CE-003: Estatística II - Turma: O, Avaliações Semanais 1^o semestre/2012

1. Foi feita uma pesquisa sobre as condições salariais de professores de um certo estado. Os dados foram organizados em uma tabela. A seguir é mostrada uma pequena fração dos dados. A descrição dos atributos anotados está na Tabela abaixo.

```
  Degree Rank Sex Year YSdeg Salary
1      1    3   0   25    35  36350
2      1    3   0   13    22  35350
3      1    3   0   10    23  28200
4      1    3   1    7    27  26775
5      0    3   0   19    30  33696
6      1    3   0   16    21  28516
```

Atributo	Descrição
Degree	Formação: 1: Doutorado, 0: Mestrado
Rank	Cargo (1: Prof Assistente, 2: Prof Associado, 3: Prof Pleno)
Sex	1: feminino, 0: masculino
Year	Anos de trabalho
YSdeg	Anos desde a obtenção da maior titulação
Salary	Salário em dolares por ano

(a) Classifique cada um dos atributos (variáveis).

(b) Esboçe um gráfico adequado para resumir cada um dos atributos individualmente

Solução:

(a) Sex: Qualitativa nominal
Degree, Rank: Qualitativa ordinal
Anos de trabalho*, tempo de titulação*: contínua (mas podendo ser tratada como discreta)
salário: contínua

(b) Esboçe um gráfico adequado para resumir cada um dos atributos individualmente

---

2. Foram registrados o tempo de execução (em segundos) de rotinas enviadas por vinte programadores.

```
10.4 13.8 51.0 17.6 18.5 23.8  5.2  9.3 11.5 22.7

18.2 10.5 24.4  2.2 28.4 11.3  6.4 14.0  6.7  8.9
```

(a) faça um histograma dos dados

(b) faça um gráfico *boxplot*

(c) faça um diagrama ramo-e-folhas

(d) obtenha a média e desvio padrão

(e) obtenha o coeficiente de variação

(f) obtenha a amplitude e a amplitude interquartílica

(g) caracterize/discuta a distribuição dos dados

(a)

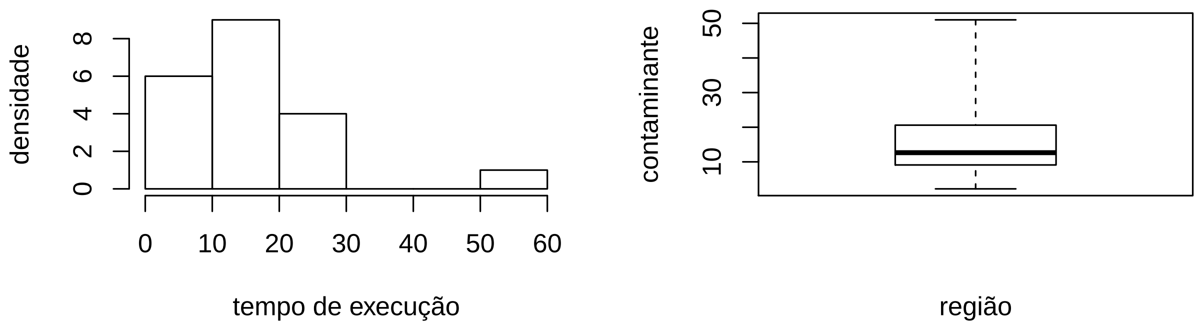

Figura 1: Histograma (esquerda) e *boxplot* (direita) dos tempos de execução.

Solução:

(b) *> stem(x)*

```
The decimal point is 1 digit(s) to the right of the |

0 | 256799
1 | 011244889
2 | 3448
3 |
4 |
5 | 1
```

(c) *> c(media = mean(x), desvioPadrao = sd(x))*

```
  media desvioPadrao
  15.74        10.91
```

(d) obtenha o coeficiente de variação

*> 100 * sd(x)/mean(x)*

```
[1] 69.31
```

(e) *> c(A = diff(range(x)), AI = unname(diff(quantile(x)[c(2,4)])))*

```
    A    AI
48.80 10.35
```

(f) *Comentar sobre: posição, variabilidade, assimetria e dados atípicos*

---

3. Uma série de características químicas foram medidas em diferentes vinhos. Os gráficos a seguir mostram quatro delas. Discuta os gráficos e suas interpretações utilizando conceitos e princípios de análise estatística descritiva/exploratória de dados. Inclua na sua discussão possíveis tratamentos dos dados.

Solução: Discussões/comentários devem incluir:

- análises univariadas de cada elemento: posição, variação, assimetria/transformação, dados discrepantes
- análises bivariadas: existência de relação, linearidade, monotonicidade, dados discrepantes, intensidade da relação, possíveis efeitos de transformações

---

4. Um algorítmo de classificação deve tentar resolver corretamente dois problemas, A e B. A probabilidade resolver A corretamente é de 0,6. Caso resolva A corretamente, a probabilidade de resolver B corretamente é de 0,85; caso contrário, essa probabilidade é de 0,25.

(a) Qual a probabilidade de ele:

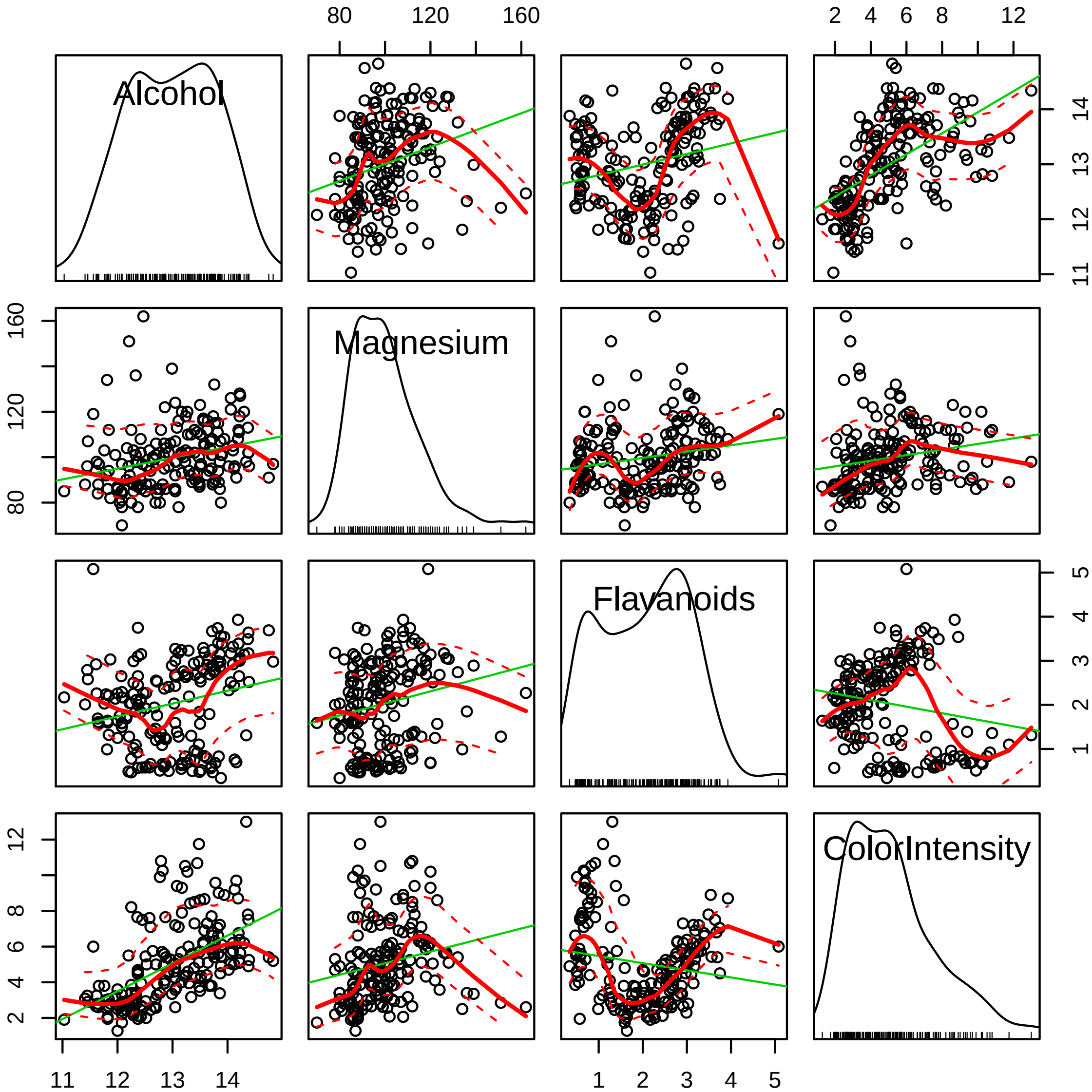

Figura 2: Algumas características de amostras de vinhos.

- resolver corretamente os dois problemas?
- resolver corretamente apenas um dos problemas?
- não resolver nenhum corretamente?

(b) os eventos "resolver corretamente A" e "resolver corretamente B",

- são independentes? (justifique)
- são mutuamente exclusivos? (justifique)

Solução:

$$A: \text{ resolver corretamente o problema A}$$
$$B: \text{ resolver corretamente o problema B}$$
$$P[A] = 0,6 \quad ; \quad P[B|A] = 0,85 \quad ; \quad P[B|\overline{A}] = 0,25$$
$$P[\overline{A}] = 0,4 \quad ; \quad P[\overline{B}|A] = 0,15 \quad ; \quad P[\overline{B}|\overline{A}] = 0,75$$

(a) • $P[A \cap B] = P[A] \cdot P[B|A] = (0,6) \cdot (0,85) = 0.51$

- $P[A \cap \overline{B}] + P[\overline{A} \cap B] = P[A] \cdot P[\overline{B}|A] + P[\overline{A}] \cdot P[B|\overline{A}] = (0,6) \cdot (0,15) + (0,4) \cdot (0,25) = 0.19$
- $P[\overline{A} \cap \overline{B}] = P[\overline{A}] \cdot P[\overline{B}|\overline{A}] = (0,4) \cdot (0,75) = 0.3$

(b)
- Não, pois $P[A \cap B] = 0.51 \neq P[A] \cdot P[B] = 0.61$,
  em que $P[B] = P[B \cap A] + P[B \cap \overline{A}] = (0,6)(0,85) + (0,4)(0,25) = 0.61$
- Não pois $P[A \cap B] \neq 0$

5. Considere o lançamento de uma moeda 10 vezes.

   (a) Qual a probabilidade de obter a face ”cara” em todos os lançamentos?
   (b) Considere agora que 1.000 pessoas fazem o mesmo. Qual a probabilidade de que alguém obtenha 10 ”caras”?
   (c) Qual(ais) a(s) suposição(ções) feita(s) nos cálculos?
   (d) Discuta e interprete os resultados.

   Solução:

   (a)

$$
\begin{aligned}
(1/2)^{10} &= 1/1024 = 0.00098 \\
&\text{ou} \\
&X : \text{número de caras em 10 lançamentos} \\
&X \sim B(n = 10, p = 1/2) \\
P[X = 10] &= \binom{10}{10}(1/2)^{10}(1 - 1/2)^{10-10} = 0.00098
\end{aligned}
$$

   (b)

$$
\begin{aligned}
1 - (1 - (1/2)^10)^{1000} &= 0.62358 \\
&\text{ou} \\
&Y : \text{número pessoas que obtém 10 caras} \\
&Y \sim B(n = 1000, p = (1/2)^10) \\
P[Y \geq 1] = 1 - P[Y = 0] = 1 - &\binom{1000}{0}[(1/2)^{10}]^0[1 - (1/2)^{10}]^{1000-0} = 0.62358
\end{aligned}
$$

   (c) Independência entre lançamentos, independência entre resultados de diferentes pessoas, probabilidades constantes de lançamentos de pessoas obterem 10 caras.
   (d)

6. Responda as questões a seguir declarando a variável aleatória e a sua distribuição.

   (a) Registros de um sistema mostram que 1 a cada 20 requisições de acesso de um determinado serviço não são completadas.
      i. Se forem feitas 15 requisições qual a probabilidade de que no máximo duas não sejam completadas.
      ii. Em um teste para avaliar o sistema requisições serão feitas sequencialmente até que a primeiro acesso não seja completado. Se este teste for feito diversas vezes, anotando-se o número de acessos a cada teste, qual deve ser o valor da média do número de acessos? Como voce calcularia a probabilidade de que o esse número de acessos não chegue a 5?
      iii. O teste anterior foi repetido porém até que o terceiro acesso não fosse completado. Em um particular ensaio foram feitas 10 análise desta forma. Qual a probabilidade desta ocorrência? Se este teste for repetido diversas vezes e o número de acessos anotado, qual deve ser a média do número de acessos?

      Solução:

$$p = 1/20$$

      i.

$$
\begin{aligned}
&X : \text{número de requisições não atendidas em 15 solicitações} \\
&X \sim B(n = 15, p = 1/20) \\
P[X \leq 2] &= P[X = 0] + P[X = 1] + P[X = 2] = 0.9638
\end{aligned}
$$

ii.

$$X : \text{número de requisições } \textbf{atendidas} \text{ até a primeira não completada}$$
$$X \sim G(p = 1/20) \quad P[X = x] = p \cdot (1-p)^x, x = 0, 1, \ldots$$
$$E[X] + 1 = \frac{1-p}{p} + 1 = \frac{1}{p} = 20$$
$$P[X < 4] = \sum_{i=0}^{4} P[X = i] = P[X = 0] + P[X = 1] + P[X = 2] + P[X = 3] + P[X = 4] = \sum_{i=0}^{4}(1/20)(1-1/20)^i = 0.2262$$

iii.

$$X : \text{número de requisições } \textbf{atendidas} \text{ até a terceira não completada}$$
$$X \sim BN(r = 3, p = 1/20) \quad ; \quad P[X = x] = \binom{x+3-1}{3}(1/20)^3 \cdot (1-1/20)^x, x = 0, 1, \ldots$$
$$10 \text{ requisições } \longrightarrow 7 \text{ atendidas}$$
$$P[X = 7] = 0.0031$$
$$E[X] + 3 = r\frac{1-p}{p} + 1 = \frac{r}{p} = 60$$

(b) Tem-se um conjunto de 40 sensores das quais 15 estão danificados. Uma transmissão é feita para 12 sensores foram selecionadas ao acaso. Qual a probabilidade da transmissão ter sido enviada para 4 ou mais sensores operantes?

$$X : \text{número de sensores operantes dentre os 12}$$
$$X \sim HG(N = 40, r = (40-15 = 25), n = 12) \quad ; \quad P[X = x] = \frac{\binom{25}{x}\binom{15}{12-x}}{\binom{40}{12}}$$
$$P[X \geq 4] = 1 - P[X \leq 3] = 1 - \sum_{i=0}^{2} P[X = i] = 1 - P[X = 0] - P[X = 1] - P[X = 2] - P[X = 3] = 0.9978$$

(c) Considere agora uma transmissão de dados que tem uma taxa de falha de 5,2 falhas por hora. Qual a probabilidade de que em um intervalos de 15 minutos não haja nenhuma falhas de transmissão? E de que seja registradas mais do que 4 falhas?

$$X : \text{número de falhas em 15 minutos}$$
$$X \sim P(\lambda = 5,2/4 = 1,3) \quad P[X = x] = \frac{e^{-1,3} 1,3^x}{x!}, x = 0, 1, \ldots$$
$$P[X = 0] = 0.2725$$
$$P[X > 4] = 1 - P[X \leq 3] = 1 - (P[X = 0] + P[X = 1] + P[X = 2] + P[X = 3] + P[X = 4]) = 1 - \sum_{i=0}^{4} \frac{e^{-1,3} 1,3^i}{i!} = 0.0107$$

---

7. (a) O peso de um tênis de corrida sofisticado é normalmente distribuído com média de 12 onças (onça é uma unidade de peso) e desvio padrão de 0,5 onças.

   i. qual a probabilidade de um tênis pesar mais que 13,2 onças?
   ii. qual a probabilidade de um tênis pesar entre 11,6 e 12,7 onças?
   iii. quanto deveria ser o desvio padrão para que 99,9% dos tênis tenham menos que 13 onças?
   iv. se o desvio padrão se mantiver em 0,5, quanto deveria ser a média para que 99,9% dos tênis tenham menos que 13 onças?

Solução:

$$X \sim N(12, 0.5^2)$$

i. $P[X > 13,2] = P[Z > \frac{13,2-12}{0,5}] = P[Z > 2.4] = 0.0082$

ii. $P[11,6 < X < 12,7] = P[\frac{11,6-12}{0,5} < Z < \frac{12,7-12}{0,5}] = P[-0.8 < Z < 1.4] = 0.707$

iii.

$$\begin{aligned}
X &\sim N(12, \sigma^2) \\
P[X < 13] &= 0,999 \ \ ; \ \ \sigma = ? \\
P[Z < \frac{13-12}{\sigma}] &= 0,999 \\
z &= 3.09 \\
\frac{13-12}{\sigma} &= 3.09 \\
\sigma &= 0.324
\end{aligned}$$

iv. se o desvio padrão se mantiver em 0,5, quanto deveria ser a média para que 99,9% dos tênis tenham menos que 13 onças?

$$\begin{aligned}
X &\sim N(\mu, 0.5^2) \\
P[X < 13] &= 0,999 \ \ ; \ \ \mu = ? \\
P[Z < \frac{13-\mu}{0,5}] &= 0,999 \\
z &= 3.09 \\
\frac{13-\mu}{0,5} &= 3.09 \\
\mu &= 13 - 0,5(3.09) \\
\mu &= 11.455
\end{aligned}$$

(b) Um teste de aptidão feito por pilotos de aeronaves em treinamento inicial requer que uma série de operações seja realizada em uma rápida sucessão. Suponha que o tempo necessário para completar o teste seja distribuído de acordo com o modelo normal, com média de 90 minutos e desvio padrão de 20 minutos.

i. Para passar no teste, o candidato deve completá-lo em menos de 80 minutos. Se 65 candidatos tomam o teste, quantos são esperados passar?

ii. Se os 5% melhores candidatos são alocados para aeronaves maiores, quão rápido deve ter sido o candidato para que obtenha esta posição?

iii. Se forem sorteados ao acaso 5 candidatos, qual a probabilidade de tomar ao menos 2 aprovados?

Solução:

$$X \sim N(90, 20^2)$$

i.

$$\begin{aligned}
P[X < 80] = P[Z < \frac{80-90}{20}] &= P[Z < -0.5] = 0.309 \\
65 \cdot P[X < 80] &= 20 \text{ candidatos.}
\end{aligned}$$

ii.

$$\begin{aligned}
P[X < t] &= 0,05 \ \ ; \ \ t = ? \\
P[Z < \frac{t-90}{20}] &= 0,05 \\
z &= -1.645 \\
\frac{t-90}{20} &= -1.645 \\
t &= 90 + 20(-1.645) \\
t &= 57.103
\end{aligned}$$

iii.

$$\begin{aligned}
Y &: \text{ número de aprovados dentre 5 candidatos} \\
Y &\sim B(n = 5, p = P[X < 80] = 0.309) \\
P[Y \geq 2] &= 1 - P[Y = 0] - P[Y = 1] = 0.489
\end{aligned}$$

8. Um máquina enche latas com um volume nominal de certo material, entretanto há variações de volume de uma lata para outra. A fim de monitorar e controlar o processo foi tomada uma amostra de latas e medido o volume individualmente (em ml). fornecendo os dados mostrados a seguir.

```
[1] 362 357 359 358 365 362 355 360 363 364 360 360
```

(a) Caracterize o volume de enchimento das latas através de um resumo estatístico adequado dos dados.

(b) Qual medida amostral voce escolheria para estimar o real volume de enchimento do equipamento?

(c) Qual a estimativa deste valor? Como voce representaria a incerteza sobre este estimativa?

(d) A especificação de volume é de 350 ml e caso o valor seja ultrapassado ou esteja abaixo do especificado, a processo deve ser parado para a máquina ser regulada. Baseando-se nos dados, voce indicaria a parada para regulagem?

(e) Idem anterior para volume de 360 ml.

---

9. Considere novamente o problema apresentado na avaliação da semana passada.

Um máquina enche latas com um volume nominal de certo material, entretanto há variações de volume de uma lata para outra. A fim de monitorar e controlar o processo foi tomada uma amostra de latas e medido o volume individualmente (em ml). fornecendo os dados mostrados a seguir. Se o volume for diferente de 350 ml o processo deve ser parado para verificações.

```
[1] 362 357 359 358 365 362 355 360 363 364 360 360
```

Utilizando os conceitos e resultados discutidos nas últimas aulas responda aos seguintes items. Faça a suposição de que os volumes das latas possuem uma variância de 9 unidades.

(a) Obtenha estimativas pontual e intervalar para o volume das latas.

(b) Quais as suposições que foram feitas na obtenção destas estimativas?

(c) Utilize os resultados para indicar se recomenda ou não uma parada para regulagem, justificando sua resposta.

(d) Qual deveria ser o tamanho mínimo da amostra para que a margem de erro da estimativa fosse de no máximo de 1 unidade (com nível de confiança de 95%).

(e) Para amostra do tamanho atual $n = 12$, qual deveria ser a variância dos volumes da lata para que a margem de erro não ultrapasse 0,5 unidades?

Solução:

$X \sim N(\mu \;\; ; \;\; \sigma^2 = 9 = 3^2)$

(a)

i. Estimador Pontual: média amostral, $\overline{X} = \frac{\sum_{1=1}^{n} X_i}{n}$
   Estimativa pontual: $\overline{x} = 360.42$

ii. Estimador Intervalar: $\overline{X} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$
   Estimativa Intervalar (para confiança de 90%): $360.42 \pm 1.645 \cdot \frac{3}{\sqrt{12}} = (359.94; 360.89)$

(b) amostra aleatória (independência), variância populacional conhecida (σ^2=9), confiança de 90%, média amostral como estimador adequado.

(c) Um possível critério é: verificar se o valor de referência (350) está ou não contido no intervalo; não estando rocomenda-se a parada. Respostas e justificativas serão analisadas.

(d) Vamos assumir nível de confiança de 90%

$$M.E. = z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}}$$

$$1 = 1.645 \cdot \frac{3}{\sqrt{n_0}}$$

$$n_0 = \left(\frac{1.645 \cdot 3}{1}\right)^2 = 25$$

(e)

$$M.E. = z_{(1-\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}}$$
$$1,5 = 1.645 \cdot \frac{\sigma}{\sqrt{12}}$$
$$\sigma = \frac{1.5^2 \cdot 12}{1.645} = 16.41$$

---

10. (a) Considere uma pesquisa para estimar a proporção de votos de um determinado candidato.
   i. Qual o tamanho necessário de uma amostra para uma margem de erro de 2,5% e confiança de 95% ?
   ii. Comente sobre as suposições feitas para o cálculo anterior.

Solução:

$$X \sim B(\theta)$$
$$\mu_x = E[X] = \theta$$
$$\sigma^2_X = Var[X] = \theta(1-\theta)$$

i.

$$\overline{\theta} \approx N(\mu_{\overline{X}} = \mu_X = \theta; \sigma^2_{\overline{X}} = \sigma^2_X/n = \frac{\theta(1-\theta)}{n})$$
$$M.E. = z\sqrt{\frac{\theta(1-\theta)}{n}} = 0,025$$
$$\text{assumindo } \theta = 1/2$$
$$M.E. = 1.96\frac{1}{\sqrt{4 \cdot n}} = 0,025$$
$$n = \frac{1,96^2}{4 \cdot 0,025^2} = 1537$$

ii.

(b) Assume-se que o tempo de votação em uma urna eletrônica possui distribuição uniforme com valores entre 5 e 30 segundos. Considere um grupo de 50 votantes.
   i. Qual a probabilidade do último votante gastar mais que 20 segundos?
   ii. Qual a probabilidade do tempo de votação de todo o grupo ser superior a 15 minutos?
   iii. Comente sobre as suposições feitas para o cálculo anterior.

Solução:

$$X \sim U(5,30)$$
$$\mu_x = E[X] = \frac{a+b}{2} = \frac{5+30}{2} = 17,5$$
$$\sigma^2_X = Var[X] = \frac{(b-a)^2}{12} = \frac{(30-5)^2}{12} = 52.08$$

i. $P[X > 20] = \frac{30-20}{30-5} = 0.4$

ii.

$$\overline{X} \approx N(\mu_{\overline{X}} = \mu_X = 17,5 \;\; ; \;\; \sigma^2_{\overline{X}} = \sigma^2_X/n = 1.04)$$
$$P[\sum_{i=1}^{50} X_i > 15' \cdot 60] = P[\frac{\sum_{i=1}^{50} X_i}{50} > \frac{15 \cdot 60''}{50}] = P[\overline{X} > 18] = 0.31$$

iii.
