۱-

a. نادرست – F-statistic در برابر نقض شرط برابری واریانس‌ها چندان robust نیست. اما در برابر نقض شرط normality مقاوم‌تر است. در واقع توجه بیشتری باید به عدم برابری واریانس‌ها نسبت به عدم نرمال بودن دیتا پرداخته شود. چرا که F-statistic در برابر نرمال نبودن دیتا مقاوم است به شرط اینکه:
- هر یک از گروه‌ها دارای توزیع متقارن و uni-modal باشند
- اندازه هریک از گروه‌ها بزرگترمساوی ۱۰ باشد

b. درست – از آنجایی که F-ratio از تقسیم دوتا variability بدست می‌آید(variability بین گروهی تقسیم بر variability درون گروهی)و variability مقدار منفی نمی‌گیرد، همواره بزرگتر مساوی صفر است. همچنین توزیع F یک توزیع right-skewed می‌باشد، پس چولگی مثبت دارد.

c. درست – اختلاف زیاد بین مقادیر میانگین گروه‌های مختلف باعث variability بالا بین گروهی می‌شود و مقادیر واریانس کوچک در سمپل نشان‌دهنده‌ی variability کم درون گروهی است. و طبق رابطه‌ی F-ratio که در قسمت (b) هم معرفی شده، صورت کسر بزرگ و مخرج کسر کوچک منجر به مقدار F-ratio بالایی می‌شود.

d. درست -

e. نادرست – تعداد افراد شرکت کننده در این آزمایش برابر $n = 3 + 36 + 1 = 40$ می‌باشد.

f. نادرست – یک bootstrap sample نمونه‌ای است که با نمونه‌برداری با جایگذاری به طور تصادفی از original sample بدست می‌آید و اندازه‌ی آن با سایز original sample یکسان است. (از آنجایی که نمونه‌برداری از sample اولیه با جایگذاری انجام می‌شود امکان تکرار نقاط نمونه sample اولیه در bootstrap sample وجود دارد.) ولی در این bootstrap sample داده‌ی ۵۶ وجود دارد در حالیکه در original sample موجود نیست.

g. درست- مطابق محاسبات زیر:

$$SE = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = \sqrt{\frac{5.25251^2}{20}+\frac{2.33735^2}{25}} \approx \sqrt{1.38+0.22} = \sqrt{1.6} = 1.26$$

$$Df = \min(20-1, 25-1) = 19 \qquad CL = 95\% \rightarrow t^*_{df=19} = 2.09$$

$$(\bar{x}_2 - \bar{x}_1) \pm t^*_{df=19} \times SE = (5.25 - 3.75) \pm 2.09 \times 1.26 = 1.5 \pm 2.09 \times 1.26$$

h. نادرست – برای مشخص کردن چنین چیزی می‌توان از یک unpaired T-student test استفاده کرد نه paired. مگر اینکه گروه دخترها و پسرها به نوعی بهم وابسته باشند. مثلا تست روی یک سری خواهر و برادر انجام شده باشد.

i. نادرست –

j. نادرست – برای مقایسه‌ی میانگین دو گروه زمانی که سایز نمونه‌مون کوچک می‌باشد از T-student test استفاده می‌کنیم. در چنین حالتی در درجه‌ی آزادی توزیع T از رابطه‌ی $DF = \min(n_1 - 1, n_2 - 1)$ بدست می‌آید. پس DF = 19 می‌شود. توزیع T با درجه‌ی آزادی کوچک(مثل ۱۹) با توزیع نرمال فاصله نسبتاً زیادی دارد. پس نمی‌توان از توزیع نرمال استفاده کرد.

k. درست – البته در صورتی که با استفاده از یکی از روشهای تصحیح α مثل(The Bonferroni Correction) از مقدار تصحیح شده‌ی α برای هر کدام از pairwise test ها استفاده کنیم تا مشکل inflated Type 1 error رخ ندهد و باعث اشتباه شدن نتیجه‌ی تست‌ها نشود.

۲-

	SS	DF	MS	F
(GROUP)BETWEEN	**20**	4	**5**	**1**
(ERROR)WITHIN	**100**	20	5	
TOTAL	120	**24**		

$$df_{Total} = n - 1 = 25 - 1 = \mathbf{24}$$

$$MSE = \frac{SSE}{df_E} \rightarrow SSE = MSE \times df_E = 5 \times 20 = \mathbf{100}$$

$$SSE = SST - SSG \rightarrow SSG = SST - SSE = 120 - 100 = \mathbf{20}$$

$$MSG = \frac{SSG}{df_G} = \frac{20}{4} = \mathbf{5}$$

$$F = \frac{MSG}{MSE} = \frac{5}{5} = \mathbf{1}$$

$$p - value = P(F > 1) = 0.4306816$$

```
> pf(1,4,20,lower.tail = FALSE)
[1] 0.4306816
```

با فرض یافتن سطح اطمینان 95% و $\alpha = 0.05$ داریم:

$$p - value = 0.4306816 > 0.05 \rightarrow \boldsymbol{Fail\ to\ Reject\ H_0}$$

نمی‌توان گفت که اختلاف قابل توجهی بین میانگین گروه‌ها وجود دارد. یعنی می‌گوییم گروه‌ها میانگین تقریبا برابر هم دارند.

۳-

a) یک دیتافریم با ستون‌های مشخص شده از دیتاست penguins به نام penguins_2col ساختم:

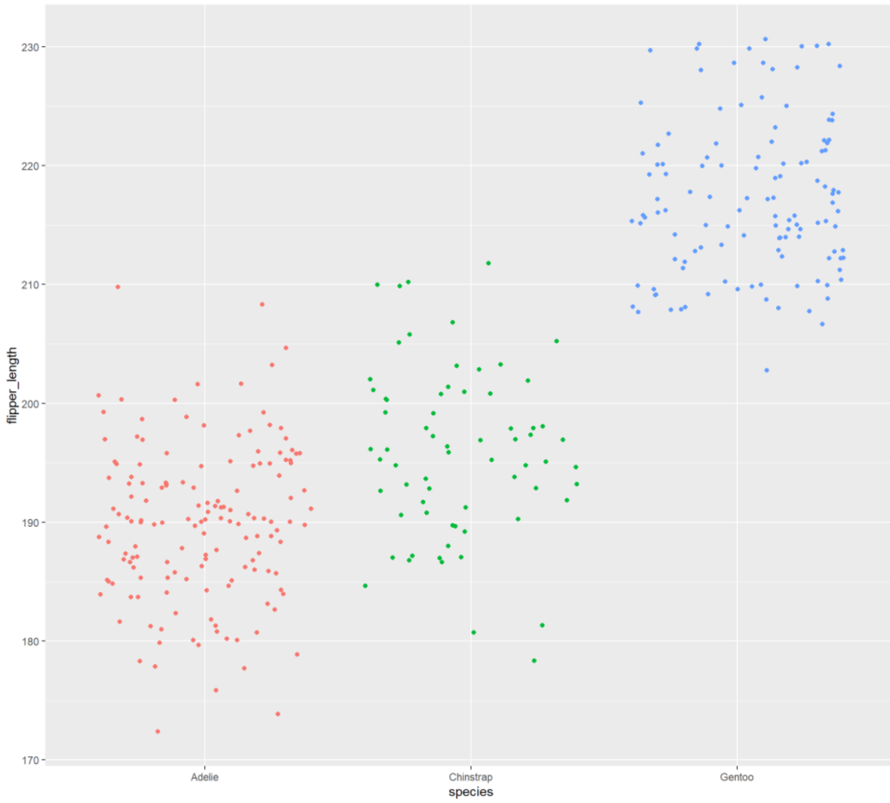
```
#a)construscting a dataframe with only "flipper_length_mm" & "species" columns of the penguin dataset
penguins_2col <- data.frame(flipper_length=penguins$flipper_length_mm,species=penguins$species)
```

---

b)

```
#b)plotting flipper length of each species
ggplot(penguins_2col) +
  aes(x = species, y = flipper_length, color = species) +
  geom_jitter() +
  theme(legend.position = "none")
```

The result:



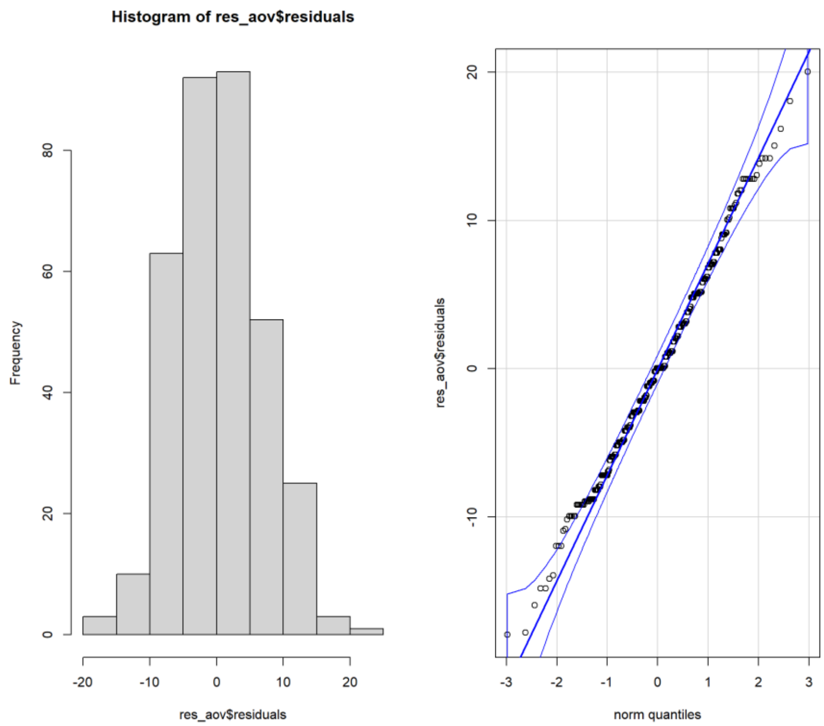
---

c) شروط تست ANOVA به شرح زیر می‌باشند:

- استقلال مشاهدات: فرض می‌کنیم پنگوئن‌های دیتاست به طور تصادفی انتخاب شده و در نتیجه نقاط نمونه مستقل از هم هستند.
- داده‌های هر گروه باید توزیع تقریبا نرمال داشته باشند: این شرط را می‌توان با رسم QQ-plot و هیستوگرام چک کرد. (با توجه به شکل هیستوگرام میبینیم که توزیع تقریبا متقارن و نزدیک نرمال داریم. همچنین از روی QQ-plot که توزیع داده‌ها با توزیع نرمال را مقایسه می‌کند، می‌توان دید که توزیع داده‌ها و توزیع نرمال رابطه‌ی تقریبا خطی دارند، پس می‌توان نتیجه گرفت که داده‌ها توزیع تقریباً نرمال دارند)
- واریانس گروه‌های مختلف باید تقریبا باهم برابر باشد: این شرط را می‌توان با رسم box plot یا dot plot بررسی کرد که در اینجا با رسم dot plot اینکار را انجام دادیم.

```
#c)
res_aov <- aov(flipper_length ~ species,data = penguins_2col)
##checking the "Normality" assumption
par(mfrow = c(1, 2))
# histogram
hist(res_aov$residuals)
# QQ-plot
qqPlot(res_aov$residuals,id = FALSE)
```

The result:

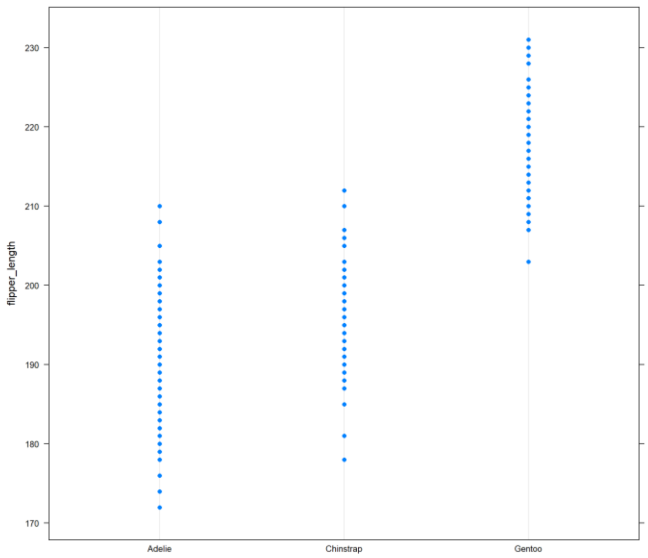


---

d) بررسی شرط برابری واریانس‌ها با استفاده از Dot plot: با توجه به dot plot می‌توان دید که مقدار واریانس گروه‌ها بسیار نزدیک به هم می‌باشد.

```
#d)checking the "equality of variances" assumption
# Dot plot
dotplot(flipper_length ~ species,data = penguins_2col)
```

The result:



---

e)

```
#e)mean and Sd of each group
aggregate(flipper_length ~ species,
          data = penguins_2col,
          function(x) round(c(mean = mean(x), sd = sd(x)), 2)
)
```

The result:

```
     species flipper_length.mean flipper_length.sd
1     Adelie              189.95              6.54
2  Chinstrap              195.82              7.13
3     Gentoo              217.19              6.48
>
```

---

f) مطابق نتیجه‌ای که تابع report داده: تأثیر نوع گونه‌ی پنگوئن بر طول باله‌ی پنگوئن‌ها قابل توجه است. یعنی از آنجایی که $p-value<0.001$، فرض null (H_0) رد می‌شود. (H_0: نوع گونه‌ی پنگوئن‌ها تأثیری در طول باله‌ی آنها ندارد)

$p-value=P(F>594.80)<0.001$

```
#f)Doing ANOVA test and showing the results
res_aov <- aov(flipper_length ~ species,data = penguins_2col)
report(res_aov)
```

The result:

```
The ANOVA (formula: flipper_length ~ species) suggests that:

  - The main effect of species is statistically significant and large (F(2, 339) = 594.80, p < .001; Eta2 = 0.78, 95% CI [0.75, 1.00])

Effect sizes were labelled following Field's (2013) recommendations.
>
```

(g

$$number\ of\ pairwise\ comparisons = K = \binom{\#\ of\ Groups}{2} = \binom{3}{2} = 3$$

$$\alpha^* = \frac{\alpha}{K} = \frac{\alpha}{3} \leftarrow \boldsymbol{Bonferroni\ correction}$$

-۴

a) تیم ملی بسکتبال ایران و تیم ایده‌آل سنی معرفی شده از هم مستقل هستند. پس برای اجرای آزمون فرض از T test استفاده می‌کنیم.

$$Df = \min(n_{national-team} - 1, n_{model} - 1) = \min(50 - 1, 40 - 1) = 39$$

$$H_0: \mu_{national\_team} - \mu_{model} = 0$$

$$H_A: \mu_{national\_team} - \mu_{model} \neq 0$$

$$\bar{x}_{national\_team} - \bar{x}_{model} = 26.9 - 29 = -2.1$$

$$SE_{(\bar{x}_{national\_team} - \bar{x}_{model})} = \sqrt{\frac{s^2_{national\_team}}{n_{national\_team}} + \frac{s^2_{model}}{n_{model}}} = \sqrt{\frac{5.6^2}{50} + \frac{2.4^2}{40}} = \sqrt{0.6272 + 0.144}$$
$$= \sqrt{0.7712} \approx 0.88$$

برای محاسبه‌ی p-value با استفاده از R داریم:

```
> 2*pt(2.38,39,lower.tail = FALSE)
[1] 0.02229791
>
```

$$p - value = P\left(|\bar{x}_{national\_team} - \bar{x}_{model}| > T^* | H_0\right) = P\left(|t^*_{39}| > \frac{2.1-0}{0.88}\right) = P(|t^*_{39}| > 2.38) \approx 0.022$$

$$p - value \approx 0.022 < \alpha = 0.05 \rightarrow \boldsymbol{Reject\ H_0}$$

پس میانگین سن باریکنان بسکتبال تیم ملی ایران با میانگین سن تیم ایده‌آل سنی بسکتبال اختلاف دارد.

(b

$$\mu_a = 28$$

```
> qt(0.025,39,lower.tail = FALSE)
[1] 2.022691
> 
```

اگر اختلاف بین میانگین سن باریکنان بسکتبال تیم ملی ایران و میانگین سن تیم ایده‌آل بسکتبال از $2.02\ SE$ بیشتر یا از $-2.02\ SE$ کمتر باشد، H_0 را رد می‌کنیم.

$$2.02 \times SE = 2.02 \times 0.88 = 1.7776$$

$$\mu_{national\_team} - \mu_{model} = 26.9 - 28 = -1.1$$

برای محاسبه‌ی *power* با استفاده از *R* داریم:

```
> 2*pt(0.77,39,lower.tail = FALSE)
[1] 0.4459391
> 
```

$$Power = P\left(\left|\bar{x}_{national_{team}} - \bar{x}_{model}\right| > T^*\right) = P\left(|t^*_{39}| > \frac{1.7776 - 1.1}{0.88}\right)$$
$$= P(t^*_{39} < -0.77) + P(t^*_{39} < 0.77) = 0.4459391$$

$$Power = 1 - \beta = 0.4459391 \rightarrow \beta = 0.5540609 = P(Type\ 2\ error)$$

(c

۵-

۶- از R فقط برای محاسبه میانگین و انحراف معیار استفاده شده(به جای ماشین حساب)

	n	mean	sd
Catering	6	1263.67	15.56
No Catering	11	1249.82	27.72

```
> catering <- c(1284,1272,1256,1254,1242,1274)
> mean(catering)
[1] 1263.667
> sd(catering)
[1] 15.56492
```

```
> NoCatering <- c(1294,1279,1274,1264,1263,1254,1240,1232,1220,1218,1210)
> mean(NoCatering)
[1] 1249.818
> sd(NoCatering)
[1] 27.71577
```

```
> all <- c(catering,NoCatering)
> mean(all)
[1] 1254.706
> sd(all)
[1] 24.54273
```

می‌توانیم از T test برای اینکار استفاده کنیم(سایز نمونه کوچکتر از ۳۰ است. پس نمی‌توانیم از توزیع نرمال برای آزمون فرض استفاده کنیم):

$$Df = \min(n_1 - 1, n_2 - 1) = \min(5,10) = 5$$

$$SE = \sqrt{\frac{s_{catering}^2}{n_{catering}} + \frac{s_{NoCatering}^2}{n_{NoCatering}}} = \sqrt{\frac{15.56^2}{6} + \frac{27.72^2}{11}} = \sqrt{110.2} \approx 10.5$$

حال آزمون فرض را می‌نویسیم:

$$H_0: \mu_{catering} - \mu_{NoCatering} = 0$$

$$H_A: \mu_{catering} - \mu_{NoCatering} \neq 0$$

$$\bar{x}_{catering} - \bar{x}_{NoCatering} = 1263.67 - 1249.82 = 13.85$$

$$T_5 = \frac{observation - null}{SE} = \frac{13.85 - 0}{10.5} \approx 1.32$$

برای محاسبه‌ی p-value با استفاده از R داریم:

```
> 2*pt(1.32,5,lower.tail = FALSE)
[1] 0.2440351
```

$$p - value = P(|t_{df=5}| > 1.32) = P(t_{df=5} < -1.32) + P(t_{df=5} > 1.32) = 0.244$$

$$p - value = 0.244 > \alpha = 0.05 \quad \rightarrow \boldsymbol{fail\ to\ Reject\ H_0}$$

تا زمانی که مقدار *p-value* از α بیشتر باشد، H_0 را نمی‌توان *reject* کرد، پس:

$$\alpha_{max} = 0.244 \quad \rightarrow 0.244 - 0.05 = 0.194$$

پس α را می‌توان حدوداً به اندازه‌ی 0.194 افزایش داد.

۷- شرایط استفاده از T test برای مقایسه‌ی دوتا میانگین عبارتند از:

- استقلال(استقلال بین گروهی و استقلال درون گروهی)
- Sample size/ skew

از شرایط استقلال می‌بینیم که شرط استقلال بین گروهی را نداریم. چرا که از افراد یکسانی برای هر دو آزمایش استفاده کردیم. و اصطلاحاً گفته می‌شود که paired data داریم. در چنین مواقعی باید از paired test برای بررسی تأثیر متغیر categorical مربوطه(استفاده یا عدم استفاده از AI) بر روی متغیر numerical خود(نمره داده شده از سوی کاربر) استفاده کنیم. برای این کار از متغیر جدید diff استفاده می‌کنیم که مقدار آن برای هر user برای تفاضل نمره‌ی user به محصول با AI و محصول بدون AI می‌باشد. سپس مثل T test عادی عمل می‌کنیم.(دقت شود چون اندازه سمپل کوچکتر از ۳۰ است، نمی‌توانیم از توزیع نرمال استفاده کنیم.

user	Without AI	With AI	diff
1	210	197	13
2	205	195	10
3	193	191	2
4	182	174	8
5	259	236	23
6	239	226	13
7	164	157	7
8	197	196	1
9	222	201	21
10	211	196	15
11	187	181	6
12	175	164	11
13	186	181	5
14	243	229	14
15	246	231	15

```
> diff <- c(13,10,2,8,23,13,7,1,21,15,6,11,5,14,15)
> mean(diff)
[1] 10.93333
> sd(diff)
[1] 6.329824
```

$\bar{x}_{diff} = 10.93 \qquad s_{diff} = 6.33 \qquad n_{diff} = 15 \qquad SE = \frac{s_{diff}}{\sqrt{n_{diff}}} = \frac{6.33}{\sqrt{15}} = \frac{6.33}{3.87} = 1.63$

حال آزمون فرض را طراحی می‌کنیم:

$H_0 : \mu_{diff} = 0$

$H_A : \mu_{diff} \neq 0$

$Df = 15 - 1 = 14 \qquad T_{14} = \frac{10.93 - 0}{1.63} = 6.71$

برای محاسبه‌ی p-value با استفاده از R داریم:

```
> 2*pt(6.71,14,lower.tail = FALSE)
[1] 9.942629e-06
```

$p - value = P(|t_{df=14}| > 6.71) = P(t_{df=14} < -6.71) + P(t_{df=14} > 6.71) \approx 0$

$p - value \approx 0 < \alpha = 0.05 \rightarrow$ ***Reject H_0*** $\rightarrow$ دو روش برابر نیستند

٨-

۹-

a.

$H_0: \mu = 35$
$H_A: \mu > 35$

$$\bar{x} = \frac{25 + 27 + 35 + 42 + 28 + 37 + 40 + 31 + 29 + 33 + 30 + 26 + 31 + 28 + 30 + 15}{16} = \frac{487}{16} = 30.4375$$

$$SE = \frac{s}{\sqrt{16}} = \frac{6.376715}{4} = 1.59 \;, Df = 16 - 1 = 15$$

```
> pt(-2.87,15, lower.tail = FALSE)
[1] 0.9941581
```

$$p - value = P(\bar{x} > 35|H_0) = P\left(t_{df=15} > \frac{30.4375 - 35}{1.59}\right) = P\left(t_{df=15} > -2.87\right) = 0.99 > \alpha = 0.05$$

$p - value = 0.99 > \alpha = 0.05 \quad \rightarrow \boldsymbol{Fail\ to\ Reject\ H_0}$

Computing the Confidence Interval:

$CL = 95\% \quad \rightarrow \quad t^*_{df=15} = 2.13$

$\bar{x} \pm t^*_{df=15} \times SE = 30.4375 \pm 2.13 \times 1.59 = (-27.0508\,, 33.8242)$

۱۰-

a. متغیر price از دیتاست car_train را انتخاب کردم:

i. برای انجام آزمون فرض از یک نمونه با سایز ۲۵ استفاده کردم پس باید از توزیع T-student استفاده کرد.فرض می‌کنیم میانگین کل ستون price برابر با 22000 می‌باشد و آن را به عنوان null value در نظر می‌گیریم. سپس یک نمونه با سایز ۲۵ از ستون price انتخاب می‌کنیم. حال با استفاده از آزمون فرض می‌خواهیم ببینیم آیا اختلاف قابل توجهی از نظر آماری بین میانگین سمپل ما و مقدار ۲۲۰۰۰ وجود دارد یا اینکه اختلاف مشاهده شده بین مقادیر دو میانگین صرفاً تصادفی بوده و به علت sampling variability می‌باشد. (از confidence level = 95% استفاده می‌کنیم)

```
> print(sample_mean_price)
[1] 21910.04
```

$H_0: \mu = 22000$
$H_A: \mu \neq 22000$

```
##i)
#H0:mu = 22000
#HA:mu != 22000
null_value <- 22000
sample_t <- sample(car_train$price,25)
sample_mean_price <- mean(sample_t)
t_24 <- abs((sample_mean_price - null_value)/(sqrt(var(sample_t))/(25^0.5)))
p_value <- 2 * pt(t_24 , df = 24 , lower.tail = FALSE)
if(p_value < 0.05){
  cat("Reject H0 -> there is a statisticaly significant difference between\n
        the mean value of 22000 and the sample mean value")
}else{
  cat("Fail to Reject H0 -> the difference between the means is simply due
      to sampling variability.\nwe can't say the sample mean is not equal to 22000")
}
```

The result:

```
Fail to Reject H0 -> the difference between the means is simply due to sampling variability.
we can't say the sample mean is not equal to 22000
```

همانطور که از نتیجه کد بالا مشخص است، *P-value* بزرگتر از *0.05* شده. پس نمی‌توانیم *H0* را *reject* کنیم. یعنی مقدار میانگین*22000* برای متغیر *price* درست بوده و اختلاف میانگین سمپل ما با *22000* تصادفی بوده و صرفاً به علت *sampling variability* می‌باشد.

---

ii. با توجه به نتایج کد که در ادامه آمده، می‌بینیم که *22000*(*null value*) در بازه‌ی اطمینان بدست آمده می‌باشد. پس نمی‌توانیم *null hypothesis* را رد کنیم و نتیجه بدست آمده از هر دو روش باهم همخوانی دارد.

```
##ii)making the 95% confidence interval (from,to)
t_star <- abs(qt(0.025 , df = 24))
from <- sample_mean_price - abs(t_star) * (sqrt(var(sample_t))/(25^0.5))
to <- sample_mean_price + abs(t_star) * (sqrt(var(sample_t))/(25^0.5))
if(null_value < from || null_value > to){
  print("Reject H0 -> there is a statisticaly significant difference between
        the actual mean value and the sample mean value")
  sprintf("%s is not in the confidence interval range(%s,%s)",null_value,from,to)
}else{
  print("Fail to Reject H0 -> the difference between the means is simply due to sampling variability")
  sprintf("%s is in the confidence interval range(%s,%s)",null_value,from,to)
}
```

The result:

```
[1] "Fail to Reject H0 -> the difference between the means is simply due to sampling variability"
[1] "22000 is in the confidence interval range(18906.1432481189,28683.7767518811)"
```

---

iii. برای محاسبه *Type 2 error rate* به *power* نیاز داریم. برای محاسبه‌ی *power* به یک *actual mean* نیاز داریم. برای این کار از مقدار میانگین واقعی ستون *price* استفاده می‌کنیم.

$$P(Type\ 2\ error) = \beta \quad , \qquad Power = 1 - \beta \quad \rightarrow \quad P(Type\ 2\ error) = 1 - Power$$

```
actual_mean_price <- mean(car_train$price,na.rm = TRUE)
print(actual_mean_price)
```

The result:

```
> print(actual_mean_price)
[1] 22170.48
```

$$\mu_a = 22170.48$$

.iv

.b

.c