Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Segundo Semestre 2020

EYP1113 - Interrogación 2
Pauta

Pregunta 1 **[1.0 ptos]**

Este domingo se llevara a cabo el plebiscito donde se consultara si se aprueba o rechaza la redacción de una nueva Constitución. Las mesas recibirán votantes a partir de las 08.00 horas. Estudios epidemiológicos establecen que el 7 % de los votante serán portadores de Covid este domingo. Si los votantes llegan a la mesa de manera independiente y suponemos que la proporción de Covid se mantiene constante ese día. ¿Cuál es la probabilidad que el sexto votante Covid llegue a la mesa entre los primeros 50 votantes?

```
## X ~ BinNeg(6, p)
p = 0.07
pnbinom(50-6, size = 6, prob = p)
[1] 0.1350464

## Y ~ Binomial(50, p)
1-pbinom(5, size = 50, prob = p)
[1] 0.1350464
```

Los alumnos que utilizaron `dnbinom` y pusieron como respuesta `0.0092` se les asignó **[0.2 ptos]** ya que tienen un error del 93 % con respecto a la respuesta correcta.

Pregunta 2 **[1.0 ptos]**

Este domingo se llevara a cabo el plebiscito donde se consultara si se aprueba o rechaza la redacción de una nueva Constitución. Las mesas recibirán votantes a partir de las 08.00 horas. Estudios indican que el 75 % del padrón electoral asistirá a votar. Una mesa tiene registrado 350 electores, ¿cuál es la probabilidad que lleguen más del 80 % de los electores de esta mesa a votar? Considere que el padrón electoral para este domingo es de 14.5 millones electores

```
N = 14500000
m = N*0.75
n = 350

## X ~ Hipergeomerica(n, N, m)
1-phyper(n*.8, m = m, n = N-m, k = n)
[1] 0.01166939

## X ~ Binomial(n, p = m/N)
1-pbinom(n*.8, size = n, prob = m/N)
[1] 0.0116702
```

Los alumnos que utilizaron `1-phyper(n*.8-1,...)` o `1-pbinom(n*.8-1,...)` y pusieron como respuesta `0.0163` se les asignó **[0.8 ptos]**, este pequeño detalle tiene un error de 39 % con respecto a la respuesta correcta.

Los alumnos que utilizaron `1-dhyper(...)` o `1-dbinom(...)` tienen un error de un 750 % con respecto a la respuesta correcta por eso no tienen puntaje parcial.

Pregunta 3 [1.0 ptos]

El tiempo de traslado de los electores registrados en un centro de votación se comportaran según una variable aleatoria Weibull(η, β). Simulaciones sitúan el IQR en 20 minutos y el 75 % le tomará al menos 10 min en llegar. ¿Qué porcentaje de los electores de dicho centro de votación les tomará más de 45 min de traslado?

```
## X ~ Weibull(eta, beta)
## ln(xp) = ln(eta) + (1/beta) * ln(-ln(1-p))
iqr = 20
x25 = 10
x75 = iqr+x25
beta = (log(-log(0.25)) - log(-log(0.75)) )/(log(x75)-log(x25))
eta = exp(log(x75)-(1/beta)*log(-log(0.25)))
x = 45
1-pweibull(x, shape = beta, scale = eta)
[1] 0.08400285
(1-pweibull(x, shape = beta, scale = eta))*100
[1] 8.400285
```

Los alumnos que utilizaron `1-dweibull(...)` tienen un error del 1082 % con respecto a la respuesta correcta, así que no tienen puntaje parcial.

Los alumnos que utilizaron `1-pweibull(...)` pero invirtieron los parámetros tienen un error prácticamente infinito con respecto a la respuesta correcta, así que no tienen puntaje parcial.

El tiempo de traslado de los electores registrados en un centro de votación se comportaran según una variable aleatoria Log-Logistica(μ, σ). Simulaciones sitúan el IQR en 20 minutos y el 75 % le tomará al menos 10 min en llegar. ¿Qué porcentaje de los electores de dicho centro de votación les tomará más de 45 min de traslado?

```
## X ~ Log-Logistica(mu, sigma)
## ln(xp) = mu + sigma * ln(p/(1-p))
iqr = 20
x25 = 10
x75 = iqr+x25
sigma = (log(x75)-log(x25))/(log(0.75/0.25) - log(0.25/0.75) )
mu = log(x75)-sigma*log(0.75/0.25)
x = 45
1-plogis(log(x), location = mu, scale = sigma)
[1] 0.1290323
(1-plogis(log(x), location = mu, scale = sigma))*100
[1] 12.90323
```

Los alumnos que utilizaron `1-dlnom(...)` tienen un error del 600 % con respecto a la respuesta correcta, así que no tienen puntaje parcial.

Pregunta 4 **[1.0 ptos]**

Estudios estiman que la tasa de llegada a un centro de votación rural entre 10.00-11.00 horas será de 5 electores por minuto. Si los electores llegan de manera independiente según un proceso de Poisson, ¿cuál es la probabilidad que entre las 10:15:00 y 10:17:00 horas, es decir 2 min, ingresen más de 12 electores?

```
## Xt ~ Poisson(nu*t)
nu = 5
1-ppois(12, lambda = nu*2)
[1] 0.2084435
## T13 ~ Gamma(k = 13, nu)
pgamma(2, shape = 13, rate = nu)
[1] 0.2084435
```

Los alumnos que utilizaron `1-ppois(11,..)` o `pgamma(..., shape = 12,...)`, cuya respuesta fue `0.3032` tienen un error del 45 % con respecto a la respuesta correcta, se les asignó **[0.8 ptos]**.

Estudios estiman que la tasa de llegada a un centro de votación rural entre 10.00-11.00 horas será de 4 electores por minuto. Si los electores llegan de manera independiente según un proceso de Poisson, ¿cuál es la probabilidad que entre las 10:15:00 y 10:18:00 horas, es decir 3 min, ingresen más de 12 electores?

```
## Xt ~ Poisson(nu*t)
nu = 4
1-ppois(12, lambda = nu*3)
[1] 0.4240348
## T12 ~ Gamma(k = 13, nu)
pgamma(3, shape = 13, rate = nu)
[1] 0.4240348
```

Los alumnos que utilizaron `1-ppois(11,..)` o `pgamma(..., shape = 12,...)`, cuya respuesta fue `0.5384` tienen un error del 27 % con respecto a la respuesta correcta, se les asignó **[0.8 ptos]**.

Pregunta 5 [1.0 ptos]

En la industria salmonera uno de los aspectos más importantes y que es un % importante del costo del proceso de crianza de las piezas de Salmones es el alimento. Las piezas que se exportan en formato HG (sin cabeza) presentan variaciones en su peso esperado, debido principalmente a ineficiencia en la alimentación, ya que los Salmones comen hasta saciarse y eso implicará en el futuro un exceso de grasa, la cual al entrar al proceso de preparación y limpieza se desecha, afectando el margen económico y las estimaciones de los Kg comprometidos para la venta.

Se tomaron 200 muestras en un centro de cultivo, registrando en terreno el peso fuera del agua (en Kg) y su largo (en cm). Posteriormente en la planta durante el procesamiento de las piezas, se registro el peso de la grasa y de la pieza en formato HG lista para exportación.

Bajo el supuesto que el comportamiento entre el peso HG (Y) y peso fuera del agua (X), ambos en kilos, es Normal bivariado, ¿cuál sería su pronóstico para el peso HG, si el peso fuera del agua es de 3.05 kilos?

```
mean(Y) mean(X)  sd(Y)  sd(X) cov(X,Y)
 2.7052  3.8774 0.3976 0.5603   0.2144

mu.y = 2.7052
mu.x = 3.8774
sigma.y = 0.3976
sigma.x = 0.5603
rho = 0.2144/(sigma.x*sigma.y)
a = mu.y-mu.x*rho*sigma.y/sigma.x
b = rho*sigma.y/sigma.x
a+b*3.05
[1] 2.140134
```

```
mean(Y) mean(X)  sd(Y)  sd(X) cov(X,Y)
2.6052 3.9774  0.3976 0.5630   0.2414

mu.y = 2.6052
mu.x = 3.9774
sigma.y = 0.3976
sigma.x = 0.5630
rho = 0.2414/(sigma.x*sigma.y)
a = mu.y-mu.x*rho*sigma.y/sigma.x
b = rho*sigma.y/sigma.x
a+b*3.05
[1] 1.898903
```

```
mean(Y) mean(X)   sd(Y)  sd(X) cov(X,Y)
2.6052  3.7974 0.3796  0.5630   0.2414

mu.y = 2.6052
mu.x = 3.7974
sigma.y = 0.3796
sigma.x = 0.5630
rho = 0.2414/(sigma.x*sigma.y)
a = mu.y-mu.x*rho*sigma.y/sigma.x
b = rho*sigma.y/sigma.x
a+b*3.05
[1] 2.035989
```

Pregunta 6 [1.0 ptos]

Pokémon es una franquicia de medios que originalmente comenzó como un videojuego RPG, pero debido a su popularidad ha logrado expandirse a otros medios de entretenimiento como series de televisión, películas, juegos de cartas, ropa, entre otros, convirtiéndose en una marca que es reconocida mundialmente.

Los pokémones son personajes de ficción que se encuentran clasificados en el Pokédex según múltiples atributos. La base de datos Pokedex.xlsx tiene en su 1ra hoja la base de datos y en la 2da la descripción de las variables.

Para la variable velocidad, ajuste mediante un gráfico de probabilidad una distribución: Weibull, Logística y Normal. Según el coeficiente r^2 ajustado elija el mejor modelo y calcule el rango intercuartílico de la velocidad.

```
library(rio)
Data = import("Pokedex.xlsx")
xp   = sort(Data$velocidad)
N    = length(xp)
p    = 1:N/(N+1)
fit1 = lm(log(xp) ~ log(-log(1-p)))
fit2 = lm(xp ~ qlogis(p))
fit3 = lm(xp ~ qnorm(p))
summary(fit1)$adj.r.squared
[1] 0.9936802
summary(fit2)$adj.r.squared
[1] 0.9712406
summary(fit3)$adj.r.squared
[1] 0.9847523
qweibull(0.75, shape = 1/fit1$coef[2], scale = exp(fit1$coef[1]))-
+ qweibull(0.25, shape = 1/fit1$coef[2], scale = exp(fit1$coef[1]))
[1] 40.11137
```

Si el alumno se equivoca de modelo, pero hace correctamente el cálculo tiene **[0.5 ptos]**.

Para la variable HP, ajuste mediante un gráfico de probabilidad una distribución: Weibull, Logística y Log-Logística. Según el coeficiente r^2 ajustado elija el mejor modelo y calcule la probabilidad de que al elegir un pokémon al azar este tenga HP mayor que 100.

```
library(rio)
Data = import("Pokedex.xlsx")
xp   = sort(Data$HP)
N    = length(xp)
p    = 1:N/(N+1)
fit1 = lm(log(xp) ~ log(-log(1-p)))
fit2 = lm(xp ~ qlogis(p))
fit3 = lm(log(xp) ~ qlogis(p))
summary(fit1)$adj.r.squared
[1] 0.9187706
summary(fit2)$adj.r.squared
[1] 0.9262711
summary(fit3)$adj.r.squared
[1] 0.9223675
1-plogis(100, location = fit2$coef[1], scale = fit2$coef[2])
[1] 0.0962334
```

Si el alumno se equivoca de modelo, pero hace correctamente el cálculo tiene **[0.5 ptos]**.

Para la variable defensa, ajuste mediante un gráfico de probabilidad una distribución: Weibull, Log-Logística y Log-Normal. Según el coeficiente r^2 ajustado elija el mejor modelo y calcule la probabilidad de que la defensa sea menor que 75.

```
library(rio)
Data = import("Pokedex.xlsx")
xp   = sort(Data$defensa)
N    = length(xp)
p    = 1:N/(N+1)
fit1 = lm(log(xp) ~ log(-log(1-p)))
fit2 = lm(log(xp) ~ qlogis(p))
fit3 = lm(log(xp) ~ qnorm(p))
summary(fit1)$adj.r.squared
[1] 0.9670822
summary(fit2)$adj.r.squared
[1] 0.971127
summary(fit3)$adj.r.squared
[1] 0.9611112
plogis(log(75), location = fit2$coef[1], scale = fit2$coef[2])
[1] 0.6065904
```

Si el alumno se equivoca de modelo, pro hace correctamente el cálculo tiene **[0.5 ptos]**.

Pregunta 7 **[1.0 ptos]**

Hoy es común ver fuera de muchos establecimientos comerciales a repartidores de delivery esperando que los atiendan. Suponga que en uno de estos establecimientos, hay un encargado que exclusivamente atiende a delivery.

El número de repartidores X que al encargado le toca atender durante su jornada laboral se comporta como una variable aleatoria Binomial-Negativa$(r,\, p)$ y el tiempo dedicado a la atención, por ejemplo, de x clientes se comporta como una Gamma$(x,\, \nu)$. ¿Cuál sería el coeficiente de variación del tiempo que dedicará a la atención de los delivery en la jornada de mañana viernes?

Tenemos que

$$X \sim \text{Bin-Neg}(r,\, p) \quad \text{y} \quad T \,|\, X = x \sim \text{Gamma}(x,\, \nu)$$

Se pide $\delta_T = \dfrac{\sigma_T}{\mu_T}$.

$$\begin{aligned}
\mu_T &= \text{E}[\text{E}(T \,|\, X)] = \frac{r}{\nu\, p} \\
\sigma_T^2 &= \text{E}[\text{Var}(T \,|\, X)] + \text{Var}[\text{E}(T \,|\, X)] = \frac{r}{\nu^2\, p^2}
\end{aligned}$$

Reemplazando se tiene que $\delta_T = \dfrac{1}{\sqrt{r}}$.

Pregunta 8 [1.0 ptos]

Suponga que el tiempo en recorrer una distancia en vehículo, condicionado al número de semáforos x en rojo que se le presentan en el trayecto distribuye Gamma(x, ν). Si el número de semáforos en rojo que se le presentan en el trayecto distribuye Geométrica(p). Entonces la distribución incondicional al número de semáforos en rojo del tiempo que toma en recorrer una distancia en vehículo se comporta como una variable aleatoria:

Tenemos que

$$X \sim \text{Geometrica}(p) \quad \text{y} \quad T \,|\, X = x \sim \text{Gamma}(x,\, \nu)$$

Se pide $f_T(t)$.

$$\begin{aligned} f_T(t) &= \sum_{x=1}^{\infty} (1-p)^{x-1}\, p \cdot \frac{\nu^x}{\Gamma(x)}\, t^{x-1}\, e^{-\nu\, t} \\ &= \nu\, p\, e^{-\nu\, t} \sum_{y=0}^{\infty} \frac{[(1-p)\,\nu\, t]^y}{y!} \\ &= \nu\, p\, e^{-\nu\, p\, t}, \quad t > 0 \end{aligned}$$

Por lo tanto

$$T \sim \text{Exponencial}(\nu\, p)$$