبدیل روی داده‌ها اعمال شد. پس باتوجه به اینکه به صورت پیشفرض توزیع داده‌ها مشخص نیست پس بهتر است از این نظر از تست‌های nonparametric استفاده شود. از نظر تعداد داده‌ها نیز اگرچه مقدار متوسطی دارند ولی به اندازه خوبی زیاد نیستند که از با اطمینان بالا از قضیه حد مرکزی بهره گرفته شود. پس باتوجه به موارد ذکرشده تست‌های nonparametric که وابسته به توزیع نیستند احتمالا در این مسئله بهتر هستند.

۴_پاسخ مسئله شماره ۴

از contingency table برای بررسی استقلال دو ویژگی استفاده می‌شود. در این مسئله جدول زیر داده‌شده است:

0.15	0.09	0.06
0.15	0.09	0.06
0.20	0.12	0.08

۴-۱_ پاسخ قسمت ۱

برای اثبات مستقل بودن آزمون فرض زیر برقرار می‌شود:

$H0 : p_{ij} = p_i + p_j$

$H1 : The\ hypothesis\ H0\ is\ not\ true.$

پس باید برای تمام r*c این مقادیر چک شوند. کد مربوطه برای چک کردن پیوست شد. مشاهده شد که مقادیر جمع شده در هر خانه برابر جمع احتمال سطر و ستون $p_{ij} = p_i + p_j$ هستند. پس فرض صفر رد نشده و می‌توان نتیجه گرفت که استقلال دارند.

به این منظور ابتدا مقدار هر ردیف محاسبه شد تا احتمال هر خانه محاسبه شود. سپس مقادیر هر ستون نیز محاسبه شد تا احتمالات marginal بدست آیند. برای مثال:

$$pi = \sum_{i=1}^{n} pij$$

سپس مقدار هر ستون محاسبه شده و با کمک تابع allclose اگر که مقادیر برابر با جدول اولیه باشند خروجی برابر با مستقل بودن داده‌هاست.

۴-۲_ پاسخ قسمت ۲

به این منظور ۳۰۰ نمونه تصادفی رندم ایجاد شد و باتوجه به توضیحات مسئله در جدول تقسیم شد. نتیجه حاصل به شرح زیر است:

جدول فراوانی خروجی:

39	32	25
40	20	18
69	36	21

کد مربوطه پیوست شد.

۴-۳_ پاسخ قسمت ۳

اگر فرض شود که مقدار pij نامعلوم است، و هدف این است که متوجه شدیم مدلی که داده‌ها را تولید کرده، مدل مورد نظر است یا خیر از goodness of fit استفاده می‌شود که بررسی می‌شود که آن چیزی که در جدول است آن چیزی است که انتظار می‌رود یا خیر:

$$X^2 = \sum^{M} \frac{(O_I - E_I)^2}{E_I}$$

در این نوع سوال آماره حاصل از این تست از توزیع خی دو با s-t-1 درجه آزادی است:

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{\left(f_{ij} - e_{ij}\right)^2}{e_{ij}}$$

پس مقدار *test statistics* محاسبه شده و خروجی به صورت زیر خواهدبود:

```
Statistics: 5.59448601440789
Pvalue : 0.23154808227461396
```

محاسبه درجه آزادی

$s = rc = 9$

$t = r + c - 2 = 4$

$df = s - t - 1 = 9 - 4 - 1 = 4$

هرچند که خروجی کد میزان pvalue هم گزارش کرده است ولی به صورت جداگانه هم برای این مقدار و درجه آزادی ۴ برابر 0.231555 محاسبه شد که باتوجه به این موضوع که بزرگ تر از آلفای برابر 0.05 است پس فرض صفر رد نمی‌شود و استقلال داده‌ها نتیجه گرفته می‌شود.

۴-۴_ پاسخ قسمت ۴

در این قسمت بررسی شده که فرض شده که تمامی دانشجویان این سوال را حل کرده و باتوجه به اینکه اعداد تصادفی متفاوتی ارائه شده پس مقدار statistics حاصل از توزیع خی دو احتمالا تفاوت خواهدداشت. حال باید آزمون فرض به نحوی باشد که بتوان نتیجه گرفت تمامی این این نمونه‌های تصادفی از توزیع مورد نظر هستند. این مورد در قسمت قبلی سوال نیز اشاره شد که از goodness of fit استفاده می‌شود.

پس فرض صفر به صورت زیر تعریف می‌شود:

$H0 : observed\ data\ is\ from\ X^2\ distribution\ df = 4$

$H1 : The\ hypothesis\ H0\ is\ not\ true$

به این منظور ما یک سری داده مشاهده کرده‌ایم که می‌دانیم این داده‌ها توزیعشان یک خی دو با درجه آزای ۴ است.

$$X^2 = \sum^{M} \frac{(O_I - E_I)^2}{E_I}$$

در این مسئله فراوانی داده‌های مشاهده شده وجود دارد. پس برای محاسبه O احتمال در فراوانی ضرب می‌شود:

$O_i = f_i * p$

برای محاسبه E باتوجه به اینکه توزیع خی دو است راه‌های مختلفی وجود دارد. در این مسئله فراوانی و میانگین برای محاسبه نیاز است. که میانگین برابر با درجه آزادی خواهدبود.

به این ترتیب X^2 محاسبه شده و پس از آن با بدست آوردن pvalue و مقایسه با significance مثلا برابر با 0.05 اگر که pvalue بیشتر باشد، می‌توان نتیجه گرفت که داده‌ها از همین توزیع هستند.

۵_پاسخ مسئله شماره ۵

در این مسئله ارتباط بین داشتن سیبیل در مردان و سن آن‌ها بررسی می‌شود. به این منظور اگر که ارتباط معناداری وجود دارد نتیجه گرفته‌شود. به این منظور نمونه ۱۰۰ تایی از مردان بالای ۱۸ سال سن به صورت تصادفی انتخاب شده است. در جدول زیر هر مرد در این نمونه‌برداری بر اساس اینکه سن بین ۱۸ تا ۳۰ سال داشته و اینکه آیا سیبیل دارد یا خیر دسته‌بندی شده است.

سیبیل دارد	بدون سیبیل	
۱۲	۲۸	بین ۱۸ تا ۳۰ سال
۸	۵۲	بیش از ۳۰ سال

آزمون فرض به صورت زیر طراحی می‌شود:

فرض صفر به این صورت است که ارتباطی بین داشتن سیبیل و سن فرد وجود ندارد و فرض مقابل نیز به صورت عکس فرض صفر برقرار می‌شود. یعنی فرض صفر بیان می‌کند داشتن سیبیل و سن مستقل از هم هستند.

$H0 : haveing\ moustache\ and\ age\ are\ independent$

$H1 : The\ hypothesis\ H0\ is\ not\ true \rightarrow dependent$

به این منظور از تست goodness of fit استفاده می‌شود تا استقلال خانه‌های جدول بررسی شود:

$$= \sum^{M} \frac{(O_I - E_I)^2}{E_I}$$

در این نوع سوال آماره حاصل از این تست از توزیع خی دو با s-t-1 درجه آزادی است:

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$$

یعنی برای هر یک از خانه‌های جدول مقدار مشاهده‌شده از مقدار مورد انتظار کم شده و به توان ۲ می‌رسد. سپس با مقدار مشاهده شده نرمال‌سازی می‌شود:

$$e_{ij} = \frac{sum\ of\ row * sum\ of\ column}{sum\ of\ table = 100}$$

$$e_{11} = \frac{40*80}{100} = 32$$

$$e_{12} = \frac{40*20}{100} = 8$$

$$e_{21} = \frac{60*80}{100} = 48$$

$$e_{22} = \frac{60*20}{100} = 12$$

$$X^2 = \frac{(28-32)^2}{32} + \frac{(12-8)^2}{8} + \frac{(52-48)^2}{48} + \frac{(8-12)^2}{12} = \frac{16}{32} + \frac{16}{8} + \frac{16}{12} + \frac{4}{12} = 4.166$$

محاسبه درجه آزادی

$$s = rc = 4$$

$$t = r + c - 2 = 2$$

$$df = s - t - 1 = 4 - 2 - 1 = 1$$

برای statistics برابر با ۴.۱۶ و درجه آزادی برابر با ۱ مقدار pvalue برابر با 0.041389 بوده و باتوجه به اینکه از مقدار significance level برابر با 0.05 کمتر است پس فرض صفر رد شده و اثبات می‌شود که ارتباطی بین داشتن سیبیل و سن وجود دارد.

۶_پاسخ مسئله شماره ۶

این مسئله نیز مانند مسئله شماره ۵ به بررسی استقلال در دو احتمال می‌پردازد. به این منظور نمونه ۳۰۰ تایی از افراد هم از نظر نوع گروه خونی و هم از نظر منفی یا مثبت بودن آن یعنی فاکتور rh تقسیم‌بندی شده‌اند. مقادیر در جدول زیر مشاهده می‌شوند:

	O	A	B	AB
Rh positive	۸۲	۸۹	۵۴	۱۹
Rh negative	۱۳	۲۷	۷	۹

آزمون فرض به صورت زیر طراحی می‌شود:

فرض صفر به این صورت است که ارتباطی بین داشتن گروه خونی و Rh فرد وجود ندارد و فرض مقابل نیز به صورت عکس فرض صفر برقرار می‌شود. یعنی فرض صفر بیان می‌کند گروه خونی و فاکتور RH مستقل از هم هستند.

H0 :Blood type and rh are independent

H1 : The hypothesis H0 is not true → dependent

به این منظور از تست goodness of fit استفاده می‌شود تا استقلال خانه‌های جدول بررسی شود:

$$= \sum^{M} \frac{(O_I - E_I)^2}{E_I}$$

در این نوع سوال آماره حاصل از این تست از توزیع خی دو با s-t-1 درجه آزادی است:

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(f_{ij} - e_{ij}\right)^2}{e_{ij}}$$

یعنی برای هر یک از خانه‌های جدول مقدار مشاهده‌شده از مقدار مورد انتظار کم شده و به توان ۲ می‌رسد. سپس با مقدار مشاهده شده نرمال‌سازی می‌شود:

$$e_{ij} = \frac{sum\ of\ row * sum\ of\ column}{sum\ of\ table = 300}$$

$$e_{11} = \frac{244 * 95}{300} = 77.26$$

$$e_{12} = \frac{244 * 116}{300} = 94.34$$

$$e_{13} = \frac{244 * 61}{300} = 49.61$$

$$e_{14} = \frac{244 * 28}{300} = 22.77$$

$$e_{21} = \frac{56 * 95}{300} = 17.73$$

$$e_{22} = \frac{56 * 116}{300} = 21.65$$

$$e_{23} = \frac{56 * 61}{300} = 11.38$$

$$e_{24} = \frac{56 * 28}{300} = 5.22$$

$$X^2 = \frac{(82 - 77.26)^2}{77.26} + \frac{(89 - 94.34)^2}{94.34} + \frac{(54 - 49.61)^2}{49.61} + \frac{(19 - 22.77)^2}{22.77} +$$

$$+ \frac{(13 - 17.73)^2}{17.73} + \frac{(27 - 21.65)^2}{21.65} + \frac{(7 - 11.38)^2}{11.38} + \frac{(9 - 5.22)^2}{5.22} = 8.61$$

محاسبه درجه آزادی

$$s = rc = 8$$

$$t = r + c - 2 = 4$$

$$df = s - t - 1 = 8 - 4 - 1 = 3$$

برای statistics برابر با 8.61 و درجه آزادی برابر با 3 مقدار pvalue برابر با 0.034952 بوده و باتوجه به اینکه از مقدار significance level برابر با 0.05 کمتر است پس فرض صفر رد شده و اثبات می‌شود که ارتباطی بین گروه خونی و rh افراد وجود دارد.

۷_پاسخ مسئله شماره ۷

در این مسئله مقادیر مرتب شده از یک نمونه تصادفی به صورت زیر هستند:

$$y_1 < y_2 < y_3 < y_4 < y_5$$

$\mathrm{F_n}(x)$ برابر است با sample cdf ساخته‌شده در مقادیر این نمونه.

D_n به صورت زیر تعریف شده‌است:

$$D_n = \sup\left|Fn_{(x)} - F(x)\right|$$

در واقع sup تفاضل بین cdf نمونه و $F(x)$ که cdf متغیر تصادفی پیوسته است برابر با D است.

۱-۷_ پاسخ قسمت ۱

این سوال حول مبحث Kolmogorov Smirnov می‌باشد که در این سوال به بررسی آن پرداخته می‌شود.

باتوجه به اینکه ۵ مقدار مشاهده شده در هر مرحله مقدار emprical cdf $\mathrm{F_n}(x)$ $\frac{1}{5}$ افزایش پیدا می‌کند. یعنی در هر مرحله 0.2 اضافه می‌شود. حال به بررسی کمترین مقداری که ممکن است D_i بپذیرد می‌پردازیم. $F(x)$ پیوسته است. پس بیشترین اختلاف بین این دو مقدار قبل از هر y_i رخ می‌دهد. زیرا که در حالت بزرگترمساوی آن، مقدار یک پنجم اضافه شده به $\mathrm{F_n}(x)$ منجر به کم شدن اختلاف می‌شود. مینیمم این ماکزیمم‌ها زمانی است که که این دو به یکدیگر به قدری نزدیک‌ند ولی هنوز مقدار y_i بعدی به $\mathrm{F_n}(x)$ اضافه نشده‌است. پس در این حالت $F(x)$ باید کمترمساوی از 0. 1 باشد. در حالت قسمت اول این مسئله در حالتی که تابع $F(x)$ برابر با مقادیر داده‌شده است، شرط کافی وجود دارد که $\mathrm{F_n}(x)$ با همه مقادیر برقرار است و بیشترین مقدار برابر *0.01* خواهدبود. از طرفی اگر D برابر با ۰.۱ باشد زمانی برقرار است که $\mathrm{F_n}(x)$ یک تابع پله‌ای بوده که در هر y_i به اندازه ۰.۱ اضافه می‌شود. بنابراین حداقل ممکن برای D در این حالت همان ۰.۰۱ خواهدبود.

۲-۷_ پاسخ قسمت ۲

اگر که شرایط ورودی این مسئله برقرار باشد، آنگاه لازم است که $\mathrm{F_n}(x)$ به صورتی باشد که در هر گام به اندازه ۰.۰۲ اضافه شود. که در این حالت منجر می‌شود که D کوچکتر مساوی ۰.۰۲ باشد. در این صورت بیشترین اختلاف بین $F(x)$و $\mathrm{F_n}(x)$ برابر با ۰.۰۲ خواهدشد. که این حالت زمانی اتفاق می‌افتد که $\mathrm{F_n}(x)$ به صورت پله‌ای هربار ۰.۰۲ اضافه شود و بین این حالات مقدارش ثابت باشد. که در این حالت مطلوب مسئله اثبات می‌شود.

۸_پاسخ مسئله شماره ۸

در این مسئله برای تشخیص اینکه داده‌های ارائه‌شده از چه توزیعی هستند باید از تست Kolmogorov Smirnov استفاده کرد. در مسائل قبلی از تست Goodness of fit استفاده شد که حالتی پارامتریک دارد. ولی در این مسئله فقط داده‌های مسئله ارائه شده‌اند.

۱-۸_ پاسخ قسمت ۱

برای حل این مسئله مراحل زیر باید طی شوند:

۱. ابتدا داده‌ها به ترتیب مرتب شوند.
۲. برای هر داده empirical CDF محاسبه شود.

۳. اختلاف با CDF محاسبه شده تا D_n بدست آید.
۴. محاسبه Pvalue از روی D_n و مقایسه با Dcritical و در نهایت بررسی فرض صفر.

باتوجه به اینکه در گروه درسی اعلام شد این سوال با کد هم قابل قبول است، نوت بوک مربوطه پیوست شد.

به این منظور از کتابخانه kstest از پکیج scipy استفاده کرده و جدول مسئله را به ورودی آن می‌دهیم. فرض صفر این است که داده‌ها توزیع یکنواخت دارند پس پارامتر تابع uniform خواهدبود.

نتیجه خروجی به شرح زیر است:

```
Dn: 0.18
Pvalue : 0.3501198034535574
```

باتوجه به اینکه مقدار pvalue بیشتر از ۰.۰۵ است پس فرض صفر رد نمی‌شود و نتیجه می‌شود که داده‌ها توزیع یکنواخت دارند.

۲-۸_ پاسخ قسمت ۲

باتوجه به تغییر سوال مجددا کد تغییر کرد:

رویکرد این قسمت از مسئله نیز همانند قسمت ۱ است با این تفاوت که توزیعی که ارائه شده توزیع شناخته‌شده‌ای مثل یکنواخت نیست. تابع pdf این توزیع به شکل زیر است:

$$f(x)\begin{cases}\frac{3}{2} & for\ 0 < x \le \frac{1}{2}\\ \frac{1}{2} & for\ \frac{1}{2} < x < 1\\ 0 & otherwise\end{cases}$$

برای حل این مسئله ابتدا cdf مربوط به این سوال باید محاسبه شود. به این منظور pdf بالا را به صورت یک تابع تعریف کرده و سپس یک تابع برای محاسبه cdf تعریف می‌شود. این تابع باتوجه به اینکه ورودی چه مقداری باشد بررسی می‌کند که چه مقداری قبل از آن وجود داشته است. اگر مقدار بین صفر تا ۰.۰۵ باشد پس کافی است خروجی توزیع که ۳/۲ است در ورودی ضرب شود. ولی اگر بیشتر از ۰.۰۵ باشد باید مقادیر کمتر از نیم در ۳/۲ ضرب شده و با مقادیر بعد از آن که در ۱/۲ ضرب می‌شوند جمع شود.

پس از این مرحله، cdf توزیع را برای جدول ورودی محاسبه کرده و به تابع kstest هم ورودی و هم cdf جدید را ارائه می‌دهیم. خروجی به شکل زیر خواهدبود:

```
Dn: 0.15000000000000002
Pvalue : 0.5758089853851963
```

باتوجه به اینکه pvalue بیشتر از ۰.۰۵ است پس فرض صفر رد نشده و نتیجه می‌شود که داده‌ها از این توزیع هستند.

۳-۸_ پاسخ قسمت ۳

این قسمت از رویکرد بیزی استفاده می‌کند که هنوز تدریس نشده‌است. ولی باتوجه به راهنمایی‌های انجام‌شده در گروه درسی نتیجه گیری زیر انجام شد:

در این مسئله احتمال پسین باید محاسبه شود. باتوجه به اینکه در مسئله توزیع ارائه شده و احتمال پیشین نیز مشخص است می‌توان احتمال پسین که داده‌ها از توزیع یکنواخت باشند را محاسبه کرد.

احتمال پیشین

احتمال پیشین برای این آزمون فرض برابر است با:

$$P(Uniform) = \frac{1}{2}$$

$$P(pdf\ of\ problem) = \frac{1}{2}$$

محاسبه Likelihoods

$$L(uniform) = 1^{25} = 1$$

برای pdf این مسئله ، Likelihood برابر با ضرب مقادیر pdf برای تمامی داده‌هاست:

$$L(pdf\ of\ problem) = \prod_{i=1}^{25} f(x_i)$$

که pdf در این مسئله برابر است با:

$$f(x)\begin{cases} \frac{3}{2} & for\ 0 < x \leq \frac{1}{2} \\ \frac{1}{2} & for\ \frac{1}{2} < x < 1 \\ 0 & otherwise \end{cases}$$

محاسبه Marginal likelihood

$$P(data) = L(Uniform) * P(Uniform) + L(pdf\ of\ problem) * P(pdf\ of\ problem)$$

پس از محاسبه پارامترهای لازم، این مقادیر در فرمول اصلی جایگذاری می‌شوند تا احتمال پسین محاسیه شود:

محاسبه احتمال پسین برای توزیع یکنواخت

برای محاسبه این احتمال، از رویکرد بیزی استفاده می‌شود:

$$P(Uniform|data) = \frac{L(Uniform) * P(Unifrom)}{P(data)}$$

برای انجام محاسبات بالا کد مربوطه پیوست شد. نتیجه به صورت زیر است:

```
Posterior Probability for Uniform Distribution : 0.4380408216570839
```

۹_پاسخ مسئله شماره ۹

در این مسئله ۲ توزیع ناشناخته مورد بررسی قرار می‌گیرد. این دو توزیع $F(x)$ و $G(x)$ هستند. کد مربوطه پیوست شد.

۱-۹_ مقایسه دو توزیع

برای مقایسه این دو توزیع باتوجه به اینکه ناشناخته بوده و پارامتری ارائه نشده پس نمی‌توان از goodness of fit استفاده کرد. از تست Kolmogorov Smirnov استفاده می‌شود. رویکرد حل این تست به این صورت است که ۲ گروه وجود دارند. پس two sample است. داده‌های هر گروه مرتب شده و empirical cdf برای هر کدام از آن‌ها محاسبه می‌شود. سپس مقدار Dn در اینجا برابر است با بیشترین اختلافی که این دو cdf بدست آمده باهم دارند. سپس مقدار Dcritical محاسبه شده و بر اساس آن و مقایسه با Dn برای رد فرض صفر تصمیم‌گیری می‌شود.

برای حل این مسئله در پایتون، ابتدا داده‌های مشاهده شده از هر دو توزیع به ورودی تابع kstest داده شده و نتیجه زیر حاصل می‌شود:

```
Emprical CDF Of F: [0.04 0.08 0.12 0.16 0.2  0.2  0.24 0.28 0.32 0.32 0
.36 0.36 0.36 0.36
 0.4  0.48 0.48 0.48 0.52 0.56 0.56 0.56 0.56 0.6  0.6  0.64 0.68 0.72
 0.72 0.72 0.72 0.76 0.8  0.84 0.88 0.88 0.88 0.88 0.92 0.92 0.96 0.96
 0.96 1.   1.  ]
Emprical CDF Of G: [0.   0.   0.   0.   0.   0.05 0.05 0.05 0.05 0.1  0
.1  0.15 0.2  0.25
 0.25 0.25 0.25 0.3  0.3  0.3  0.35 0.4  0.45 0.45 0.5  0.5  0.5  0.5
 0.55 0.6  0.65 0.65 0.65 0.65 0.65 0.7  0.75 0.8  0.8  0.85 0.85 0.9
 0.95 0.95 1.   ]
Dn From Calculations: 0.27
Dn of kstest function: 0.27
Pvalue : 0.33570035263225584
```

اگر هردو cdf هم خروجی گرفته شود این اختلاف مشخص است.

باتوجه به اینکه pvalue بیشتر از ۰.۰۵ است پس فرض صفر رد نمی‌شود و نتیجه می‌شود که هردو داده از یک توزیع هستند و می‌توان نتیجه گرفت که توابع یکسانی هستند.

۹-۲_ مقایسه با توزیع Shift شده

در این قسمت از سوال مجدد تست Kolmogorov Smirnov برای دو گروه اجرا می‌شود با این تفاوت که در این قسمت از سوال توزیع اولیه را با توزیعی که شیفت شده مقایسه می‌کنیم. یعنی یک مقدار ثابت به تمامی مقادیر آن اضافه می‌شود.

به این منظور به داده‌های مشاهده شده جدول f مقدار ۲ اضافه شده و سپس مجددا مراحل قبلی برای انجام تست انجام می‌شود. خروجی به صورت زیر خواهدبود.

```
Emprical CDF Of Shifted F: [0.   0.   0.   0.   0.   0.   0.   0.04 0.0
4 0.04 0.04 0.04 0.04 0.04
 0.04 0.04 0.04 0.04 0.08 0.08 0.08 0.08 0.08 0.08 0.08 0.08 0.08 0.08
 0.08 0.08 0.12 0.12 0.12 0.12 0.2  0.2  0.24 0.24 0.32 0.32 0.36 0.48
 0.56 0.56 0.84]
Emprical CDF Of G: [0.   0.   0.   0.   0.   0.05 0.05 0.05 0.05 0.1  0
.1  0.15 0.2  0.25
 0.25 0.25 0.25 0.3  0.3  0.3  0.35 0.4  0.45 0.45 0.5  0.5  0.5  0.5
 0.55 0.6  0.65 0.65 0.65 0.65 0.65 0.7  0.75 0.8  0.8  0.85 0.85 0.9
 0.95 0.95 1.  ]
Dn From Calculations: 0.56
Dn of kstest function: 0.56
Pvalue : 0.000995587375361461
```

باتوجه به اینکه مقدار pvalue کوچکتر از ۰.۰۵ است پس فرض صفر رد می‌شود. پس وقتی داده‌های جدول اول بعلاوه ۲ شوند نتیجه می‌شود که توزیع جدید با قبلی متفاوت خواهدبود و $F(x)$ و $G(x)$ جدید از یک توزیع نیستند.

۹-۳_ مقایسه با توزیع Scale شده

در این قسمت از سوال مجدد تست Kolmogorov Smirnov برای دو گروه اجرا می‌شود با این تفاوت که در این قسمت از سوال توزیع اولیه را با توزیعی که اسکیل شده مقایسه می‌کنیم. یعنی یک مقدار ثابت به تمامی مقادیر آن ضرب می‌شود.

به این منظور به داده‌های مشاهده شده جدول g مقدار ۳ ضرب شده و سپس مجددا مراحل قبلی برای انجام تست انجام می‌شود. خروجی به صورت زیر خواهدبود.

```
Emprical CDF Of F: [0.04 0.08 0.12 0.16 0.2  0.2  0.24 0.28 0.32 0.32 0
.36 0.36 0.36 0.36
 0.4  0.48 0.48 0.48 0.52 0.56 0.56 0.56 0.56 0.6  0.6  0.64 0.68 0.72
 0.72 0.72 0.72 0.76 0.8  0.84 0.88 0.88 0.88 0.88 0.92 0.92 0.96 0.96
 0.96 1.   1.  ]
Emprical CDF Of Scaled G: [0.   0.05 0.05 0.15 0.2  0.2  0.2  0.2  0.2
0.2  0.2  0.2  0.2  0.25
 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.3  0.3
 0.3  0.3  0.3  0.3  0.3  0.35 0.4  0.45 0.45 0.5  0.5  0.5  0.5  0.6
 0.6  0.6  0.65]
Dn From Calculations: 0.5
Dn of kstest function: 0.5
Pvalue : 0.004861780555388566
```

باتوجه به اینکه مقدار pvalue کوچکتر از ۰.۰۵ است پس فرض صفر رد می‌شود. پس وقتی داده‌های جدول دوم در ۳ ضرب شوند نتیجه می‌شود که توزیع جدید با قبلی متفاوت خواهدبود و $F(x)$ و $G(x)$ جدید از یک توزیع نیستند.

۱۰_ پاسخ مسئله شماره ۱۰

یک دیتاست به ورودی داده‌شده است و هدف یافتن ارتباط بین جنسیت و افرادی است که بازمانده از کشتی تایتانیک هستند. کد مربوطه پیوست شد. مراحل زیر به ترتیب طبق مطلوب مسئله انجام شد:

۱-۱۰_ قسمت ۱ مسئله

بخش a

با کمک تابع pd.read_csv فایل مربوطه بارگذاری شد و متد head() برای نمایش ۵ ردیف اول استفاده شد.

بخش b

contingency table برای مشاهده ارتباط بین ستون sex و survive با کمک تابع crosstab از کتابخانه pandas قابل دسترس است. به این منظور دو ستون مد نظر را به ورودی می‌دهیم. خروجی جدول به صورت زیر است:

survive	no	yes
sex		
female	156	307
male	708	142

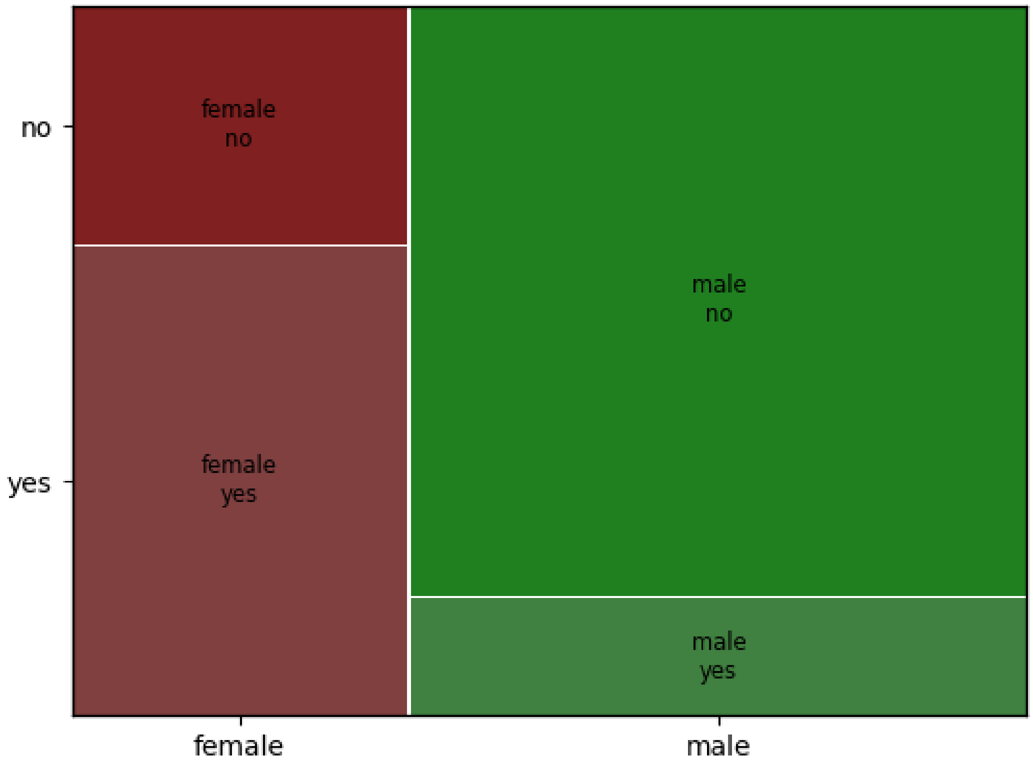
شکل ۳ contigency table برای دیتاست تایتانیک

بخش c

با کمک متد tocsv جدول قسمت قبلی ذخیره شد.

بخش d

ارتباط بین دو متغیر مورد نظر جنسیت و بازماندگان با نمودار mosaic نمایش داده‌شد.



شکل ۴ نمودار mosaic برای بررسی ارتباط بین جنسیت و بازماندگان تایتانیک

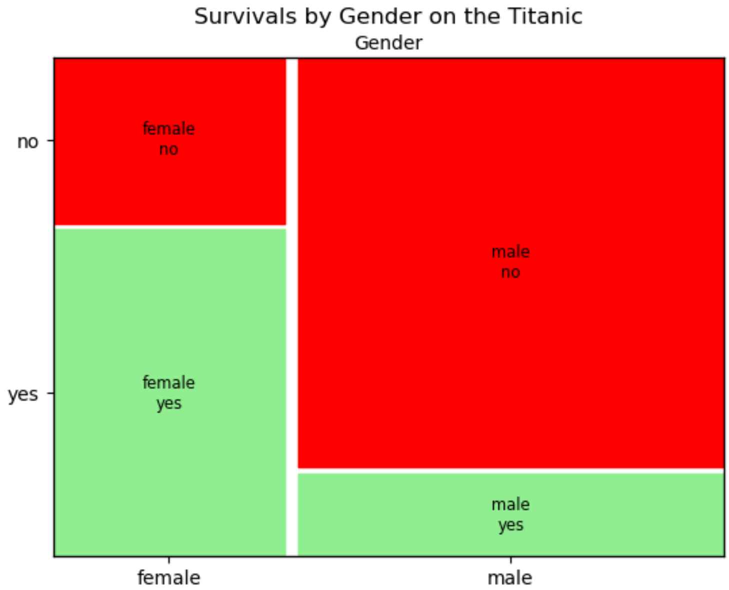
مشخص است که ارتباطی وجود داشته و تعداد بازماندگان زن بیشتر است.

بخش e

survive	no	yes
sex		
female	156	307
male	708	142

شکل ۵ contigency table برای دیتاست تایتانیک

بخش f



شکل ۶نمودار mosaic برای بررسی ارتباط بین جنسیت و بازماندگان تایتانیک

۲-۱۰_ قسمت ۲ مسئله

در این قسمت از سوال تست‌های خی دو بر روی contingency table انجام می‌شود تا ارتباط بین جنسیت و بازماندگان بررسی شود. فرض صفر بر این اساس است که ارتباطی بین جنسیت و بازماندگان وجود ندارد.

بخش a

خروجی تست به شرح زیر است:

```
Statistics: 325.5037787069806
Degree Of Freedom: 1
Pvalue : 9.164113332735093e-73
```

باتوجه به اینکه مقدار pvalue بسیار کم بوده و از ۰.۰۵ کمتر است، فرض صفر قویا رد می‌شود و مشخص است که ارتباطی بین جنسیت افراد و بازماندگان وجود دارد.

بخش b

فرضیاتی که آزمون chi squre دارد به شرح زیر است که برای هرکدام از آن‌ها با شرایط آزمون فعلی چک می‌شود.

فرض استقلال : در این جدول جنسیت زن و و بازماندگان نیز مستقل هستند و این شرط برقرار است.

فرض تصادفی بودن نمونه‌برداری: مشخص است که این نمونه‌برداری از مسافران کشتی تایتانیک است که بعضی از آنان نجات یافته و بعضی نجات نیافتند. نمونه برداری تصادفی است و این شرط برقرار است.

فرض Categorical بودن متغیرها: در این مسئله هم جنسیت به صورت زن و مرد تعریف شده و هم بازماندگان به صورت yes و no پس این شرط هم برقرار است.

بخش c

در این قسمت از سوال تست Fisher's exact بر روی جدول contingency table اجرا می‌شود. به این منظور از تابع fisher_exact از کتابخانه scipy استفاده می‌شود. نتیجه به شرح زیر است:

```
Statistics: 0.10191575111798155
Pvalue : 5.187445473452701e-73
```

باتوجه به اینکه مقدار pvalue بسیار کم بوده و از ۰.۰۵ کمتر است، فرض صفر قویا رد می‌شود و مشخص است که ارتباطی بین جنسیت افراد و بازماندگان وجود دارد.

بخش d

این خروجی در قسمت‌های قبلی نیز محاسبه شد:

```
Chi2 Statistics: 325.5037787069806
Degree Of Freedom: 1
Pvalue : 9.164113332735093e-73
```

تحلیل:

میزان آماره‌ای که تست می‌دهد بسیار بالاست و این نشان می‌دهد که اختلاف بین مقادیر مورد انتظار و مقادیر مشاهده شده زیاد است. و همین اختلاف نشان‌دهنده موضوع وجود شواهی خلاف بر چیزی است که انتظار می‌رود برقرار باشد. زیرا که فرض صفر بر اساس این موضوع است که جنسیت تاثیری بر روی اینکه شخصی بازمانده باشد یا نه ندارد.

درجه آزادی در این تست از روش زیر محاسبه می‌شود:

$s = rc = 4$

$t = r + c - 2 = 2$

$df = s - t - 1 = 4 - 2 - 1 = 1$

مقدار pvalue نیز در قسمت قبلی تحلیل شد. فرض صفر در این مسئله رد شده است .باتوجه به اینکه مقدار آن بسیار کم است، شواهد قوی وجود دارد که ارتباطی بین اینکه فردی چه جنسیتی داشته که نجات یافته است وجود دارد.

بخش e

نتایج این تست به شرح زیر است:

```
Statistics: 0.10191575111798155
Pvalue : 5.187445473452701e-73
```

مقدار آماره‌ای که این تست می‌دهد نشان‌دهنده ارتباط قوی بین بین دو متغیر است. زیرا که نه تنها از ۱ کمتر است بلکه نزدیک به صفر است. پس میزان شانسی بودن تا حد بسیار زیادی رد می‌شود.

مقدار pvalue نیز همانند تست قبلی خی دو بسیار کم و نزدیک به صفر است و به صورت قوی فرض صفر را رد می‌کند. پس شواهد قوی وجود دارد که ارتباطی بین اینکه فردی چه جنسیتی داشته که نجات یافته است وجود دارد.

۱۱_ پاسخ مسئله شماره ۱۱

۱-۱۱_ پرسش‌های با پاسخ کوتاه

a **درست** ممکن است، در اصل ما در نتیجه تست بیان می‌کنیم که با چه شدتی فرض صفر رد شده و نتیجه فرض مقابل پذیرفته می‌شود. حال هرچه شدت بیشتر باشد، متوجه می‌شویم که نتیجه مورد نظر احتمال کمتری دارد که شانسی باشد.

b **نادرست** ارتباطی بیت یک عدد بزرگ و اینکه significant بودن محرز شود وجود ندارد و ممکن است برعکس باشد.

c **درست** اگرچه اختلاف این دو p-value باهم کم است؛ ولی متفاوت بوده و در مقابل تست‌های مختلف ممکن است نتایج مختلفی برای رد فرض صفر مشاهده شود.

۲-۱۱_ پرسش‌های قسمت ۲

a بله شانس و احتمالات در نتیجه‌گیری تأثیر دارد و در نتیجه تست هم قدرت رد فرض صفر بیان می‌شود و ممکن است تفاوتی که با فرض صفر مشاهده‌شده بر اثر شانس باشد و سعی داریم که عنصر شانس و رندم بودن تاثیر کمتری در نتیجه‌گیری تست داشته باشد. در اصل هدف تست این است که متوجه شویم تفاوت بر اساس شانس بوده یا اینکه الگوی خاصی وجود دارد.

b بله هرچه تفاوت بیشتر باشد با اطمینان بیشتری می‌توان فرض صفر را رد کرد و پایه تست‌های استنباط آماری بر این است که تفاوت را بیان می‌کنند. حال اینکه میزان تفاوت چقدر باشد نتایج تغییر می‌کنند.

c خیر تست‌های آماری درباره اینکه تفاوت‌ها چه چیزی بیان می‌کنند توضیحی ارائه نمی‌دهند.

d خیر استنباط آماری درباره کیفیت آزمایش نظری نمی‌دهد. اگر کیفیت پایین باشد نتیجه استنباط آماری نیز مطلوب نخواهدبود. ولی این مورد را نتایج استنباط آماری بیان نمی‌کنند. بلکه مربوط به حوزه‌های دیگری از حوزه آمار است.

۳-۱۱_ پرسش قسمت ۳

باتوجه به پرسش ارائه شده یک فرض صفر توسط ۲ نفر بررسی می‌شود. این فرض صفر بیان می‌کند که میانگین برابر ۵۰ است. بر اساس فرض مقابل نیز میانگین برابر ۵۰ نیست. یعنی کوچکتر یا مساوی ۵۰ . هردو نفر نیز با نمونه‌برداری ۱۰۰ تایی و ۹۰۰ تایی از جعبه، از نوع با جایگذاری میزان انحراف معیار ۱۰ را نتیجه می‌گیرند. حال پرسش مطرح شده که آیا فردی که pvalue کمتری گزارش کرده کسی است که میانگین بدست آمده از نمونه‌اش از ۵۰ فاصله بیشتری دارد؟

خیر. باتوجه به فرمول برای بدست آوردن test statistics تست از نوع z test که تمامی پارامترهای میانگین، انحراف معیار و اندازه نمونه تاثیر دارند. در اینجا اندازه نمونه متفاوت است. ممکن است که فرد اول با نمونه کوچکتر متفاوت تر از میانگین باشد. ولی به طور کلی هرچه که pvalue کمتر باشد، احتمال بیشتری وجود دارد که میانگین به ۵۰ نزدیک تر است.

۴-۱۱_ پرسش قسمت ۴

در این پرسش بیان شده که مطابث نتیجه یک z-test ، تفاوت معناداری بین درصد کارکنان سفیدپوست یک شرکت و درصد سفیدپوستان نسبت به دیگر نژادها در آن منطقه وجود دارد و حال هدف این تست این است که آیا تبعیض نژادی صورت گرفته یا خیر.

فرض شده که ۱۰ درصد جمعیت یک شهر سفیدپوست هستند و فرض شده که استخدام کارگران به صورت نمونه‌برداری تصادفی بر اساس فاکتور نژاد صورت گرفته است. باتوجه به اینکه با کمک تست z می‌توانیم بررسی کنیم که میانگین استخدام شدگان سفید برابر مقدار خاصی باشد، در صورتی که در نمونه‌برداری از این کارخانه‌ها، نتیجه‌ای حاصل شود که نشان دهد تفاوت معناداری میان میانگین نژاد مشاهده‌شده در شهر و میانگین مشاهده‌شده در نمونه‌برداری از کارخانه وجود دارد(رد فرض صفر اگر فرض صفر را بر این اساس در

نظر گرفته شود که میانگین نژاد در کارخانه برابر میانگین در شهر است) می‌توان نتیجه گرفت که در فرآیند استخدامی تبعیض نژادی صورت گرفته است.

۱۲_ پاسخ مسئله شماره ۱۲

در این مسئله برای تست سالم بودن تاس، ۱۰۰ بار تاس انداخته شده و در این آزمایش اگر بین ۴۰ تا ۶۰ بار شیر بیاید، سالم‌بودن تاس نتیجه گرفته می‌شود که همان فرض صفر است. در غیر این صورت فرض صفر رد شده و نتیجه تاس ناسالم خواهدبود.

۱-۱۲_ پاسخ قسمت ۱

در این قسمت حالتی مد نظر است که با اینکه فرض صفر صحیح بوده، رد شود. یعنی خطای نوع ۱ رخ دهد. P(Reject H0 | H0)

در این مسئله این حالت برابر است با حالتی که در خارج از محدوده مد نظر شیر بیاید. یعنی کمتر از ۴۰ و بیشتر از ۶۰ بار شیر بیاید. پس باید این احتمال محاسبه شود. باتوجه به اینکه توزیع پرتاب تاس یک توزیع دوجمله ای است می‌توان برای n های بزرگ آن را به نرمال تقریب زد. پس ابتدا میانگین و انحراف معیار آن را محاسبه کرده و به نرمال استاندارد تبدیل می‌شود.

برای توزیع دوجمله‌ای:

$$\mu = np$$

$$\sigma = \sqrt{np(1-p)}$$

در این آزمایش *n=100* و *p=0.5* :

$$\mu = np = 100 * 0.5 = 50$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{50(0.5)} = 5$$

تبدیل به نرمال استاندارد برای ناحیه بحرانی کمتر از ۴۰ و بیشتر از ۵۰:

$$\mu < 40 \rightarrow P(x < 40) = p\left(z < \frac{40-\mu}{\sigma}\right) = \frac{40-50}{5} = -\frac{10}{5} = -2$$

$$= \varphi(-2) = 0.02275$$

$$\mu > 60 \rightarrow P(x < 60) = p\left(z < \frac{60-\mu}{\sigma}\right) = \frac{60-50}{5} = \frac{10}{5} = 2$$

$$= 1 - \varphi(2) = 1 - 0.97725 = 0.02275$$

هرکدام در دم توزیع 0.02 احتمال دارند که جمع آن‌ها برابر با حالتی است که فرض صفر به اشتباه رد شده است. که این مقدار برابر است با :

$$P(\text{Reject H0} \mid \text{H0}) = 0.0455$$

۲-۱۲_ پاسخ قسمت ۲

باتوجه به مطالب درس، حالتی که فرض صفر به اشتباه رد شده یعنی فرض صفر صحیح بوده ولی رد شده است برابر با مقدار significance level یا همان α است که در این مسئله مقدار آن برابر 0.0455 محاسبه شد. اگر محافظه‌کارانه عمل شود باید مقدار α کمتر هم شود ولی فرض همان 0.0455 است. می‌توان آزادانه‌تر هم برخورد کرد و $\alpha = 0.05$ در نظر گرفت.

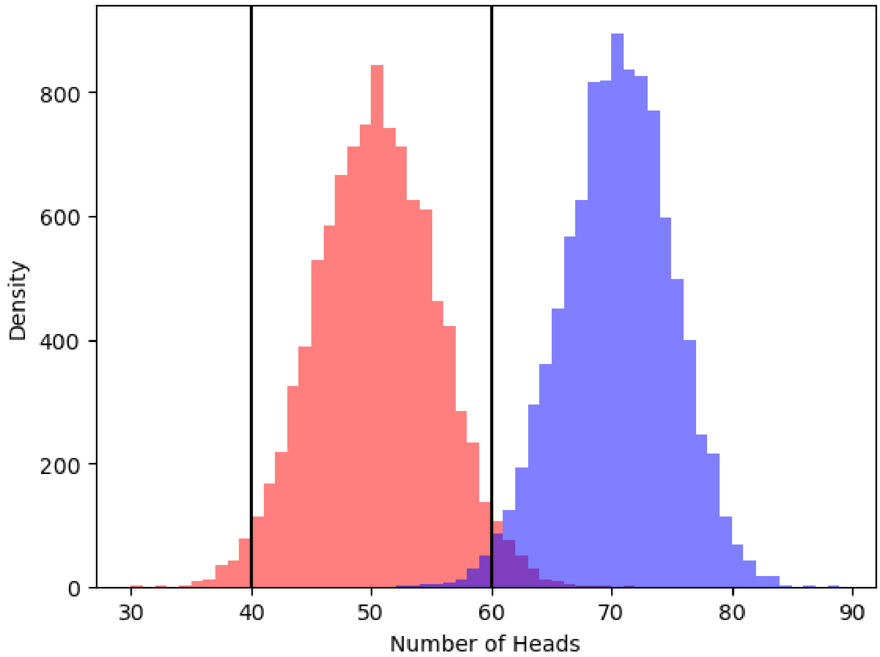
۳-۱۲_ پاسخ قسمت ۳

فایل شبیه‌سازی پیوست شد. در این مسئله همانطور که در شکل زیر مشاهده می‌شود، برای توزیع‌های دوجمله‌ای با p=0.5 و p=0.7 با دو رنگ آبی و قرمز مشخص شده است. مقدار میانگین برابر ۴۰ و ۶۰ نیز با خط مشکی مشخص شده است. ابتدا دو توزیع دوجمله‌ای تصادفی با کمک تابع np.random.binomial تشکیل داده‌شده و سپس نمودارهرکدام در یک هیستوگرام واحد نمایش داده‌شده است. مقداری که باعث رد فرض صفر شده و نتیجه می‌دهد که سکه سالم نبوده و unfair است زمانی است که میانگین آمدن شیر بیشتر از ۶۰ و یا کمتر از ۴۰ بوده که برابر است با مساحت زیر نمودار قرمز رنگ بعد و قبل از خط مشکی که در قسمت قبل سوال با کمک تقریب نرمال محاسبه شد. اگر که مقدار بین دو خط مشکی در نمودار قرمز باشد، در این حالت فرض صفر رد نشده و نتیجه گرفته می‌شود که سکه سالم است.

مقادیر α و β در شکل زیر به شرح زیر هستند(برای مشخص کردن این نواحی مقدار alpha=0.5 قرار داده شده تا همپوشانی هر دو توزیع مشاهده شود.):

α ناحیه بنفش سمت راست خط مشکی در قسمت میانگین برابر ۶۰ که در صورتی که حاصل test statistics در این ناحیه باشد، فرض صفر رد می‌شود.(در صورتی که واقعا فرض صفر صحیح باشد به اشتباه رد شده و خطای نوع ۱ رخ می‌دهد)

β ناحیه بنفش سمت چپ خط مشکی در قسمت میانگین برابر ۶۰ که در صورتی که مقادیر در این ناحیه باشند، فرض صفر رد نمی‌شود. ولی ممکن است که فرض صفر اشتباه بوده و نمودار آبی در این حوزه متمرکز باشد و خطای نوع دو رخ دهد.



شکل ۷ توزیع دوجمله برای ۱۰۰ بار پرتاب تاس سالم و ناسالم

برای محاسبه قدرت تست باید ابتدا احتمال خطای نوع دو محاسبه شده و از ۱ کم شود و ناحیه‌ی قرمز بجز قسمت بنفش حاصل این تفاضل خواهدبود.

خروجی کد:

```
Beta: 0.9813
Power = 1 - Beta: 0.018700000000000005
```

حل مسئله:

$$\beta = P(Accept\ H0\ |H1) = P(x < 60) = 0.02$$

$Power = 1 - \beta = 0.02$

باتوجه به مقدار کمی که β دارد نتیجه گرفته می‌شود که احتمال اینکه سکه ناسالم باشد و در تست نتیجه گرفته شود که سکه سالم است کم بوده و تست قدرت بالایی دارد.

حال وقتی که مقدار تکرار را کم می‌کنیم ناحیه β در برخی اجراها بیشتر می‌شود که نتیجه گرفته می‌شود که شانس دخالت بیشتری دارد و هرچه بیشتر می‌شود به توزیع نرمال نزدیک‌تر خواهدبود.

۱۳_ پاسخ مسئله شماره ۱۳

در این مسئله یک نمونه تصادفی ۵۰ تایی از دانش‌آموزان مدرسه عمومی و ۶۵ تایی از دانش‌آموزان مدارس کاتولیک در نظر گرفته شده است. میانگین نمره در نمونه مدرسه عمومی ۷۰ با انحراف معیار ۴ و در مدرسه کاتولیک ۷۴ با انحراف معیار ۶ است.

۱۳-۱_ نوع تست

در این مسئله میانگین دو توزیع نرمال مد نظر بوده پس تست از نوع Parametric خواهدبود. همچنین هدف مقایسه میانگین دو نمونه مستقل از هم بوده و برای هر دو گروه مقدار n از ۳۰ بیشتر است. از طرفی واریانس نیز معلوم است پس از تست دو گروه مستقل z-test استفاده می‌شود. اگر واریانس معلوم نبود و مقدار n کم باشد از two sample t-testاستفاده می‌شد. ولی اینجا از انحراف معیار نمونه برای این مورد استفاده می‌شود زیرا که n به اندازه کافی بزرگ است.

۱۳-۲_ تست تفاوت حول میانگین دو گروه

$Significance\ level(\alpha) = 0.05 \rightarrow 2sided \rightarrow z_{0.025} = 1.96$

$H0 : \mu_1 = \mu_2$

$H1 : \mu_1 \neq \mu_2$

$\overline{x_1} = 70 \qquad s_1 = 4$

$\overline{x_2} = 74 \qquad s_2 = 6$

$\delta = \mu_1 - \mu_2$

$$z = \frac{\overline{x_1} - \overline{x_2} - \delta}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} = \frac{70 - 74}{\sqrt{\frac{16}{50} + \frac{36}{65}}} = -4.27$$

باتوجه به اینکه مقدار *test statistics* که از z بدست آمده از *-1.96* کمتر است پس فرض صفر رد می‌شود. مقدار *pvalue* نیز برابر ۰.۰۰۰۰۱ که از α کمتر است. پس مشخص است که فرض صفر رد می‌شود. و نتیجه می‌شود که میانگین این دو گروه تفاوت معناداری با یکدیگر دارند.

۱۴_ پاسخ مسئله شماره ۱۴

در این تست هدف بررسی تفاوت بین افرادی است که بازی‌های ویدئویی انجام داده و افرادی است که انجام نمی‌دهند(یا کمتر انجام می‌دهند) و این تفاوت را از نظر ادراک فضایی (spatial perceptions به اختصار sp) بررسی می‌کند. ۲۰ نفر از نمونه افرادی هستند که کمتر از یک ساعت در هفته بازی می‌کنند و میانگین sp برابر ۱۲۰ و انحراف معیار ۲۰ دارند و نمونه دوم ۱۵ نفر از افرادی هستند که حداقل ۱۰ ساعت در هفته بازی می‌کنند و میانگین sp آن‌ها ۱۰۰ بوده و انحراف معیار ۵۰ است.

۱-۱۴_ پاسخ قسمت ۱

واضح است که این آزمایش از نوع observational است زیرا که دو گروه فقط بر اساس ویژگی‌های خودشان است و محقق دخالتی در آن نداشته و تغییری در ویژگی‌ها یا رفتار افراد انجام نداده است.

۲-۱۴_ پاسخ قسمت ۲

در این مسئله برای مقایسه میانگین‌ها باتوجه به اینکه مقدار n کم است ای تقریب نرمال استفاده نمی‌شود و از two-sample t test استفاده می‌شود.

Degree of freedom = m+n-2 = 33

$H0 : \mu_1 = \mu_2 \rightarrow \delta = \mu_1 - \mu_2 = 0$

$H1 : \mu_1 > \mu_2 \rightarrow \delta = \mu_1 - \mu_2 > 0$

$\overline{x_1} = 120 \quad s_1 = 20 \quad n_1 = 20$

$\overline{x_2} = 100 \quad s_2 = 50 \quad n_2 = 15$

$$sp = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(20 - 1)400 + (15 - 1)2500}{20 + 50 - 2}} =$$

$$\sqrt{\frac{19 * 400 + 14 * 2500}{68}} = 25.02$$

$$t = \frac{\overline{x_1} - \overline{x_2} - \delta}{sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{120 - 100 - 0}{25\sqrt{\frac{1}{20} + \frac{1}{50}}} = \frac{20}{25\sqrt{\frac{1}{20} + \frac{1}{50}}} = 3.02$$

$Significance\ level$ با $(\alpha) = 0.05$

$\alpha = 0.05\, , df = 33 \rightarrow t_{0.05,33} = 1.69$

باتوجه به اینکه آماره t از ناحیه بحرانی بیشتر است پس فرض صفر رد می‌شود.

$Significance\ level$ با $(\alpha) = 0.01$

$\alpha = 0.01 , df = 33 \rightarrow t_{0.01,33} = 2.44$

باتوجه به اینکه آماره t از ناحیه بحرانی بیشتر است پس فرض صفر رد می‌شود. پس فرض صفر نتیجه گیری می‌شود و میانگین spatial perceptions در گروه اول بیشتر است.

۳-۱۴_ پاسخ قسمت ۳

در این تست ارتباط بین spatial perceptions و انجام بازی‌های ویدئویی نتیجه گرفته‌شد. ولی اینکه یک correlation بین این دو باشد به این سادگی قابل نتیجه نیست. عناصر بسیار زیادی در ارتباط بین دو عامل موثراند و باتوجه به اندازه نمونه کوچک احتمال تاثیر شانس نیز تاثیر زیادی دارد. و به طور کلی عوامل بسیار زیادی می‌توانند بر نتیجه این تست تاثیر بگذارند. برخی از این عوامل:

ممکن است نتیجه برعکس باشد، یعنی افرادی که spatial perceptions بالا دارند علاقه دارند بیشتر بازی کنند.

عوامل ژنتیکی نیز بر نتایج آزمایش تاثیر دارند. عوامل دیگر که بر عامل spatial perceptions تاثیر دارند در اینجا بررسی نشده‌اند. برای نمونه مثال دارو ارائه شده در کلاس درس که عوامل مختلفی ممکن است تاثیرگذار باشند که برای سنجش تاثیر همه آن‌ها آزمایش دقیق تر و کنترل شده نیاز است. در این آزمایش که Observational است عامل برقراری مورد بحث قرار نگرفته و فقط بیان شده که افرادی که بیشتر بازی می‌کنند چه خصوصیتی دارند.

۱۵_ پاسخ مسئله شماره ۱۵

در این مسئله فرض شده که در یک تست یک طرفه حول میانگین مقدار power برابر است با:

$$power = \Phi\left(z_\alpha + \frac{n^{\frac{1}{2}}(\mu_A - \mu_0)}{\sigma}\right)$$

۱-۱۵_ پاسخ قسمت ۱

میزان قوی بودن یک تست اگر که از نمونه به اندازه ۴۹ توزیع نرمال بوده که واریانس آن ۴۹ بوده و مقدار تفاوت بررسی شده میان میانگین فرض صفر و فرض مقابل برابر ۰.۵ است. مقدار خطای نوع ۱ یا α نیز برابر با ۰.۰۵ می‌باشد.

حال با جایگذاری در فرمول میزان power تست محاسبه می‌شود:

$$power = \Phi\left(z_\alpha + \frac{n^{\frac{1}{2}}(\mu_A - \mu_0)}{\sigma}\right) = \Phi\left(1.645 + \frac{49^{\frac{1}{2}}(0.5)}{\sqrt{49}}\right) =$$

$$\Phi(2.145) = 0.98319$$

پس میزان قدرت تست برابر است با ۰.۹۸۳۱۹ یعنی در بیش از ۹۸ درصد مواقع وقتی که فرض صفر رد شده و فرض مقابل پذیرفته می‌شود، فرض مقابل واقعا درست است.

۱۵-۲_ پاسخ قسمت ۲

میزان خطای نوع ۲ یعنی زمانی که فرض صفر به اشتباه پذیرفته شده است برابر مکمل power است که در قسمت قبل محاسبات آن انجام شد.

$$power = 1 - \beta$$

$$Type\ 2\ Error = \beta = 1 - power = 1 - \ 0.98319 = 0.01681$$

۱۵-۳_ پاسخ قسمت ۳

در این قسمت معلوم و مجهول جابجا شده‌اند و کافی است که n را به عنوان مجهول قرار داده و میزان power برابر ۰.۹۹ قرار داده‌شود:

$$power = \Phi\left(z_\alpha + \frac{n^{\frac{1}{2}}(\mu_A - \mu_0)}{\sigma}\right) = 0.99 \rightarrow \Phi\left(1.645 + \frac{n^{\frac{1}{2}}(0.5)}{7}\right) = 0.99$$

از طرفی

$$\Phi^{-1}(0.99) = 2.326$$

$$1.645 + \frac{n^{\frac{1}{2}}(0.5)}{7} = 2.326 \rightarrow \frac{n^{\frac{1}{2}}(0.5)}{7} = 0.684 \rightarrow n^{\frac{1}{2}} = 9.567 \rightarrow n \approx 91.527489$$

پس n باید حداقل ۹۲ باشد.

۱۶_ پاسخ مسئله شماره ۱۶

X یک متغیر تصادفی با توزیع دوجمله‌ای با n=100 می‌باشد. آزمون فرض به صورت زیر تنظیم شده است:

$$H0 : p = 0.5$$

$$H1 : p \neq 0.5$$

اگر که شرایط زیر برقرار شود، فرض صفر رد می‌شود:

$|X-50|>10$

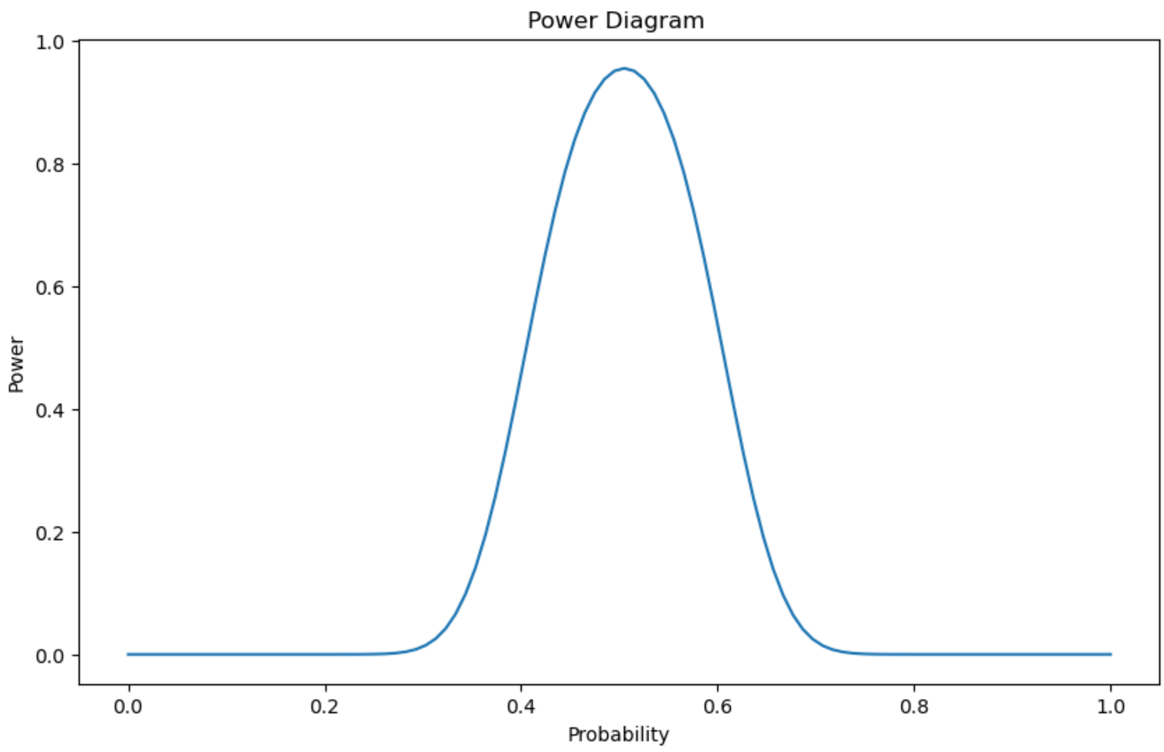
باتوجه به اینکه توزیع دوجمله‌ای است واریانس و میانگین آن مشخص است مقدار آماره‌ای که به ازای آن فرض صفر رد می‌شود محاسبه می‌شود. این حالتی است که اختلاف با میانگین ۱۰ باشد. مثلا ۶۰:

$\overline{mean}=np=100p \quad var=np(1-p)\rightarrow s=\sqrt{np(1-p)} \quad n=100$

باتوجه به اینکه قدرمطلق استفاده‌شده پس تست دو طرفه است و باید مقدار $z_{\frac{\alpha}{2}}$ محاسبه شود:

$$z_{\frac{\alpha}{2}}=\frac{x-\mu}{\sigma}=\frac{x-100p}{\sqrt{np(1-p)}}=\frac{60-100*0.5}{\sqrt{100*0.5*0.5}}=2$$

$$\Phi\ (2)=\frac{\alpha}{2}=0.02275\rightarrow\alpha=0.0455$$



شکل ۸ نمودار تقریب قدرت تست

۱۷_ پاسخ مسئله شماره ۱۷

در این مسئله تابع likelihood برای یک توزیع نمایی که تست حول میانگین آن برقرار شده ارائه شده‌است:

$f(x|\theta) = \theta e^{-\theta x}$

$H0 : \theta = \theta 0$

$H1 : \theta \neq 0.5$

حال مطوب مسئله محاسبه likelihood ratio و rejection region می‌باشد:

برای این مسئله تابع likelihood وقتی که نمونه تصادفی برابر X1, ...,Xn به صورت زیر خواهدبود:

$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^{n} x_I}$

محاسبه likelihood ratio

$$\Lambda = \frac{\text{MAX } \theta \epsilon \theta_0 \ \ L(\theta)}{\text{MAX } \theta \epsilon \theta \ \ L(\theta)} = \frac{L(\theta_0)}{L(\hat{\theta})}$$

پس تخمین $\hat{\theta}$ که برابر با MLE θ محاسبه می‌شود:

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$$

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})} = \frac{\theta_0^n e^{-\theta_0 \sum_{i=1}^{n} x_i}}{\hat{\theta}^n e^{-\hat{\theta} \sum_{i=1}^{n} x_i}} = \frac{\theta_0^n e^{-\theta_0 \bar{x}}}{\left(\frac{1}{\bar{\bar{x}}}\right)^n e^{-\frac{1}{\bar{\bar{x}}}\bar{x}}} = \bar{x}^{\,n}\ \theta_0^n\ e^{n(1-\theta_0 \bar{x})}$$

ناحیه‌ای که رد که همان ناحیه‌ای است که باعث رد فرض صفر می‌شود:

$Reject\ H0 <\!\!\to\ \ \Lambda(X) < critical\ value$

حال باید معادله‌ای که در آن $\Lambda(X)$ در صورت بزرگتر بودن از آن فرض صفر رد می‌شود پیدا شود که آن ناحیه همان rejection region خواهدبود یعنی مقادیری که وقتی لاندا از آن کمتر است فرض صفر رد می‌شود:

$-2\log(\Lambda) = X_1^2$

$-2\log\left(\bar{x}^{\,n}\ \theta_0^n\ e^{n(1-\theta_0\bar{x})}\right) = -2n\log(\bar{x}) - 2n\log(\theta_0) + 2n - 2n\theta_0\bar{x} \ \ = X_1^2$

وقتی که مقدار بالا از $X_1^2(\alpha)$ بیشتر باشد فرض صفر رد می‌شود.

۱۸_ پاسخ مسئله شماره ۱۸

در این مسئله میانگین وزن نوزادان تازه‌متولدشده در بیمارستان ارائه شده است. یک نمونه ۵۰تایی از این جامعه برداشته شده که میانگین نمونه برابر ۲۵.۹ و خطای استاندارد آن ۵.۶ است. باتوجه به اینکه تست حول میانگین است و تعداد نمونه بیشتر از ۳۰ است پس می‌توان از توزیع z که همان نرمال استاندارد است استفاده شود. فرض صفر به عنوان میانگین برابر ۲۸ و فرض مقابل به عنوان کوچکتربودن میانگین از ۲۸ در نظر گرفته می‌شود.

۱-۱۸_ آزمون فرض

$Significance\ level(\alpha) = 0.5 \to 1.64$

$\bar{x} = 25.9$

$\mu = 28$

$s = 5.6$

$n=50$

$$z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{25.9 - 28}{\frac{5.6}{\sqrt{50}}} = -2.65$$

حال باید مقدار محاسبه شده در تابع نرمال استاندارد محاسبه شود. هرچند که در این پرسش مقدار α برابر با 0.05 بیان شده است که مقدار z در جدول برابر با 1.64 برای تست یک طرفه می‌باشد. پس برای مقادیر بزرگتر از 1.64و کمتر از 1.64- فرض صفر رد می‌شود. باتوجه به اینکه -2.65 از 1.64- کمتر است پس فرض صفر رد شده و متوجه می‌شویم که میانگین وزن نوزادان تفاوت معناداری با میانگین برابر ۲۸ دارد و از آن کمتر است.

روش دیگر نیز این است که مقدار p-value برای 2.65- برابر با 0.004 است که باتوجه به کمتر بودن نسبت به α فرض صفر رد می‌شود.

۲-۱۸_ خطای نوع ۲

در این قسمت باید خطای نوع ۲ در حالتی بررسی شود که مقدار میانگین جامعه بیشتر از ۲۷ نبوده و کمترمساوی ۲۷ است. خطای نوع دو یا β برابر با حالتی که فرض مقابل به اشتباه رد شده‌است یا به معنای دیگر فرض صفر به اشتباه رد نشده‌است. حال این احتمال را محاسبه می‌کنیم.

$\alpha = 0.05$

$s = 5.6$

$\mu \leq 27 \rightarrow true\ mean$

$Ha \rightarrow \mu_a < 28$

$\beta = P\big(Accept\ H0(cant\ Reject\ H0)\ \big|H1) = ?$

ابتدا test statistic با فرض اینکه فرض صفر صحیح باشد را در نظر می‌گیریم.

$$z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

حال با کمک معادله بالا کمترین میانگین نمونه‌ای که نمی‌توانیم با آن فرض صفر را رد کنیم محاسبه می‌کنیم:

$$\bar{x} = \frac{s}{\sqrt{n}} z + \mu_a = \frac{5.6}{\sqrt{50}} * -1.64 + 28 = 26.70$$

حال احتمال اینکه این مقدار میانگین نمونه در ناحیه‌ای باشد که نتوانیم فرض صفر را رد کنیم محاسبه می‌کنیم. کافی است کمترین مقدار ممکن برای میانگین نمونه‌ای را باتوجه‌به میانگین واقعی به نرمال استاندارد تبدیل کرده و مقدار موردنظر را از جدول بیابیم.

$$P\left(Z \geq \frac{26.70 - 27}{\frac{5.6}{\sqrt{50}}}\right) = P(Z \geq -0.37) = 1 - P(Z \leq\ -0.37) = 0.64$$

پس به‌احتمال ۶۴ درصد در این حالت فرض صفر به اشتباه رد نمی‌شود.(خطای نوع ۲)

۱۹_ پاسخ مسئله شماره ۱۹

در این مسئله ۴۰ نمونه از جراحی‌ها را در نظر گرفته که در این نمونه‌ها، میانگین نمونه‌ای برابر ۱۳۰ دقیقه و انحراف معیار برابر با ۵ دقیقه استبازه اطمینان ۹۵ درصدی برای میانگین زمان جراحی‌ها با کمک روش Bootstrap با ۱۰۰۰ نمونه تولیدشده بازه‌ی ۱۲۸.۵ تا ۱۳۱.۵ دقیقه بدست آمده است.

۱-۱۹_ پاسخ قسمت ۱

در این پژوهش باتوجه‌به نمونه‌برداری انجام شده متوجه شده‌ایم که میانگین زمان عمل‌ها در نمونه برابر با ۱۳۰ دقیقه است. و انحراف معیار در این نمونه برابر ۵ دقیقه است. .هرچند باتوجه به خطای نمونه‌برداری نمیتوان گفت که میانگین جامعه هم قطعا برابر ۱۳۰ خواهدبود. و همچنین با کمک روش Bootstrap یک بازه اطمینان برای پوشش میانگین بدست آمده‌است. اگرچه که این بازه اطمینان دقت کافی به اندازه روش

قبلی با کمک قضیه حد مرکزی را ندارد. ولی می‌توان گفت که ۹۵درصد اطمینان داریم که میانگین جامعه در بازه بین ۱۲۸.۵ تا ۱۳۱.۵ دقیقه باشد.

۱۹-۲_ پاسخ قسمت ۲

برای به‌دست‌آوردن بازه اطمینان برای نمونه روش normal approximation ابتدا از جامعه نمونه‌برداری کرده و با از نمونه، نمونه‌برداری می‌کنیم و دنیای bootstrap را تشکیل می‌دهیم. برای مثال ۱۰۰۰ نمونه جدید با جایگذاری با کمک نمونه‌برداری از جامعه تشکیل می‌شود. سپس میانه هرکدام محاسبه شده و سپس از یکی از روش‌های تشکیل بازه اطمینان استفاده می‌کنیم.

مثلاً می‌توانیم میانه‌ها را مرتب کرده و از ۱۰۰۰ تا B ۲۵ تای اول و آخر را حذف کرده و در نتیجه یک بازه اطمینان ۹۵ درصدی تشکیل دهیم. یا از نمونه‌های bootstrap تخمین میانگین از جامعه و SD را محاسبه کرده و با کمک فرمول زیر یک بازه اطمینان برای پارامتر مورد نظر که در اینجا میانه است تشکیل می‌دهیم:

$$\hat{\theta} = Z_{\frac{\alpha}{2}}\widehat{SE}^{*}$$

۱۹-۳_ پاسخ قسمت ۳

مزیت این روش این است که وابسته به این نیست که توزیع داده‌های اصلی مشخص شده باشد و می‌توان برای داده‌هایی که توزیع آن مشخص نیست و نرمال نبوده و یا توزیع پیچیده است نیز استفاده شود. همچنین برای شرایطی که اندازه نمونه کوچک است نیز روش Bootstrap کارایی دارد و می‌توان از روی نمونه کوچک، نمونه‌های Bootstrap تولید کرده و محاسبات خود را انجام دهیم.

۱۹-۴_ پاسخ قسمت ۴

خیر. روش Bootstrap همیشه به‌عنوان بهترین روش برای تخمین پارامترها نیست. اگر که ما توزیع داده‌ها را بدانیم و از روش‌های پارامتریک استفاده کنیم، قطعاً بادقت بالاتری می‌توان پارامترهای توزیع را تخمین زد. همچنین محاسبات این روش نسبت به روش‌های دیگر ممکن است بیشتر باشد و به‌طورکلی بسته به شرایط مسئله روش موردنظر را انتخاب می‌کنیم.

۱۹-۵_ پاسخ قسمت ۵

bootstrap statistics از روی نمونه‌های تشکیل شده bootstrap به دست می‌آیند که مراحل تشکیل آن در قسمت‌های قبلی این مسئله توضیح داده شده است. پس از نمونه‌برداری با جایگزینی از داده‌های مشاهده‌شده، و محاسبه آماره (مثلاً میانگین) برای هر نمونه و تکرار آن به تعداد خاصی مثلاً ۱۰۰۰ توزیع

آماره‌های بدست آمده را تشکیل داده و با کمک این توزیع می‌توانیم برای مثال میانگین جامعه را تخمین بزنیم و یا برای پارامتر مورد نظر بازه اطمینان تشکیل دهیم.

۲۰_ پاسخ مسئله شماره ۲۰

در این پرسش برنامه‌نویسی در نوت‌بوک انجام‌شده و در فایل پیوست قرار گرفته است. باتوجه‌به مقادیر مسئله با کمک تابع stats.wilcoxon مقدار test_statistic محاسبه شده و سپس با تقریب نرمال مقدار power تست اندازه گرفته شده است.

در قسمت اول باتوجه pvalue=.99 فرض صفر رد نمی‌شود.

باتوجه‌به نتایج خروجی و پاور نزدیک به ۱، پس دو توزیع فاصله زیادی از هم دارند و وقتی فرض صفر رد می‌شود به احتمال بالایی فرض مقابل درست است.

وقتی که مقدار b=.4 بود power کمتری نسبت به حالتی که برابر 0.6 باشد دارد. می‌توان نتیجه گرفت که مقادیر در قسمت سوم که b=0.6 بود به میانه نزدیک‌تر هستند. پس سخت‌تر فرض صفر رد می‌شود.

۲۱_ پاسخ مسئله شماره ۲۱

در این پرسش انحراف معیار دو گروه مختلف مقایسه شده است پس تست از نوع two sample است و باتوجه‌به اینکه هردو توزیع مشخص هستند پس می‌توان از تست parametric استفاده کرد. کافی است که از تستی حول واریانس مثل تست F برای مقایسه انحراف معیارها استفاده شود تا مشخص شود آیا دو انحراف معیار به‌صورت تصادفی برابر هم هستند یا اینکه می‌توانیم باقدرت این احتمال را رد کنیم. باتوجه‌به اینکه تفاضل واریانس بی‌معنی است باید نسبت واریانس دو توزیع مشخص شود. اگر که این نسبت برابر ۱ باشد؛ یعنی این پارامتر در هر دوبرابر است. تست‌های مقایسه واریانس نیز به نرمال وابسته هستند پس‌شرط مسئله برقرار است؛ زیرا که هردو توزیع نرمال هستند.

باتوجه‌به شکل توزیع F، تست باید برای هر دو طرف انجام شود و به طور مجرا تست یک‌طرفه است.

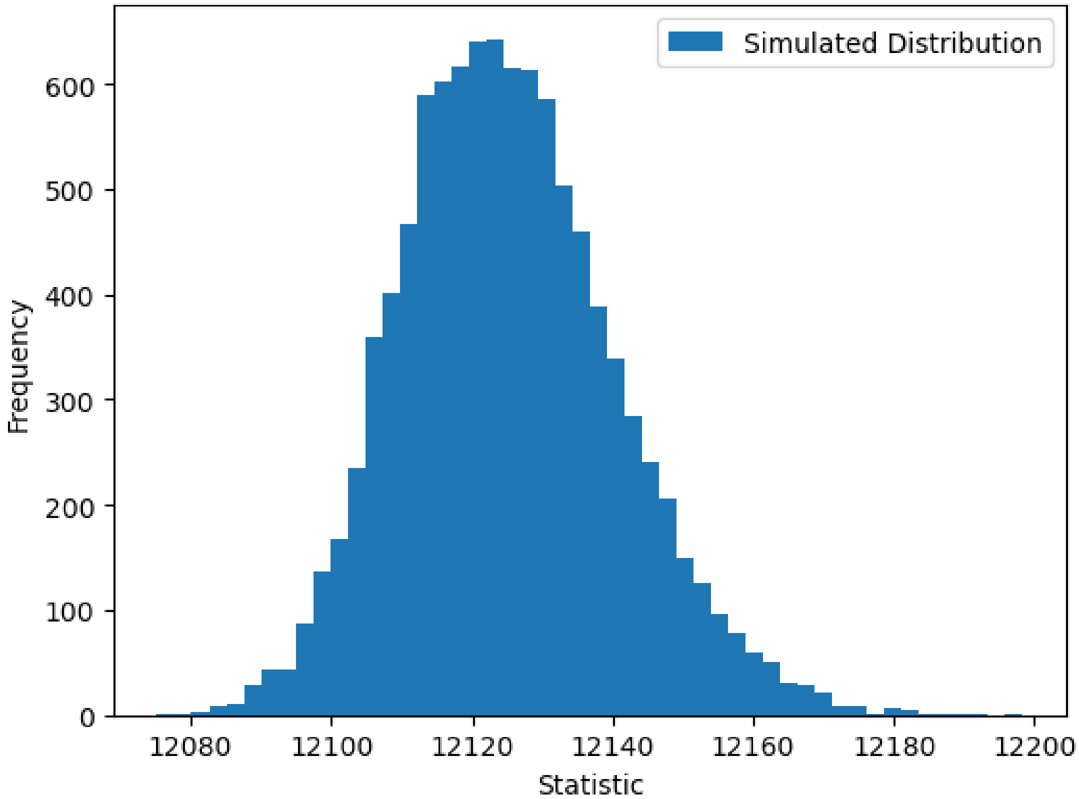
برای حل این سؤال ابتدا دو نمونه ۱۰۰۰تایی نرمال با واریانس ۳ و ۱۰ تولید کرده و سپس واریانس هر دو به صورت unbiased محاسبه می‌شود. سپس باتوجه به اینکه اندازه هر دو نمونه تعداد برابری دارند پس درجه آزادی هردو برابر است پس نسبت Y به X را محاسبه کرده و توزیع F تشکیل می‌شود. در این سوال مقدار آلفا را برابر 0.05 در نظر گرفته و نتیجه را در خروجی چاپ کرده‌ایم. در این مسئله فرض صفر رد می‌شود و نتیجه می‌شود که انحراف معیارها تفاوت معناداری با یکدیگر دارند.

۲۲_ پاسخ مسئله شماره ۲۲

برای حل این مسئله که یک جدول به همراه فراوانی ارائه شده و فرض صفر به‌عنوان یک توزیع بیان می‌شود از تست *goodness of fit* استفاده می‌شود. در این تست هدف این است که متوجه شویم داده‌های مشاهده با فراوانی مذکور با داده‌های مورد انتظار از توزیع فرض صفر پیروی می‌کنند یا خیر پس مقدار تفاضل این دو مقدار را با مقدار مورد انتظار نرمال‌سازی کرده و *test statistics* محاسبه می‌شود. این مقدار از توزیع خی دو با *n-1* درجه آزادی پیروی می‌کند که در اینجا درجه آزادی برابر با ۶۱۱۴ خواهد بود. مقدار آن محاسبه شده و در خروجی چاپ شد:

```
Chi-square Test Ttatistic=  249.19544266399507
```

در قسمت بعدی یک شبیه‌سازی برای آماره ارائه شده انجام شده و نمودار هیستوگرام آن نیز در زیر رسم شده است. تعداد شبیه‌سازی‌ها ۱۰۰۰ بوده و مشاهده می‌شود که توزیع آماره از نرمال پیروی می‌کند.



شکل ۹ توزیع شبیه‌سازی توزیع آماره

در قسمت بعدی با توجه به آماره بدست آمده در قسمت قبل که از توزیع خی دو با ۶۱۱۴ درجه آزادی است، مقدار p_value محاسبه می‌شود. درجه آزادی برابر با تعداد خانواده‌ها منهای یک است. پس از

مقدار cdf در آن ناحیه محاسبه شده و از ۱ کم می‌شود. زیرا که مقدار p_value احتمال نواحی بزرگ‌تر از *test statistics* است. مقدار p_value برابر ۱ بوده و نمی‌توان فرض صفر را رد کرد؛ زیرا که از 0.05 کمتر نیست.