

Probabilidad y estadística: PEC2

UOC

NOMBRE: Juan Luis Acebal Rico

Indique claramente su nombre y apellidos en la primera página del informe. Resuelva los ejercicios utilizando R siempre que sea necesario, es importante mostrar cada uno de los pasos realizados para llegar al resultado final. Indique en cada caso qué instrucciones de R ha utilizado y las salidas que se obtienen.

Importante: el informe final se entregará a través del moodle en un archivo con formato “pdf” o “html”. No se debe entregar el archivo con formato “Rmd”.

Esta PEC debe realizarse de forma estrictamente individual, quedando totalmente prohibido el uso de herramientas de IA. Cualquier indicio de copia será penalizado con un suspenso (D) para todas las partes implicadas y la posible evaluación negativa de la asignatura de forma íntegra.

```
# Cuestionario  
pnorm(239, mean = 235, sd = 8)
```

```
## [1] 0.6914625
```

```
1 - pnorm(231.8, mean = 235, sd = 8)
```

```
## [1] 0.6554217
```

```
pnorm(242.2, mean = 235, sd = 8) - pnorm(237.4, mean = 235, sd = 8)
```

```
## [1] 0.1980285
```

```
qnorm(0.81, mean = 235, sd = 8)
```

```
## [1] 242.0232
```

```
qnorm(0.44, mean = 235, sd = 8)
```

```
## [1] 233.7922
```

Problema 1

(4 puntos) Los datos del archivo *company_workers.csv* contienen información sobre los trabajadores de una empresa. Importe los datos al programa R y observe los nombres de las variables.

```
# leo csv, con separador ; y muestro un resumen ,cabecera y nombres
trabajadores_compania <- read.csv("company_workers.csv", sep=";")
summary(trabajadores_compania)
```

```
##      EmployeeId      Department      Building      Months_at_company
## Min.   : 1.0      Length:427      Length:427      Min.   : 2.0
## 1st Qu.:107.5     Class :character      Class :character  1st Qu.:102.0
## Median :214.0     Mode  :character      Mode  :character  Median :193.0
## Mean   :214.0
## 3rd Qu.:320.5
## Max.   :427.0
##                                     Mean   :188.7
##                                     3rd Qu.:271.0
##                                     Max.   :360.0
```

```
head(trabajadores_compania)
```

```
##      EmployeeId Department Building Months_at_company
## 1             1         T         B             289
## 2             2         T         A             273
## 3             3        RH         B             249
## 4             4        RH         C             164
## 5             5         C         A             168
## 6             6        RH         C             309
```

```
names(trabajadores_compania)
```

```
## [1] "EmployeeId"      "Department"      "Building"
## [4] "Months_at_company"
```

- a) La variable *Department* indica el departamento en el que trabaja el trabajador, clasificado en *Comercial* (C), *Recursos Humanos* (RH) o *Tecnología* (T). La variable *Building* toma los valores A, B o C, según el edificio donde trabaja el trabajador. Elabore una tabla de contingencia entre las variables *Department* y *Building*:

```
#Creo la tabla de contingencia entre las dos variables pedidas, ello hace que se pueda
tabla_contingencia <- table(trabajadores_compania$Department, trabajadores_compania$Build)
print(tabla_contingencia)
```

```
##
##      A  B  C
##  C  44 53 49
##  RH 48 47 40
##  T   57 51 38
```

- b) Si escogemos un trabajador de la base de datos al azar, ¿cuál es la probabilidad de que trabaje en el departamento Comercial y en el edificio A?

```
# Igual pero entre departamento comercial y edificio A
trabajadores_totales <- nrow(trabajadores_compania)
comercial_A <- tabla_contingencia["C", "A"]
prob_comercial_A <- comercial_A / trabajadores_totales
print(prob_comercial_A)
```

```
## [1] 0.1030445
```

- c) ¿Cuál es la probabilidad de que un trabajador que trabaja en el edificio B esté en el departamento de Recursos Humanos?

```
# Entre Edificio B y recursos humanos
total_B <- sum(tabla_contingencia[, "B"])
rh_B <- tabla_contingencia["RH", "B"]
probabilidad_rh_dado_B <- rh_B / total_B
print(probabilidad_rh_dado_B)
```

```
## [1] 0.3112583
```

- d) ¿Cuál es la probabilidad de que un trabajador del departamento de Tecnología trabaje en el edificio C?

```
# Entre tecnologia y edificio C
total_tecnologia <- sum(tabla_contingencia["T", ])
tecnologia_C <- tabla_contingencia["T", "C"]
probabilidad_C_dado_tecnologia <- tecnologia_C / total_tecnologia
print(probabilidad_C_dado_tecnologia)
```

```
## [1] 0.260274
```

Problema 2

(3 puntos) Continuamos con los datos del problema anterior.

- a) Si escogemos un trabajador de la base de datos al azar, ¿cuál es la probabilidad de que trabaje en el departamento Comercial?

```
# Probabilidad que sea del departamento comercial
total_comercial <- sum(tabla_contingencia["C", ])
probabilidad_comercial <- total_comercial / trabajadores_totales
print(probabilidad_comercial)
```

```
## [1] 0.3419204
```

- b) Escogemos al azar 20 trabajadores de la base de datos, con reposición. Consideramos la variable que nos indica el número de trabajadores, entre los 20, que trabajan en el departamento Comercial.

- i.) ¿Qué distribución sigue esta variable? ¿De qué parámetros?

```
# Es una distribucion binomial con parametros n = 20 y p = probabilidad_comercial
```

- ii.) ¿Cuál es la probabilidad de que exactamente 7 de los 20 trabajen en el departamento Comercial?

```
# Aqui es calcular la probabilidad usando dbinom
probabilidad_exactamente_7 <- dbinom(7, 20, probabilidad_comercial)
print(probabilidad_exactamente_7)
```

```
## [1] 0.183869
```

- c) Supongamos ahora que sabemos que el 60% de los trabajadores del departamento Comercial trabajan en proyectos estratégicos. También sabemos que un 40% de los trabajadores que trabajan en proyectos estratégicos están en el departamento Comercial. ¿Cuál es la probabilidad de que un trabajador esté en un proyecto estratégico?

```
# formula probabilidad total
P_E_dado_C <- 0.60
P_C_dado_E <- 0.40
P_C <- probabilidad_comercial
P_E <- (P_E_dado_C * P_C) / P_C_dado_E
print(P_E)
```

```
## [1] 0.5128806
```

Problema 3

(3 puntos) Consideremos ahora la variable *Months_at_company* (meses en la empresa). Calcule m , la media de los meses que llevan los trabajadores en la empresa, y d , la desviación típica **poblacional**.

```
#media
media_meses <- mean(trabajadores_compania$Months_at_company)
print(media_meses)
```

```
## [1] 188.6956
```

```
# calculo desviacion (standard deviation: sd)
desviacion_meses <- sd(trabajadores_compania$Months_at_company)
print(desviacion_meses)
```

```
## [1] 103.3548
```

Supongamos ahora que sabemos que los meses que llevan los trabajadores en la empresa se distribuyen siguiendo una distribución normal de media m y desviación típica d .

Escogemos un trabajador al azar:

a) ¿Cuál es la probabilidad de que lleve más de 300 meses en la empresa?

```
# uso pnorm para calcular la probabilidad a la derecha de 300 meses (por eso 1 menos p)
probabilidad_mas_300 <- 1 - pnorm(300, mean = media_meses, sd = desviacion_meses)
print(probabilidad_mas_300)
```

```
## [1] 0.1407588
```

b) ¿Cuál es la probabilidad de que lleve menos de 100 meses?

```
# igual, te cda lo que sea a la izquierda de 100 meses
probabilidad_menos_100 <- pnorm(100, mean = media_meses, sd = desviacion_meses)
print(probabilidad_menos_100)
```

```
## [1] 0.1954004
```

c) Encuentre un número de meses tal que el 75% de los trabajadores hayan estado menos tiempo en la empresa que ese valor y el 25% hayan estado más tiempo.

```
# aqui calculo con qnorm el percentil 0.75  
percentil_75 <- qnorm(0.75, mean = media_meses, sd = desviacion_meses)  
print(percentil_75)
```

```
## [1] 258.4073
```
