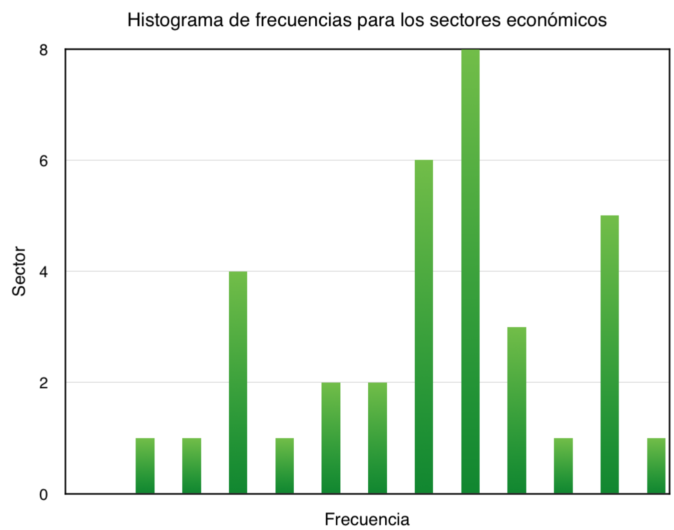
Sebastián Valencia Calderón 201111578
Edgar Daniel Gómez 201317343

Punto 1

(a) **(4 puntos)** Construya un histograma de frecuencias de las 35 empresas de acuerdo con el sector al que pertenece cada una. Muestre la tabla de frecuencias y la gráfica del histograma. ¿Cuáles son los 3 sectores de mayor frecuencia? ¿Qué porcentaje de la muestra es representado por las empresas que pertenecen a estos sectores? ¿Cuáles son los sectores de menor frecuencia? ¿Qué porcentaje de la muestra es representado por las empresas que pertenecen a estos sectores?

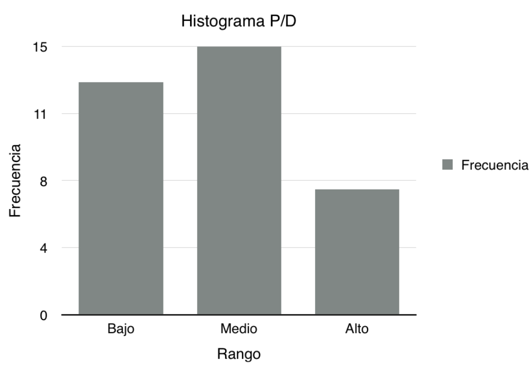


Los sectores de mayor frecuencia son el financiero, el petrolero y el energético cada uno con 8, 5, y 6 empresas representativas. Cada una con 22.85 %, 14.28 % y 17.14 % respectivamente. Los sectores de menos frecuencia son Aeronáutico, Aeronáutico, Comunicación, Comunicación y Textil. Cada uno con el 2.85 %.

(b) **(4 puntos)** Construya un diagrama de caja que represente la distribución del precio promedio de la acción de las 35 empresas de la muestra. Comente sobre todos los elementos que representa el diagrama de caja.

Se observa que la mayoría de los precios están sobre el delimitador de la caja.

(c) **(3 puntos)** Se desea analizar una nueva variable correspondiente al cociente entre el precio de la acción y los dividendos que promete (P/D). De acuerdo con este indicador se clasifica cada empresa de la siguiente manera:



(d) **(2 puntos)** Los inversionistas están interesados en conocer cuáles son los tres sectores que prometen en promedio mayores dividendos. Para ello construya y presente una tabla dinámica que le permita comparar cada sector en términos de sus dividendos promedio.

Etiquetas de fila	Promedio de Dividendo
Comercio	531
Alimentación	396
Financiero	326,925
Inversiones	289,3766667
Empaques	229,425
Minero	170
Petrolero	165,108
Energético	138,4433333
Cemento	123,7333333
Aeronáutico	50
Textil	45
Construcción	24,87
Comunicación	19
Total general	**206,5862857**

(e) **(2 puntos)** Los inversionistas han decidido establecer la siguiente clasificación partir de la ubicación de las empresas en el ranking Merco:

(f) **(3 puntos)** Obtenga las estadísticas descriptivas de la variable precio/dividendo. ¿Qué puede concluir sobre la simetría y la curtosis de la distribución de los datos para la variable precio/dividendo? ¿Cuál es la diferencia conceptual entre el error típico y la desviación estándar?

(g) **(4 puntos)** Presente la tabla de percentiles del precio promedio de las acciones.

(I) ¿A qué compañía pertenece el percentil 50?
Cartón de Colombia

(II) ¿Cuál es el precio de la acción y los dividendos asociados con esta compañía?
4500 COP y dividendos de 323.65' COP

(III) ¿Cómo interpreta este resultado? Que este dato separa los datos en dos, es decir, e 50 % de los datos están arriba y los otros abajo

Punto 2

(a) **(3 puntos)** Calcule el sesgo para cada uno de los estimadores planteados. ¿Cuáles de estos estimadores son insesgados?

Para que los estimadores planteados sean insesgados, debe ser cierto que $\mathbb{E}(\hat{\mu}_{1,2,3}) = \mu$.

Para cada uno de los $\hat{\mu}_i \forall i \in 1,2,3$, se verifica la propiedad:

$$\hat{\mu}_1 \Rightarrow \mathbb{E}(\hat{\mu}_1) = \bar{X} = \mathbb{E}\left(\frac{1}{n} \times \sum_{i=1}^{n} X_i\right) = \frac{1}{n} \times \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n} \times (n \times \mu) = \mu$$

$$\hat{\mu}_2 \Rightarrow \mathbb{E}(\hat{\mu}_2) = \mathbb{E}\left(\frac{X_1 + 2X_2 + 3X_3}{6}\right) = \mathbb{E}\left(\frac{1}{6} \times \sum_{i=1}^{3} iX_i\right) = \frac{1}{6} \times \sum_{i=1}^{3} iX_i$$

$$= \frac{\mu}{6} \times \sum_{i=1}^{3} i = \frac{\mu}{6} \times (1+2+3) = \frac{\mu}{6} \times 6 = \mu$$

$$\hat{\mu}_3 \Rightarrow \mathbb{E}(\hat{\mu}_3) = \mathbb{E}\left(\frac{X_1 + X_6 + X_{10}}{4}\right) = \frac{1}{4} \times (\mathbb{E}(X_1) + \mathbb{E}(X_6) + \mathbb{E}(X_{10})) = \frac{3 \times \mu}{4}$$

Como se puede ver los dos primeros estimadores son insesgados, mientras que el último es sesgado. El sesgo de cada estimador, se muestra en las ultimas partes de las igualdades.

(b) **(3 puntos)** Calcule el error cuadrático medio de los tres estimadores.

$$\text{ECM}(\mu_{1,2,3}) = \text{Var}(\mu_{1,2,3})) + \text{Sesgo}^2(\mu_{1,2,3}))$$

$$\text{Var}(\mu_{1,2,3})) \Rightarrow$$

$$\text{Var}(\mu_1) = \text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \times \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \times \sum_{i=1}^{n} \text{Var}\left(X_i\right) = \frac{\sigma^2}{n}$$

$$\Rightarrow \text{ECM}(\mu_1) = \frac{\sigma^2}{n}$$

$$\text{Var}(\mu_2) = \text{Var}(\bar{X}) = \text{Var}\left(\sum_{i=1}^{3} i \times X_i\right) = \frac{1}{36} \times \left(\sum_{i=1}^{3} i \times \text{Var}(X_i)\right)$$

$$= \frac{\sigma^2}{36} \times \left(\sum_{i=1}^{3} i^2 \right) = \frac{\sigma^2}{36} \times (9+4+1) = \frac{7}{18} \times \sigma^2$$

$$\Rightarrow \text{ECM}(\mu_2) = \frac{7}{18} \times \sigma^2$$

$$\text{Var}(\mu_3) = \text{Var}(\frac{X_1 + X_6 + X_{10}}{4}) = \frac{1}{16} \times \left(\sum_{i \in \{1,6,10\}} \text{Var}(X_i) \right) = \frac{3 \times \sigma^2}{16}$$

$$\Rightarrow \text{ECM}(\mu_3) = \frac{3 \times \sigma^2}{16} + \left(\frac{\mu}{4}\right)^2$$

(c) **(3 puntos)** Para cada uno de los estimadores, evalúe si es consistente o si no lo es.

Para cada uno de los $\hat{\mu}_i \forall i \in 1, 2, 3$, se verifican las propiedades:

$$\lim_{n\to\infty} \mathbb{E}(\mu_i) = \mu \wedge \lim_{n\to\infty} \text{Var}(\mu_i) = 0$$

Para μ_1 :

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}(\mu_i) &= \lim_{n\to\infty} \mathbb{E}(\bar{X}) \\
\lim_{n\to\infty} \mathbb{E}(\mu_i) &= \lim_{n\to\infty} \mu \\
\lim_{n\to\infty} \mathbb{E}(\mu_i) &= \mu \\
\lim_{n\to\infty} \text{Var}(\mu_1) &= \lim_{n\to\infty} \text{Var}(\bar{X}) \\
\lim_{n\to\infty} \text{Var}(\mu_1) &= \lim_{n\to\infty} \frac{\sigma^2}{n} \\
\lim_{n\to\infty} \text{Var}(\mu_1) &= 0
\end{aligned} \tag{1}$$

Por la verificación de las propiedades, se sabe que μ_1 es consistente.

Para μ_2 :

$$
\begin{aligned}
\lim_{n\to\infty} \mathbb{E}(\mu_i) = \lim_{n\to\infty} \mathbb{E}\left(\frac{X_1 + 2X_2 + 3X_3}{6}\right) \\
\lim_{n\to\infty} \mathbb{E}(\mu_i) = \lim_{n\to\infty} \mu \\
\lim_{n\to\infty} \mathbb{E}(\mu_i) = \mu \\
\lim_{n\to\infty} \operatorname{Var}(\mu_1) = \lim_{n\to\infty} \operatorname{Var}\left(\frac{X_1 + 2X_2 + 3X_3}{6}\right) \\
\lim_{n\to\infty} \operatorname{Var}(\mu_1) = \lim_{n\to\infty} \frac{7 \times \sigma^2}{18} \\
\lim_{n\to\infty} \operatorname{Var}(\mu_1) = \frac{7 \times \sigma^2}{18} \neq 0
\end{aligned} \tag{2}
$$

Por la verificación de las propiedades, se sabe que μ_2 no es consistente.

Para μ_3 :

$$
\begin{aligned}
\lim_{n\to\infty} \mathbb{E}(\mu_i) = \lim_{n\to\infty} \mathbb{E}\left(\frac{X_1 + X_6 + X_{10}}{4}\right) \\
\lim_{n\to\infty} \mathbb{E}(\mu_i) = \lim_{n\to\infty} \frac{3 \times \mu}{4} \\
\lim_{n\to\infty} \mathbb{E}(\mu_i) = \frac{3 \times \mu}{4} \neq \mu \\
\lim_{n\to\infty} \operatorname{Var}(\mu_1) = \lim_{n\to\infty} \operatorname{Var}\left(\frac{X_1 + X_6 + X_{10}}{4}\right) \\
\lim_{n\to\infty} \operatorname{Var}(\mu_1) = \lim_{n\to\infty} \frac{7 \times \sigma^2}{18} \\
\lim_{n\to\infty} \operatorname{Var}(\mu_1) = \frac{7 \times \sigma^2}{18} \neq 0
\end{aligned} \tag{3}
$$

Por la verificación de las propiedades, se sabe que μ_3 no es consistente.

(d) **(3 puntos)** De acuerdo a las conclusiones obtenidas en los literales anteriores, ¿cuál de los tres estimadores sería el más adecuado? Justifique su respuesta.

Es necesario un estimador insesgado y consistente, además, debe minimizar el error cuadrático medio. Considerando los anteriores resultados, el mejor estimador es $\hat{\mu}_1$.

Punto 3

(a) **(6 puntos)** Encuentre el estimador de máxima verosimilitud para dichos parámetros. Para esto siga los siguientes pasos:

(I) Plantee la función de verosimilitud:

$$
\begin{aligned}
L(x_1,\ldots,x_p;\mu,\sigma^2) &= \prod_{i=1}^{n} f_X(x_i) \\
&= \prod_{i=1}^{n}\left[\frac{1}{\sigma\times\sqrt{2\pi}}\exp -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \\
&= \left(\frac{1}{\sigma\times\sqrt{2\pi}}\right)^n \prod_{i=1}^{n}\left[\exp -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \\
&= \frac{1}{\sigma^n\times\sqrt{(2\pi)}^n}\times\exp -\frac{1}{2}\sum_{i=2}^{n}\left(\frac{x-\mu}{\sigma}\right)^2
\end{aligned}
\tag{4}
$$

(II) Halle el logaritmo natural de la función de verosimilitud.

$$
\begin{aligned}
ln(L(x_1,\ldots,x_p;\mu,\sigma^2)) &= ln\left(\frac{1}{\sigma^n\times\sqrt{(2\pi)^n}}\times\exp -\frac{1}{2}\sum_{i=2}^{n}\left(\frac{x-\mu}{\sigma}\right)^2\right) \\
&= \left(0-ln(2\pi^{n/2})\right)+\left(0-ln((\sigma^2)^{n/2})\right)+\left(-\frac{1}{2}\sum_{i=2}^{n}\left(\frac{x-\mu}{\sigma}\right)^2\right) \\
&= -\frac{n\times ln(2\pi)}{2}-\frac{n\times ln(\sigma^2)}{2}-\frac{1}{2}\times\sum_{i=2}^{n}\left(\frac{x-\mu}{\sigma}\right)^2
\end{aligned}
\tag{5}
$$

(III) Derive el resultado anterior respecto al parámetro correspondiente.
Para μ :

$$
\begin{aligned}
\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right) &= 0 \\
\sum_{i=1}^{n} x_i - n\mu &= 0 \\
\mu &= \frac{\sum_{i=1}^{n} x_i}{n}
\end{aligned}
\tag{6}
$$

Para σ :

$$
\begin{aligned}
-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^n)^2} \times \sum_{i=1}^{n}(x_i - \mu)^2 = 0 \\
\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}
\end{aligned}
\tag{7}
$$

(IV) Iguale la derivada a cero para encontrar el estimador de máxima verosimilitud para cada parámetro. Para μ :

$$
\frac{\partial ln(L(x_1, \ldots, x_p; \mu, \sigma^2))}{\partial \mu} = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma} \right)
$$

Para σ :

$$
\frac{\partial ln(L(x_1, \ldots, x_p; \mu, \sigma^2))}{\partial \sigma} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^n)^2} \times \sum_{i=1}^{n}(x_i - \mu)^2
$$

(b) **(4 puntos)** Determine si el estimador de máxima verosimilitud para el parámetro μ es eficiente.

Para determinar a eficiencia de un estimador, la varianza de los mismos debe ser igual a la cota de Rao-Cramer.

$$
\begin{aligned}
\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n} \\
ln(L(x_1, \ldots, x_p; \mu, \sigma^2)) = -ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}(x - \mu)^2 \\
\frac{\partial ln(L(x_1, \ldots, x_p; \mu, \sigma^2))}{\partial \mu} = \frac{x}{\sigma^2} - \frac{\mu}{\sigma} \\
\frac{\partial^2 ln(L(x_1, \ldots, x_p; \mu, \sigma^2))}{\partial \mu^2} = -\frac{1}{\sigma^2}
\end{aligned}
\tag{8}
$$

$$
\begin{aligned}
\mathrm{Var}(\bar{X}) \geq \left(-n \times \mathbb{E} \left[\frac{\partial^2 \times ln(L(x_1, \ldots, x_p; \mu, \sigma^2))}{\partial \sigma^2} \right] \right)^{-1} \\
\frac{\sigma^2}{n} \geq \left(-n \times \mathbb{E} \left[-\frac{1}{\sigma^2} \right] \right)^{-1} = \left(\frac{n}{\sigma^2} \right)^{-1} \Rightarrow \frac{\sigma^2}{n} \geq \frac{\sigma^2}{n}
\end{aligned}
\tag{9}
$$

Por la anterior deducción, el estimador para μ es eficiente.

Punto 4

(a) **(4 puntos)** Calcule un intervalo de confianza del 95 % para la media del tiempo de procesamiento de cada uno de los tornos. Interprete cada uno de los intervalos de confianza. Interprete este resultado.

Es necesario obtener los valores para S_Y y H de manera que sea posible construir el intervalo de confianza.

$$S_X = \sqrt{\frac{\sum(x_i - \bar{X})}{n-1}} \quad \bar{X} = \frac{1}{n} \times \sum x_i \Rightarrow$$

$$S_{XA} = 1{,}1 \wedge \bar{X}_A = 1{,}413 \wedge S_{XB} = 0{,}8 \wedge \bar{X}_B = 1{,}68$$
$$t_{(9{,}975)} = 2{,}086$$

El intervalo para el torno A es:

$$\left(1{,}413 - t_{(9{,}975)} \times \frac{S_{XA}}{\sqrt{21}} \leq 1{,}413 \leq 1{,}413 + t_{(9{,}975)} \times \frac{S_{XA}}{\sqrt{21}}\right)$$
$$(0{,}91 \leq 1{,}413 \leq 1{,}91)$$

El intervalo para el torno B es:

$$\left(1{,}68 - t_{(9{,}975)} \times \frac{S_{XB}}{\sqrt{21}} \leq 1{,}68 \leq 1{,}68 + t_{(9{,}975)} \times \frac{S_{XB}}{\sqrt{21}}\right)$$
$$(1{,}35 \leq 1{,}68 \leq 2{,}04)$$

(b) **(4 puntos)** Calcule un intervalo de confianza del 90 % para la varianza del tiempo de procesamiento de cada uno de los tornos. Interprete este resultado.

Dado que la confianza es del 90 %, se tienen los siguientes parámetros:

$$\alpha = 0{,}1 \Rightarrow \mathcal{X}_{(0{,}95)} = 31{,}41 \wedge \mathcal{X}_{(0{,}05)} = 10{,}85$$

El intervalo para el torno A es:

$$\left(\frac{(1{,}1)^2(20)}{(31{,}41)^2} \leq \sigma^2 \leq \frac{(1{,}1)^2(20)}{10{,}85}\right) \Rightarrow \left(0{,}77 \leq \sigma^2 \leq 2{,}23\right)$$

El intervalo para el torno B es:

$$\left(\frac{(0{,}8)^2(20)}{(31{,}41)^2} \leq \sigma^2 \leq \frac{(0{,}8)^2(20)}{10{,}85}\right) \Rightarrow \left(0{,}407 \leq \sigma^2 \leq 1{,}18\right)$$

(c) **(4 puntos)** Construya un intervalo de confianza del 90 % que le permita evaluar si las varianzas del tiempo de procesamiento de cada uno de los tornos son estadísticamente diferentes. ¿Qué puede concluir?

Para evaluar la diferencia estadística de la diferencia de dos varianzas, se debe construir un intervalo de confianza para su cociente. $\alpha = 0{,}1 \Rightarrow \mathcal{F}_{((\alpha/2)=0{,}05)} = 0{,}47 \wedge \mathcal{F}_{((1-\alpha/2)=0{,}95)} = 2{,}12$.

$$\left[0{,}65 \leq \frac{\sigma_A^2}{\sigma_B^2} \leq 2{,}91\right]$$

Dado que uno no está en el intervalo, con un 95 % de confianza, las varianzas son distintas en el intervalo de confianza.

(d) **(4 puntos)** Teniendo en cuenta la conclusión obtenida en el literal anterior, construya un intervalo de confianza del 95 % para la diferencia de las medias de procesamiento de ambos tornos ($\mu_X - \mu_Y$). ¿Qué puede concluir?

$t_{(1-\alpha/2)=0{,}975} = 2{,}02$

$$1{,}68 - 1{,}413 - \sqrt{\frac{2}{21} \times \left[\frac{(0{,}8)^2(20) + (1{,}1)^2(20)}{21 + 21 - 2}\right]} \leq \mu_B - \mu_A$$

$$\mu_B - \mu_A \leq 1{,}68 - 1{,}413 + \sqrt{\frac{2}{21} \times \left[\frac{(0{,}8)^2(20) + (1{,}1)^2(20)}{21 + 21 - 2}\right]}$$

$$(-0{,}33 \leq \mu_B - \mu_A \leq 0{,}866)$$

Con confianza del 95 %, no es posible afirmar si un tiempo es ayor al otro.

Punto 5

(a) **(4 puntos)** Calcule un intervalo de confianza del 98 % para la proporción de automóviles diésel que no cumplen con los estándares de emisiones durante revisión técnico-mecánica.

Para el desarrollo de los siguientes literales, sea A la cantidad de vehículos.

$X \sim Bernoulli(p), p = 1$ if no presentan daño $\quad p = 0$ if presentan daño, $P_A = 23/75 = 0{,}307$. Dado el 98 % de confiabilidad, $\mathcal{Z}_{0,99} = 2{,}32$, entonces:

$$0{,}307 - 2{,}32 \times \sqrt{\frac{0{,}307(1-0{,}307)}{75}} \leq p \leq 0{,}307 + 2{,}32 \times \sqrt{\frac{0{,}307(1-0{,}307)}{75}}$$

$$0{,}183 \leq p \leq 0{,}43$$

(b) **(4 puntos)** Calcule un intervalo de confianza del 98 % para la proporción de automóviles diésel que cumplen con los estándares de emisiones en condiciones normales de uso.

$P_B = (83 - 17)/83 = 0{,}79$. Dado el 98 % de confiabilidad, $\mathcal{Z}_{0,99} = 2{,}32$, entonces:

$$0{,}79 - 2{,}32 \times \sqrt{\frac{0{,}79(1-0{,}79)}{81}} \leq p \leq 0{,}79 + 2{,}32 \times \sqrt{\frac{0{,}79(1-0{,}79)}{81}}$$

$$0{,}68 \leq p \leq 0{,}89$$

(c) **(4 puntos)** Interprete los dos intervalos de confianza construidos en los puntos anteriores e indique si estos son comparables. Justifique su respuesta.

De acuerdo a los intervalos de confianza, se afirma que 98 de cada 100 carros presentan la característica relacionada a su estudio en el intervalo relacionado. Dado que el segundo está contenido en el primero de forma parcial, si son comparables.

(d) **(4 puntos)** Calcule un intervalo de confianza del 95 % que le permita establecer si existe diferencia entre la proporción de vehículos que cumplen con las emisiones mínimas de gases en condiciones normales de uso y la proporción de vehículos que cumplen durante revisión técnico-mecánica. ¿Qué puede concluir?

El estadístico relacionado es:

$$\frac{\hat{p}_y - \hat{p}_x - (\hat{p}_x - \hat{p}_y)}{\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}} \sim \mathcal{N}(0,1)$$

El intervalo de confianza es generado por:

$$\left[\hat{p}_x - \hat{p}_x \pm \sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}\right] \Rightarrow \mathbf{IC} = [-0{,}03, 0{,}23]$$

Como el intervalo incluye el número cero, no es posible afirmar que ambas proporciones son distintas.

Punto 6

(a) **(4 puntos)** Halle el Error Tipo I, el Error Tipo II y la potencia de la prueba 1. Utilice como hipótesis alterna $\mu = 7.8$ metros.

Error tipo I, Prueba I: $\hat{\alpha} = \mathbb{P}(\bar{X} \leq 7{,}77 \mid \mu = 8)$

$$\mu = 8 \wedge \sigma = 1{,}6 \Rightarrow \sigma^2 = 4{,}56 \ \alpha = \mathbb{P}(z \leq \tfrac{7{,}77-8}{2{,}56/80}) = 1 - \mathbb{P}(z \leq 7{,}18) = 0{,}099$$

Error tipo II, Prueba I: $\hat{\beta} = \mathbb{P}(\bar{X} > 7{,}77 \mid \mu = 7{,}8)$

$$\mu = 7{,}8 \wedge \sigma = 1{,}6 \Rightarrow \sigma^2 = 4{,}56 \ \beta = 1 - \mathbb{P}(z \leq \tfrac{7{,}77-7{,}8}{\sqrt{2{,}56/80}}) = 0{,}57$$

Potencia de la prueba I: $1 - \beta_1 = 0{,}43$

(b) **(4 puntos)** Halle el Error Tipo I, el Error Tipo II y la potencia de la prueba 2. Utilice como hipótesis alterna $\mu = 7.8$ metros.

Error tipo I, Prueba II: $\hat{\alpha} = \mathbb{P}(\bar{X} \leq 7{,}75 \mid \mu = 8)$

$$\mu = 8 \wedge \sigma = 1{,}6 \Rightarrow \sigma^2 = 4{,}56 \ \alpha = \mathbb{P}(z \leq \tfrac{7{,}75-8}{2{,}56/80}) = 1 - \mathbb{P}(z \leq 1{,}39) = 0{,}081$$

Error tipo II, Prueba II: $\hat{\beta} = \mathbb{P}(\bar{X} > 7{,}75 \mid \mu = 7{,}8)$

$$\mu = 7{,}8 \wedge \sigma = 1{,}6 \Rightarrow \sigma^2 = 4{,}56 \ \beta = 1 - \mathbb{P}(z \leq \tfrac{7{,}75-7{,}8}{\sqrt{2{,}56/80}}) = 0{,}61$$

Potencia de la prueba II: $1 - \beta_2 = 0{,}39$

(c) **(2 puntos)** Si la compañía desea evaluar la hipótesis planteada con un nivel de significancia máximo del 0.1, ¿cuál de los dos pruebas establecidas, le recomendaría usar a la compañía?

Dado que la probabilidad de obtener errores tipo I en ambas pruebas es menor a la tolerancia o nivel de significancia, se elije la prueba de mayor potencia, por lo tanto debe ser seleccionada la prueba 1.

Punto 7

(a) **(4 puntos)** Construya una prueba de hipótesis para determinar si las varianzas poblacionales de los tiempos de producción de las dos líneas son diferentes. Plantee la hipótesis nula y alterna, el estadístico de prueba, la región de rechazo y concluya en términos del problema. Utilice un nivel de significancia del 5 %.

Dado que es necesario evaluar si las varianzas son distintas, para esto, en pertnente plantear las siguienets hipótesis:

$$H_0 : \frac{\sigma_X^2}{\sigma_Y^2} = 1 \wedge H_1 : \frac{\sigma_X^2}{\sigma_Y^2} \neq 1$$

Para la prueba de hipótesis, se dispone un estadístico con distribución $\mathcal{F}$, usando el mejor estimador de las varianzas.

$$\mathbf{EP} = \frac{S_X^2}{S_Y^2} = \frac{2{,}35^2}{3{,}21^2} = 0{,}536 \Rightarrow \mathcal{F}_{(32,41)}$$

La región de rechazo para evaluar **EP** es $\mathcal{F}_{(95\,\%;32,41)} = 1{,}72$. La región crítica es $\alpha = 0{,}05$, el p-valor es de 0.008 %. Con un 95 % de confiabilidad no se rechaza la hipótesis nula. Luego, las varianzas son iguales.

(b) **(4 puntos)** Teniendo en cuenta el resultado anterior, plantee una prueba de hipótesis para comprobar si el jefe de producción tiene razón sobre la diferencia del tiempo entre las líneas de producción. Escriba las hipótesis nula y alterna, el estadístico de prueba y concluya en términos del problema. Utilice un nivel del significancia del 10 %, y utilice el criterio del p-valor para evaluar la prueba. Adicionalmente, el jefe de planta desea evaluar la línea de producción de botellas de vidrio únicamente, ya que considera que ésta línea presenta un tiempo de producción mayor a 22 segundos/botella, lo cual representaría una reducción considerable en la cantidad de botellas de vidrio producidas para suplir la demanda del próximo mes.

Para evaluar la diferencia entre tiempos, se plantea la siguiente prueba de hipótesis: $H_0 : \mu_X - \mu_Y = 10 \wedge H_1 : \mu_X - \mu_Y \neq 10$. Usando los mejores estimadores, se plantea el estadístico de prueba.

$$\mathbf{EP} = \frac{22{,}45 - 11{,}98 - 10}{2{,}86 \times \sqrt{\frac{1}{33} - \frac{1}{42}}} = 0{,}676$$

. Con los p-values, se tiene como criterio de rechazo $2 \times pvalor < 0{,}1$, para calcular el $pvalor$:

$$\begin{aligned}
pvalue &= \mathbb{P}(\mathcal{T}_{n_X+n_Y-2} \geq E_p) = 1 - \mathbb{P}(\mathcal{T}_{n_X+n_Y-2} \leq E_p) \\
pvalue &= \mathbb{P}(\mathcal{T}_{73} \geq 0{,}676) = 1 - 0{,}4562 = 0{,}5437 \\
2 \times pvalue &= 1{,}087 \Rightarrow 2 \times pvalue > \alpha \Rightarrow \\
&\textbf{No rechazar la hipótesis nula}
\end{aligned} \tag{10}$$

No existe evidencia para afirmar que la diferencia de medias de tiempo de procesamiento, sea diferente de 10 segundos/botella.

(c) **(4 puntos)** Suponga que se conoce la varianza poblacional del tiempo que se demora un producto en salir de la línea de producción de botellas de vidrio ($\sigma^2 = 2{,}4^2$). Plantee una prueba de hipótesis para comprobar si el jefe de planta tiene razón sobre el tiempo de producción en la línea de botellas de vidrio. Plantee las hipótesis nula y alterna, el estadístico de prueba, la región de rechazo y concluya en términos del problema. Utilice un nivel de significancia del 10 %.

$$H_0 : \mu = 22 \wedge H_1 : \mu > 22$$

$$\mathbf{EP} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = 1{,}029 \sim \mathcal{N}(0,1)$$

La región crítica es $\mathcal{Z}_{(0,90)} = 1{,}28 \Rightarrow$ no hay evidencia estadísitica con un nivel de confianza de 90 % para rechazar la hipótesis nula.

(d) **(4 puntos)** El jefe de planta tiene dudas sobre la varianza poblacional de la línea de producción de botellas de vidrio ya que considera que es mayor a la establecida en el punto anterior. Plantee las hipótesis nula y alterna, el estadístico de prueba, la región de rechazo y concluya en términos del problema. Utilice un nivel de significancia del 5 %.

$$H_0 : \sigma^2 = 2{,}4^2 \wedge H_1 : \sigma^2 > 2{,}4^2$$

$$\mathbf{EP} = \frac{(33-1) \times (2{,}35^2)}{2{,}4^2} = 30{,}68 \sim \mathcal{X}^2(32)$$

La región crítica es $\mathcal{X}^2_{(0,95;32)} = 46{,}19 \Rightarrow$ no hay evidencia estadísitica con un nivel de confianza de 90 % para rechazar la hipótesis nula, es decir, se acepta.

(e) **(4 puntos)** Asuma ahora que la información acerca de la varianza poblacional del tiempo que se demora un producto en salir de la línea de producción de botellas de vidrio es descartada. Teniendo esto en cuenta, plantee nuevamente una prueba de hipótesis para probar si el jefe de planta tiene razón sobre el tiempo de producción en dicha línea. Escriba las hipótesis nula y alterna, el estadístico de prueba, la región de rechazo y concluya en términos del problema. Utilice un nivel de significancia del 10 %.

$$H_0 : \mu_X = 22 \wedge H_1 : \mu > 22$$

$$\bar{X} \sim \mathcal{N}(\mu_X, \sigma_X^2/n_X)$$

$$E_p = \frac{\bar{X} - \mu_X}{\frac{S_X}{\sqrt{n_X}}} \sim t(n_X - 1) \Rightarrow E_p = 1{,}0511$$

La región crítica: $\alpha = 0{,}1 \wedge \Rightarrow v = 1{,}30 \Rightarrow E_p \leq v \Rightarrow$ no se rechaza la hipótesis nula. Es decir, no hay suficiente evidencia estadística para afirmar que el tiempo de procesamiento de P_1 es mayor a 22.

Punto 8

(a) **(4 puntos)** El grupo de analistas asegura que la proporción de estudiantes de la Universidad A con un promedio inferior a 4.0 es menor a 0.65. Construya una prueba de hipótesis que le permita evaluar si la afirmación es verdadera. Concluya a partir del cálculo del p-valor y utilice un nivel de significancia del 5 %.

$$H_0 : P = 0{,}65 \wedge H_1 : P < 0{,}65 \wedge \hat{P} = 0{,}63$$
$$\mathbf{EP} = \frac{0{,}63 - 0{,}65}{\sqrt{\frac{0{,}65(1-0{,}65)}{105}}} = -0{,}43$$

La región crítica es: $\mathcal{Z}_{0,95} = 1{,}64 \Rightarrow pvalue = 0{,}66$. Con un nivel de confianza del 95 % no se rechaza la hipótesis nula, no existe mucha evidencia.

(b) **(4 puntos)** Así mismo, el grupo de analistas asegura que la proporción de estudiantes de la Universidad B con un promedio superior a 4.0 es mayor a 0.4. Construya una prueba de hipótesis que le permita evaluar si la afirmación es verdadera. Concluya a partir del cálculo del p-valor y utilice un nivel de significancia del 5 %.

$$H_0 : P = 0{,}4 \wedge H_1 : P > 0{,}4 \wedge \hat{P} = 0{,}36$$
$$\mathbf{EP} = \frac{0{,}36 - 0{,}4}{\sqrt{\frac{0{,}4(1-0{,}4)}{110}}} = -0{,}86$$

La región crítica es: $\mathcal{Z}_{0,95} = 1{,}64$. Con un nivel de confianza del 95 % no se rechaza la hipótesis nula, no existe mucha evidencia par afirmar que P es mayor a 0.4.

(c) **(4 puntos)** Históricamente la proporción de alumnos de la universidad A que se gradúan con un promedio superior a 4.0 es mayor que los alumnos que se gradúan de manera similar de la universidad B. Construya una prueba que le permita evaluar si esta promoción cumple con la afirmación planteada. Concluya a partir del cálculo del p-valor y utilice un nivel de significancia del 5 %.

$$H_0 : P_A - P_B = 0 \wedge H_1 : P_A - P_B > 0$$
$$\mathbf{EP} = \frac{0{,}63 - 0{,}36}{\sqrt{0{,}49 \times (1 - 0{,}49) \times \left(\frac{1}{105} + \frac{1}{110}\right)}} = 3{,}96$$

La región crítica es: $\mathcal{Z}_{0,95} = 1{,}64$. Con un nivel de confianza del 95 % no se rechaza la hipótesis nula, no existe mucha evidencia par afirmar que $P_A > P_B$.

Punto 9

(a) **(2 puntos)** Plantee la hipótesis nula y alterna que le permite al CEO de la compañía probar la afirmación sobre el precio de venta del paquete.

Para el desarrollo de todos los literales, sean:

X : precio en dólares de la tabletas. Y : precio en dólares de los computadores.

$X \sim \mathcal{N}(\mu_t, \sigma^2 = 1080) \wedge Y \sim \mathcal{N}(\mu_t, \sigma^2 = 2020)$

Las hipótesis nula y altera según los requerimientos de la compañía (es alto el costo de cinco tabletas y tres computadores) son:

$H_0 : 5\mu x + 3\mu y = 5500 \wedge H_1 : 5\mu x + 3\mu y < 5500$

(b) **7 puntos)** Construya y calcule el estadístico de prueba apropiado que le permita evaluar la hipótesis definida en el numeral anterior. Defina paso a paso la construcción del estadístico y no olvide determinar su distribución.

$$\begin{aligned} &X \sim \mathcal{N}(\mu_t, \sigma^2 = 1080) \wedge Y \sim \mathcal{N}(\mu_t, \sigma^2 = 2020) \Rightarrow \\ &H_0 : 5\mu x + 3\mu y \sim 5 \times (\mathcal{N}(\mu_t, \sigma^2 = 1080)) + 3 \times (\mathcal{N}(\mu_t, \sigma^2 = 2020)) \sim \\ &\qquad \mathcal{N}\left(5\mu_t + 3\mu_c, \frac{25 \times 1080}{|t|} + \frac{9 \times 2020}{|c|}\right) \end{aligned} \tag{11}$$

Entonces, el estadístico de prueba será:

$$\mathbf{EP} = \frac{5\bar{X} + 3\bar{Y} - 5\mu_t - 3\mu_c}{\sqrt{\frac{25\times1080}{|t|} + + \frac{9\times2020}{|c|}}} = \frac{5\bar{X} + 3\bar{Y} - 5500}{\sqrt{\frac{25\times1080}{|t|} + + \frac{9\times2020}{|c|}}} = -1{,}49$$

(c) **(2 puntos)** Especifique la región de rechazo y concluya usando un nivel de significancia del 5 %.

$$\{\alpha \mid \mathcal{Z}(\mathbf{EP}) = \alpha\} = \{-1{,}645\} \Rightarrow -\mathcal{Z}_{1-\alpha} = -1{,}645$$

$$-1{,}49 > -1{,}645 \Rightarrow \textbf{Se debe rechazar la hipótesis nula.}$$

(d) **(6 puntos)** Construya y calcule un intervalo de confianza del 95 % para el precio del paquete promocional a partir de la media de los precios de cada producto por separado.

$$
\begin{aligned}
5X + 3Y - \mathcal{N}(5\bar{X} + 3\bar{Y}, \frac{25(1080)}{|t|} + \frac{9(2020)}{|c|}) \\
\mathbf{IC}(Q) = \{q \mid q_1 < q < q_2\} \Rightarrow q = \frac{5\bar{X} + 3\bar{Y} - 5\mu_t - 3\mu_c}{\sqrt{\frac{25\times 1080}{|t|} + + \frac{9\times 2020}{|c|}}} \\
\mathbf{A} = \frac{25(1080)}{|t|} + \frac{9(2020)}{|c|} \\
\mathbb{P}\left(-q1\sqrt{\mathbf{A}} + 5\bar{X} + 3\bar{Y} > \mu > 5\bar{X} + 3\bar{Y} - q_2\sqrt{\mathbf{A}}\right) \Rightarrow q \sim \mathcal{N}(0,1)
\end{aligned}
\tag{12}
$$

$$q_1 = \mathcal{Z}_{(1-\alpha/2)} \wedge q_1 \approx q_2$$

Point	Precio	Rank	Percent
24	59880	1	100,00%
30	38580	2	97,00%
18	38200	3	94,10%
21	38000	4	91,10%
26	36200	5	88,20%
32	35780	6	85,20%
22	25300	7	82,30%
20	24280	8	79,40%
2	21040	9	76,40%
6	14400	10	73,50%
3	11480	11	70,50%
5	9500	12	67,60%
27	9170	13	64,70%
14	9000	14	61,70%
31	7220	15	58,80%
15	7080	16	55,80%
34	7000	17	52,90%
11	4550	18	50,00%
1	4340	19	47,00%
16	3350	20	44,10%
17	3035	21	41,10%
12	2600	22	38,20%
29	1955	23	35,20%
13	1720	24	32,30%
10	1600	25	29,40%
8	1480	26	26,40%
33	1460	27	23,50%
23	1400	28	20,50%
9	1220	29	17,60%
25	1190	30	14,70%
7	600	31	11,70%
19	505	32	8,80%
28	469	33	5,80%
35	14,3	34	2,90%
4	7,5	35	0,00%