Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Segundo Semestre 2020

# EYP1113 - Interrogación 3
# Pauta

Digimon es una franquicia de medios creado por Akiyoshi Hongo, que incluye productos como mascotas vituales, videojuegos, películas, animes, mangas, etc.

El archivo Digimon.xlsx contiene una lista de todos los Digimon del videojuego “Digimon Digimon Story: Cyber Sleuth”, junto con su ID, Nombre, Etapa, Tipo, Atributo, Memoria, Espacios de Equipo.

Todas las estadísticas (HP, SP, Ataque, Defensa, Inteligencia y Velocidad) son para ese Digimon en el nivel 50.

A continuación, responda las siguientes preguntas

**Pregunta 1** [1.0 ptos]

Asumiendo normalidad, ¿existe evidencia para afirmar que la media de la Inteligencia de los digimones de Tipo Virus en Etapa Mega difiere de 160?

Realice la prueba de hipótesis correspondiente y responda:

```
Estadístico de Prueba = -2.213228
Valor-p = 0.03586381
¿Existe evidencia al considerar un nivel de significancia del 3%? NO
```

**Pregunta 2** [1.0 ptos]

Asumiendo normalidad, ¿existe evidencia para afirmar que la varianza de la Velocidad de los digimones de Atributo Fuego es superior a 400?

Realice la prueba de hipótesis correspondiente y responda:

```
Estadístico de Prueba = 46.7047
Valor-p = 0.04506355
¿Existe evidencia al considerar un nivel de significancia del 5%? SI
```

**Pregunta 3** [1.0 ptos]

Asumiendo normalidad, ¿existe evidencia para afirmar que la media de la suma entre Ataque y Defensa de los digimones en Etapa Champion es inferior a 220? Asuma independencia entre Ataque y Defensa.

Realice la prueba de hipótesis correspondiente y responda:

```
Estadístico de Prueba = -1.38693
Valor-p = 0.08563442
¿Existe evidencia al considerar un nivel de significancia del 10%? SI
```

**Pregunta 4** [1.0 ptos]

Para la variable HP, ajuste por máxima verosimilitud una Log-Normal y por método de momentos una Gamma. En ambos casos estime la probabilidad que HP > 1500.

$$\text{HP} \sim \text{Log-Norma}(\lambda = \texttt{7.0603369},\, \zeta = \texttt{0.2855065}) \to P(\text{HP} > 1500) = \texttt{0.1878798}$$
$$\text{HP} \sim \text{Gamma}(k = \texttt{13.84345416},\, \nu = \texttt{0.01143252}) \to P(\text{HP} > 1500) = \texttt{0.1803019}$$

¿Cuál estima mejor la probabilidad empírica de observar HP > 1500? `Log-Normal`

```
mean(X>1500)
[1] 0.2008032
```

**Pregunta 5** [1.0 ptos]

Una distribución popularmente utilizada para modelar velocidades máximas de viento tiene una función de distribución de probabilidad acumulada dada por:

$$F_X(x) = 1 - \exp[-\alpha \exp(\beta x)] = 1 - e^{-\alpha e^{\beta x}},$$

con $\alpha > 0$, $\beta > 0$ y $x \in \mathbb{R}$.

Defina

$$Z = \min\left\{e^{\beta X_1}, \ldots, e^{\beta X_n}\right\}$$

¿Cuál es la probabilidad que $P\left(Z > \mu_Z\right)$?, con $\mu_Z = \text{E}(Z)$

Tenemos que $Y = e^{\beta X} \to X = \ln\left(Y^{1/\beta}\right) \to F_Y(y) = F_X\left(\ln\left(y^{1/\beta}\right)\right) = 1 - e^{-\alpha\, y}$, es decir,

$$Y \sim \text{Exponencial}(\alpha) \to Z \sim \text{Exponencial}(n\,\alpha) = \text{Gamma}(1,\, n\,\alpha)$$

Se pide

$$P\left(Z > \frac{1}{n\,\alpha}\right) = P(n\,\alpha\, Z > 1) = e^{-1} = \texttt{0.3678794}$$

ya que

$$n\,\alpha\, Z \sim \text{Gamma}(1, 1) = \text{Exponential}(1)$$

**Pregunta 6** [1.0 ptos]

Recurrentemente se dan a conocer índices que buscan ilustrar el potencial social/económico de los países. Dentro de estos destacan dos: IDH-indice de desarrollo humano y el ICH-Indice de capital humano, índices que están compuesto por variables que buscan reflejar aspectos cuantificables. Suponga que

$$\text{IDH} = X + \frac{1}{2} \cdot Y \quad \text{e} \quad \text{ICH} = 2 \cdot Z + X$$

Si las variables $X$ e $Y$ distribuyen conjuntamente según una $\text{Normal}_{\,2}\left(\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho\right)$ y $Z \sim \text{Gamma}(k, \nu)$ independiente de $X$ e $Y$.

Evalúe la covarianza entre IDH e ICH.

$$\text{Cov(IDH, ICH)} = \sigma_X^2 + \frac{1}{2}\,\rho\,\sigma_X\,\sigma_Y$$

**Pregunta 7** [1.0 ptos]

Suponga que $X_1, \ldots, X_n$ es una muestra iid proveniente de una distribución Weibull $(\eta, 2)$. Se propone como estimador de $\eta$ el siguientes estadístico

$$\hat{\eta} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^{1/2}$$

Obtenga la varianza aproximada de 1er orden de $\hat{\eta}$ considerando que

$$\hat{\eta} = g\left(X_1^2, \ldots, X_n^2\right)$$

Por iid se tiene que

$$\text{Var}(\hat{\eta}) \approx n\,\sigma_{X^2}^2 \left(\frac{1}{2}\,\frac{1}{\sqrt{\mu_{X^2}}}\,\frac{1}{n}\right)^2 = \frac{\eta^2}{4\,n} = \texttt{0.01}$$

**Pregunta 8** [1.0 ptos]

Obtenga la Cota de Cramer Rao a partir de la muestra aleatoria iid proveniente de una distribución Weibull($\eta$, 2) analizada en el ejercicio anterior. Evalúe para $\eta = 2$ y $n = 10$.

Por iid se tiene que

$$\begin{aligned}
L(\eta) &= 2^n \cdot \eta^{-2\,n} \cdot \left(\prod_{i=1}^{n} X_i\right) \cdot \exp\left\{-\frac{1}{\eta^2}\sum_{i=1}^{n} X_i^2\right\} \\
\ln\, L(\eta) &= n\ln(2) - 2\,n\ln(\eta) + \sum_{i=1}^{n}\ln(X_i) - \frac{1}{\eta^2}\sum_{i=1}^{n} X_i^2 \\
\frac{\partial}{\partial\,\eta}\ln L(\eta) &= -\frac{2n}{\eta} + \frac{2}{\eta^3}\sum_{i=1}^{n} X_i^2 \\
\frac{\partial^2}{\partial\,\eta^2}\ln L(\eta) &= \frac{2n}{\eta^2} - \frac{6}{\eta^4}\sum_{i=1}^{n} X_i^2
\end{aligned}$$

Luego

$$I(\eta) = -\text{E}\left(\frac{2n}{\eta^2} - \frac{6}{\eta^4}\sum_{i=1}^{n} X_i^2\right) = \frac{4n}{\eta^2} \to \text{CCR}(\eta) = \frac{\eta^2}{4\,n} = \texttt{0.1}$$

**Pregunta 9** [1.0 ptos]

Como han de saber, este domingo se realizarán primarias para gobernador, entre otras. Un politólogo vaticina que el nivel de participación será inferior a un 30 %. Usted, lleva a cabo un estudio que permita validar la afirmación realizada por este, y para ello entrevista a 120 santiaguinos de los cuales solo 30 manifiestan su interés en participar.

```
Estadístico de Prueba = -1.195229 (z.test) / 1.428571 (prop.test)
Valor-p = 0.1159989
¿hay evidencia que permita validar la afirmación realizada? Utilice un nivel de
significancia del 5%. NO
```

**Pregunta 10** [1.0 ptos]

El periodo de atención de una clase zoom ha sido estudiada por expertos en neurolingüística, quienes afirman que, en promedio, antes de los 20 min se producen desatenciones. Para validar la afirmación, se lleva a cabo mediciones en un grupo de 100 estudiantes de un curso, observando un tiempo promedio de atención de 17 min y coeficiente de variación igual a 1.

Asumiendo que el tiempo tiene un comportamiento exponencial. ¿Existe evidencia que permita validar la afirmación de los expertos en neurolingüística? Realice la prueba de hipótesis correspondiente y responda:

```
Estadístico de Prueba = 1.764706 / -1.5
Valor-p = 0.0388066 / 0.0668072
¿hay evidencia que permita validar la afirmación realizada? Utilice un nivel de
significancia del 10%. SI
```

**Pregunta 11** [1.0 ptos]

Suponga que los tiempos que les toma a los votantes, desde que se presentan en la mesa y se retiran, se comportan como variables aleatorias independientes con media de 2 min y una desviación estándar de 1 min.

Bajo este supuesto, ¿cuál es la probabilidad aproximada qué el tiempo de atención para 60 votantes sea inferior a 130 min?

Suponga que la mesa solo puede atender a un votante a la vez, producto de la pandemia.

```
n = 60
mu = 2
sigma = 1
pnorm(130, mean = n*mu, sd = sqrt(n)*sigma)
[1] 0.9016472
```

**Pregunta 12** [1.0 ptos]

Un caso poco frecuente es que en una muestra aleatoria Normal($\mu$, $\sigma$), el parámetro $\mu$ sea conocido. En este escenario, ¿cuál sería la probabilidad exacta que en una muestra iid de tamaño 100 el estimador máximo verosímil $\hat{\sigma}^2$ supere al parámetro que está estimando?

Tenemos que

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \to \frac{n\,\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n) = \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$$

Se pide

$$P(\hat{\sigma}^2 > \sigma^2) = P\left(\frac{n\,\hat{\sigma}^2}{\sigma^2} > n\right) = \texttt{1-pchisq(100, df = 100) = 0.4811917}$$