

**Curso** : Probabilidad y Estadística  
**Sigla** : EYP1113  
**Profesores** : Ricardo Aravena C., Ricardo Olea O. y Felipe Ossa M.

### PAUTA INTERROGACIÓN 4

En el archivo `salario2022.xlsx` contiene información de una muestra de 52 postulantes a un puesto de trabajo:

- **SALARIO**: expectativa en miles CL\$.
- **EXPER**: experiencia laboral en años.
- **ANTIG**: antigüedad (en años) último trabajo.
- **SEXO**: 0 = Masculino, 1 = Femenino.

En base a esta información responda las siguientes nueve preguntas.

```
Data <- rio::import("salario2022.xlsx")
head(Data)
```

#### Pregunta 1

¿Existe evidencia que permita afirmar que menos de un tercio de los postulantes son de sexo femenino? Considerando un nivel de significancia del 10 %, realice la prueba de hipótesis correspondiente y responda:

Valor-p:

Respuesta (Si o No):

#### Solución

```
## H0: p = p0 vs Ha: p < p0
## alpha = 10%
p0 = 1/3
X <- Data$SEXO
prop.test(x = sum(X), n = length(X), alternative = "less", p = p0, correct = F)$p.value
TeachingDemos::z.test(x = X, mu = p0, alternative = "less", stdev = sqrt(p0*(1-p0)))$p.value
## Valor-p: 0.1633998 [0.8 Puntos]
## Respuesta: NO [0.2 Puntos]

## Alternativas parcialmente correctas
TeachingDemos::z.test(x = X, mu = p0, alternative = "less", stdev = sqrt(mean(X)*(1-mean(X))))$p.value
## Valor-p: 0.1486742 [0.6 Puntos]
## Respuesta: NO [0.2 Puntos]

t.test(x = X, mu = p0, alternative = "less")$p.value
## Valor-p: 0.1534551 [0.6 Puntos]
## Respuesta: NO [0.2 Puntos]

prop.test(x = sum(X), n = length(X), alternative = "less", p = p0, correct = T)$p.value
## Valor-p: 0.2022832 [0.6 Puntos]
## Respuesta: NO [0.2 Puntos]
```

*Observación: Las respuestas SI o NO en línea con valores-p incorrectos tienen [0.2 Puntos].*

## Pregunta 2

¿Existe evidencia que permita afirmar que la experiencia media de los postulantes, variable EXPER, es mayor a 13 años? Asumiendo normalidad y un nivel de significancia del 5%, realice la prueba de hipótesis correspondiente y responda:

Valor-p:

Respuesta (Si o No):

## Solución

```
## H0: mu = mu0 vs Ha: mu > mu0
## alpha = 5%
mu0 = 13
X <- Data$EXPER
t.test(x = X, mu = mu0, alternative = "greater")$p.value
## Valor-p: 0.0162694 [0.8 Puntos]
## Respuesta: SI      [0.2 Puntos]

## Aproximación t-Student --> Normal(0,1)
TeachingDemos::z.test(x = X, mu = mu0, stdev = sd(X), alternative = "greater")$p.value
## Valor-p: 0.01398622 [0.8 Puntos]
## Respuesta: SI      [0.2 Puntos]
```

*Observación: Las respuestas SI o NO en línea con valores-p incorrectos tienen [0.2 Puntos].*

### Pregunta 3

¿Es válido afirmar que antigüedad media (esperada) de los postulantes, variable ANTIG, difiere entre sexos? Asumiendo normalidad y un nivel de significancia del 1 %, realice la prueba de hipótesis correspondiente y responda:

Valor-p:

Respuesta (Si o No):

### Solución

```
X <- dplyr::filter(Data, SEXO == 0)$ANTIG
Y <- dplyr::filter(Data, SEXO == 1)$ANTIG
## alpha = 1%
var.test(x = X, y = Y)$p.value
## ## H0: mu.X = mu.Y vs Ha: mu.X != mu.Y, con varianzas desconocidas e iguales
t.test(x = X, y = Y, var.equal = T)$p.value
## Valor-p: 0.005537842 [0.8 Puntos]
## Respuesta: SI [0.2 Puntos]

## Alternativa asumiendo varianzas desconocidas y distintas
t.test(x = X, y = Y, var.equal = F)$p.value
## Valor-p: 0.00070774 [0.4 Puntos]
## Respuesta: SI [0.2 Puntos]
```

*Observación: Las respuestas SI o NO en línea con valores-p incorrectos tienen [0.2 Puntos].*

#### Pregunta 4

¿Los salarios de los postulantes se ajustan a una distribución Log-Normal( $\lambda = 7$ ,  $\zeta = 0.6$ )? Considerando los intervalos

(0,600], (600, 1200], (1200, 2000], (2000, Inf),

y un nivel de significancia del 5 %, realice la prueba de hipótesis  $\chi^2$  de bondad de ajuste y responda:

Valor estadístico de prueba:

Valor-p:

Respuesta (Si o No):

#### Solución

```
## H0: X ~ Log-Normal(7, 0.6) vs H0: X !~ Log-Normal(7, 0.6)
## alpha = 5%
X <- Data$SALARIO
O <- table(cut(X, c(0, 600, 1200, 2000, Inf)))
p <- diff(plnorm(c(0, 600, 1200, 2000, Inf), 7, 0.6))
chisq.test(x = O, p = p)$statistic
chisq.test(x = O, p = p)$p.value
## Valor estadístico de prueba: 7.023873 [0.4 Puntos]
## Valor-p: 0.07114075 [0.4 Puntos]
## Respuesta: SI [0.2 Puntos]

## Alternativa asumiendo que X2 ~ chi2(4-1-2)
1-pchisq(chisq.test(x = O, p = p)$statistic, df = 4-1-2)
## Valor estadístico de prueba: 7.023873 [0.4 Puntos]
## Valor-p: 0.008043009 [0.2 Puntos]
## Respuesta: NO [0.2 Puntos]
```

*Observación: Las respuestas SI o NO en línea con valores-p incorrectos tienen [0.2 Puntos].*

### Pregunta 5

Basado en los estimadores de máxima verosimilitud y la prueba de hipótesis de bondad de ajuste KS, ¿que modelo ajusta mejor a los salarios de los postulantes: Log-Normal o Gamma?

Mejor ajuste (Log-Normal o Gamma):

Valor-p (mejor ajuste):

### Solución

```
## EMV Log-Normal:
par.ln <- fitdistrplus::fitdist(data = X, distr = "lnorm", method = "mle")$estimate
## EMV Gamma:
par.gamma <- fitdistrplus::fitdist(data = X, distr = "gamma", method = "mle")$estimate

ks.test(x = X, "plnorm", meanlog = par.ln[1], sdlog = par.ln[2])$p.value
ks.test(x = X, "pgamma", shape = par.gamma[1], rate = par.gamma[2])$p.value

## Mejor Ajuste: Gamma + Valor-p (KS): 0.1832536      [1.0 Puntos]

## Respuesta incorrecta
## Mejor Ajuste: Log-Normal + Valor-p (KS): 0.1234018 [0.5 Puntos]
```

## Pregunta 6

A partir de modelos de regresión lineal simple, se le pide a usted que evalúe qué variable explica de mejor manera la expectativa de salario de un postulante: experiencia o antigüedad.

¿Qué medida utilizará para comparar?:

Mejor variable explicativa (EXPER o ANTIG):

Valor medida utilizada del mejor ajuste:

## Solución

```
## Coeficiente de determinación (mejor modelo)
mod01 <- lm(SALARIO ~ EXPER, data = Data)
mod02 <- lm(SALARIO ~ ANTIG, data = Data)
summary(mod01)$r.squared
summary(mod02)$r.squared

summary(mod01)$adj.r.squared
summary(mod02)$adj.r.squared

anova(mod01)$"Pr(>F)"[1]
anova(mod02)$"Pr(>F)"[1]

## Mejor variable: EXPER [0.5 Puntos]
## Valor medida: R2 = 0.4599438, r2 = 0.4491427, p.value = 3.315735e-08 [0.5 Puntos]
```

### Pregunta 7

Evalúe si el aporte **conjunto** de las variables SEXO y ANTIG, aportan significativamente al modelo de regresión simple basado sólo en la experiencia (EXPER) para explicar la expectativa de salario de los postulantes. Lleve a cabo el test correspondiente y responda considerando un nivel de significancia del 5%:

Valor-p:

Aporte conjunto es significativo (Si o NO):

### Solución

```
## alpha = 5%  
mod03 = lm(SALARIO ~ EXPER+SEXO+ANTIG, data = Data)  
anova(mod01,mod03)  
## Valor-p = 0.04864 [0.8 Puntos]  
## Respuesta = SI      [0.2 Puntos]
```

*Observación: Las respuestas SI o NO en línea con valores-p incorrectos tienen [0.2 Puntos].*

## Pregunta 8

Para un estudio futuro se piensa tomar esta información como referencia. ¿Cuál es tamaño de muestra necesario para estimar la proporción de postulantes de sexo masculino con un error de estimación de a lo más siete puntos porcentuales?

Tamaño muestra:

## Solución

```
p = mean(Data$SEX0 == 0)
(qnorm(0.975)*sqrt(p*(1-p)))/0.07)^2
## n = 154.2429 --> 155 [1.0 Puntos]

## Alternativas parcialmente correctas [0.8 Puntos]
(qnorm(0.975)*sqrt(0.25)/0.07)^2
## n = 195.9928 --> 196
(qnorm(0.975)*sd(Data$SEX0==0)/0.07)^2
## n = 157.2672 --> 158

## Error 1: no elevar al cuadrado [0.5 Puntos]
(qnorm(0.975)*sqrt(p*(1-p)))/0.07)
## n = 12.41945 --> 13
(qnorm(0.975)*sqrt(0.25)/0.07)
## n = 13.99974 --> 14

## Error 2: Utilizar var en vez de sd [0.2 Puntos]
(qnorm(0.975)*(p*(1-p))/0.07)^2
## n = 30.3466 --> 31
(qnorm(0.975)*(0.25)/0.07)^2
## n = 48.9982 --> 49
```



## Pregunta 9

Bajo el supuesto que las expectativas de salarios entre los postulantes distribuyen Log-Normal, construya un intervalo de confianza al 95 % para la mediana de estas expectativas.

$$P(\text{mediana} \in [\quad, \quad]) = 0.95$$

## Solución

```
X <- Data$SALARIO
lambda <- fitdistrplus::fitdist(data = X, distr = "lnorm")$estimate[1]
CCR <- fitdistrplus::fitdist(data = X, distr = "lnorm")$sd[1]^2

exp(lambda)-qnorm(0.975)*sqrt(exp(2*lambda)*CCR)
## LI = 900.3426 [0.5 Puntos]
exp(lambda)+qnorm(0.975)*sqrt(exp(2*lambda)*CCR)
## LS = 1234.078 [0.5 Puntos]

## Alternativa (invarianza)
exp(lambda-qnorm(0.975)*sqrt(CCR))
## LI = 912.734 [0.5 Puntos]
exp(lambda+qnorm(0.975)*sqrt(CCR))
## LS = 1247.831 [0.5 Puntos]
```