9231_41_Summer_2020_Q1

Solution

To determine whether the grades awarded are independent of the presence of background music, we perform a **Chi-square test for independence**.

1. Hypotheses and Significance Level

- **Null Hypothesis** (H_0): Grade awarded and background music are independent.
- **Alternative Hypothesis** (H_1): Grade awarded and background music are not independent.
- Significance level: $\alpha = 0.10$.

2. Observed Frequencies and Totals The observed frequencies (O_{ij}) are provided in the contingency table. We first calculate the row totals, column totals, and the grand total (N).

	Grade A	Grade B	Grade C	Row Total
Music	49	51	40	140
Silence	93	68	49	210
Column Total	142	119	89	350

3. Expected Frequencies Under the assumption of independence, the **expected frequency** (E_{ij}) for each cell is calculated as:

$$E_{ij} = \frac{(\text{Row Total}) \times (\text{Column Total})}{\text{Grand Total}}$$

- For Music and Grade A: $E_{11} = \frac{140\times142}{350} = 56.8$
- For Music and Grade B: $E_{12} = \frac{140\times119}{350} = 47.6$
- For Music and Grade C: $E_{13} = \frac{140\times89}{350} = 35.6$
- For Silence and Grade A: $E_{21} = \frac{210\times142}{350} = 85.2$
- For Silence and Grade B: $E_{22} = \frac{210\times119}{350} = 71.4$
- For Silence and Grade C: $E_{23} = \frac{210\times89}{350} = 53.4$

4. Test Statistic Calculation The **Chi-square test statistic** (χ^2) is given by:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

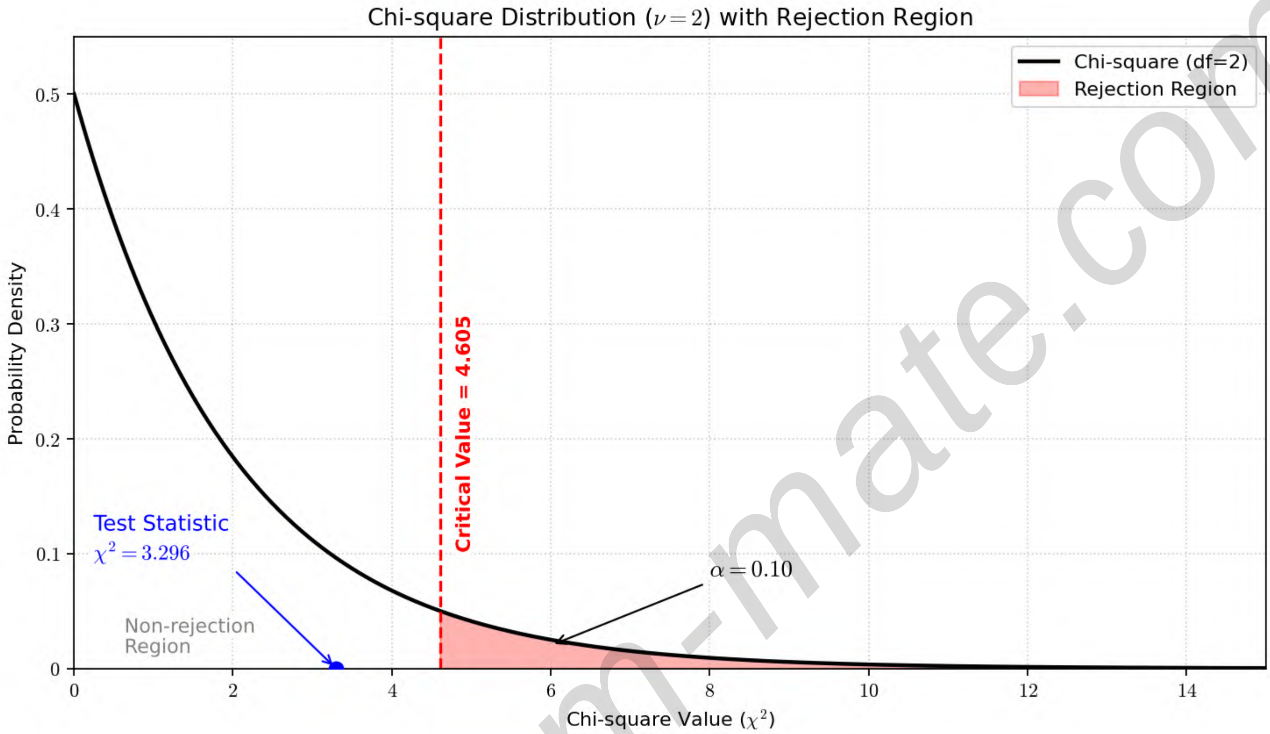
$$\begin{aligned}\chi^2 &= \frac{(49-56.8)^2}{56.8} + \frac{(51-47.6)^2}{47.6} + \frac{(40-35.6)^2}{35.6} + \frac{(93-85.2)^2}{85.2} + \frac{(68-71.4)^2}{71.4} + \frac{(49-53.4)^2}{53.4} \\ &= \frac{(-7.8)^2}{56.8} + \frac{3.4^2}{47.6} + \frac{4.4^2}{35.6} + \frac{7.8^2}{85.2} + \frac{(-3.4)^2}{71.4} + \frac{(-4.4)^2}{53.4} \\ &\approx 1.0711 + 0.2429 + 0.5438 + 0.7141 + 0.1619 + 0.3625 \\ &\approx 3.2963\end{aligned}$$

5. Degrees of Freedom and Critical Value The **degrees of freedom** (df) for a contingency table with r rows and c columns is:

$$df = (r-1)(c-1) = (2-1)(3-1) = 2$$

Using the **Chi-square distribution** table for $df = 2$ at $\alpha = 0.10$:

$$\chi^2_{0.10,2} = 4.605$$



6. Conclusion Since the calculated test statistic $\chi^2 \approx 3.296$ is less than the critical value 4.605, we fail to reject the null hypothesis H_0.

There is insufficient evidence at the 10% significance level to suggest that the grades awarded are dependent on whether background music is playing. We conclude that the grades and the presence of music are independent.

$$\boxed{\text{Fail to reject } H_0; \text{ grades and music are independent at the } 10\%\text{level.}}$$

9231_41_Summer_2020_Q2

Solution

The problem asks for a **Wilcoxon signed-rank test** to evaluate a claim about the median of a population based on a sample of 10 observations.

1. Hypotheses and Significance Level Let m be the population median time to complete the task.

- **Null Hypothesis** $H_0 : m = 6.4\,\mathrm{ms}$
- **Alternative Hypothesis** $H_1 : m \neq 6.4\,\mathrm{ms}$ (two-tailed test)
- **Significance Level** $\alpha = 0.05$

2. Calculation of Differences and Ranks We calculate the difference $d_i = x_i - 6.4$ for each observation x_i, then rank the absolute differences $\mid d_i \mid$.

x_i	$d_i = x_i - 6.4$	$\mid d_i \mid$	Rank	Signed Rank
6.44	$+0.04$	0.04	1	$+1$
6.16	-0.24	0.24	7	-7
5.62	-0.78	0.78	10	-10
5.82	-0.58	0.58	9	-9
6.51	$+0.11$	0.11	4	$+4$
6.62	$+0.22$	0.22	6	$+6$
6.19	-0.21	0.21	5	-5
6.42	$+0.02$	0.02	2	$+2$
6.34	-0.06	0.06	3	-3
6.28	-0.12	0.12	8	-8

3. Test Statistic Calculation The test statistic W is the smaller of the sum of positive ranks (W^+) and the sum of negative ranks (W^-).

- Sum of positive ranks:

$$W^+ = 1 + 4 + 6 + 2 = 13$$

- Sum of negative ranks:

$$W^- = 7 + 10 + 9 + 5 + 3 + 8 = 42$$

- Test statistic:

$$T = \min(W^+, W^-) = 13$$

4. Critical Value and Conclusion For a two-tailed test with $n = 10$ at $\alpha = 0.05$:

- From the **Wilcoxon signed-rank table**, the critical value T_{crit} is 8.
- The decision rule for this test is to reject H_0 if $T \leq T_{\text{crit}}$.
- Since $13 > 8$, we fail to reject the null hypothesis.

There is insufficient evidence at the 5% significance level to suggest that the median time to complete the task is not $6.4\,\mathrm{ms}$. The claim is supported by the data.

5. Underlying Assumption The **Wilcoxon signed-rank test** requires that the population distribution from which the sample is drawn is **symmetrical** about its median.

Final Answers:

(a)

$$\begin{aligned} H_0 &: m = 6.4 \\ H_1 &: m \neq 6.4 \\ \text{Sum of positive ranks } (W^+) &= 13 \\ \text{Sum of negative ranks } (W^-) &= 42 \\ \text{Test statistic } T &= 13 \\ \text{Critical value } (n = 10,\ \text{two-tailed},\ 5\%) &= 8 \end{aligned}$$

Since $13 > 8$, do not reject H_0. There is no significant evidence to reject the claim that the median is $6.4\,\mathrm{ms}$.

(b) The distribution of the population must be **symmetrical**.

9231_41_Summer_2020_Q3

Solution

The continuous random variable X is defined by the **probability density function** (PDF) $f(x)$:

$$f(x) = \begin{cases} \frac{3}{16}(2-\sqrt{x}) & 0 \le x < 1 \\ \frac{3}{16\sqrt{x}} & 1 \le x \le 9 \\ 0 & \text{otherwise} \end{cases}$$

1. Calculation of the Expected Value $E(X)$

The **expected value** of a continuous random variable is defined by the integral:

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

Given the piecewise nature of $f(x)$, we split the integral into two non-zero regions:

- **Region 1: $0 \le x < 1$**

$$\begin{aligned} I_1 &= \int_0^1 x \cdot \frac{3}{16}(2-\sqrt{x})\, dx \\ &= \frac{3}{16}\int_0^1 (2x - x^{3/2})\, dx \\ &= \frac{3}{16}\left[x^2 - \frac{2}{5}x^{5/2}\right]_0^1 \\ &= \frac{3}{16}\left(1-\frac{2}{5}\right) = \frac{3}{16}\cdot\frac{3}{5} = \frac{9}{80} \end{aligned}$$

- **Region 2: $1 \le x \le 9$**

$$\begin{aligned} I_2 &= \int_1^9 x \cdot \frac{3}{16\sqrt{x}}\, dx \\ &= \frac{3}{16}\int_1^9 x^{1/2}\, dx \\ &= \frac{3}{16}\left[\frac{2}{3}x^{3/2}\right]_1^9 \\ &= \frac{1}{8}\left(9^{3/2} - 1^{3/2}\right) \\ &= \frac{1}{8}(27-1) = \frac{26}{8} = \frac{13}{4} \end{aligned}$$

- **Total Expected Value:**

$$
\begin{aligned}
E(X) &= I_1 + I_2 \\
&= \frac{9}{80} + \frac{13}{4} \\
&= \frac{9}{80} + \frac{260}{80} = \frac{269}{80} = 3.3625
\end{aligned}
$$

2. Probability Density Function of $Y = \sqrt{X}$

To find the PDF of Y, we use the **method of transformations**. Let $g(x) = \sqrt{x}$. Since $g(x)$ is strictly increasing for $x \geq 0$, we have $X = Y^2$. The PDF of Y is given by:

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$$

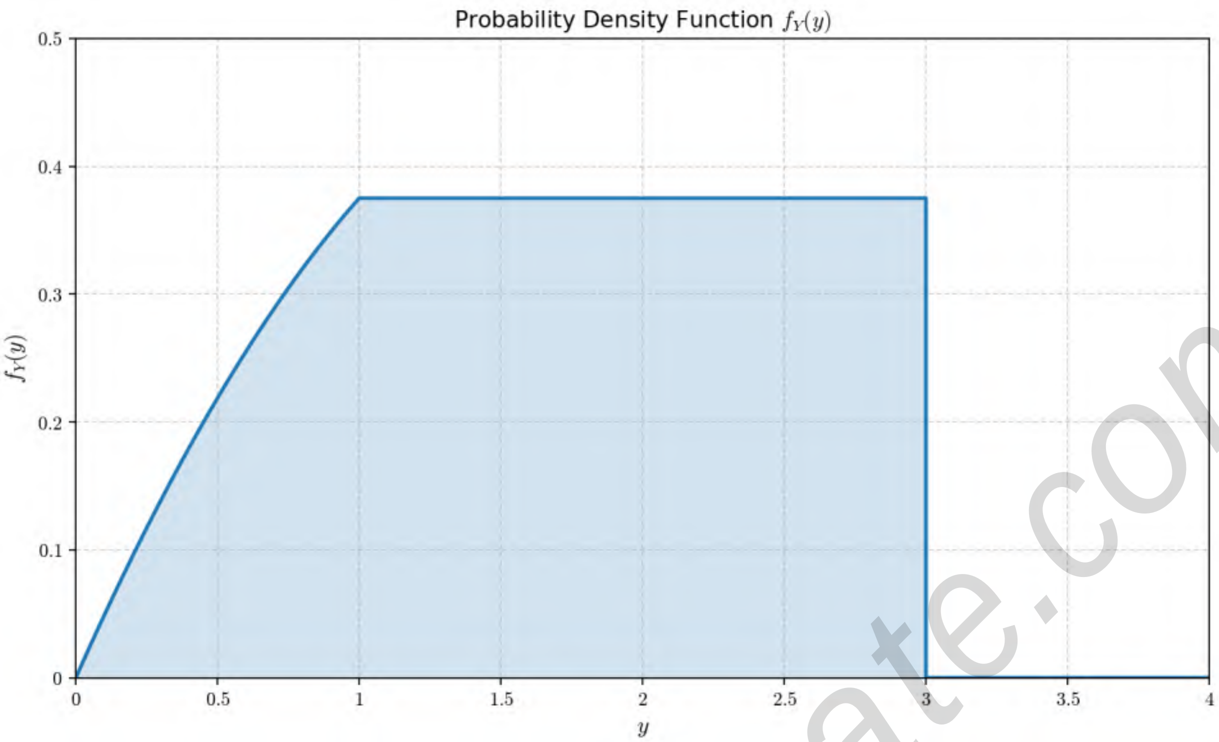
where $x = y^2$ and $\frac{dx}{dy} = 2y$.

- **Determine the range of Y:**
  - For $0 \leq x < 1$, $0 \leq y < 1$.
  - For $1 \leq x \leq 9$, $1 \leq y \leq 3$.
- **Transforming the PDF:**
  - For $0 \leq y < 1$:

$$
\begin{aligned}
f_Y(y) &= f_X(y^2) \cdot 2y \\
&= \frac{3}{16}(2 - \sqrt{y^2}) \cdot 2y \\
&= \frac{3}{16}(2 - y) \cdot 2y = \frac{3y}{8}(2 - y)
\end{aligned}
$$

  - For $1 \leq y \leq 3$:

$$
\begin{aligned}
f_Y(y) &= f_X(y^2) \cdot 2y \\
&= \frac{3}{16\sqrt{y^2}} \cdot 2y \\
&= \frac{3}{16y} \cdot 2y = \frac{3}{8}
\end{aligned}
$$



The final PDF for Y is:

$$f_Y(y) = \begin{cases} \frac{3y}{8}(2-y) & 0 \le y < 1 \\ \frac{3}{8} & 1 \le y \le 3 \\ 0 & \text{otherwise} \end{cases}$$

Final Answers:

(a) $\boxed{3.3625}$

(b) $f_Y(y) = \begin{cases} \frac{3y}{8}(2-y) & 0 \le y < 1 \\ \frac{3}{8} & 1 \le y \le 3 \\ 0 & \text{otherwise} \end{cases}$

9231_41_Summer_2020_Q4

Solution

To test the manager's claim that there is no difference between the mean volume of coffee filled by machine X and machine Y, we perform a **two-sample t-test** (or z-test, given the large sample sizes).

1. State the Hypotheses Let μ_x and μ_y be the population mean volumes for machines X and Y, respectively.

- **Null Hypothesis** (H_0): $\mu_x = \mu_y$ (or $\mu_x - \mu_y = 0$)
- **Alternative Hypothesis** (H_1): $\mu_x \neq \mu_y$ (Two-tailed test)
- Significance level: $\alpha = 0.10$

2. Calculate Sample Statistics For Machine X ($n_x = 50$):

- Sample mean: $\bar{x} = \frac{\sum x}{n_x} = \frac{15.2}{50} = 0.304$
- Unbiased estimate of variance (s_x^2):

$$s_x^2 = \frac{1}{n_x - 1}\left(\sum x^2 - \frac{(\sum x)^2}{n_x}\right)$$
$$= \frac{1}{49}\left(5.1 - \frac{15.2^2}{50}\right)$$
$$= \frac{1}{49}(5.1 - 4.6208)$$
$$= \frac{0.4792}{49} \approx 0.00977959$$

For Machine Y ($n_y = 40$):

- Sample mean: $\bar{y} = \frac{\sum y}{n_y} = \frac{13.4}{40} = 0.335$
- Unbiased estimate of variance (s_y^2):

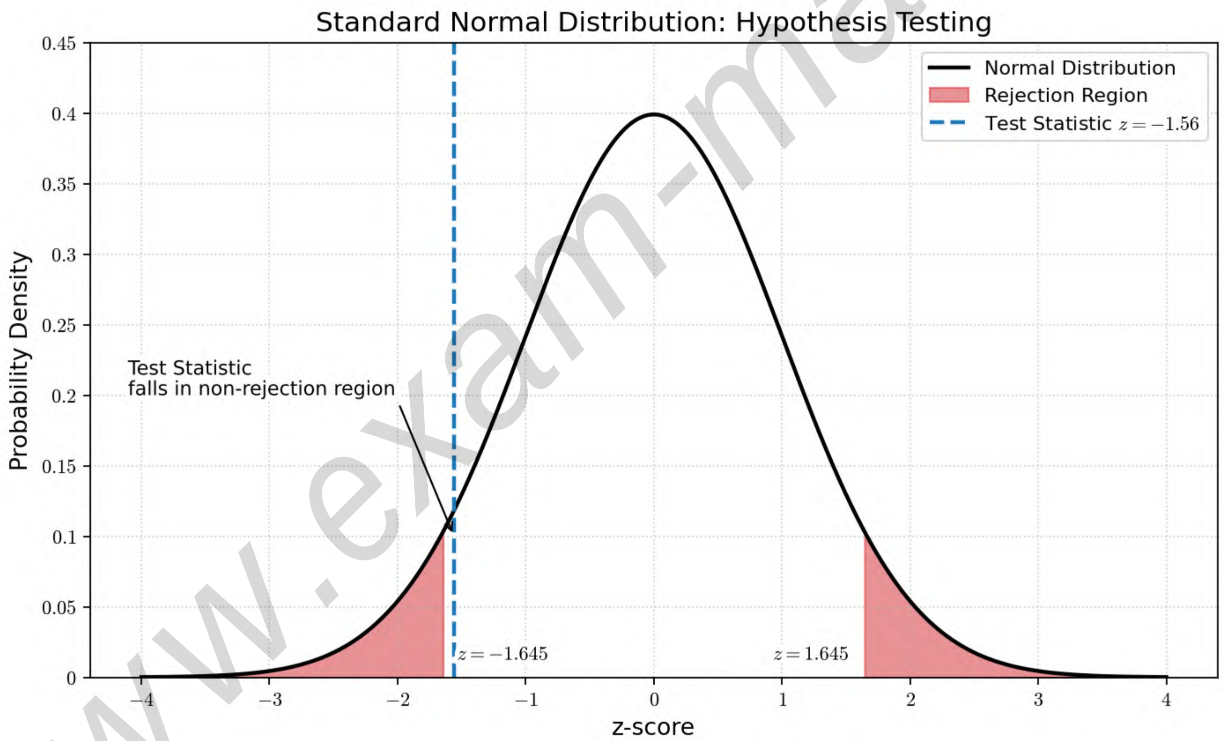
$$s_y^2 = \frac{1}{n_y - 1}\left(\sum y^2 - \frac{(\sum y)^2}{n_y}\right)$$
$$= \frac{1}{39}\left(4.8 - \frac{13.4^2}{40}\right)$$
$$= \frac{1}{39}(4.8 - 4.489)$$
$$= \frac{0.311}{39} \approx 0.00797436$$

3. Calculate the Test Statistic Since the sample sizes are large ($n_x, n_y \geq 30$), we use the **Central Limit Theorem** to justify the use of the **standard normal distribution** (Z). The test statistic is:

$$
\begin{aligned}
z &= \frac{(\bar{x}-\bar{y})-0}{\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}} \\
&= \frac{0.304-0.335}{\sqrt{\frac{0.00977959}{50}+\frac{0.00797436}{40}}} \\
&= \frac{-0.031}{\sqrt{0.00019559+0.00019936}} \\
&= \frac{-0.031}{\sqrt{0.00039495}} \\
&\approx \frac{-0.031}{0.019873} \\
&\approx -1.5599
\end{aligned}
$$

4. Determine the Critical Value and Conclusion For a two-tailed test at the 10% significance level ($\alpha = 0.10$), the critical values are obtained from the standard normal table:

- $P(Z > z_c) = 0.05 \implies z_c = 1.645$
- The **rejection region** is $|\ z\ | > 1.645$.



Comparing the calculated value to the critical value:

$$|-1.560| = 1.560 < 1.645$$

Since the test statistic does not fall in the rejection region, we fail to reject the null hypothesis. There is insufficient evidence at the 10% significance level to suggest a difference between the mean volumes of coffee filled by the two machines. The manager's claim is supported by the data.

The calculated test statistic is $z \approx -1.56$. Since $|-1.56| < 1.645$, we **accept H_0**. There is no significant evidence of a difference in means.

9231_41_Summer_2020_Q5

Solution

1. Descriptive Statistics

First, we calculate the sample mean and sample standard deviation from the given data set of $n = 8$ distances: $\{19.8, 22.1, 24.4, 21.5, 20.8, 26.3, 23.7, 25.0\}$.

- The **sample mean** ($\bar{x}$) is:

$$\begin{aligned}\bar{x} &= \frac{\sum x_i}{n} \\ &= \frac{19.8 + 22.1 + 24.4 + 21.5 + 20.8 + 26.3 + 23.7 + 25.0}{8} \\ &= \frac{183.6}{8} \\ &= 22.95\,\mathrm{m}\end{aligned}$$

- The **unbiased sample variance** (s^2) is:

$$\begin{aligned}s^2 &= \frac{\sum(x_i - \bar{x})^2}{n-1} \\ &= \frac{(19.8 - 22.95)^2 + ... + (25.0 - 22.95)^2}{7} \\ &= \frac{35.42}{7} \\ &= 5.06\end{aligned}$$

- The **sample standard deviation** (s) is:

$$s = \sqrt{5.06} \approx 2.24944\,\mathrm{m}$$

2. Hypothesis Test (Part a)

We perform a one-sample **t-test** because the population variance is unknown and the sample size is small ($n < 30$).

- **Hypotheses:**
  - **Null Hypothesis** $H_0 : \mu = 22.0$
  - **Alternative Hypothesis** $H_1 : \mu > 22.0$ (one-tailed test)
- **Test Statistic:**

$$\begin{aligned}t &= \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \\ &= \frac{22.95 - 22.0}{2.24944/\sqrt{8}} \\ &= \frac{0.95}{0.7953} \\ &\approx 1.1945\end{aligned}$$

- **Critical Value:** For a significance level $\alpha = 0.05$ and **degrees of freedom** $df = n - 1 = 7$, the critical value from the **Student's t-distribution** table is:

$$t_{0.05,7} = 1.895$$

- **Conclusion:** Since the calculated t-value (1.1945) is less than the critical value (1.895), we fail to reject the null hypothesis. There is insufficient evidence at the 5% significance level to suggest that the population mean distance is more than 22.0 metres.

3. 95% Confidence Interval (Part b)

The formula for the **confidence interval** of the mean is:

$$\bar{x} \pm t_{\alpha/2,df} \cdot \left(\frac{s}{\sqrt{n}}\right)$$

- For a 95% confidence level, $\alpha = 0.05$, so $\alpha/2 = 0.025$.
- The critical value for $df = 7$ is $t_{0.025,7} = 2.365$.
- **Calculation:**

$$\begin{aligned} \text{CI} &= 22.95 \pm 2.365 \cdot \left(\frac{2.24944}{\sqrt{8}}\right) \\ &= 22.95 \pm 2.365 \cdot (0.7953) \\ &= 22.95 \pm 1.8809 \end{aligned}$$

- **Lower Bound:** $22.95 - 1.8809 = 21.0691$
- **Upper Bound:** $22.95 + 1.8809 = 24.8309$

Rounding to three significant figures:

$$\boxed{[21.1, 24.8]}$$

9231_41_Summer_2020_Q6

Solution

1. Probability Generating Function of X

The random variable X represents the number of red balls selected when 2 balls are chosen from a bag containing 4 red and 6 blue balls without replacement. This follows a **hypergeometric distribution**. The possible values for X are $\{0, 1, 2\}$.

- The total number of ways to choose 2 balls from 10 is $\binom{10}{2} = 45$.
- $P(X=0) = \frac{\binom{4}{0}\binom{6}{2}}{45} = \frac{1\times15}{45} = \frac{1}{3}$
- $P(X=1) = \frac{\binom{4}{1}\binom{6}{1}}{45} = \frac{4\times6}{45} = \frac{24}{45} = \frac{8}{15}$
- $P(X=2) = \frac{\binom{4}{2}\binom{6}{0}}{45} = \frac{6\times1}{45} = \frac{6}{45} = \frac{2}{15}$

The **probability generating function** (PGF) is defined as $G_X(t) = \sum P(X=x)t^x$:

$$\begin{aligned} G_X(t) &= \frac{1}{3}t^0 + \frac{8}{15}t^1 + \frac{2}{15}t^2 \\ &= \frac{5}{15} + \frac{8}{15}t + \frac{2}{15}t^2 \end{aligned}$$

$$\boxed{G_X(t) = \frac{1}{15}(5 + 8t + 2t^2)}$$

2. Probability Generating Function of Y

Y is the sum of two independent Bernoulli trials (coin tosses). Let Y_1 be the result of the first coin ($P(\text{Head}) = 2/3$) and Y_2 be the result of the second coin ($P(\text{Head}) = p$). The PGF of a sum of independent variables is the product of their individual PGFs:

- $G_{Y_1}(t) = \frac{1}{3} + \frac{2}{3}t$
- $G_{Y_2}(t) = (1-p) + pt$
- $G_Y(t) = G_{Y_1}(t)G_{Y_2}(t) = \left(\frac{1}{3} + \frac{2}{3}t\right)((1-p) + pt)$

Expanding $G_Y(t)$:

$$G_Y(t) = \frac{1}{3}(1-p) + \left[\frac{1}{3}p + \frac{2}{3}(1-p)\right]t + \frac{2}{3}pt^2$$

The coefficient of t is given as $7/12$:

$$\begin{aligned} \frac{1}{3}p + \frac{2}{3} - \frac{2}{3}p &= \frac{7}{12} \\ \frac{2}{3} - \frac{1}{3}p &= \frac{7}{12} \\ \frac{1}{3}p &= \frac{8}{12} - \frac{7}{12} = \frac{1}{12} \\ p &= \frac{1}{4} \end{aligned}$$

Substituting $p = 1/4$ back into $G_Y(t)$:

$$\begin{aligned} G_Y(t) &= \left(\frac{1}{3} + \frac{2}{3}t\right)\left(\frac{3}{4} + \frac{1}{4}t\right) \\ &= \frac{3}{12} + \frac{1}{12}t + \frac{6}{12}t + \frac{2}{12}t^2 \\ &= \frac{1}{4} + \frac{7}{12}t + \frac{1}{6}t^2 \end{aligned}$$

$$\boxed{G_Y(t) = \frac{1}{12}(3 + 7t + 2t^2)}$$

3. Probability Generating Function of Z

Since $Z = X + Y$ and X, Y are independent, $G_Z(t) = G_X(t) \cdot G_Y(t)$:

$$\begin{aligned} G_Z(t) &= \left[\frac{1}{15}(5 + 8t + 2t^2)\right]\left[\frac{1}{12}(3 + 7t + 2t^2)\right] \\ &= \frac{1}{180}(5 + 8t + 2t^2)(3 + 7t + 2t^2) \\ &= \frac{1}{180}[5(3 + 7t + 2t^2) + 8t(3 + 7t + 2t^2) + 2t^2(3 + 7t + 2t^2)] \\ &= \frac{1}{180}[15 + 35t + 10t^2 + 24t + 56t^2 + 16t^3 + 6t^2 + 14t^3 + 4t^4] \\ &= \frac{1}{180}(15 + 59t + 72t^2 + 30t^3 + 4t^4) \end{aligned}$$

$$\boxed{G_Z(t) = \frac{1}{12} + \frac{59}{180}t + \frac{2}{5}t^2 + \frac{1}{6}t^3 + \frac{1}{45}t^4}$$

4. Expected Value of Z

The **expected value** can be found using the property $E(Z) = G_{Z'}(1)$. Alternatively, by the **linearity of expectation**, $E(Z) = E(X) + E(Y)$.

- For X: $E(X) = G_{X'}(1)$.

$$G_{X'}(t) = \frac{1}{15}(8 + 4t)$$

$$E(X) = G_{X'}(1) = \frac{12}{15} = 0.80$$

- For Y: $E(Y) = G_{Y'}(1)$.

$$G_{Y'}(t) = \frac{1}{12}(7 + 4t)$$

$$E(Y) = G_{Y'}(1) = \frac{11}{12} \approx 0.9167$$

- For Z:

$$\begin{aligned} E(Z) &= \frac{12}{15} + \frac{11}{12} \\ &= \frac{48}{60} + \frac{55}{60} \\ &= \frac{103}{60} \end{aligned}$$

$$\boxed{E(Z) = \frac{103}{60} \approx 1.717}$$