Statistical Inference

Lecturer: Abdol-Hossein Vahabie

Spring Semester 1401-1402

Marzieh Alidadi_810101236 Writing Assignment III Deadline 1402/02/18

۱ درست یا غلط؟

۱-۱ زیربخش a

این گزاره درست است.

با توجه به این‌که توزیع نمونه مشخص نیست، افراد به صورت مستقل نمونه برداری شده و در گروه‌ها قرار گرفته‌اند، و همچنین سایز نمونه در مقابل تعداد کل افراد دارای فشار خون، کوچک است و برای بررسی مستقیم مناسب نیست، استفاده از توزیع bootstrap برای نمونه برداری دوباره از بیماران حاضر در هر یک از گروه‌ها، کار مناسبی است.

۱-۲ زیربخش b

این گزاره غلط است.

وجود تفاوت significant در میانگین فشار خون بیماران حاضر در دو گروه، لزوماً به دلیل بهتر بودن درمان استفاده شده در یکی از این دو گروه نیست. عوامل دیگری از جمله confounding variable ها و blocking variable ها نیز باید بررسی شود و تاثیر آن‌ها از آزمایش حذف شود و فقط تاثیر درمان‌های استفاده شده در دو گروه باقی بماند، تا بتوان در این مورد نتیجه‌گیری کرد.

۱-۳ زیربخش c

این گزاره غلط است.

اگر چه در بسیاری از موارد، افزایش سایز نمونه موجب بهبود آزمون T می‌شود؛ ولی

به طور کلی پارامترهایی برای تعیین سایز مناسب برای نمونه وجود دارد. گاهاً ممکن است با افزایش سایز نمونه، تنوع داده‌ها افزایش یابد. که این قضیه موجب کاهش دقت آزمون T می‌شود. همچنین، افزایش سایز نمونه موجب می‌شود تا تفاوت‌هایی به عنوان significant شناخته شوند، که واقعاً significant نیستند. به علاوه، بهتر است که سایز نمونه‌ها حداکثر برابر ۱۰ درصد تعداد کل جمعیت باشد.

۱-۴ زیربخش d

این گزاره درست است.
در آزمون T ای که به صورت paired انجام شود، برای هر یک از گروه‌ها، میانگین فشار خون آن‌ها قبل و بعد از انجام درمان محاسبه می‌شود. و فرض صفر در این آزمون، برابر خواهدبود با این‌که تفاوت میانگین فشار خون بیماران حاضر در هر یک از گروه‌ها قبل و بعد از درمان، برابر با صفر باشد. که این معادل این است که به طور کلی، میانگین فشار خون قبل و بعد از درمان برای تمام بیماران برابر صفر باشد.

۱-۵ زیربخش e

این گزاره درست است.
اگر تنوع در هر یک از گروه‌ها زیاد باشد، پراکندگی داده‌ها نیز زیاد خواهدبود. در این حالت، برای این‌که تفاوت میانگین فشار خون بیماران دو گروه، significant به حساب آید، لازم است این تفاوت میانگین، از پراکندگی داده‌های هر یک از این دو گروه بیشتر باشد. لزوم این بررسی، موجب می‌شود تا تشخیص تفاوت‌های significant در میانگین فشار خون مربوط به دو گروه، دشوارتر شود.

۱-۶ زیربخش f

این گزاره غلط است.
اگر تفاوت میانگین فشار خون بیماران قبل و بعد از درمان، در یکی از گروه‌ها، بیشتر از تفاوت میانگین فشار خون بیماران قبل و بعد از درمان، در گروه دیگر باشد، نمی‌توان لزوماً این نتیجه‌گیری را داشت که درمان استفاده شده در گروهِ با تفاوت میانگین

بیشتر، موثرتر است. عوامل احتمالی دیگری از جمله confounding variable ها و blocking variable ها نیز باید بررسی شود و تاثیر آن‌ها از آزمایش حذف شود و فقط تاثیر درمان‌های استفاده شده در دو گروه باقی بماند، تا بتوان در این مورد نتیجه‌گیری کرد. همچنین، باید از آزمون‌های آماری (مانند آزمون T)، برای بررسی significant بودن/نبودن تفاوت میانگین فشار خون بیماران در دو گروه، قبل و بعد از درمان، استفاده کرد. تا در صورتی که تفاوت، significant بود، بتوان نتیجه‌گیری کرد.

۷-۱ زیربخش g

این گزاره درست است.
confidence interval محدوده‌ی قابل قبولی از مقادیر برای پارامتر موردنظر را نشان می‌دهد. برای محاسبه‌ی آن، به standard error میانگین فشار خون بیماران، قبل و بعد از انجام درمان، نیاز داریم؛ تا اختلاف آن‌ها را حساب کنیم و در محاسبه‌ی فرمول confidence interval از آن استفاده کنیم.

۸-۱ زیربخش h

این گزاره غلط است.
قدرت آزمون فرضیه در این است که اگر فرض صفر غلط باشد، با احتمال بالایی به درستی رد شود. سطح significance، این را تعیین می‌کند که چه موقع فرض صفر رد شود یا رد نشود؛ به این صورت که اگر احتمال پایین‌تر از آن باشد، رد می‌شود، و اگر احتمال بالاتر از آن باشد، رد نمی‌شود. کاهش آن، موجب می‌شود که احتمال فرض صفر را با عدد کوچک‌تری مقایسه کنیم و در نتیجه با احتمال کم‌تری آن را رد کنیم. این موجب می‌شود تا احتمال رد کردن فرض صفر کاهش یابد و در نتیجه احتمال این‌که یک فرض صفر غلط را به اشتباه رد نکنیم، افزایش یابد. یعنی با افزایش سطح significance، احتمال رد نکردن یک فرض صفر غلط بیشتر می‌شود؛ که این یعنی خطای نوع دوم افزایش می‌یابد. و در نتیجه قدرت آزمون فرضیه کاهش می‌یابد.

۲ آزمون فرضیه

بررسی شرایط لازم برای انجام آزمون فرضیه:
نمونه به صورت تصادفی از جامعه آماری انتخاب شده‌است. و همچنین، با توجه به این‌که سایز نمونه برابر ۵۲ است، قطعاً از ۱۰ درصد تعداد افراد جامعه آماری کم‌تر است. بنابراین، شرط مستقل بودن افراد درون نمونه، برقرار است.
باید بررسی شود که یا توزیع افراد جامعه آماری به صورت نرمال باشد، یا اگر توزیع را نمی‌دانیم و یا دارای چولگی است، باید نمونه‌ی تهیه شده از جامعه، به اندازه‌ی کافی بزرگ باشد. این‌جا سایز نمونه از ۳۰ بزرگ‌تر است. پس این مسئله رعایت شده و سایز نمونه مناسب است.

۲-۱ زیربخش ۱

فرض صفر، این را بیان می‌کند که میانگین درآمد سالانه‌ی فارغ التحصیلان رشته‌های برق و کامپیوتر، برابر ۹۸.۶ هزار دلار است.

$$H_0 : \mu = \$98.6K \tag{۱}$$

فرض جایگزین، این را بیان می‌کند که میانگین درآمد سالانه‌ی فارغ التحصیلان رشته‌های برق و کامپیوتر، بیشتر از ۹۸.۶ هزار دلار است.

$$H_A : \mu > \$98.6K \tag{۲}$$

۲-۲ زیربخش ۲

با توجه به این‌که اطمینان ۹۸ درصد مدنظر است، از جدول آماره‌ی آزمون T برای زمانی که درجه آزادی برابر ۱-۵۲ است و مجموع دو سمت خارج از بازه اطمینان برابر ۲ درصد است، استفاده می‌شود:

$$t^*_{n-1}(Two-tails=0.02) = t^*_{52-1}(Two-tails=0.02) = t^*_{51}(Two-tails=0.02) = 2.403 \tag{۳}$$

در این حالت، بازه‌ی اطمینان به شرح زیر است:

$$\begin{aligned}
\bar{x} - t^*_{n-1} \cdot SE \leq CI \leq \bar{x} + t^*_{n-1} \cdot SE \\
\bar{x} - t^*_{n-1} \cdot (\frac{s}{\sqrt{n}}) \leq CI \leq \bar{x} + t^*_{n-1} \cdot (\frac{s}{\sqrt{n}}) \\
98.2932 - t^*_{51} \cdot (\frac{0.6865}{\sqrt{52}}) \leq CI \leq 98.2932 + t^*_{51} \cdot (\frac{0.6865}{\sqrt{52}}) \\
98.2932 - 2.403 * (\frac{0.6865}{7.2111}) \leq CI \leq 98.2932 + 2.403 * (\frac{0.6865}{7.2111}) \\
98.2932 - 2.403 * 0.0952 \leq CI \leq 98.2932 + 2.403 * 0.0952 \\
98.2932 - 0.2287 \leq CI \leq 98.2932 + 0.2287 \\
98.0645 \leq CI \leq 98.5219
\end{aligned} \tag{۴}$$

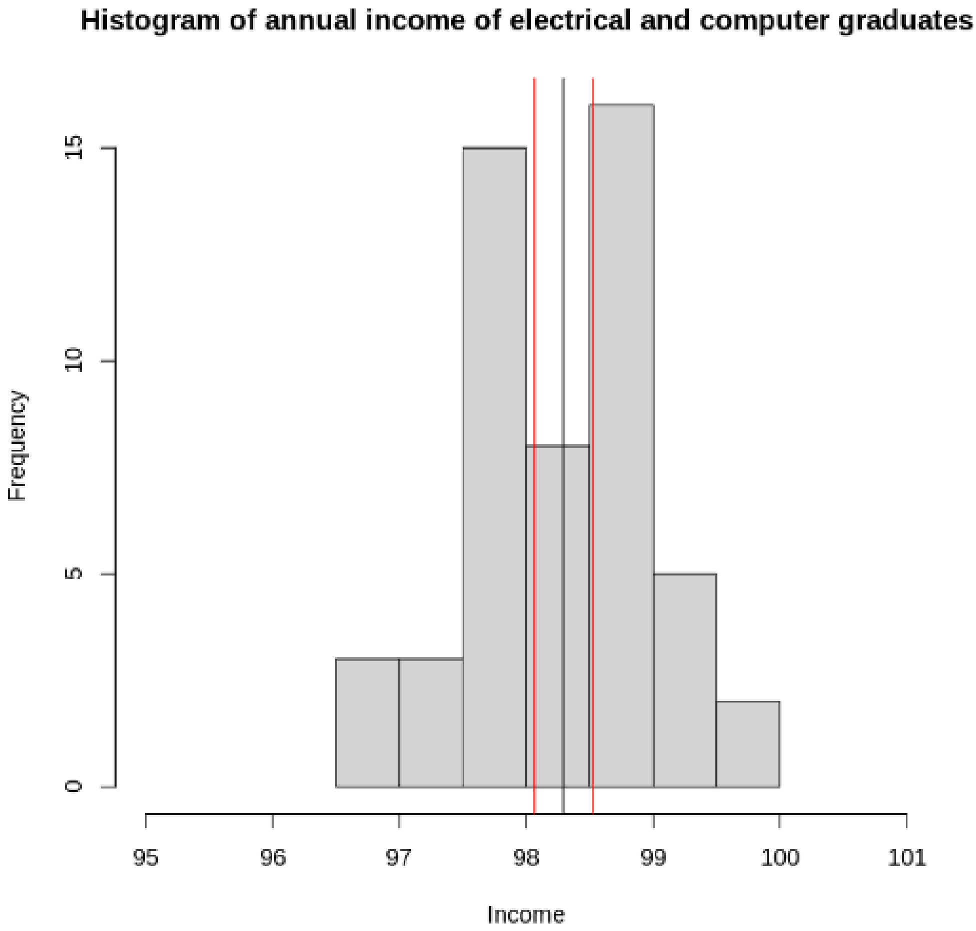
بنابراین، بازه‌ی اطمینان ۹۸ درصد عبارت است از:

$$[98.0645, 98.5219] \tag{۵}$$

این یعنی با احتمال ۹۸ درصد، میانگین درآمد سالانه‌ی فارغ التحصیلان رشته‌های برق و کامپیوتر، در بازه‌ی محاسبه شده در بالا، قرار دارد. به عبارتی، ۹۸ درصد نمونه‌هایی که با سایز ۵۲ از جامعه آماری ساخته می‌شوند، میانگین جامعه آماری را در بر می‌گیرند.

۲-۳ زیربخش ۳

نمودار میله‌ای خواسته شده، به شکل زیر رسم شده و کد R مربوط به آن ضمیمه شده‌است. میانگین درآمد با خط عمودی قرمز نمایش داده شده است و حد بالا و پایین بازه‌ی اطمینان ۹۸ درصد نیز با دو خط عمودی سیاه مشخص شده‌اند:



۴-۲ زیربخش ۴

برای این حالت، دوباره فرض صفر و فرض جایگزین را تعریف می‌کنیم:
فرض صفر، این را بیان می‌کند که میانگین درآمد سالانه‌ی فارغ التحصیلان رشته‌های برق و کامپیوتر، برابر ۹۸.۶ هزار دلار است.

$$H_0 : \mu = \$98.6K \tag{۶}$$

فرض جایگزین، این را بیان می‌کند که میانگین درآمد سالانه‌ی فارغ التحصیلان رشته‌های برق و کامپیوتر، برابر ۹۸.۶ هزار دلار نیست.

$$H_A : \mu \neq \$98.6K \tag{۷}$$

با توجه به این‌که سایز نمونه از ۳۰ بیشتر است و توزیع نمونه مشخص نیست، می‌توان از آزمون T استفاده کرد. آماره‌ی T به شرح زیر محاسبه می‌شود:

$$T = \frac{observation - null}{SE} = \frac{observation - null}{\frac{s}{\sqrt{n}}} = \frac{98.2932 - 98.6}{\frac{0.6865}{\sqrt{52}}} = \frac{-0.3068}{\frac{0.6865}{7.2111}} = \frac{-0.3068}{0.0952} = -3.2226 \tag{۸}$$

درجه‌ی آزادی در این آزمون:

$$df = n - 1 = 52 - 1 = 51 \tag{۹}$$

با توجه به این‌که می‌دانیم درجه آزادی برابر ۵۱ است و قدرمطلق مقدار آماره‌ی T برابر ۳.۲۲۲۶ است، محدوده‌ی مقدار p-value را با استفاده از جدول آماره‌ی T بدست می‌آوریم. مقدار p-value در حالت تست دو طرفه، از ۰.۰۰۱ کم‌تر است. با توجه به این‌که سطح significance برابر ۰.۰۲ است، و مقدار p-value از آن کم‌تر است، پس فرض صفر رد می‌شود. و مشخص می‌شود که نمی‌توان ادعا کرد که میانگین درآمد سالانه‌ی فارغ التحصیلان رشته‌های برق و کامپیوتر، برابر ۹۸.۶ است.

۳ آزمون تحلیل واریانس (ANOVA)

فرض صفر، این را بیان می‌کند که میانگین زمان سفر خودروهای خودران آموزش دیده با استفاده از این الگوریتم‌ها، با هم برابر است.

$$H_0 : \mu_{Q-learning} = \mu_{Actor-Critic} = \mu_{DQN} \tag{۱۰}$$

فرض جایگزین، این را بیان می‌کند که میانگین زمان سفر خودروهای خودران آموزش دیده با استفاده از این الگوریتم‌ها، حداقل برای ۲ تا از الگوریتم‌ها با هم برابر نیست.

$$\begin{aligned} H_A : \mu_{Q-learning} &\neq \mu_{Actor-Critic} \\ or \\ \mu_{Q-learning} &\neq \mu_{DQN} \\ or \\ \mu_{Actor-Critic} &\neq \mu_{DQN} \end{aligned} \tag{۱۱}$$

برای انجام آزمون ANOVA باید ابتدا یک سری پارامتر محاسبه شود:
میانگین زمان سفر خودروهای خودران آموزش دیده با استفاده از الگوریتم Q-learning:

$$\bar{y}_{Q-learning} = \frac{1000+1100+700+800+500+700}{6} = \frac{4800}{6} = 800 \tag{۱۲}$$

میانگین زمان سفر خودروهای خودران آموزش دیده با استفاده از الگوریتم Actor-Critic:

$$\bar{y}_{Actor-Critic} = \frac{890+650+1100+900+400+350}{6} = \frac{4290}{6} = 715 \tag{۱۳}$$

میانگین زمان سفر خودروهای خودران آموزش دیده با استفاده از الگوریتم DQN:

$$\bar{y}_{DQN} = \frac{1200+1000+980+900+750+800}{6} = \frac{5630}{6} = 938.33 \tag{۱۴}$$

میانگین زمان سفر خودروهای خودران آموزش دیده با استفاده از هر سه الگوریتم:

$$\bar{y} = \frac{14720}{18} = 817.778 \tag{۱۵}$$

Sum of Squares Total که تنوع کلی داده‌ها را اندازه‌گیری می‌کند:

$$
\begin{aligned}
SST &= \sum_{i=1}^{18}(y_i - \bar{y})^2 \\
&= (1000 - 817.778)^2 + (1100 - 817.778)^2 + (700 - 817.778)^2 \\
&+(800 - 817.778)^2 + (500 - 817.778)^2 + (700 - 817.778)^2 \\
&+(890 - 817.778)^2 + (650 - 817.778)^2 + (1100 - 817.778)^2 \\
&+(900 - 817.778)^2 + (400 - 817.778)^2 + (350 - 817.778)^2 \\
&+(1200 - 817.778)^2 + (1000 - 817.778)^2 + (980 - 817.778)^2 \\
&+(900 - 817.778)^2 + (750 - 817.778)^2 + (800 - 817.778)^2 \\
&= (182.222)^2 + (282.222)^2 + (-117.778)^2 \\
&+(-17.778)^2 + (-317.778)^2 + (-117.778)^2 \\
&+(72.222)^2 + (-167.778)^2 + (282.222)^2 \\
&+(82.222)^2 + (-417.778)^2 + (-467.778)^2 \\
&+(382.222)^2 + (182.222)^2 + (162.222)^2 \\
&+(82.222)^2 + (-67.778)^2 + (-17.778)^2 \\
&= 33204.857 + 79649.257 + 13871.657 + 316.057 \\
&+100982.857 + 13871.657 + 5216.017 + 28149.457 \\
&+79649.257 + 6760.457 + 174538.457 + 218816.257 \\
&+146093.657 + 33204.857 + 26315.977 + 6760.457 \\
&+4593.857 + 316.057 = 972311.106
\end{aligned} \tag{۱۶}
$$

Sum of Squares Groups که تنوع بین گروهی را اندازه‌گیری می‌کند:

$$\begin{aligned} SSG &= \sum_{j=1}^{3} n_j(\bar{y}_j - \bar{y})^2 \\ &= 6(800 - 817.778)^2 + 6(715 - 817.778)^2 + 6(938.33 - 817.778)^2 \\ &= 6(-17.778)^2 + 6(-102.778)^2 + 6(120.552)^2 \\ &= 6(316.057) + 6(10563.317) + 6(14532.784) \\ &= 1896.342 + 63379.902 + 87196.704 = 152472.948 \end{aligned} \tag{۱۷}$$

:Sum of Squares Error

$$SSE = SST - SSG = 972311.106 - 152472.948 = 819838.158 \tag{۱۸}$$

درجه آزادی کل:

$$df_T = n - 1 = 18 - 1 = 17 \tag{۱۹}$$

درجه آزادی گروه:

$$df_G = K - 1 = 3 - 1 = 2 \tag{۲۰}$$

درجه آزادی خطا:

$$df_E = df_T - df_G = 17 - 2 = 15 \tag{۲۱}$$

mean squares گروه:

$$MSG = \frac{SSG}{df_G} = \frac{152472.948}{2} = 76236.474 \tag{۲۲}$$

mean squares خطا:

$$MSE = \frac{SSE}{df_E} = \frac{819838.158}{15} = 54655.8772 \qquad (۲۳)$$

مقدار آماره‌ی F:

$$F = \frac{MSG}{MSE} = \frac{76236.474}{54655.8772} = 1.394 \qquad (۲۴)$$

جدول ANOVA به فرم زیر است:

		DF	Sum sq	Mean sq	F value	Pr (>F)
Group	Class	DFG=k-۱	SSG	MSG=SSG/DFG	MSG/MSE	؟
Error	Residuals	DFE=DFT-DFG	SSE	MSE=SSE/DFE		
	Total	DFT=n-۱	SST			

با جایگذاری به فرم زیر حاصل می‌شود:

		DF	Sum sq	Mean sq	F value	Pr (>F)
Group	Class	۲	۱۵۲۴۷۲.۹۴۸	۷۶۲۳۶.۴۷۴	۱.۳۹۴	۰.۲۷۸
Error	Residuals	۱۵	۸۱۹۸۳۸.۱۵۸	۵۴۶۵۵.۸۷۷۲		
	Total	۱۷	۹۷۲۳۱۱.۱۰۶			

به کمک یک ماشین حساب آنلاین مربوط به آماره‌ی F، برای توزیع F با درجه آزادی‌های ۲ و ۱۵، احتمال این‌که مقدارش از ۱.۳۹۴ بیشتر شود، برابر ۰.۲۷۸ بدست آمد. با توجه به این‌که این مقدار، از سطح significance که برابر ۰.۰۵ است، بیشتر است، نمی‌توان فرض صفر را رد کرد. و نمی‌توان ادعا کرد که هیچ تفاوت significant ای بین میانگین زمان سفر خودروهای خودران آموزش دیده با استفاده از این الگوریتم‌ها وجود ندارد.

۴ آزمون فرضیه

۴-۱ زیربخش ۱

با توجه به این‌که قصد داریم ارتباط بین دو متغیر را بدست آوریم، می‌توان از آزمون chi-square استفاده کرد.
فرض صفر و فرض جایگزین به شرح زیر تعریف می‌شوند:

- فرض صفر، این را بیان می‌کند که ارتباطی بین زمان روز و نوع نوشیدنی وجود ندارد.
- فرض جایگزین، این را بیان می‌کند که که ارتباطی بین زمان روز و نوع نوشیدنی وجود دارد.

برای محاسبه‌ی chi-square، به expected value مربوط به هر یک از خانه‌های جدول نیاز داریم:

$$E[morning][coffee] = \frac{120 * 165}{330} = \frac{19800}{330} = 60 \tag{۲۵}$$

$$E[morning][tea] = \frac{120 * 90}{330} = \frac{10800}{330} = 32.73 \tag{۲۶}$$

$$E[morning][smoothie] = \frac{120 * 75}{330} = \frac{9000}{330} = 27.27 \tag{۲۷}$$

$$E[afternoon][coffee] = \frac{120 * 165}{330} = \frac{19800}{330} = 60 \tag{۲۸}$$

$$E[afternoon][tea] = \frac{120 * 90}{330} = \frac{10800}{330} = 32.73 \tag{۲۹}$$

$$E[afternoon][smoothie] = \frac{120 * 75}{330} = \frac{9000}{330} = 27.27 \tag{۳۰}$$

$$E[evening][coffee] = \frac{90 * 165}{330} = \frac{14850}{330} = 45 \tag{۳۱}$$

$$E[evening][tea] = \frac{90 * 90}{330} = \frac{8100}{330} = 24.55 \tag{۳۲}$$

$$E[evening][smoothie] = \frac{90 * 75}{330} = \frac{6750}{330} = 20.45 \tag{۳۳}$$

expected value ها به جدول اضافه می‌شوند:

	coffee	tea	smoothie
morning	۷۵ (۶۰)	۳۰ (۳۲.۷۳)	۱۵ (۲۷.۲۷)
afternoon	۶۰ (۶۰)	۴۵ (۳۲.۷۳)	۱۵ (۲۷.۲۷)
evening	۳۰ (۴۵)	۱۵ (۲۴.۵۵)	۴۵ (۲۰.۴۵)

مقدار chi-square به شکل زیر محاسبه می‌شود:

$$\begin{aligned}
\chi^2 = \sum \frac{(observed - expected)^2}{expected} &= \frac{(75-60)^2}{60} + \frac{(30-32.73)^2}{32.73} + \frac{(15-27.27)^2}{27.27} \\
&+ \frac{(60-60)^2}{60} + \frac{(45-32.73)^2}{32.73} + \frac{(15-27.27)^2}{27.27} + \frac{(30-45)^2}{45} \\
&+ \frac{(15-24.55)^2}{24.55} + \frac{(45-20.45)^2}{20.45} = \frac{(15)^2}{60} + \frac{(-2.73)^2}{32.73} + \frac{(-12.27)^2}{27.27} \\
&+ \frac{(0)^2}{60} + \frac{(12.27)^2}{32.73} + \frac{(-12.27)^2}{27.27} + \frac{(-15)^2}{45} + \frac{(-9.55)^2}{24.55} + \frac{(24.55)^2}{20.45} \\
&= \frac{225}{60} + \frac{7.45}{32.73} + \frac{150.55}{27.27} + \frac{0}{60} + \frac{150.55}{32.73} + \frac{150.55}{27.27} + \frac{225}{45} + \frac{91.2}{24.55} + \frac{602.7}{20.45} \\
&= 3.75 + 0.23 + 5.52 + 0 + 4.60 + 5.52 + 5 + 3.71 + 29.47 = 57.8
\end{aligned} \tag{۳۴}$$

درجه آزادی به شرح زیر است:

$$\begin{aligned}
DegreeOfFreedom &= (NumberOfRows - 1) * (NumberOfColumns - 1) \\
&= (3-1) * (3-1) = 4
\end{aligned} \tag{۳۵}$$

مقدار بحرانی برای توزیع chi-square با درجه آزادی ۴ و سطح significance برابر ۰.۰۵، با توجه به جدول این توزیع، برابر ۹.۴۸۸ است. با توجه به این‌که مقدار chi-square در این سوال برابر با ۵۷.۸ به دست آمد، و این مقدار از ۹.۴۸۸ بیشتر است، فرض صفر رد می‌شود. و نمی‌توان ادعا کرد که ارتباطی بین زمان روز و نوع نوشیدنی وجود ندارد.

۴-۲ زیربخش ۲

۴-۲-۱ (a

فرض صفر، این را بیان می‌کند که میانگین کالری دریافتی روزانه افرادی که به طور منظم در رستوران‌های فست فود غذا می خورند، ۲۵۰۰ کالری در روز است.

$$H_0 : \mu = 2500 \tag{۳۶}$$

فرض جایگزین، این را بیان می‌کند که میانگین کالری دریافتی روزانه افرادی که به طور منظم در رستوران‌های فست فود غذا می خورند، ۲۵۰۰ کالری در روز نیست.

$$H_A : \mu \neq 2500 \qquad (۳۷)$$

۴-۲-۲ (b

بررسی شرایط لازم برای انجام آزمون فرضیه:
نمونه به صورت تصادفی از جامعه آماری انتخاب شده‌است. و همچنین، با توجه به این‌که سایز نمونه برابر ۵۰ است، قطعاً از ۱۰ درصد تعداد افراد جامعه آماری کم‌تر است. بنابراین، شرط مستقل بودن افراد درون نمونه، برقرار است.
باید بررسی شود که یا توزیع افراد جامعه آماری به صورت نرمال باشد، یا اگر توزیع را نمی‌دانیم و یا دارای چولگی است، باید نمونه‌ی تهیه شده از جامعه، به اندازه‌ی کافی بزرگ باشد. این‌جا سایز نمونه از ۳۰ بزرگ‌تر است. پس این مسئله رعایت شده و سایز نمونه مناسب است.
آماره‌ی T به شرح زیر محاسبه می‌شود:

$$\begin{aligned} T = \frac{observation - null}{SE} &= \frac{observation - null}{\frac{s}{\sqrt{n}}} = \frac{2100 - 2500}{\frac{500}{\sqrt{50}}} \\ &= \frac{-400}{\frac{500}{7.07}} = \frac{-400}{70.72} = -5.66 \end{aligned} \qquad (۳۸)$$

درجه‌ی آزادی در این آزمون:

$$df = n - 1 = 50 - 1 = 49 \qquad (۳۹)$$

۴-۲-۳ (c

p-value، احتمال به دست آوردن مقدار مشاهده شده در نمونه برای میانگین کالری دریافتی روزانه افرادی که به طور منظم در رستوران‌های فست فود غذا می‌خورند، یا بیشتر از این مقدار، به شرط این‌که فرض صفر صحیح باشد، را بیان می‌کند. در واقع در این سوال، p-value، احتمال به دست آوردن مقدار ۲۱۰۰ یا بیشتر برای میانگین

کالری دریافتی روزانه افرادی که به طور منظم در رستوران‌های فست فود غذا می‌خورند را، به شرط این‌که میانگین کالری دریافتی روزانه افرادی که به طور منظم در رستورانهای فست فود غذا می خورند، ۲۵۰۰ کالری در روز باشد، بیان می‌کند.
با توجه به این‌که می‌دانیم درجه آزادی برابر ۴۹ است و قدرمطلق مقدار آماره‌ی T برابر ۵.۶۶ است، محدوده‌ی مقدار p-value را با استفاده از جدول آماره‌ی T بدست می‌آوریم. مقدار p-value در حالت تست دو طرفه، از ۰.۰۰۱ کم‌تر است. با توجه به این‌که سطح significance برابر ۰.۰۵ است، و مقدار p-value از آن کم‌تر است، پس فرض صفر رد می‌شود. و مشخص می‌شود که نمی‌توان ادعا کرد که میانگین کالری دریافتی روزانه افرادی که به طور منظم در رستوران‌های فست فود غذا می خورند، ۲۵۰۰ کالری در روز است.

۴-۲-۴ (d

همانطور که در بخش قبل بیان شد، p-value در حالت تست دو طرفه، از ۰.۰۰۱ کم‌تر است. با توجه به این‌که سطح significance برابر ۰.۰۵ است، و مقدار p-value از آن کم‌تر است، پس فرض صفر رد می‌شود. و مشخص می‌شود که نمی‌توان ادعا کرد که میانگین کالری دریافتی روزانه افرادی که به طور منظم در رستوران‌های فست فود غذا می خورند، ۲۵۰۰ کالری در روز است.

۵-۲-۴ (e

از جدول آماره‌ی آزمون T برای زمانی که درجه آزادی برابر ۴۹ است و مجموع دو سمت خارج از بازه اطمینان برابر ۵ درصد است، استفاده می‌شود:

$$t^*_{n-1}(Two-tails=0.05)=t^*_{50-1}(Two-tails=0.05)=t^*_{49}(Two-tails=0.05)=2.009 \quad (۴۰)$$

در این حالت، بازه‌ی اطمینان ۹۵ درصد به شرح زیر است:

$$\begin{aligned}
\bar{x} - t^*_{n-1} \cdot SE \leq CI \leq \bar{x} + t^*_{n-1} \cdot SE \\
\bar{x} - t^*_{n-1} \cdot (\frac{s}{\sqrt{n}}) \leq CI \leq \bar{x} + t^*_{n-1} \cdot (\frac{s}{\sqrt{n}}) \\
2100 - t^*_{49} \cdot (\frac{500}{\sqrt{50}}) \leq CI \leq 2100 + t^*_{49} \cdot (\frac{500}{\sqrt{50}}) \\
2100 - 2.009 * (\frac{500}{7.07}) \leq CI \leq 2100 + 2.009 * (\frac{500}{7.07}) \\
2100 - 2.009 * 70.72 \leq CI \leq 2100 + 2.009 * 70.72 \\
2100 - 142.07 \leq CI \leq 2100 + 142.07 \\
1957.93 \leq CI \leq 2242.07
\end{aligned} \tag{۴۱}$$

بنابراین، بازه‌ی اطمینان ۹۵ درصد عبارت است از:

$$[1957.93, 2242.07] \tag{۴۲}$$

این یعنی با احتمال ۹۵ درصد، میانگین کالری دریافتی روزانه افرادی که به طور منظم در رستوران‌های فست فود غذا می خورند، در بازه‌ی محاسبه شده در بالا، قرار دارد. عدد ۲۵۰۰ در این بازه قرار ندارد. در نتیجه، با احتمال ۹۵ درصد، میانگین واقعی جامعه آماری برای کالری دریافتی روزانه افرادی که به طور منظم در رستوران‌های فست فود غذا می‌خورند، برابر ۲۵۰۰ نخواهد بود. یعنی این اتفاق که این میانگین برای جامعه آماری برابر ۲۵۰۰ شود، بسیار غیرمحتمل است.

۵ برنامه‌نویسی R

(از میان دو سوال ۵ و ۶، سوال ۶ پاسخ داده شده است.)

۶ برنامه‌نویسی R

(از میان دو سوال ۵ و ۶، سوال ۶ پاسخ داده شده است.)
با استفاده از یک مدل deep learning که با استفاده از train set مربوط به mnist آموزش داده شده، برچسب ۵۰۰ نمونه از عکس‌های مربوط به test set مربوط به mnist پیشبینی شد. و دقت عملکرد این مدل با استفاده از تست chi-squared ارزیابی شد. مقدار آماره‌ی chi-squared برابر ۱۰.۹۶ بدست آمد. با توجه به این مقدار، و درجه آزادی آن که برابر ۹ است، p-value برابر ۰.۲۷۸۵ است. که این مقدار، از ۰.۰۵ کوچک‌تر است. همچنین با توجه به مقدار درجه آزادی که برابر ۹ است، و سطح significance که برابر ۰.۰۵ است، مقدار بحرانی توزیع chi-squared برابر ۱۶.۹۱۸ است. و عدد آماره که برابر ۱۰.۹۶ بود، از آن کوچک‌تر است.
با توجه به این دو بررسی، می‌توان فهمید که عملکرد این مدل بر روی داده‌های جدید، خیلی تفاوتی با حالت شانسی ندارد. و این مدل، دقت خوبی در پیشبینی اعداد درون تصاویر جدید نداشته است.
(کد مربوطه ضمیمه شده‌است.)