سؤال اول) a) غلط : با افزایش تعداد گروه‌ها خطای نوع اول افزایش می‌یابد و سطح معناداری اصلاح شده باید کاهش یابد

b) درست، با توجه به وابستگی درجه آزادی در Anova به تعداد I، افزایش می‌یابد

غلط c) توزیع F مقادیر مثبت دارد و توزیع متمایل به راست است، چون میانگین مربع مقادیر مثبت (غلط)

d) غلط : با استفاده از آزمون Anova می‌توان نتیجه گرفت که حداقل یکی از میانگین‌ها متفاوت است اما نمی‌توان نتیجه گرفت کدام متفاوت هستند

e) درست، اگر فرض اولیه رد شود تغییرات بین گروه‌ها بیشتر است از تغییرات درون گروه‌ها.

سؤال سوم) a) H_0 : هیچ تفاوتی بین میانگین‌های ۳ درمان وجود ندارد $\mu_1 = \mu_2 = \mu_3$

H_1 : حداقل یکی از میانگین‌های درمان‌ها از دیگران متفاوت باشد

b) نتیجه آزمون : مقدار p-value برای آزمون Anova برابر است با ۰٫۰۴۶۱ که با توجه به اینکه از سطح معنادار ۰٫۰۵ کمتر است در نتیجه فرض صفر رد می‌شود.

c) $Tr۱ , Tr۲ :$ $d = ۶٫۲۱ - ۲٫۸۶ = ۳٫۳۵$ $SE = \sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}} = \sqrt{\frac{(۱۲٫۳)^2}{۱۴} + \frac{(۷٫۹۴)^2}{۱۴}} = \sqrt{\frac{۱۵۱٫۲۹}{۱۴} + \frac{۶۳٫۰۴}{۱۴}} = \sqrt{۱۵٫۳} \approx ۳٫۹۱$

$t = \frac{\bar{d}}{SE} = \frac{۳٫۳۵}{۳٫۹۱} = ۰٫۸۶$

با توجه به اینکه طبق جدول t با درجه‌ی آزادی ۲۶=۱۴+۱۴-۲ داریم ۲٫۰۵۶ و این مقدار بیشتر است از ۰٫۸۶ در نتیجه تفاوت معناداری بین این ۲ گروه نیست

$Tr۱ , Tr۳$ $d = ۶٫۲۱ - (-۳٫۲۱) = ۹٫۴۲$ $SE = \sqrt{\frac{(۱۲٫۳)^2}{۱۴} + \frac{(۸٫۵۷)^2}{۱۴}} = \sqrt{۱۶٫۰۴} \approx ۴$ $t = \frac{۹٫۴۲}{۴} = ۲٫۳۶$

با توجه به مقدار t طبق جدول با $\alpha = ۰٫۰۵$ و $df = ۲۶$ داریم ۲٫۰۵۶ و این مقدار t بدست آمده بیشتر از این ۲٫۰۵۶ است در نتیجه تفاوت معناداری بین این ۲ گروه است

$Tr۲ , Tr۳ :$ $d = ۲٫۸۶ - (-۳٫۲۱) = ۶٫۰۷$ $SE = \sqrt{\frac{۷٫۹۴^2}{۱۴} + \frac{(۸٫۵۷)^2}{۱۴}} = \sqrt{۹٫۷۵} \approx ۳٫۱۲$ $t = \frac{۶٫۰۷}{۳٫۱۲} \simeq ۱٫۹۵$

با توجه به مقدار t در جدول با $\alpha = ۰٫۰۵$ و $df = ۲۶$ مقدار $t_\alpha = ۲٫۰۵۶$ و با توجه به مقدار بدست آمده طبق فرمول ۱٫۹۵ که کمتر است در نتیجه فرض ~~صفر~~ ~~رد نمی‌شود~~ و تفاوت معناداری بین این ۲ گروه نداریم

سؤال چهارم : a) این مطالعه در واقع یک مطالعه تجربی است زیرا در آن از درمان واقعی و درمان با دارونما استفاده شده است در واقع در قالب گروه‌های درمانی، کنترل برای مقایسه نتایج استفاده می‌شود.

b) درمان تجربی : مصرف ۲۵ گرم دانه چیا ۲ بار در روز درمان کنترل : استفاده از دارونما (داروی جعلی)

c) برای کنترل اثر متغیرها از بلوک‌بندی استفاده شده است و متغیر بلوک در اینجا جنسیت (زن و مرد) می‌باشد که برای جلوگیری از تاثیرات جنسیتی، مردان و زنان جدا بلوک می‌شوند

d) بله چون شرکت کنندگان نمی‌دانند دارو را مصرف می‌کنند یا دارونما

e) توی سؤال گفته نشده محققان می‌دانند یا خیر اگر محققان ندانند بله از double blinding استفاده شده است در غیر این صورت خیر

f) با توجه به اینکه هر دو گروه از درمان کنترل و تجربی استفاده شده، در واقع sample assigment داشتیم می‌توانیم نتیجه‌گیری causal داشته باشیم. منظور از جذب مرد و زن به صورت تصادفی بوده با ما که هم نمی‌توان در این به منظور از جذب کننده شده انگاری انجام نمود

سوال هشتم: a) تحت فرض صفر که می‌گوید تابع چگالی f متقارن حول صفر است در نتیجه می‌توان گفت مقادیر $|D_i|$ احتمال‌های مشابهی برای هر i ایجاد می‌کنند و در واقع مقادیر D_i به طور متقارن حول صفر توزیع شده‌اند در نتیجه علامت D_i با احتمال $\frac{1}{2}$ مثبت و با احتمال $\frac{1}{2}$ منفی است و در نتیجه این توزیع آماره T وابسته به علامت D_i ها بوده و این توزیع یک توزیع جایگشتی از تمامی ترکیب‌های ممکن علامت D_i ها می‌باشد، تعداد حالات ممکن 2^n، هر ترکیب با احتمال 2^{-n} رخ می‌دهد

b) شبیه‌سازی با استفاده از مونت کارلو:

1. مقدار T را برای داده‌های اصلی محاسبه می‌کنیم 2. به صورت تصادفی علامت D_i ها را انتخاب کنیم 3. مقدار T را برای هر ترکیب علامتی محاسبه می‌کنیم 4. از توزیع تجربی به دست آمده از شبیه‌سازی‌ها، α را به دست می‌آوریم 5. مقایسه با مقدار بحرانی

c) اگر $D_i = X_i - Y_i$ باشد فرض صفر به این صورت است: توزیع D_i متقارن حول صفر است

در شبیه‌سازی مقدار T را برای D_i جدید محاسبه کنیم، برای هر جفت (X_i, Y_i) به صورت تصادفی جای X_i و Y_i را عوض می‌کنیم و این کار را به تعداد زیاد تکرار می‌کنیم تا توزیع T دقیق‌تر به دست آید و در نهایت آن را با α مقایسه می‌کنیم

سوال دهم:

$$\operatorname{cov}(\beta_0, \beta_1) \xlongequal{\beta_0 = \bar{Y} - \beta_1 \bar{x}} \operatorname{cov}(\bar{Y} - \beta_1 \bar{x}, \beta_1) = \underbrace{\operatorname{cov}(\bar{Y}, \beta_1)}_{\text{۰ چون } \bar{Y} \text{ عدد ثابت است}} - \operatorname{cov}(\beta_1 \bar{X}, \beta_1) = 0 - \bar{X}\operatorname{Var}(\beta_1) \quad (1)$$

$$= \bar{X}\operatorname{cov}(\beta_1, \beta_1) = \bar{X}\operatorname{Var}(\beta_1)$$

(1) با توجه به عبارت به دست آمده برای اینکه $-\bar{X}\operatorname{Var}(\beta_1) = 0$ باشد تا β_1 و β_0 مستقل باشند باید یا $\bar{X}$ صفر شود یا $\operatorname{Var}(\beta_1)$

اگر $\bar{X} = 0$ باشد $\operatorname{cov}(\beta_0, \beta_1)$ نیز صفر می‌شود و ۲ متغیر مستقل هستند

اگر $\operatorname{Var}(\beta_1) = 0$ شود یعنی باید β_1 یک عدد ثابت باشد در شرایط عادی رگرسیون اتفاق نمی‌افتد مگر این داده‌ها هیچ تغییر و واریانسی نداشته باشد

به طور کلی β_1 و β_0 مستقل هستند که $\bar{X} = 0$ باشد

سوال یازدهم: i)

$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad Y = X\beta + \varepsilon \qquad Y = [Y_1, \dots, Y_n]^T \qquad \beta = [\beta_0, \beta_1]^T \qquad \varepsilon = [\varepsilon_1, \dots, \varepsilon_n]^T$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \longrightarrow$$

ستون اول شامل 1 (برای β_0)
و ستون دوم x_i که مقادیر مستقل هستند

ii)

$$Y \sim N_n(X\beta, \sigma^2 I_n) \xrightarrow[(X'X)^{-1}X']{\text{ضرب ۲ طرف در}} (X'X)^{-1}X'Y \sim N_2\left((X'X)^{-1}X'X\beta, (X'X)^{-1}X'\sigma^2 I_n X (X'X)^{-1}\right)$$

$$\xrightarrow{(X'X)^{-1}X'X = I} (X'X)^{-1}X'Y \sim N_2(\beta_1, \sigma^2 (X'X)^{-1})$$

$$E\left[(X'X)^{-1}X'Y\right] = \beta \Rightarrow \text{unbiased نتیجه}$$

iii)

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad \hat{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad X'X = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \Rightarrow$$

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x}) Y_i}{\sum (x_i - \bar{x})^2} = \left[\frac{\sum (x_i - \bar{x})}{\sum (x_i - \bar{x})^2}, \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2}\right] Y \xrightarrow[\text{به تابع خطی از } Y]{\text{نوشتن } \hat{\beta}_1 \text{ به صورت}} \hat{\beta}_1 = \left[0, \frac{\sum (x_i - \bar{X}) y_i}{\sum (x_i - \bar{x})^2}\right] Y$$

از آنجایی که $Y = \beta X + \varepsilon$ و ε دارای توزیع نرمال با میانگین صفر است و واریانس σ^2 می‌توانیم از ویژگی‌های رگرسیون خطی برای به دست آوردن توزیع $\hat{\beta}_1$ استفاده کنیم در واقع $\hat{\beta}_1$ دارای توزیع نرمال با میانگین β_1 (شیب واقعی) و واریانس $\frac{\sigma^2}{\sum (x_i - \bar{x})^2}$ است

ادامه سوال یازدهم: iii)

$$\Rightarrow \hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum(x_i-\bar{x})^2}\right)$$

iv)

$$\hat{\beta}_1 \pm z_{\alpha/2}\,\sigma\sqrt{\frac{1}{\sum(x_i-\bar{x})^2}} \quad , \quad \frac{\hat{\beta}_1-\beta_1}{\frac{\sigma}{\sqrt{\sum(x_i-\bar{x})^2}}} \sim N(0,1) \quad , \quad P\left(-z_{\alpha/2} < \frac{\hat{\beta}_1-\beta_1}{\frac{\sigma}{\sqrt{\sum(x_i-\bar{x})^2}}} < z_{\alpha/2}\right) = 0.9$$

$$\alpha = 1-0.9 = 0.1 \Rightarrow \alpha/2 = 0.05 \ , \ z_{0.05} = 1.645$$

$$\Rightarrow -1.645 < \frac{\hat{\beta}_1-\beta_1}{\frac{\sigma}{\sqrt{\sum(x_i-\bar{x})^2}}} < 1.645 \Rightarrow -1.645 \times \frac{\sigma}{\sqrt{\sum(x_i-\bar{x})^2}} < \hat{\beta}_1-\beta_1 < 1.645 \times \frac{\sigma}{\sqrt{\sum(x_i-\bar{x})^2}}$$

$$\Rightarrow \hat{\beta}_1 - 1.645\frac{\sigma}{\sqrt{\sum(x_i-\bar{x})^2}} < \beta_1 < \hat{\beta}_1 + 1.645\frac{\sigma}{\sqrt{\sum(x_i-\bar{x})^2}}$$

سوال دوازدهم: i)

MS model: ? mean square

$$SS_{model} = F \times MS_{residual} \times DF_{model} = 15.12 \times 0.59770344 \times 3 = 27.1335$$

$$SS_{total} = SS_{model} + SS_{Residual} = 27.1335 + 7.1592394 = 34.2927$$

$$MS_{model} = \frac{SS_{model}}{df_{model}} = \frac{27.1335}{3} = 9.0445$$

ii)

$$R^2 = \frac{SS_{model}}{SS_{total}} = \frac{27.1335}{34.2927} = 0.791$$

iii) با توجه به اینکه $p<0.001$ ، $\alpha=0.01$ می باشد ، $p<\alpha$ درنتیجه فرض صفر رد می شود

iv) $\hat{Y} = \beta_0 + \beta_1(\text{Size}) + \beta_2(\text{Parking}) + \beta_3(\text{Income}) + \varepsilon$ ، $\beta_1, \beta_2, \beta_3 =$ coefficient های size، Parking، income

v) coefficient پارکینگ نشان دهنده تغییرات مورد انتظار در میزان فروش به ازای هر فضای پارکینگ اضافی (در صورت ثابت بودن بقیه)

1) اگر $\beta_2>0$ افزایش تعداد پارکینگ باعث افزایش فروش می شود

2) اگر $\beta_2<0$ افزایش تعداد پارکینگ باعث کاهش فروش می شود

3) اگر $\beta_2=0$ تعداد پارکینگ تاثیر قابل توجهی در فروش ندارد

vi)

$$CI = \hat{\beta}_0 \mp t_{\alpha/2,\underset{12}{df}} \times SE(\hat{\beta}_0) \Rightarrow CI = \hat{\beta}_0 \mp (2.179 \times SE(\hat{\beta}_0))$$

$$t_{\alpha/2,12} = 2.179$$

سوال سیزدهم: i)

$$L(\mu,\sigma^2|x_1 \dots x_n) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2$$

$$\frac{\partial L}{\partial \mu} = \frac{1}{\sigma^2}\sum(x_i-\mu) = 0 \Rightarrow \hat{\mu} = \frac{1}{n}\sum x_i = \bar{x} \quad , \quad \frac{\partial L}{\partial \sigma^2} \xrightarrow{\text{مشابه}} \hat{\sigma}^2 = \frac{1}{n}\sum(x_i-\hat{\mu})^2$$

$$I(\mu,\sigma^2) = -E\left[\frac{\partial^2 L(\mu,\sigma^2)}{\partial\mu^2}\right] \quad , \quad I(\sigma^2,\sigma^2) = -E\left[\frac{\partial^2 L(\mu,\sigma^2)}{\partial(\sigma^2)^2}\right] \Rightarrow \frac{\partial^2 L}{\partial\mu^2} = -\frac{n}{\sigma^2} \Rightarrow I(\mu,\mu) = \frac{n}{\sigma^2}$$

$$\frac{\partial^2 L}{\partial(\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} = I(\sigma^2,\sigma^2)$$

$$\frac{\partial^2 L}{\partial\mu\,\partial\sigma^2} = 0 \Rightarrow I(\mu,\sigma^2) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2(\sigma^2)^2} \end{bmatrix} \Rightarrow \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}, \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}\right)$$

ii) $L(\sigma^2|x_1 \dots x_n) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum x_i^2 = 0$ ، $\hat{\sigma}^2 = \frac{1}{n}\sum x_i^2 \Rightarrow$ در ماتریس I عناصر خارج از قطر اصلی صفر هستند دقیق $\mu=0 \Rightarrow$ $\hat{\sigma}^2$ و $\hat{\mu}$ در n های بزرگ مستقل اند. اگر $\mu=0$ با ... نه $X_i \sim N(0,\sigma^2)$

ادامه سوال رو نزدیم .

$$\Rightarrow Var(\hat{\sigma}^2) = \frac{2\sigma^4}{n} \quad , Var(\hat{\mu}) = \frac{\sigma^2}{n}$$

مشابه‌ها، برای : i)

$$E[X] = 0 \cdot P[X=0] + 1 \times P[X=1] + 2 \times P[X=2] \times 3 \times P[X=3] = \frac{1}{3}\theta + \frac{4}{3}(1-\theta) + 3 \times \frac{1}{3}(1-\theta)$$

$$\Rightarrow E[X] = \frac{1}{3}\theta + \frac{4}{3} - \frac{4}{3}\theta + \frac{3}{3} - \frac{3}{3}\theta = \frac{7}{3} - 2\theta = \bar{X} = \frac{3+0+2+1+3+2+1+0+2+1}{10} = 1.5$$

$$\Rightarrow \theta = \frac{2.5}{6} = \frac{5}{12} \approx 0.417$$

$$Var(\hat{\theta}) = \frac{Var(X)}{n\left(\frac{dE[X]}{d\theta}\right)^2} \qquad \frac{dE(X)}{d\theta} = -2 \Rightarrow Var(\hat{\theta}) = \frac{Var(X)}{10 \times 4} \Rightarrow Var(X) = \frac{1}{10}\sum(x_i - \bar{x})^2$$

ii)

$$L(\theta) = \prod P(x_i) \Rightarrow \log L(\theta) = \sum \log P(x_i)$$

$$\Rightarrow L(\theta) = \left(\frac{2}{3}\theta\right)^2 \times \left(\frac{1}{3}\theta\right)^2 \times \left(\frac{2}{3}(1-\theta)\right)^3 \times \left(\frac{1}{3}(1-\theta)\right)^2$$

$$\log(L(\theta)) = 2\log\frac{2}{3}\theta + 3\log\frac{1}{3}\theta + 3\log\frac{2}{3}(1-\theta) + 2\log\left(\frac{1}{3}(1-\theta)\right)$$

$$\frac{dL}{d\theta} = 0 \Rightarrow \frac{2}{\theta} + \frac{2}{\theta} - \frac{2 \times 3}{1-\theta} - \frac{2}{1-\theta} \Rightarrow \hat{\theta}_{MLE} = \frac{2+2}{2+2+(2\times3)+(3\times2)}$$

$$I(\theta) = -E\left[\frac{d^2L(\theta)}{d\theta^2}\right] \quad , SE(\hat{\theta}_{MLE}) = \frac{1}{\sqrt{I(\theta)}}$$