Zeinab Hosseini HW2 - SI

401130596 5008

Question 1:

$A = (1,3), (2,2), (3,1) \Rightarrow P(A) = \frac{3}{36}$

$B = (1,5), (2,4), (3,3), (4,2), (5,1) = \frac{5}{36}$

$C = (1,2), (2,2), (2,1), (2,3), (3,2), (2,4), (4,2), (2,5),$
$(5,2), (6,2), (2,6) = \frac{11}{36}$

$P(A \cap C) = \frac{1}{36}$

a) $P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{1/36}{11/36} = \frac{1}{11}$

برای نشان دادن اینکه A , C از هم مستقل هستند باید نشان دهیم که

$P(A \cap C) = P(A) * P(C)$

$\Rightarrow \frac{3}{36} * \frac{11}{36} \neq \frac{1}{36}$

و چون این تساوی برقرار نیست، این دو از هم مستقل نیستند.

b) $P(B|C) = \frac{P(B \cap C)}{P(C)} = \frac{2/36}{11/36} = \frac{2}{11}$

$P(B \cap C) = P(B) * P(C) = \frac{5}{36} * \frac{11}{36} \neq \frac{2}{11}$

پس از هم مستقل نیستند.

* روش استفاده شده در اسلاید 19 فایل 2 (probability) گفته شده است.

Question 2 :

$P(F) = \frac{55}{100}$ $\quad P(CS) = \frac{8}{100}$ $\quad P(F \cap CS) = \frac{3}{100}$

a) $P(F \mid CS) = \frac{P(F \cap CS)}{P(CS)} = \frac{3/100}{8/100} = 3/8 = 0{,}375\%$

b) $P(CS \mid F) = \frac{P(F \cap CS)}{P(F)} = \frac{3/100}{55/100} = 3/55 = 0{,}0545\%$

Question 3 :

a) یک متغیر دوجمله‌ای است. زیرا مردم یا رئیس را تایید می‌کنند یا رد می‌کنند و انتخاب آنها مستقل از انتخاب های دیگران است. بنابراین تنها دو انتخاب وجود دارد (رد یا تایید، صفر یا یک) و یک متغیر دوجمله‌ای محسوب می‌شود.

b)

$n = 400$ تعداد کل افراد نمونه

$p = 0{,}92$ احتمال پاسخ بله

$q = 1 - p = 0{,}08$

$$P(x) = \binom{n}{x} \left(\frac{92}{100}\right)^{n} \left(\frac{8}{100}\right)^{n-x} = 0.044$$

```
> pbinom(358,400,0.92)
[1] 0.04410268
```

c) expected number $= np = 400 \times \frac{92}{100} = 368$

standard Deviation $= \sqrt{npq} = \sqrt{400 \times \frac{92}{100} \times \frac{8}{100}} = 5.42$

d) $P(x \leq 358) \approx P\left(Z \leq \frac{۳۵۸ - np}{\sigma}\right) =$

$P\left(Z \leq \frac{۳۵۸ - ۳۶۸}{۵,۴۲}\right) = P(Z \leq -1,8) = ۰,۰۳۵۹۳۰۳۲$

* این عدد از این کد R به دست آورده:

```
> pnorm(1.8,mean = 0,sd= 1, lower.tail = FALSE )
[1] 0.03593032
```

بله، به خاطر اینکه جواب به دست آورده به مقدار واقعی آن که توی بخش b به دست آورده، نزدیک است پس میشه گفت این تقریب کاربرد و رضایت بخش.

Question 4:

$P(win \mid T_1) = ۰,۳$: احتمال برنده شدن در یک بازی در برابر شخصی از بازیکنان (نوع 1)

$P(win \mid T_2) = ۰,۴$: ~ یک چهارم ~ ~ ~ ~ (نوع 2)

$P(win \mid T_3) = ۰,۵$: ~ یک چهارم باقی مانده ~ ~ ~ ~ (نوع 3)

$P(T_1) = 1/2 \qquad P(T_2) = 1/4 \qquad P(T_3) = 1/4$

$$P(\omega) = ۰,3 \times 1/2 + ۰,۴ \times \frac{1}{۴} + ۰,۵ \times \frac{1}{۴} = \frac{3}{8}$$

Question 5:

از اونجایی که هر بار run کردن این پروژه مستقل از دفعات قبله است و موفقیت و عدم موفقیت تنها تابع این آزمایش هستند، یعنی احتمال موفقیت برای p و احتمال شکست (1-p) پس با توجه به متن جمله می‌توان گفت: Geometric distribution describes the waiting time until a success for independent and identically distributed (iid) Bernoulli random variables.

independence: outcomes of trials don't affect each other

identical: the probability of success is the same for each trial

از این توزیع پیروی می‌کند پس: $\mu = 1/p$ و $\sigma = \sqrt{\frac{1-p}{p^2}}$.

Question 6:

روش اول:

mean = 40

$\sigma^2 = 7 \rightarrow \sigma = \sqrt{7} = 2.6457$ در R این مقدار را محاسبه می‌کنیم:

```
qnorm ( 0.95, mean = 40, sd = 2.6457 )
```

$\rightarrow 44.35179$

و برای اینکه آخرین زمان ممکن بخواهد حرکت کند باید از ساعت 1 این مقدار را کم کنیم:

$13 - 44.35179 \approx 12.15$

نهایتاً حدود ساعت 12:15 باید حرکت کند.

- روش دیگر هم این است که برای احتمال 95 درصد مقدار 2 را از جدول نرمال توزیع استاندارد پیدا کنیم که می‌شود $2 = 1.64$ پس مقدار x را محاسبه می‌کنیم:

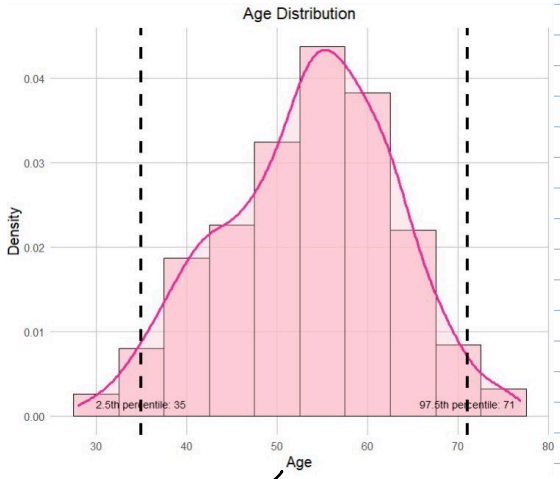
$2 = \frac{x - 40}{2.6457} \rightarrow x = 1.42 \times 2.6457 + 40 \approx 44.34$

← که مجدداً باید حدود ساعت 12:15 حرکت کند.

Question 7:

a) از آنجا که در این سوال ما به دنبال تجزیه و تحلیل اثرات سنی بر شرایط قلبی هستیم، اندازه bin = 5 سال، مقدار منطقی‌ای است:

```
#a)
# Calculate 2.5% and 97.5% quantiles
q025 <- quantile(Heart$age, 0.025)
q975 <- quantile(Heart$age, 0.975)

# Plot histogram and density curve
ggplot(Heart, aes(x = age)) +
  geom_histogram(aes(y = ..density..), binwidth = 5 , fill = "pink", color = "black", alpha = 0.7) +
  geom_density(color = "deeppink", alpha = .3,fill = "pink", size = 1) +
  geom_vline(xintercept = q025, linetype = "dashed", color = "black", size = 1.2) +
  geom_vline(xintercept = q975, linetype = "dashed", color = "black", size = 1.2) +
  annotate("text", x = q025, y = 0, label = paste0("2.5th percentile: ", round(q025, 2)), vjust = -1, color = "black", size = 3) +
  annotate("text", x = q975, y = 0, label = paste0("97.5th percentile: ", round(q975, 2)), vjust = -1, color = "black", size = 3) +
  labs(title = "Age Distribution", x = "Age", y = "Density") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, size = 13),
        axis.title = element_text(size = 12),
        panel.grid.major = element_line(color = "gray"),
        panel.grid.minor =element_blank())
```



→ bin = 5 برای

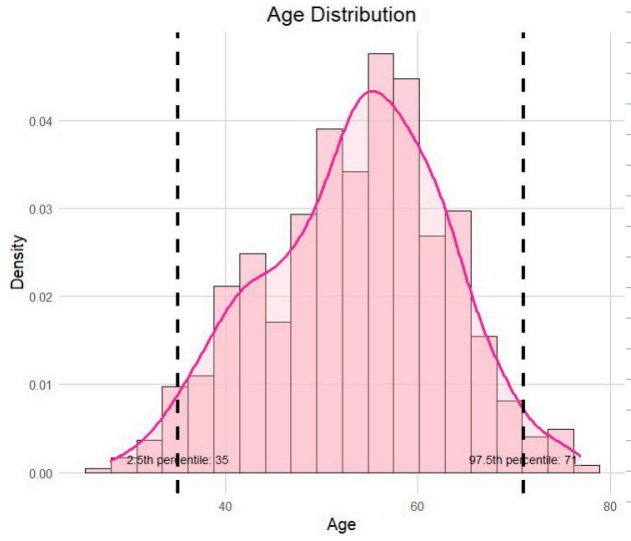
در ویدیوهای درس این قاعده گفته شده است.

برای این نمودار می‌توانیم از قاعده‌ی Freedman - Diaconis هم استفاده کرد و بنابراین bin را با فرمول :

$$\text{bin width} = \frac{2 * IQR}{n^3}$$

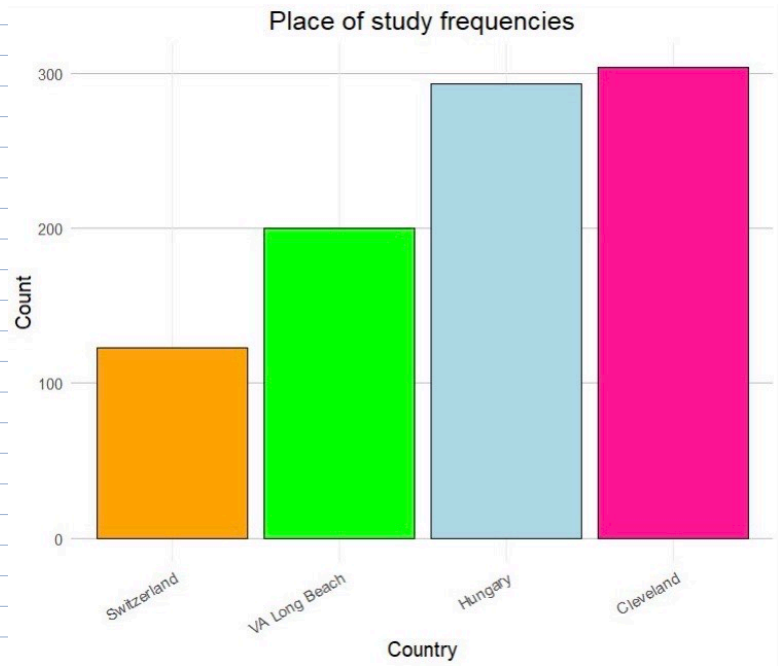
به دست آورد ؛ با این bin ها هیستوگرام به شکل زیر خواهد بود:

```
bw <- 2 * IQR(Heart$age) / length(Heart$age)^(1/3)

q025 <- quantile(Heart$age, 0.025)
q975 <- quantile(Heart$age, 0.975)

ggplot(Heart, aes(x = age)) +
  geom_histogram(aes(y = ..density..), binwidth =bw , fill = "pink", color = "black", alpha = 0.7) +
  geom_density(color = "deeppink", alpha = .3,fill = "pink", size = 1) +
  geom_vline(xintercept = q025, linetype = "dashed", color = "black", size = 1.2) +
  geom_vline(xintercept = q975, linetype = "dashed", color = "black", size = 1.2) +
  annotate("text", x = q025, y = 0, label = paste0("2.5th percentile: ", round(q025, 2)), vjust = -1, color = "black", size = 3) +
  annotate("text", x = q975, y = 0, label = paste0("97.5th percentile: ", round(q975, 2)), vjust = -1, color = "black", size = 3) +
  labs(title = "Age Distribution", x = "Age", y = "Density") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, size = 15),
        axis.title = element_text(size = 12),
        panel.grid.major = element_line(color = "gray"),
        panel.grid.minor =element_blank())
```



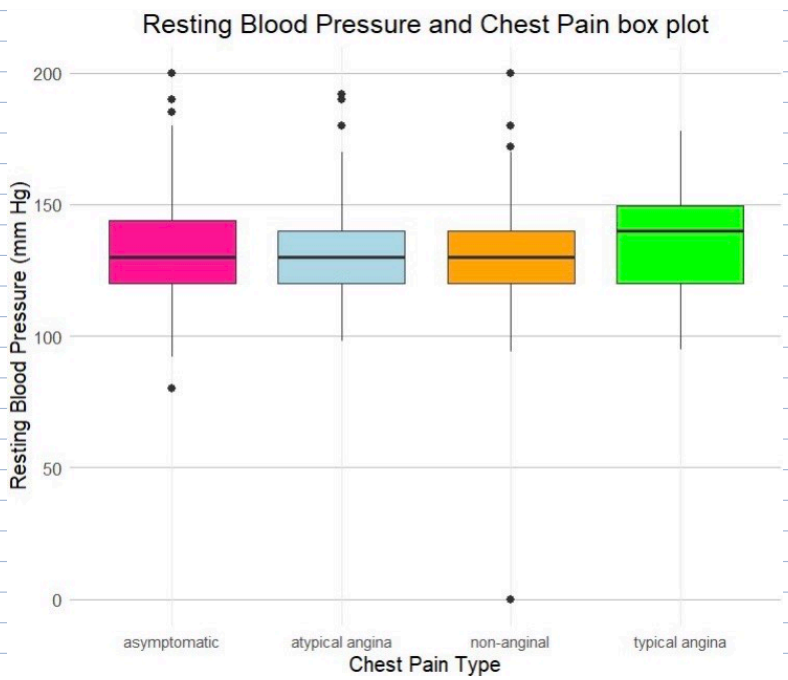
c)

```r
#c)
# Sort the categories in origin by their frequencies
dataset_freq <- sort(table(Heart$dataset), decreasing = TRUE)

# Create a data frame with sorted categories and frequencies
df <- data.frame(dataset = names(dataset_freq),
                 freq = as.vector(dataset_freq))

# Draw the horizontal bar plot with modifications
ggplot(df, aes(x = freq, y = reorder(dataset, freq), fill = dataset)) +
  geom_col(color = "black", size = 0.1) +
  coord_flip() +
  labs(title = "Place of study frequencies",
       x = "Count",
       y = "Country") +
  scale_fill_manual(values=c("deeppink","lightblue","orange", "green"), guide = "none") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, size = 15),
        axis.title = element_text(size = 12),
        axis.text.x = element_text(angle = 30, hjust = 1),
        panel.grid.major.y = element_line(color = "gray"),
        panel.grid.minor = element_blank())
```



d)

```
ggplot(Heart, aes(x = trestbps, y = factor(cp), fill = factor(cp))) +
  geom_boxplot(outlier.shape = 16, outlier.size = 2) +
  coord_flip() +
  labs(title = "Resting Blood Pressure and Chest Pain box plot",
       x = "Resting Blood Pressure (mm Hg)",
       y = "Chest Pain Type") +
  scale_fill_manual(values=c("deeppink","lightblue","orange", "green"), guide = "none") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, size = 15),
        axis.title = element_text(size = 12),
        axis.text.y = element_text(size = 10),
        legend.title = element_text(size = 12),
        legend.text = element_text(size = 10),
        panel.grid.major.y = element_line(color = "gray"),
        panel.grid.minor = element_blank())
```



Question 8 :

همانطور که خواسته شده است به طور تصادفی

نقاطی در یک مربع ایجاد می‌کنیم و حساب می‌کنیم که چه تعداد نقطه در دایره‌ای که در

درون مربع قرار دارد قرار می‌گیرد، سپس از مقایسه‌ی تعداد نقاط درون دایره و

نقاطی که در حالت کلی ایجاد کرده‌ایم، مقدار π به صورت تقریبی به دست می‌آید.

نسبت نقاط درون دایره به مربع، که تخمین از نسبت مساحت دایره به مساحت مربع (π/4)

است. با ضرب این نسبت به 4 می‌توانیم مقدار تقریبی π را تخمین زد. ضمنا این که

هر چقدر هم تعداد نقاط تصادفی بیشتر باشد، تخمین ما دقیق‌تر خواهد بود.

تعداد نقاطی که در نظر می‌گیریم ۱۰۰۰، ۱۰۰۰۰، ۱۰۰۰۰۰ و ۱۰۰۰۰۰۰ می‌باشد.

```r
estimate_pi <- function(num_points) {
  inside_circle <- 0
  for (i in 1:num_points) {
    x <- runif(1, -1, 1)
    y <- runif(1, -1, 1)
    if (x^2 + y^2 <= 1) {
      inside_circle <- inside_circle + 1
    }
  }
  pi_estimate <- 4 * inside_circle / num_points
  return(pi_estimate)
}

print(paste("Estimate of pi with 1000 points:", estimate_pi(1000)))
print(paste("Estimate of pi with 10000 points:", estimate_pi(10000)))
print(paste("Estimate of pi with 100000 points:", estimate_pi(100000)))
print(paste("Estimate of pi with 1000000 points:", estimate_pi(1000000)))
print(paste("True value of pi:", pi))
```

```
> print(paste("Estimate of pi with 1000 points:", estimate_pi(1000)))
[1] "Estimate of pi with 1000 points: 3.092"
> print(paste("Estimate of pi with 10000 points:", estimate_pi(10000)))
[1] "Estimate of pi with 10000 points: 3.1268"
> print(paste("Estimate of pi with 100000 points:", estimate_pi(100000)))
[1] "Estimate of pi with 100000 points: 3.13728"
> print(paste("Estimate of pi with 1000000 points:", estimate_pi(1000000)))
[1] "Estimate of pi with 1000000 points: 3.14028"
> print(paste("True value of pi:", pi))
[1] "True value of pi: 3.14159265358979"
```