

Índice general

Índice general	i
Preámbulo	v
1 Generación de conocimiento a partir de datos	1
1.1 Datos, información y conocimiento	1
1.2 Métodos estadísticos	2
1.3 Inferencia estadística: población, muestra e incertidumbre asociada	3
2 Instalación y primeros pasos en R	5
2.1 R	5
2.2 Instalación	6
2.2.1 Instalación de R	6
2.2.2 Instalación e introducción a RStudio	7
2.2.3 Instalando paquetes	9
2.3 Importación y exportación de datos	11
2.3.1 Utilizando el importador	11
2.3.2 Utilizando comandos	14
2.3.3 Datos online	15
2.3.4 Datos precargados	15
2.3.5 Exportación	16
2.4 tidyverse	16
3 Análisis exploratorio de datos	19
3.1 Análisis exploratorio de datos	19

3.2	Primer vistazo al conjunto de datos	20
3.3	Conceptos básicos	22
3.4	Crear subconjuntos de datos	25
3.4.1	Subconjuntos de variables	25
3.4.2	Subconjuntos de registros	26
3.5	El concepto de frecuencia	28
3.5.1	Conceptos	28
3.5.2	Tablas de frecuencia	29
3.5.3	Histogramas	31
3.5.4	Funciones de densidad	34
3.6	Describiendo las distribuciones	35
3.6.1	Campana de Gauss	35
3.6.2	Medidas de posición	36
3.6.3	Medidas de variabilidad	39
3.6.4	Medidas de forma	41
3.7	Visualización de datos	44
3.7.1	Gráfico de barras	45
3.7.2	<i>Boxplot</i>	45
3.7.3	<i>Scatterplot</i>	48
3.7.4	<i>Pie chart</i>	50
3.7.5	Combinar gráficos	51
3.7.6	Exportar gráficos	52
4	La distribución normal	53
4.1	Las distribuciones de probabilidad de los datos	53
4.2	La distribución normal	54
4.3	La distribución normal tipificada	56
4.4	Análisis de normalidad	58
4.4.1	<i>Q-Q plot</i>	58
4.5	Otras distribuciones	59
4.5.1	Distribuciones discretas	60
4.5.2	Distribuciones continuas	63
5	Contraste de hipótesis	67
5.1	Inferencia estadística	67
5.2	Contrastes de hipótesis	69
5.3	Metodología científica	71
5.4	Definición y formulación de hipótesis	72
5.5	P-valor y nivel de significación	72
5.6	Errores tipo I y tipo II	73
5.7	Contraste unilateral y bilateral	75

5.8	Intervalos de confianza	75
5.9	Ejemplo práctico	76
5.10	Tamaño de efecto, poder estadístico y tamaño de muestra	78
5.10.1	Tamaño de efecto	79
5.10.2	Poder estadístico	79
5.10.3	Calculando el tamaño de la muestra	80
6	Test estadísticos	83
6.1	Métodos de contraste de hipótesis	83
6.2	Contrastes de una variable	86
6.2.1	Contraste para una proporción	86
6.2.2	Contraste para una media	89
6.2.3	Contraste para una varianza	92
6.3	Contraste para proporciones de más de una variable	94
6.3.1	Test de independencia Chi-cuadrado	94
6.3.2	Test McNemar	99
6.4	Contraste para medias: test de una variable y dos condiciones	101
6.4.1	t-test	101
6.4.2	t-test pareado	106
6.4.3	<i>Wilcoxon-Mann-Whitney test</i>	108
6.4.4	<i>Wilcoxon signed rank test</i> (pareado)	110
6.5	Contraste para medias: test de una variable con más de dos condiciones	112
6.5.1	ANOVA	113
6.5.2	Comparación múltiple: análisis post hoc	118
6.5.3	ANOVA de medidas repetidas	122
6.5.4	Kruskal-Wallis	125
6.5.5	Friedman	129
6.6	Contraste para medias: test de una variable y más de un factor	131
6.6.1	ANOVA (<i>two-way</i>)	131
6.7	Visualización de contraste de medias	137
6.8	Contraste para varianzas: análisis de la homocedasticidad	138
6.8.1	Test de Levene	140
6.9	Bondad de ajuste: análisis de la normalidad	141
6.9.1	Test Kolmogorov-Smirnov	142
6.9.2	Test Shapiro-Wilk	144
6.9.3	¿Qué test de bondad de ajuste utilizar?	144
6.10	Categorizando variables cuantitativas	145
7	Correlaciones y regresión lineal	149
7.1	La relación de variables cuantitativas	149

7.2	Correlación lineal	150
7.2.1	Correlación y causalidad	151
7.2.2	Covarianza y coeficientes de correlación	151
7.2.3	Contraste de correlación	152
7.2.4	Coefficiente de Pearson	153
7.2.5	Coefficiente de Spearman	153
7.2.6	Ejemplo de aplicación	154
7.3	Regresión lineal simple	157
7.3.1	Formulación	158
7.3.2	Coefficientes de regresión	158
7.3.3	Análisis de la precisión del modelo	160
7.3.4	Inferencia y estimación	164
7.3.5	Condiciones necesarias	165
7.3.6	Ejemplo de aplicación	166
7.4	Regresión lineal múltiple	169
7.4.1	Formulación	169
7.4.2	Coefficientes de regresión	170
7.4.3	Análisis de la precisión del modelo	170
7.4.4	Significación global del modelo	171
7.4.5	Condiciones necesarias	171
7.4.6	Ejemplo de aplicación	173
7.4.7	Inclusión de variables categóricas	176
7.4.8	Interacciones entre variables	181
7.4.9	Regresión polinómica	183
7.4.10	Selección de variables independientes	187
7.4.11	Validación y test	191
8	Detección de valores atípicos y faltantes	199
8.1	Valores que pueden invalidar el análisis	199
8.2	Valores atípicos u <i>outliers</i>	200
8.2.1	Detección univariante	201
8.2.2	Detección multivariante	206
8.3	Valores faltantes	209
8.3.1	Descripción y visualización	210
8.3.2	Imputación	215
	Bibliografía	219