1. Ejercicio 1

1.1. Pregunta 1

Asumiendo que las afirmaciones dichas por Rich Mantle son ciertas. ¿Para cuál empleado es más probable que la prueba de hipótesis encuentre significancia estadística de los efectos de las vacaciones en los tiempos de llamada? Justifique su respuesta.
**Piense en el poder estadístico de la prueba de hipótesis.*

Para simplificar las explicaciones se denominará "semana estándar" a todas las semanas que no corresponden ni a la semana de vacaciones ni la semana siguiente a esa. Y "semana de recuperación" a la primera semana de vuelta de vacaciones.

En esta pregunta nos piden determinar para cuál empleado existe una mayor probabilidad de rechazar la hipótesis nula de la prueba de hipótesis de diferencia de medias, es decir, rechazar que las medias son iguales, suponiendo que la hipótesis nula es falsa. Esto, por definición, es equivalente a determinar para cuál empleado existe un mayor poder estadístico para la prueba de hipótesis de diferencia de medias.

A continuación se expone la fórmula para calcular el **Poder estadístico** de una prueba de hipótesis de diferencia de medias a una cola (superior):

$$\begin{aligned} P.E. &= \mathbb{P}(t > t_\alpha - t_\gamma) \\ t_\gamma &= \frac{\mu_1 - \mu_2}{\widehat{SE(X)}} \end{aligned}$$

**Cabe destacar que para nuestro ejemplo μ_1 es la media muestral durante la "semana de recuperación" y μ_2 durante las "semanas estándar".*

Como la prueba de hipótesis para todos los empleados sería realizada al mismo nivel de significancia (α) y la cantidad de llamadas durante la "semana de recuperación" y durante el resto de "semanas estándar" es igual, entonces, los grados de libertad, y por ende, la distribución del estadístico t bajo la hipótesis nula también son iguales para todos los trabajadores. Luego, el primer término de la resta (t_α) y la distribución de t son idénticas para todos los trabajadores; por lo que el empleado que posea un valor mayor para t_γ tendrá un mayor **poder estadístico** para la prueba, y por ende, una mayor probabilidad de evidenciar efectos estadísticamente significativos del efecto perjudicial de las vacaciones sobre la duración de los llamados en la "semana de recuperación".

El término t_γ es el estadístico t de la prueba de hipótesis requerida y depende de la diferencia de las medias muestrales[1] de duración de las llamadas y del estimador del error estándar bajo la nula. Por lo que, quien posea el mayor valor para el numerador o el menor valor para el denominador de la fórmula que define a t_γ no necesariamente tendrá el mayor

[1] Semana de recuperación contra semanas estándar.

valor del estadístico. Luego, sólo resta estimar el valor de t_γ para cada empleado (Ver Tabla 1).

Nombre	t_γ
Marta	0.856
Tomas	2.137
Javier	0.681
Francisca	2.702
Felipe	1.066

Tabla 1: *Estadísticos t_γ de todos los empleados truncados al tercer decimal*

Finalmente, **Francisca** es la empleada que posée el mayor valor para t_γ, y por ende, un mayor poder estadístico para la prueba de hipótesis de diferencia de medias de una cola. Por lo que para Francisca es más probable que la prueba de hipótesis encuentre significancia estadística de los efectos de las vacaciones en los tiempos de llamada.

**Cabe destacar que los resultados se mantienen para la prueba de hipótesis que no supone igual varianza poblacional, ya que cambiarían los grados de libertad pero la posición relativa de cada empleado en relación al valor de t_γ es mantiene igual.*

1.2. Pregunta 2

Utilizando una prueba de diferencia de medias, evalúe para cada agente separadamente si la duración de su llamada incrementa en la semana después de vacaciones.

Se realiza la siguiente prueba de hipótesis de una cola:

$$H_0 : \mu_{Recuperacion} = \mu_{Estandar}$$
$$H_1 : \mu_{Recuperacion} > \mu_{Estandar}$$

Los resultados son los siguientes:

Nombre	Estadistico t	Valor-p
Marta	0.856	0.196
Tomas	2.137	0.016
Javier	0.681	0.248
Francisca	2.702	0.003
Felipe	1.066	0.143

Tabla 2: *Prueba de hipótesis de diferencia de medias de una cola*

Sólo se encuentra evidencia para rechazar[2] que la duración promedio durante las semanas "estándar" es menor o igual a la de la semana de "recuperación" para Tomas y Francisca. Para los demás, no rechazamos que las duraciones sean iguales.

1.3. Pregunta 3

Conduzca una prueba de hipótesis para evaluar si la tasa de fallo para un agente mujer es diferente a la de un agente hombre.
(*Nota: Francisca y Marta son mujeres, el resto son hombres*)

La tasa de fallo muestral de las mujeres[3] es igual a la de los hombres (0,075), por lo que, el estadístico para prueba de diferencia de proporciones será cero. En consecuencia, no se rechaza que ambas proporciones (tasa de fallo de hombres y tasa de fallo de mujeres) sean iguales.

1.4. Pregunta 4

El departamento de Marketing de *QuickCall* considera que el tiempo razonable para atender la llamada de un cliente es entre 15 segundos y 7 minutos. Encuentre, para cada agente, la probabilidad de alcanzar este estándar en una semana regular y en la semana siguiente a las vacaciones.

Podemos replantear lo que nos piden en el enunciado de la siguiente forma:

$$\mathbb{P}\left(0{,}25 \leq x_i \leq 7\right) = \mathbb{P}(x_i \leq 7) - \mathbb{P}(x_i \leq 0{,}25) = F(7) - F(0{,}25)$$

En donde $F(\cdot)$ es la función de densidad acumulada de la distribución poblacional de x_i.

La función $F(\cdot)$ dependerá de la distribución poblacional de cada serie, por lo que, realizaremos la prueba de normalidad de Shapiro-Wilk para verificar si la duración de las llamadas para cada uno de los empleados, durante las "semanas estándar" y "semana de recuperación" siguien una distribución normal (al normalizar las series).

En la Tabla 3 Podemos notar que sólo podemos rechazar la hipótesis nula de distribución normal al 5 % de significancia para Felipe durante la semana de recuperación[4]. Por lo que, para el resto de los casos utilizaremos la distribución Normal con media $\bar{x}$ y desviación estándar $\hat{\sigma_x}$.

En la Tabla 4 se exponen las probabilidades estimadas de que cada empleado alcance el estándar establecido durante la "semana estándar" o "semana de recuperación":

[2] A un nivel de significancia del 5 %.

[3] Tasa de fallo pooled, es decir, el promedio de fallos para las llamadas agrupadas de todo el género.

[4] Para Felipe en la semana de recuperación aplicaremos la transformación $\frac{(\cdot)-\bar{x}}{\widehat{SE(X)}}$ y a ese valor lo compararemos con la distribución t-student con $n-1$ grados de libertad.

Nombre	Semana estándar	Semana recuperación
Marta	0.861	0.790
Tomas	0.782	0.066
Javier	0.677	0.713
Francisca	0.794	0.229
Felipe	0.835	0.016

Tabla 3: *Valores-p prueba de normalidad Shapiro-Wilk para distintas semanas, para todos los empleados*

Nombre	Semana estándar	Semana recuperación
Marta	0.996	0.999
Tomas	0.911	0.792
Javier	0.971	0.924
Francisca	0.999	0.999
Felipe	0.999	0.999

Tabla 4: *Probabilidades estimadas de alcanzar el estándar (15 segs - 7 minutos).*

2. Ejercicio 2

2.1. Pregunta 1

Estime la duración promedio de una cirugía cardiaca. Compute el Erro Estándar de la estimaci⸮

- Promedio: $373,97$
- Error Estándar: 19,153

2.2. Pregunta 2

Provea un intervalo de confianza al 95 % para la duración promedio de una cirugía, asumiendo que la duración distribuye Normal.

- Intervalo de Confianza: [336,43 ; 411,51]

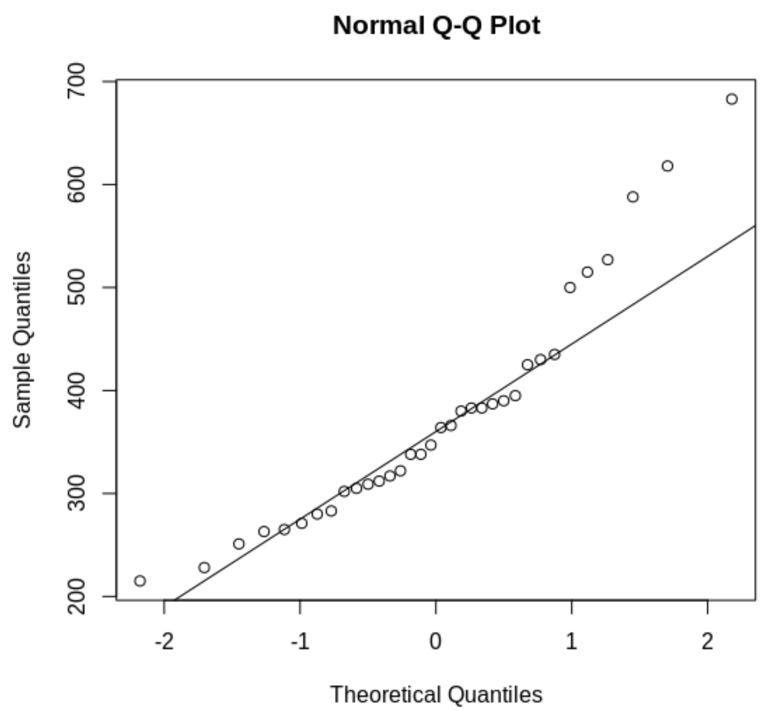
2.3. Pregunta 3

¿Cuál es la probabilidad de que una cirugía se demore más de 6 horas? ¿Y más de 8 horas? Provea una respuesta asumiendo que la duración de las cirugías siguen una distribución Normal.

- Probabilidad de que una cirugía demore más de 6 horas: 54,977 %
- Probabilidad de que una cirugía demore más de 8 horas: 17,121 %

2.4. Pregunta 4

Use un qq-plot para evaluar si las duraciones de las cirugías sigue una distribución Normal.



Podemos observar que las colas de la distribución empírica de las cirugías se alejan de las colas de una distribución normal. Esto es más evidente en la cola superior (quintiles más altos). Por lo que, visualmente descartaríamos que las duraciones siguen una distribución Normal.

2.5. Pregunta 5

Suponga que las cirugías no siguen una distribución Normal. Responda la pregunta 3 utilizando la distribución empírica de las duraciones de cirugías.

- Probabilidad de que una cirugía demore más de 6 horas: 50 %
- Probabilidad de que una cirugía demore más de 8 horas: 17,647 %

2.6. Pregunta 6

La calendarización de las cirugías es pasada por alto cuando la duración de la cirugía excede la prevista. Conduzca una prueba de hipótesis para evaluar si los casos de emergencias y de no emergencias difieren en sus probabilidades de ser pasadas por alto.

El *z-score* para la prueba de hipótesis de proporciones diferentes en este caso es igual a $-0{,}216$ (truncado a 3 decimales), lo que no supera en magnitud a los valores críticos del z-score para rechazar la hipótesis nula de igual proporción[5]. Luego, no se rechaza que la proporción de "cirugías pasadas por alto" sea igual para emergencias y no emergencias.

2.7. Pregunta 7

Los pacientes son separados en tres grupos etarios (50 o menos, 51 a 70, 71 o más). Evalúe si los tres grupos difieren en la duración promedio de sus cirugías.

Para evaluar si los promedios de duración de cirugías para los tres grupos etarios son diferentes estadísticamente se realizó una prueba ANOVA cuya hipótesis nula es igual media muestral de los tres grupos.

Los resultados son los siguientes:

Grados de Libertad	Estadístico F	Valor p
2	1.244	0.29

Tabla 5: *Resultados prueba ANOVA para media poblacional de tres muestras.*

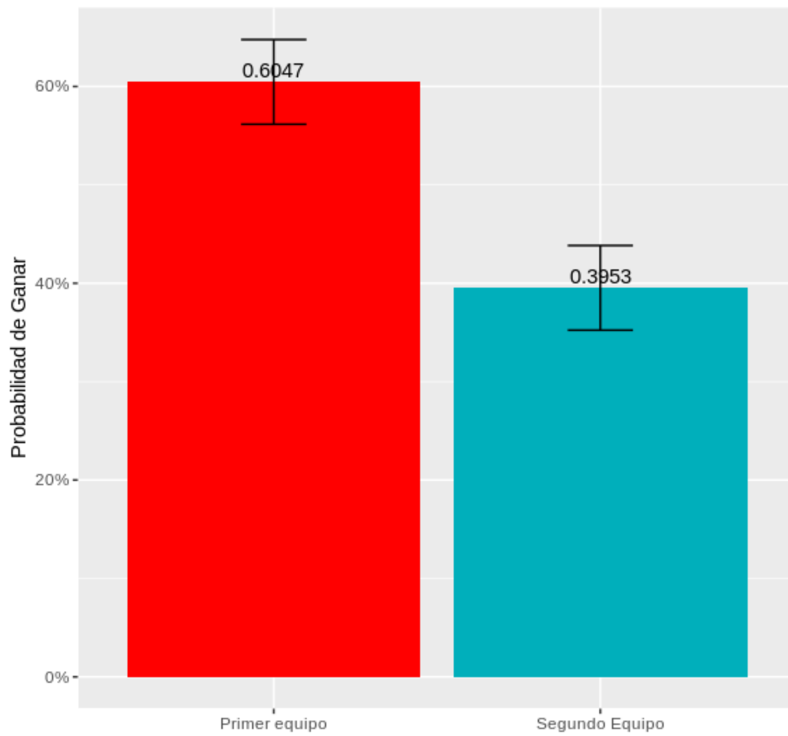
No existe suficiente evidencia para rechazar que la duración de las cirugías de los tres grupos etarios son diferentes estadísticamente[6].

[5] Los valores críticos para el *z-score* con un 95 % de significancia son 1,959964 y $-1{,}959964$.
[6] A un nivel de significancia de 95 %.

3. Ejercicio 3

3.1. Pregunta 1

Dibuje la Figura 1 del artículo e incluya intervalos de confianza para cada barra.



Intervalos de Confianza para probabilidad de ganar:

Equipo	Límite Inferior	Límite Superior
Primer Equipo	0.56	0.65
Segundo Equipo	0.35	0.44

3.2. Pregunta 2

Replique la primera prueba de la página 2554, evaluando si existe una ventaja de ser el primero en patear los penales.

Para evaluar la significancia estadística de una proporción muestral respecto a un valor teórico poblacional, hay al menos tres opciones. Se puede realizar un (1) test binomial exacto

; (2) prueba de hipótesis bajo supuesto de distribución Normal ; (3) prueba de hipótesis bajo supuesto de distribución Chi-cuadrado. Debido a que estas tres pruebas poséen diferentes supuestos, el valor-p de las tres formas podrían diferir.

Los valores-p de las tres vías son las siguientes[7]:

Prueba	Valor-p
(1)	2.17 %
(2)	1.505459 %
(3)	1.74438 %

Luego, podemos concluir que en el artículo utilizaron la prueba (3).

3.3. Pregunta 3

Asuma que la probabilidad de ganar para el equipo que comienza es 60 %, calcúle el poder estadístico de la prueba.

A continuación se exponen los poderes estadísticos de las tres pruebas expuestas anteriormente, asumiendo un $\hat{p}$ de 60 %:

Prueba	Poder Estadístico
(1)	0.6400026
(2)	0.6400026
(3)	0.3110734

**Cabe destacar que el poder estadístico de las primeras dos pruebas siempre son iguales.*

3.4. Pregunta 4

Explíque por qué el artículo sólo utiliza los datos anteriores al 2003.

El artículo sólo utiliza datos previos al 2003, debido a que durante ese año se realizó un cambio en la modalidad de elección del equipo que pateaba. Previo al año 2003 se sorteaba qué equipo patearía primero, y posteriormente se sortea qué equipo decide si patear primero o no.

Es importante realizar esta división, ya que al asignar de manear aleatoria al equipo que patea en primer lugar, se trata de evitar que no se cumpla el supuesto de "*unconfoundedness*". Este supuesto se viola cuando la esperanza del error condicional en las variables observables

[7] Para las vías (2) y (3) se realizan pruebas de dos colas.

(por las que se controla en las regresiones) es distinto de cero. Es decir, cuando existe una variable inobservable que está correlacionada con la acción de patear primero y con otras variables de los modelos, lo que genera un sesgo al estimar el coeficiente asociado a la variable de tratamiento (patear primero o no).

Un ejemplo hipotético de "confounder" que surge si se utilizan los datos posteriores al 2003, sería el caso de que los equipos más agresivos estratégicamente prefieran siempre patear primero, y los equipos menos agresivos opten por patear segundos. En este caso, "ser más agresivo" podría estar correlacionado con mayor probabilidad de ganar, lo que resultaría en una sobreestimación de la magnitud del coeficiente asociado a patear primero (si no se incluyen proxies para agresividad estratégica).