**Guía de Marketing Digital con Estadística Aplicada en R**

Introducción

El marketing digital es un campo que depende en gran medida del análisis de datos para tomar decisiones informadas. Con R, los analistas y técnicos de marketing pueden realizar estudios profundos sobre productos, precios, comunicación y estrategias de ‘retail’ o distribución utilizando la estadística aplicada. En este módulo cubriremos cómo usar R para cada uno de estos aspectos clave. El código ha sido probado y puede ser descargado del repositorio del curso en GitHub:

* Link: <https://github.com/alejandroperezpmarketing/curso_analitica_marketing_python_R.git>

Desde este punto deben tener la mente muy abierta a los cambios. Vamos a introducir aplicaciones de programación en la que, y en términos generales, un técnico de marketing no está habituado a usar, sino un técnico informático, analistas de datos, estatista, matemático, programador, o un ingeniero. No vamos a ir muy profundo en la programación, sino que nuestro objetivo en este módulo y en todos en general es de enseñarles cómo aplicar el conocimiento ya creado en proyectos del área de marketing para resolver preguntas clave del departamento.

¿Te suena conocida la siguiente frase de tu bachillerato? “Para que aprender matemáticas, sino me va a servir de nada saber cómo hacer matrices, logaritmos, o integrales en mi vida”, “Que aburridas son las matemáticas, paso de eso”. Nuestra meta en este módulo es resolver esto de una vez por todas, y aplicar las matemáticas, más concretamente la estadística en el marketing digital.

R será nuestra herramienta de uso común, y R estudio el entorno IDE (tanto para R como para Python). Se lo puede descargar usando el siguiente enlace <https://posit.co/downloads/> . En el directorio del curso podrás encontrar también un fichero requirements con todas las librerías a usar en todo el curso, listas para instalar (control + enter con el documento abierto). Para descargar el repositorio puedes usar Git desde una terminal de Windows o Linux copiando y pegando la siguiente sentencia:

“git clone <https://github.com/alejandroperezpmarketing/curso_analitica_marketing_python_R.git>”

**Primeros pasos**

**Bienvenido, necesitamos comentar un procedimiento dentro tu empresa u organización, y para eso eres sumamente necesario.** Hasta el momento se han hecho las cosas bien, pero cómo puedes saber, *las cosas que necesitan mantenimiento, si no se dañan.*

Vamos a iniciar un proceso de mucha formación, reuniones, y análisis de lo que a partir de ahora es y será nuestra materia prima para entender a nuestro cliente, y sobre eso hacer nuestro trabajo: **“los datos”.**

Un ejemplo claro, es el uso de las nuevas tecnologías (machine learning) para poder predecir cuentos clientes perderemos, y la perdida de ingresos que eso reflejaría a la empresa a través del tiempo. Esto nos ayuda a mejorar los procesos, embudos de venta, hacer estudios más detallados, y prevenir los problemas antes que sean una realidad.

**Gráfico, Gráfico de dispersión

Descripción generada automáticamente**

Ilustración : Ejemplo sobre la predicción de la tasa de pérdida de clientes a la empresa, y la pérdida de ingresos reflejados por su salida, mes a mes, a vista un año, y a vista dos años.

***Eres un dato en el mundo***

**En la actualidad, tu presencia** [**son datos**.](https://draft.blogger.com/blog/post/edit/6219196515360688826/6587716866247029378) Estos pueden ser cuantitativos como el número de mensajes enviados a un cliente, el tiempo de un usuario en nuestro sitio web, número de personas en llegar a la web por cada red social y el número que llegan desde tráfico orgánico, es decir haciendo click en alguno de los contenidos SERP, ya sean videos, imágenes, enlaces y otros (search engine results pages), el  canal utilizado, o pueden ser categóricos (o cualitativos) como si es un mensaje positivo, negativo o neutro sobre comentarios, y opiniones en entrevistas, reuniones de trabajo, negociaciones o incluso con comentarios  en redes sociales (análisis de emociones usando).

Datos como una coordenada como la latitud o la longitud o la “x” y la “y” de donde nos encontramos, así como también, donde se encuentran nuestros visitantes (nuestro mercado potencial) y nuestros seguidores, nuestros clientes finales, y quienes nos van a recomendar entre otros que intervienen en el ciclo, pueden darnos información valiosa, si la aplicamos en nuestra empresa para conocer patrones, y oportunidades de mejora, así como nuevos mercados a donde invertir.

Igualmente podemos hacer un seguimiento de nuestras ventas, y características de nuestros clientes, compararlas con diferentes variables que afectan a nuestra empresa, y nos afectan como trabajadores, y compañeros de trabajo.

Si nos ponemos a pensar, podemos hacer mucho con solo una ubicación.

**¿Y si le agregamos las ventas?** podemos hacer un seguimiento de qué cliente nos aporta más rentabilidad (quién da más…) por región, por ciudad, por provincia, etc…

Una caricatura de una persona

Descripción generada automáticamente con confianza media

Pongamos otro ejemplo. ¿Sabemos de donde vienen las ventas, si hacemos una campaña de comunicación por diferentes medios de comunicación? Podemos determinar cuánto está relacionado con las ventas de la campaña general las ventas en cada medio, respondiendo a la pregunta, ¿dónde lanzamos la campaña? ¿Facebook, Instagram, X, Pinterest…? ¿TV, Radio, periódicos? Además de eso, podemos saber el impacto que ha tenido cada red social en las demás. ¿cuándo aumenta las ventas por la campaña en TV, Aumenta en periódicos? Puede que un canal no lleve a un cliente potencial al final del embudo de ventas, pero sí que puede dejar nuestra imagen de marca presente y ser más reconocible en otros medios, haciendo más probable la venta, es el caso de la radio y los periódicos.

Imagen que contiene ventana, biombo, juego, edificio

Descripción generada automáticamente

Ilustración : Ejemplo de las correlaciones entre las ventas procedentes de una campaña de comunicación en TV, periódicos, y la radio. Esto nos ayuda a ver la interaccción entre los diferentes medios, y eleger el más optimo para nuestras próximas campañas.

Si realmente cada uno de nosotros entendemos que una empresa está en el mercado generando datos, para poder resolver los problemas de los clientes, en ese momento podremos conseguir estudiar su comportamiento y sus las demandas como fin último, y principal para darles una mejor solución, y mejorar la satisfacción de nuestro cliente.

No todo es dinero, también los **recursos humanos** generan datos, como son: datos de los empleados, material de oficina, número de bajas por enfermedad entre otros. Estos nos permiten conocer la motivación del equipo, y si lo que hacemos corresponde con nuestra existencia en un mercado.  Muchas veces ignoramos que las primeras personas que consumen nuestros productos o servicios, (o interactúan con él) son los empleados de la empresa, y hacer estudios internos de forma frecuente puede darnos buenos indicadores sobre el valor agregado de nuestros productos, así como incluir al trabajador en el proceso productivo.

Hacer nuestro trabajo como empleados “más cómodo” es trabajar para garantizar nuestra propia salud como personas, e incluso nuestra futura estancia en un mercado laboral. **Recuerda: “como el tiempo relativo también... los mercados evolucionan, y con ello sus necesidades laborales”.** Medir estos aspectos nos ayudan a evaluar la inversión en programas internos de formación, y mejoras en la materia prima que día a día tiene contacto directo, o indirectamente con el cliente.

Aquí nos centraremos a recolectar, guardar, limpiar, y analizar datos generados por cada etapa del Inbound Marketing (extraños, visitas, leads, clientes, y promotores), en sitios webs, así como los medios sociales.

Primero, hablemos de la base que usaremos para trabajar con nuestros datos en R, **la estadística univariante, y bivariante en R.**

1. Análisis de Productos

El análisis de productos busca identificar características clave que impactan el rendimiento de un producto en el mercado.

Principales objetivos:

* Identificar características más valoradas por los clientes.
* Evaluar la satisfacción del cliente.
* Analizar la canibalización entre productos.

*Ejemplo en R:*

*Análisis de satisfacción del cliente con encuestas*

*# Instalar y cargar paquetes*

*install.packages(c("dplyr", "ggplot2", "psych"))*

*library(dplyr)*

*library(ggplot2)*

*library(psych)*

*# Simular datos de encuestas*

*set.seed(123)*

*data <- data.frame(*

*Producto = sample(c("Producto A", "Producto B"), 200, replace = TRUE),*

*Satisfaccion = sample(1:5, 200, replace = TRUE)*

*)*

*# Resumen de datos*

*summary(data)*

*# Analizar distribución de satisfacción por producto*

*ggplot(data, aes(x = Satisfaccion, fill = Producto)) +*

*geom\_bar(position = "dodge") +*

*labs(title = "Distribución de Satisfacción por Producto",*

*x = "Nivel de Satisfacción",*

*y = "Frecuencia") +*

*theme\_minimal()*

*# Prueba t para comparar satisfacción entre productos*

*t.test(Satisfaccion ~ Producto, data = data)*

**2. Análisis de Precios**

El análisis de precios busca optimizar la estrategia de precios para maximizar ingresos y rentabilidad.

Principales objetivos:

* Determinar la elasticidad del precio.
* Identificar puntos de precio óptimos.
* Evaluar estrategias de descuentos.

Ejemplo en R:

Elasticidad precio-demanda

# Simular datos de precio y demanda

set.seed(123)

data <- data.frame(

Precio = seq(5, 50, by = 5),

Demanda = 1000 / seq(5, 50, by = 5) + rnorm(10, mean = 0, sd = 10)

)

# Modelo de regresión lineal

modelo <- lm(Demanda ~ Precio, data = data)

summary(modelo)

# Visualizar relación precio-demanda

ggplot(data, aes(x = Precio, y = Demanda)) +

geom\_point() +

geom\_smooth(method = "lm", se = FALSE, color = "blue") +

labs(title = "Relación Precio-Demanda",

x = "Precio",

y = "Demanda") +

theme\_minimal()

**3. Análisis de Comunicación**

El análisis de comunicación permite evaluar la efectividad de las campañas y el impacto de los mensajes publicitarios.

Principales objetivos:

* Evaluar el alcance y la impresión.
* Analizar el sentimiento de la audiencia.
* Medir el impacto en las ventas.

Ejemplo en R:

Análisis de sentimiento en redes sociales

# Instalar y cargar paquetes

install.packages(c("rtweet", "tidytext", "dplyr", "ggplot2"))

library(rtweet)

library(tidytext)

library(dplyr)

library(ggplot2)

# Obtener tweets relacionados con una marca

tweets <- search\_tweets("marca", n = 500, lang = "es")

# Análisis de sentimiento

tweets\_clean <- tweets %>%

unnest\_tokens(word, text) %>%

anti\_join(stop\_words("es"))

sentimientos <- tweets\_clean %>%

inner\_join(get\_sentiments("bing")) %>%

count(sentiment)

# Visualizar resultados

ggplot(sentimientos, aes(x = sentiment, y = n, fill = sentiment)) +

geom\_bar(stat = "identity") +

labs(title = "Análisis de Sentimiento",

x = "Sentimiento",

y = "Frecuencia") +

theme\_minimal()

**4. Análisis de Retail**

El análisis de retail se enfoca en el rendimiento de los puntos de venta y las estrategias de distribución.

Principales objetivos:

* Optimizar el surtido de productos.
* Identificar patrones de compra.
* Analizar la distribución geográfica de ventas.

Ejemplo en R:

Análisis de patrones de compra con cestas de mercado

# Instalar y cargar paquetes

install.packages("arules")

library(arules)

# Simular transacciones

data <- data.frame(

ID = 1:100,

Productos = replicate(100, paste(sample(c("Pan", "Leche", "Huevos", "Cereal"), 3), collapse = ", "))

)

# Convertir datos a transacciones

trans <- as(split(data$Productos, data$ID), "transactions")

# Reglas de asociación

reglas <- apriori(trans, parameter = list(supp = 0.01, conf = 0.8))

inspect(reglas)

Conclusión

Esta guía proporciona una introducción práctica al uso de R en marketing digital para analizar productos, precios, comunicación y retail. Implementar estas técnicas permite tomar decisiones basadas en datos y optimizar las estrategias de marketing.

**Guía: Aplicación de Contrastes de Hipótesis en Marketing Digital con Estadística en R**

Introducción

En marketing digital, los contrastes de hipótesis son herramientas fundamentales para validar supuestos y tomar decisiones basadas en datos. Desde evaluar la efectividad de una campaña hasta comparar el rendimiento entre grupos de usuarios, el uso de estadística aplicada en R permite abordar estos análisis de manera rigurosa.

En esta guía, aprenderás cómo implementar contrastes de hipótesis comunes en marketing digital con ejemplos prácticos en R.

1. Conceptos Básicos

Un contraste de hipótesis se utiliza para evaluar si existe suficiente evidencia en los datos para rechazar una hipótesis nula (¡H₀¡).

Pasos generales:

1. Formular la hipótesis:
   * Hipótesis nula (¡H₀¡): Afirma que no hay efecto o diferencia.
   * Hipótesis alternativa (¡Hₐ¡): Propone una diferencia o efecto.
2. Elegir la prueba estadística: Basada en el tipo de datos y el problema.
3. Calcular el valor p: Determina si se rechaza ¡H₀¡.
4. Tomar una decisión: Rechazar o no ¡H₀¡ según el nivel de significancia (α).

2. Tipos de Contrastes de Hipótesis en Marketing Digital

2.1 Comparación de medias

Ejemplo: Evaluar si el CTR (Click-Through Rate) de dos versiones de un anuncio es diferente.

Prueba t para muestras independientes

# Simular datos

set.seed(123)

ad\_version\_a <- rnorm(30, mean = 0.05, sd = 0.01) # CTR versión A

ad\_version\_b <- rnorm(30, mean = 0.06, sd = 0.01) # CTR versión B

# Prueba t

resultado\_t <- t.test(ad\_version\_a, ad\_version\_b, alternative = "two.sided")

print(resultado\_t)

# Interpretación

if (resultado\_t$p.value < 0.05) {

print("Hay una diferencia significativa entre las versiones del anuncio.")

} else {

print("No hay evidencia suficiente para afirmar una diferencia significativa.")

}

2.2 Proporciones

Ejemplo: Determinar si la tasa de conversión de dos páginas de aterrizaje es diferente.

Prueba de proporciones

# Datos

conversiones\_a <- c(50, 200) # 50 conversiones de 200 visitantes (página A)

conversiones\_b <- c(70, 200) # 70 conversiones de 200 visitantes (página B)

# Prueba de proporciones

resultado\_prop <- prop.test(x = c(conversiones\_a[1], conversiones\_b[1]),

n = c(conversiones\_a[2], conversiones\_b[2]),

alternative = "two.sided")

print(resultado\_prop)

# Interpretación

if (resultado\_prop$p.value < 0.05) {

print("Las tasas de conversión son significativamente diferentes.")

} else {

print("No hay evidencia suficiente para afirmar una diferencia significativa.")

}

2.3 Comparación de distribuciones

Ejemplo: Analizar si la distribución del tiempo en el sitio web difiere entre usuarios nuevos y recurrentes.

Prueba de Mann-Whitney U

# Simular datos

set.seed(123)

nuevos <- rnorm(50, mean = 300, sd = 50) # Tiempo en segundos (usuarios nuevos)

recurrentes <- rnorm(50, mean = 320, sd = 60) # Tiempo en segundos (usuarios recurrentes)

# Prueba de Mann-Whitney U

resultado\_mw <- wilcox.test(nuevos, recurrentes, alternative = "two.sided")

print(resultado\_mw)

# Interpretación

if (resultado\_mw$p.value < 0.05) {

print("La distribución del tiempo en el sitio web es significativamente diferente.")

} else {

print("No hay evidencia suficiente para afirmar una diferencia significativa.")

}

2.4 Asociación entre variables categóricas

Ejemplo: Evaluar si el dispositivo utilizado (móvil o escritorio) está asociado a la decisión de compra.

Prueba de Chi-cuadrado

# Crear tabla de contingencia

tabla <- matrix(c(30, 70, 50, 50), nrow = 2,

dimnames = list(Compra = c("Si", "No"),

Dispositivo = c("Movil", "Escritorio")))

# Prueba Chi-cuadrado

resultado\_chi <- chisq.test(tabla)

print(resultado\_chi)

# Interpretación

if (resultado\_chi$p.value < 0.05) {

print("Hay una asociación significativa entre el dispositivo y la decisión de compra.")

} else {

print("No hay evidencia suficiente para afirmar una asociación significativa.")

}

3. Conclusión

Los contrastes de hipótesis son esenciales para validar estrategias y optimizar decisiones en marketing digital. R ofrece una amplia gama de herramientas para realizar estas pruebas de manera eficiente y precisa. Con esta guía, tienes los fundamentos para implementar los contrastes más comunes y adaptarlos a tus necesidades específicas en el análisis de datos.

Guía: Aplicaciones de Distribuciones Probabilísticas en Marketing Digital con Estadística en R

Introducción

Las distribuciones probabilísticas son herramientas esenciales en el marketing digital para modelar comportamientos, predecir tendencias y tomar decisiones basadas en datos. En esta guía, exploraremos las aplicaciones de diferentes distribuciones probabilísticas comunes y cómo implementarlas en R para resolver problemas específicos en marketing digital.

1. Distribución Normal

La distribución normal es útil para modelar variables continuas como el tiempo en un sitio web o el monto promedio gastado por cliente.

Aplicación: Modelar el tiempo promedio en un sitio web

# Simular datos de tiempo en sitio web (en segundos)

set.seed(123)

tiempo\_sitio <- rnorm(1000, mean = 300, sd = 50)

# Visualizar la distribución

hist(tiempo\_sitio, breaks = 30, probability = TRUE, col = "lightblue",

main = "Distribución del Tiempo en el Sitio Web",

xlab = "Tiempo (segundos)")

curve(dnorm(x, mean = mean(tiempo\_sitio), sd = sd(tiempo\_sitio)),

col = "red", lwd = 2, add = TRUE)

# Resumen estadístico

summary(tiempo\_sitio)

2. Distribución Binomial

La distribución binomial modela eventos con dos posibles resultados, como clic o no clic en un anuncio.

Aplicación: Predicción de clics en una campaña publicitaria

# Parametros de la campaña

n\_anuncios <- 1000 # número de impresiones

p\_clic <- 0.2 # probabilidad de clic

# Simular clics

clics <- rbinom(1, size = n\_anuncios, prob = p\_clic)

# Resultados

print(paste("Número de clics esperados:", clics))

3. Distribución Poisson

La distribución Poisson es ideal para modelar eventos que ocurren en un intervalo de tiempo fijo, como el número de visitas por hora a un sitio web.

Aplicación: Modelar visitas por hora

# Tasa promedio de visitas por hora

lambda <- 30 # visitas por hora

# Simular visitas en 24 horas

visitas <- rpois(24, lambda = lambda)

# Visualizar los datos

barplot(visitas, col = "lightgreen",

main = "Visitas por Hora",

xlab = "Hora",

ylab = "Número de Visitas")

# Resumen estadístico

summary(visitas)

4. Distribución Exponencial

La distribución exponencial se utiliza para modelar tiempos entre eventos, como el tiempo entre dos conversiones consecutivas.

Aplicación: Tiempo entre conversiones

# Tasa de conversiones por minuto

lambda <- 0.1 # 1 conversión cada 10 minutos (en promedio)

# Simular tiempos entre conversiones

tiempos\_conversion <- rexp(100, rate = lambda)

# Visualizar la distribución

hist(tiempos\_conversion, breaks = 20, probability = TRUE, col = "lightpink",

main = "Tiempo entre Conversiones",

xlab = "Tiempo (minutos)")

curve(dexp(x, rate = lambda), col = "red", lwd = 2, add = TRUE)

# Resumen estadístico

summary(tiempos\_conversion)

5. Distribución Gamma

La distribución Gamma es adecuada para modelar duraciones o montos acumulativos, como la duración total de visitas en una sesión.

Aplicación: Duración total de sesiones en un día

# Parametros de la distribución Gamma

forma <- 2 # forma (shape)

escala <- 10 # escala (scale)

# Simular duraciones de sesión (en minutos)

duraciones <- rgamma(1000, shape = forma, scale = escala)

# Visualizar la distribución

hist(duraciones, breaks = 30, probability = TRUE, col = "lightblue",

main = "Duración Total de Sesiones",

xlab = "Duración (minutos)")

curve(dgamma(x, shape = forma, scale = escala), col = "red", lwd = 2, add = TRUE)

# Resumen estadístico

summary(duraciones)

6. Distribución Geométrica

La distribución geométrica modela el número de intentos necesarios para el primer éxito, como el número de correos enviados antes de una conversión.

Aplicación: Emails necesarios para la primera conversión

# Probabilidad de conversión por email

p\_conversion <- 0.05

# Simular número de emails hasta la primera conversión

emails <- rgeom(1000, prob = p\_conversion)

# Visualizar la distribución

hist(emails, breaks = 20, col = "lightcoral",

main = "Emails hasta la Primera Conversión",

xlab = "Número de Emails")

# Resumen estadístico

summary(emails)

Conclusión

El uso de distribuciones probabilísticas en marketing digital permite modelar comportamientos complejos, optimizar estrategias y predecir resultados. Con esta guía, tienes un conjunto de herramientas prácticas para aplicar estas distribuciones en tus análisis y mejorar la toma de decisiones basada en datos.

**Bibliografía**

1. Jordi Mas Elias, Análisis de datos con R en estudios internacionales, 2020
2. Cole Nussbaumer Knaflic, Storyteling con datos Visualización de datos para profesionales, 2017
3. Quintin Martín Martín, Contraste de hipótesis, 2001
4. Introducción a los SIG con R, Dominic Royé Roberto Serrano Notivoli, 2019