



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

Grado en Estadística

**RICK Y TÚ - Asistente
conversacional para aprender
estadística**

Autora:

Blanca Alonso Prieto

Tutoras:

María Aránzazu Simón Hurtado

María Borbonés García

AGRADECIMIENTOS

A Arancha, por confiar en mí cuando te conté lo que quería hacer de este trabajo y tutorizarme y a María, por haberme enseñado tantas cosas durante toda mi estancia en IBM, sin vosotras no podría haber llevado a cabo este trabajo.

A mis amigos, tanto los de toda la vida como los que he conocido a partir de mis años de carrera, ya que sin vuestro apoyo incondicional no podría haber seguido adelante, gracias por todos esos grandes momentos que hemos pasado y por no faltar ni en los malos. Quería hacer una mención especial a dos personas que han sido muy importantes en la parte académica, mi gran amiga Marina, por todos los momentos de estudio y clases que hemos vivido juntas, sin ti no me podría haber reído tanto mientras aprendía y a Adri, por haber tenido tanta paciencia a la hora de explicarme las cosas y haberme sabido transmitir esa pasión que sientes por lo que haces.

A mi gran familia, mi ejemplo a seguir, por toda su dedicación y todos esos ánimos que siempre me habéis estado dando. Gracias a vosotros he llegado hasta aquí.

A mi misma, por seguir teniendo fuerzas para levantarme pase lo que pase, continuar mejorando tanto profesional como personalmente, recordar siempre que, aunque me quede mucho por mejorar, no tengo que perder esa esencia que me caracteriza y por no dejar de cuidar nunca a todas las personas a las que quiero y a mí misma.

ÍNDICE

ÍNDICE DE FIGURAS	9
ÍNDICE DE ECUACIONES	12
RESUMEN	13
ABSTRACT	13
1. INTRODUCCIÓN	14
2. METODOLOGÍA	16
<i>2.1 Storytelling</i>	16
<i>2.2 Gamificación</i>	17
3. HERRAMIENTAS	18
<i>3.1 IBM Cloud</i>	18
<i>3.2 Watson Assistant</i>	20
<i>3.3 Node-red</i>	26
<i>3.4 Telegram</i>	27
4. ASISTENTE CONVERSACIONAL	29
<i>4.1 Icono, nombre e historia</i>	29
<i>4.2 Proceso de creación y problemas surgidos</i>	31
<i>4.3 Contenido</i>	39
5. ENCUESTA DE OPINIÓN Y RESULTADOS	123
6. CONCLUSIONES Y TRABAJO FUTURO	129
BIBLIOGRAFÍA	131
ANEXO I - CÓDIGO INICIAL DEL NODO FUNCTION EN NODE-RED	133
ANEXO II - ACCESO AL ASISTENTE RICK Y TÚ	135
ANEXO III - ENCUESTA DE OPINIÓN	144
ANEXO IV - CÓDIGO DEL ASISTENTE Y DEL FLUJO DE NODE-RED	151

ÍNDICE DEL CONTENIDO DEL ASISTENTE

HISTORIA	40
1. Estadística descriptiva	41
<i>1.1 Conceptos clave</i>	42
1.1.1 Población y Muestra	42
1.1.2 Variable	43
1.1.2.1 Variable cualitativa	43
1.1.2.2 Variable cuantitativa	44
1.1.3 Tabla de frecuencias	44
1.1.3.1 Misión 1	44
<i>1.2 Medidas de posición</i>	45
1.2.1 Media	45
1.2.1.1 Misión 2	46
1.2.2 Mediana	47
1.2.2.1 Misión 3	47
1.2.3 Moda	48
1.2.3.1 Misión 4	48
1.2.4 Cuantiles	49
1.2.4.1 Misión 5	49
<i>1.3 Medidas de dispersión</i>	50
1.3.1 Varianza, cuasivarianza y desviación típica	50
1.3.1.1 Misión 6	51
1.3.2 Recorrido, rango intercuartílico y coeficiente de variación	51
1.3.2.1 Misión 7	52
1.3.3 Tchebychev	52
1.3.3.1 Misión 8	53
<i>1.4 Medidas de forma</i>	54
1.4.1 Asimetría	54
1.4.1.1 Misión 9	54
1.4.2 Curtosis	55
1.4.2.1 Misión 10	55
<i>1.5 Representaciones gráficas</i>	56
1.5.1 Diagrama de sectores	56
1.5.2 Diagrama de barras	57
1.5.3 Diagrama de Pareto	58
1.5.4 Histograma	59
1.5.5 Polígono de frecuencias	60
1.5.6 Diagrama de tallo y hojas	61
1.5.7 Diagrama de caja	62
2. Probabilidad	63
<i>2.1 Experimento aleatorio</i>	64

<i>2.2 Suceso</i>	64
2.2.1 Operaciones	64
2.2.1.1 Misión 11	65
2.2.2 Sistema completo de sucesos	66
<i>2.3 Variables aleatorias</i>	66
2.3.1 Variable aleatoria discreta	66
2.3.1.1 Función de probabilidad	67
2.3.1.2 Función de distribución discreta	67
2.3.2 Variable aleatoria continua	68
2.3.2.1 Función de densidad	69
2.3.2.2 Función de distribución continua	70
2.3.3 Medidas de una variable aleatoria	71
2.3.3.1 Esperanza	71
2.3.3.2 Varianza	71
2.3.3.3 Mediana, cuantil y moda	72
<i>2.4 Propiedades</i>	72
2.4.1 Axiomas de Kolmogorov	73
2.4.2 Resultados	73
<i>2.5 Regla de Laplace</i>	74
2.5.1 Misión 12	74
<i>2.6 Probabilidad condicionada</i>	75
2.6.1 Misión 13	75
<i>2.7 Regla del producto</i>	76
2.7.1 Misión 14	76
<i>2.8 Independencia de sucesos</i>	77
2.8.1 Misión 15	77
<i>2.9 Teoremas</i>	78
2.9.1 Teorema de las probabilidades totales	78
2.9.1.1 Misión 16	79
2.9.2 Teorema de Bayes	79
2.9.2.1 Misión 17	81
3. Modelos probabilísticos	82
<i>3.1 Modelos discretos</i>	82
3.1.1 Bernoulli	82
3.1.2. Binomial	83
3.1.2.1 Misión 18	83
3.1.3 Geométrica	84
3.1.3.1 Misión 19	85
3.1.4 Binomial negativa	85
3.1.4.1 Misión 20	86
3.1.5 Hipergeométrica	87
3.1.5.1 Misión 21	88
3.1.6 Poisson	88
3.1.6.1 Misión 22	89
<i>3.2 Modelos continuos</i>	90
3.2.1 Uniforme	90

3.2.1.1 Misión 23	91
3.2.2 Exponencial	92
3.2.2.1 Misión 24	92
3.2.3 Normal	93
3.2.3.1 Misión 25	95
3.2.4 Chi-cuadrado	95
3.2.5 T de Student	96
3.2.6 Fisher-Snedecor	97
<i>3.3 Teorema central del límite</i>	97
4. Inferencia estadística	98
<i>4.1 Muestreo aleatorio simple</i>	98
<i>4.2 Estimación puntual</i>	98
4.2.1 Definiciones	99
4.2.2 Estimación de la media	99
4.2.3 Estimación de la varianza	99
4.2.4 Estimación de una proporción	100
4.2.5 Distribución muestral de un estimador	100
4.2.6 Calidad de un estimador	101
4.2.6.1 Sesgo de un estimador	101
4.2.6.2 Consistencia de un estimador	102
<i>4.3 Estimación por intervalos</i>	102
4.3.1 Intervalo de confianza para la media	102
4.3.1.1 Misión 26	103
4.3.2 Intervalo de confianza para la desviación típica	104
4.3.2.1 Misión 27	104
4.3.3 Intervalo de confianza para una proporción	105
4.3.3.1 Misión 28	105
4.3.3 Intervalo de confianza para la diferencia de medias	106
4.3.4 Intervalo de confianza para la razón de varianzas	108
4.3.5 Intervalo de confianza para la diferencia de proporciones	108
<i>4.4 Contraste de hipótesis paramétricas</i>	108
4.4.1 Conceptos esenciales	109
4.4.1.1 Hipótesis nula y alternativa	109
4.4.1.2 Error tipo I y error tipo II	110
4.4.1.3 Contrastos unilaterales y contrastes bilaterales	110
4.4.1.4 Región crítica de un contraste	111
4.4.1.5 P-valor	111
4.4.2 Estadísticos utilizados para hacer contrastes paramétricos	111
4.4.2.1 Estadístico para la media	111
4.4.2.2 Estadístico para la varianza	112
4.4.2.3 Estadístico para la diferencia de medias	113
4.4.2.4 Estadístico para la razón de varianzas	114
4.4.2.5 Estadístico para una proporción y su diferencia	114
4.4.3 Cómo realizar un contraste de hipótesis paramétrico	115
4.4.3.1 Misión 29	115
5. Regresión Lineal Simple	115
<i>5.1 Covarianza</i>	116

<i>5.2 Correlación lineal</i>	116
<i>5.3 Diagrama de dispersión o nube de puntos</i>	117
<i>5.4 Recta de regresión</i>	118
<i>5.5 Mínimos cuadrados</i>	119
<i>5.6 Problemas de la regresión lineal simple</i>	121
6. Otros	121
<i>6.1 Final</i>	121
<i>6.2 Siguiente</i>	121
<i>6.3 Ayuda</i>	121
<i>6.4 En otras cosas</i>	122

ÍNDICE DE FIGURAS

<i>Figura 1: Pizza como servicio</i>	18
<i>Figura 2: On-Premises, IaaS, PaaS y SaaS</i>	19
<i>Figura 3: Intenciones de Watson Assistant</i>	20
<i>Figura 4: Ejemplos de una intención en Watson Assistant</i>	21
<i>Figura 5: Entidades de Watson Assistant</i>	21
<i>Figura 6: Nodos de diálogo en Watson Assistant</i>	22
<i>Figura 7: Tipos de respuesta en Watson Assistant</i>	22
<i>Figura 8: Slots en Watson Assistant</i>	23
<i>Figura 9: Digresiones en Watson Assistant</i>	25
<i>Figura 10: Respuestas condicionadas en Watson Assistant</i>	26
<i>Figura 11: Node-red</i>	27
<i>Figura 12: Crear un bot en Telegram</i>	28
<i>Figura 13: Icono del asistente</i>	30
<i>Figura 14: Nueva estructura del asistente</i>	33
<i>Figura 15: Nodo del penúltimo nivel del asistente</i>	33
<i>Figura 16: Nodos del último nivel del asistente</i>	34
<i>Figura 17: Nodo Misiones del asistente</i>	35
<i>Figura 18: Nodo Bienvenido del asistente</i>	36
<i>Figura 19: Flujo inicial de Node-red</i>	37
<i>Figura 20: Flujo final de Node-red</i>	39
<i>Figura 21: Imagen de Rick y tú</i>	40
<i>Figura 22: Población y muestra</i>	43
<i>Figura 23: Misión 1</i>	45
<i>Figura 24: Misión 2</i>	46
<i>Figura 25: Misión 3</i>	48
<i>Figura 26: Misión 6</i>	51
<i>Figura 27: Tchebychev</i>	53
<i>Figura 28: Diagrama de sectores</i>	56
<i>Figura 29: Diagrama de barras</i>	57
<i>Figura 30: Diagrama de barras múltiple</i>	58
<i>Figura 31: Diagrama de Pareto</i>	59
<i>Figura 32: Histograma</i>	60
<i>Figura 33: Polígono de frecuencias</i>	61
<i>Figura 34: Diagrama de tallo y hojas</i>	62
<i>Figura 35: Diagrama de caja</i>	62
<i>Figura 36: Diagrama de caja múltiple</i>	63
<i>Figura 37: Operaciones de sucesos</i>	65
<i>Figura 38: Sistema completo de sucesos</i>	66
<i>Figura 39: Función de probabilidad</i>	67
<i>Figura 40: Función de distribución discreta</i>	68
<i>Figura 41: Función de densidad</i>	69
<i>Figura 42: Función de distribución continua</i>	70
<i>Figura 43: Leyes de Morgan</i>	73
<i>Figura 44: Misión 12</i>	74
<i>Figura 45: Teorema de las probabilidades totales</i>	78
<i>Figura 46: Uniforme</i>	90

<i>Figura 47: Esperanza y varianza de una uniforme</i>	91
<i>Figura 48: Exponencial</i>	92
<i>Figura 49: Normal</i>	94
<i>Figura 50: Chi-cuadrado</i>	96
<i>Figura 51: T de Student</i>	96
<i>Figura 52: Fisher-Snedecor</i>	97
<i>Figura 53: Teorema central del límite</i>	98
<i>Figura 54: Distribución muestral de un estimador</i>	101
<i>Figura 55: Intervalo de confianza para la media si conocemos σ</i>	103
<i>Figura 56: Intervalo de confianza para la media si no conocemos σ</i>	103
<i>Figura 57: Intervalo de confianza para la desviación típica si conocemos la μ</i>	104
<i>Figura 58: Intervalo de confianza para la desviación típica si conocemos la μ</i>	104
<i>Figura 59: Intervalo de confianza para una proporción</i>	105
<i>Figura 60: Misión 28</i>	106
<i>Figura 61: Intervalo de confianza para la diferencia de medias</i>	107
<i>Figura 62: Intervalo de confianza para la razón de varianzas</i>	108
<i>Figura 63: Intervalo de confianza para la diferencia de proporciones</i>	108
<i>Figura 64: Error de tipo I y error de tipo II</i>	110
<i>Figura 65: Estadístico para la media si se conoce σ</i>	112
<i>Figura 66: Estadístico para la media si no se conoce σ</i>	112
<i>Figura 67: Estadístico para la varianza si se conoce μ</i>	112
<i>Figura 68: Estadístico para la varianza si no se conoce μ</i>	112
<i>Figura 69: Estadístico para la diferencia de medias</i>	113
<i>Figura 70: Estadístico para la razón de varianzas</i>	114
<i>Figura 71: Estadístico para una proporción</i>	114
<i>Figura 72: Estadístico para la diferencia de proporciones</i>	114
<i>Figura 73: Nube de puntos</i>	118
<i>Figura 74: Recta de regresión</i>	119
<i>Figura 75: Mínimos cuadrados</i>	120
<i>Figura 76: Pregunta 1 de la encuesta</i>	124
<i>Figura 77: Pregunta 2 de la encuesta</i>	124
<i>Figura 78: Pregunta 3 de la encuesta</i>	125
<i>Figura 79: Pregunta 4 de la encuesta</i>	125
<i>Figura 80: Pregunta 5 de la encuesta</i>	126
<i>Figura 81: Pregunta 6 de la encuesta</i>	126
<i>Figura 82: Pregunta 7 de la encuesta</i>	127
<i>Figura 83: Pregunta 8 de la encuesta</i>	128
<i>Figura 84: Acceso a Rick y tú en Telegram</i>	135
<i>Figura 85: Iniciar Rick y tú en telegram</i>	135
<i>Figura 86: Mensaje de bienvenida de Rick y tú</i>	136
<i>Figura 87: El asistente guarda el nombre del usuario</i>	137
<i>Figura 88: Normas de Rick y tú en Telegram</i>	138
<i>Figura 89: Mensaje “en otras cosas” de Rick y tú</i>	139
<i>Figura 90: Link previo del asistente</i>	139
<i>Figura 91: Norma cuatro en el link previo del asistente</i>	140
<i>Figura 92: Clicar siguiente en el link previo del asistente</i>	141
<i>Figura 93: Escribir siguiente en el link previo del asistente</i>	141
<i>Figura 94: Seleccionar una opción en el link previo del asistente</i>	142
<i>Figura 95: Mensaje de “ayuda” del asistente</i>	143
<i>Figura 96: Explicación de la encuesta</i>	144
<i>Figura 97: Elegir rama de carrera en la encuesta</i>	144

<i>Figura 98: Elegir frecuencia con la que se usan asistentes en la encuesta</i>	145
<i>Figura 99: Pregunta sobre las normas de la encuesta</i>	145
<i>Figura 100: Pregunta sobre el contenido del asistente de la encuesta</i>	146
<i>Figura 101: Pregunta sobre la historia y misiones de la encuesta</i>	147
<i>Figura 102: Pregunta de usabilidad de la encuesta</i>	148
<i>Figura 103: Pregunta sobre las opciones de la encuesta</i>	149
<i>Figura 104: Pregunta sobre la opinión general de la encuesta</i>	150
<i>Figura 105: Pregunta final de la encuesta</i>	150

ÍNDICE DE ECUACIONES

<i>Ecuación 1: Media</i>	46
<i>Ecuación 2: Mediana</i>	47
<i>Ecuación 3: Varianza</i>	50
<i>Ecuación 4: Coeficiente de asimetría</i>	54
<i>Ecuación 5: Coeficiente de curtosis</i>	55
<i>Ecuación 6: La función es mayor o igual que 0</i>	69
<i>Ecuación 7: El área bajo la curva es 0</i>	69
<i>Ecuación 8: El área entre dos puntos es su probabilidad</i>	69
<i>Ecuación 9: Probabilidad de que X sea menor o igual a a</i>	70
<i>Ecuación 10: Probabilidad de que X sea mayor o igual a a</i>	70
<i>Ecuación 11: Esperanza de una variable aleatoria discreta</i>	71
<i>Ecuación 12: Esperanza de una variable aleatoria continua</i>	71
<i>Ecuación 13: Varianza de una variable aleatoria discreta</i>	72
<i>Ecuación 14: Varianza de una variable aleatoria continua</i>	72
<i>Ecuación 15: Regla de Laplace</i>	74
<i>Ecuación 16: Probabilidad condicionada</i>	75
<i>Ecuación 17: Teorema de las probabilidades totales</i>	78
<i>Ecuación 18: Regla de Bayes</i>	80
<i>Ecuación 19: Teorema de Bayes</i>	80
<i>Ecuación 20: Bernoulli</i>	82
<i>Ecuación 21: Binomial</i>	83
<i>Ecuación 22: Geométrica</i>	84
<i>Ecuación 23: Binomial negativa</i>	86
<i>Ecuación 24: Hipergeométrica</i>	87
<i>Ecuación 25: Poisson</i>	88
<i>Ecuación 26: Cálculo de probabilidades de una Normal</i>	94
<i>Ecuación 27: Varianza y cuasivarianza muestral</i>	100
<i>Ecuación 28: Proporción muestral</i>	100
<i>Ecuación 29: Sesgo de un estimador</i>	101
<i>Ecuación 30: Consistencia</i>	102
<i>Ecuación 31: Covarianza</i>	116
<i>Ecuación 32: Correlación lineal</i>	117

RESUMEN

La estadística cada vez está teniendo más presencia en todos los ámbitos de la vida actual, unido a que estamos en medio de una revolución digital, se tiene que ir enseñando cada vez más estadística desde edades más tempranas y, utilizando, estas herramientas digitales que ahora mismo tenemos al alcance de la mano.

En este trabajo se habla del desarrollo de *Rick y tú*, un asistente conversacional realizado a partir de tecnología de IBM, que quiere cumplir estos dos objetivos, ser una herramienta digital atractiva con la que se pueda empezar a aprender la base de la Estadística. Para hacer más atractiva la idea de aprender estadística, se ha elaborado una historia en la que Rick (científico de la exitosa serie *Rick y Morty*) y el usuario del asistente van viviendo una aventura en la que aprenden estadística. Los usuarios a los que está dedicado esta herramienta son personas que tengan una base matemática y se quieran iniciar en el mundo de la estadística.

Palabras clave: Estadística, Inteligencia Artificial, Enseñanza, Asistente conversacional (chatbot).

ABSTRACT

Statistics is increasingly having a presence in all areas of life today, coupled with the fact that we are in the midst of a digital revolution. Actually, statistics have to be taught from younger ages and, using these digital tools that now we have and we can use to do it.

This final project, talks about the development of *Rick y tú*, a conversational assistant made from IBM technology, who wants to meet these two objectives, be an attractive digital tool with which to start learning the basis of Statistics. To make the idea of learning statistics more attractive, a story has been created in which Rick (scientist from the successful series *Rick and Morty*) and the user of the assistant are involved in an adventure in which they learn statistics. The users to whom this tool is dedicated are people who have a mathematical base and want to start in the world of statistics.

Key words: Statistics, Artificial Intelligence, Teaching, Conversational assistant (chatbot).

1. INTRODUCCIÓN

En los tiempos que estamos viviendo la estadística se ha vuelto esencial en todas las áreas, por lo que es necesario buscar nuevas formas para poder transmitirla y que todas las personas puedan llegar a entenderla. La razón por la que este TFG nació, es que llevo escuchando toda la carrera decir a todos mis amigos y otras personas que conocía de otras carreras que dan estadística, que cómo podía hacer la carrera de estadística, si ellos que solo habían tenido una asignatura, les había parecido difícil y aburrida. Cada persona tiene diferentes gustos, pero cuando tanta gente dice que algo le parece difícil y aburrido, es que algo puede estar fallando. Por eso me planteé buscar una nueva forma de enseñar la base de la estadística con el objetivo de que la gente la pudiera aprender de una manera fácil y divertida.

Como estoy en IBM trabajando con Inteligencia Artificial, decidí utilizar el conocimiento que había aprendido sobre asistentes conversacionales para crear uno con el que la gente pudiera aprender estadística. Esta decisión se tomó debido a que las nuevas generaciones han nacido rodeadas de tecnología y todo lo que esté ligada a la misma, les parece mucho más atractivo. Se va a utilizar esta tecnología para convertirla en una especie de videojuego que tendrá una historia y misiones, con la que se quiere conseguir una herramienta digital para que personas interesadas en aprender estadística puedan empezar a hacerlo de una manera distinta que pretende ser divertida y fácil.

Así que el objetivo principal de este trabajo es conseguir un nuevo recurso para que cualquier persona con una base matemática pueda iniciarse en el mundo de la estadística, que sea fácil de usar y aporte un modo ameno y divertido de afrontar el aprendizaje de la estadística. De esta forma, se podrá fomentar una visión más positiva de la estadística entre los estudiantes.

Antes de empezar, vamos a hablar un poco de cómo se estructura este trabajo y de qué se habla en cada apartado:

Primero, en el capítulo dos, se hablará de la metodología utilizada en el asistente para enseñar estadística, storytelling (Socialmood, 2020), (Acuña, 2018) y gamificación (Martín, 2020), (Malvido, 2019), que son las dos técnicas que se han usado para poder crear la historia y las misiones (que son ejercicios para entender la teoría de estadística, se han llamado misiones para hacerlo más atractivo que si se llaman ejercicios o problemas) que tienen el propósito de hacerlo más entretenido.

Luego, en el capítulo tres, se explicarán las herramientas utilizadas para construir el asistente conversacional que se pueden encontrar en IBM Cloud (IBM, IBM Cloud, 2020). Ahora mismo, con toda la cantidad de tecnología que tenemos al alcance de la mano, es muy importante utilizarla para buscar nuevas formas de enseñanza.

Después, en el capítulo cuatro, se puede encontrar el contenido del asistente, tanto la historia como la parte teórica de estadística y las misiones. Mirando en las guías docentes de asignaturas de estadística en otras carreras (UVA, 2020) y después de haber dado toda la

carrera de estadística, me parece que, para empezar en el mundo de la estadística, la base que se tiene que enseñar es estadística descriptiva (González, 2014), probabilidad (Menéndez, 2015), (Rodríguez, Probabilidad, 2015), inferencia (Rodríguez, Inferencia, 2016) y regresión (Barba, 2016). A través de “El Estadístico Accidental” (Quintela, El estadístico accidental, 2014) (libro de Alejandro Quintela en el que se explica la estadística de una forma fácil y divertida), apuntes de la carrera e información de internet se ha ido elaborando la parte teórica. Para la historia me he inspirado en la serie Rick y Morty (Wikipedia, 2020), una serie en la que un científico que puede viajar a través del espacio vuelve a casa después de mucho tiempo y empieza a vivir aventuras con su nieto Morty. Las misiones son curiosidades de la vida del usuario del chatbot o de Rick, para hacerlas más entretenidas y prácticas.

Por último, en el capítulo cinco, se verán unas estadísticas que se han elaborado a partir de una encuesta que ha sido compartida con profesores de estadística y alumnos de la carrera de estadística, ya que ellos también tienen una visión completa de la estadística para saber si les parece adecuado el contenido y, luego, con personas de carreras de la rama de salud, económicas o ingenierías que han cursado la asignatura de estadística para saber si les parece útil para futuras personas que quieran empezar en la estadística con el objetivo, de que, en el capítulo seis, se puedan sacar unas conclusiones del trabajo realizado y una reflexión de si se han conseguido cumplir los objetivos propuestos y una línea de trabajo futuro en la que se expone todo lo que se podría hacer en un futuro para mejorar el recurso y su contenido.

2. METODOLOGÍA

A la hora de querer llevar a cabo esta idea, surgen dudas sobre cómo se podría enseñar estadística de una forma divertida. Recordando mi vida estudiantil, he decidido utilizar las técnicas que se pueden ver a continuación.

2.1 Storytelling

Para elaborar este apartado se ha consultado información en (Socialmood, 2020) y (Acuña, 2018).

Se ha decidido utilizar la técnica storytelling ya que, al recordar toda mi vida estudiantil me vino a la cabeza una clase de segundo de infantil en la que la profesora, para explicarnos las vocales, nos contó una historia, me gusto tanto que a día de hoy todavía me acuerdo de ella.

“La gente olvidará lo que dijiste, la gente olvidará lo que hiciste, pero la gente nunca olvidará cómo la hiciste sentir” – Maya Angelou (Socialmood, 2020).

Esto da la razón a la frase de Maya Angelou ya que no me acordaré exactamente de la historia pero sí de cómo me hizo sentir la misma.

Storytelling es el arte de contar historias, parte de la base de que si un mensaje genera en el receptor emociones es mucho más efectivo, no es solo importante el que se transmite sino cómo se hace. Se considera un arte ya que no es una manera de contar una simple historia si no que se busca llegar al oyente y tocarle la fibra sensible.

En los últimos años, está viéndose en la educación, no sólo en educación infantil donde ya era común su uso, ahora también está utilizándose en educaciones superiores debido a su efectividad.

Lo primero que hay que tener claro antes de elaborar la historia, es el objetivo que queremos conseguir con la misma. Una vez que se tiene claro el objetivo se puede empezar a elaborar la historia, para conseguir que sea una historia que atrape el oyente siempre hay que preparar un buen inicio que cause curiosidad, que enganche al oyente, mantener la motivación conseguida con ese inicio durante todo el desarrollo y terminar con un final consistente en el que quede claro lo que hemos querido transmitir durante el transcurso de toda la historia.

Como se quiere despertar emociones, hay que saber cuáles se quieren transmitir, en qué momento a través de qué y siempre solucionarlo de una manera positiva para que se pueda sacar el lado bueno de todo tipo de situaciones y aprender a resolver problemas.

Hay que conseguir que el oyente empatice con el/los protagonistas de la historia para que vaya viviéndola como si fuera él mismo el que la está viviendo.

2.2 Gamificación

Para elaborar este apartado se ha consultado (Martín, 2020) y (Malvido, 2019).

La gamificación es una metodología que consiste en el uso de elementos del juego en contextos no lúdicos, como en este caso es, la enseñanza para conseguir una motivación en el usuario y alcanzar mejores resultados. Elementos que son utilizados en el mundo de los juegos, como son misiones, puntos o niveles, se adaptan al mundo de la educación para hacerla más atractiva y cercana a los alumnos. Hay que pensar como un diseñador de videojuegos, pensar como la información que se quiere transmitir se puede convertir en un juego que motive y enganche al usuario. Hay que tener siempre presente el objetivo que queremos conseguir con este juego, que en este caso al estar enfocado en el mundo de la educación, es, que el oyente aprenda.

Esta técnica surge de comprobar que a la sociedad, en general, la encantan los juegos, pequeños y mayores, de una manera u otra juegan y se lo pasan bien, viendo lo efectivo que es, se adapta en el mundo de la educación para conseguir que se vea como una actividad lúdica en la que se va mejorando y consiguiendo recompensas.

Para conseguir el objetivo de que el usuario aprenda se utilizan la técnica mecánica, que es una forma de recompensar al usuario cada vez que consigue un objetivo, se puede hacer a través de misiones, obtención de puntos, recompensas o subir de nivel, entre otras. También se utiliza la técnica dinámica, que hace referencia a la motivación del usuario para jugar, motivarse y seguir cumpliendo los objetivos marcados, esto se consigue a través de competir con otros usuarios de la competición e ir alcanzando un determinado *status*.

Desde la llegada de las nuevas tecnologías y el éxito que tuvieron en la población, se vieron nuevas formas de diversión como son los videojuegos, que gustan a toda la población, pequeños y mayores, por lo que hay que aprovechar estas nuevas tecnologías para utilizarlas en el aula y conseguir que los alumnos disfruten más de la experiencia de aprender y lo vean como algo lúdico con lo que se divierten mientras aprenden.

Esta metodología pretende entender qué es lo que hace que los videojuegos consigan provocar tanta concentración, motivación y ganas de superarse a sí mismo a los usuarios de los mismos mientras juegan para aplicarlo a la enseñanza y conseguir que se vea a la misma como algo divertido y motivante.

Las razones por las que se está utilizando cada vez más es que consigue motivar al alumno, una mayor participación en la clase, simplifica actividades difíciles, crea un sentimiento de progreso al ir subiendo de niveles y recompensa por un buen trabajo al conseguir puntos, consigue promover la constancia y el triunfo, muchas veces ayuda a fomentar un ambiente de compañerismo, un sentimiento de comunidad y crear una identidad personal dentro de la misma y transforma actividades que se han visto comúnmente como aburridas en divertidas e interesantes. Pero nunca hay que olvidar el objetivo principal, que es educar al alumno y conseguir que aprenda la materia necesaria para seguir avanzando.

3. HERRAMIENTAS

En este apartado se va a hablar de la tecnología usada para construir el asistente conversacional. Esta tecnología está en la nube de IBM (IBM, IBM Cloud, 2020) en la cual cualquier persona puede entrar (sin meter número de tarjeta) y utilizar muchos servicios de manera gratuita. Todo lo que se ha desarrollado ha sido con servicios gratuitos, con las funcionalidades gratuitas, así que cualquier persona puede probar esta tecnología si se hace una cuenta. Para elaborar este apartado se ha necesitado realizar cursos sobre los servicios utilizados e IBM Cloud en (IBM, Cognitive class, 2020) y consultar (Telegram, 2020).

3.1 IBM Cloud

IBM Cloud es la plataforma de IBM que contiene más de 190 servicios, tanto de la parte de infraestructura, como de plataforma y de software (Figura 2). Vamos a explicar un poco estos conceptos a través de una metáfora con la pizza (Figura 1).

Pizza as a Service

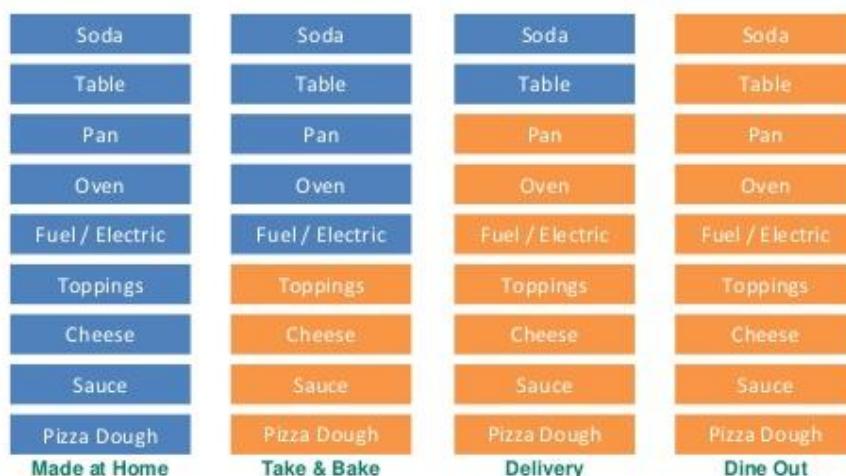


Figura 1: Pizza como servicio

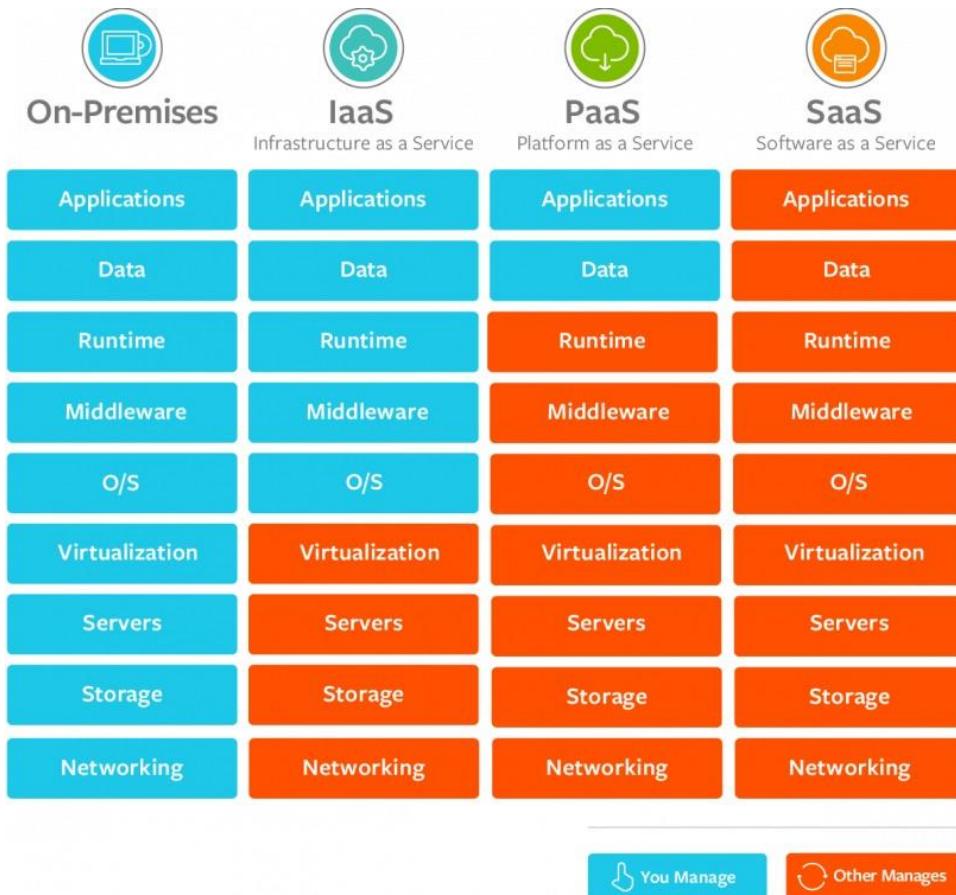


Figura 2: On-Premises, IaaS, PaaS y SaaS

Si te haces una pizza en casa casera, compras los ingredientes, haces la masa, elaboras la pizza, luego la cocinas, te la comes y limpias todos los instrumentos. Esto en el mundo de la informática es como si tuvieras todo on-premises, es decir, tú tienes todo lo necesario, tienes tus propios ordenadores, centros de datos y tú te encargas de mantenerlos.

Si te compras una pizza hecha, ya te la dan hecha, solo la cocinas, te la comes y limpias los cacharros. Aquí ya entramos en el concepto de la nube, la empresa a la que contratas los servicios es la que tiene toda la parte física, tu no te preocupas de lo físico. Se conoce como IAAS (Infrastructure as a service), manejas toda la parte de infraestructura como son las bases de datos, la seguridad, el sistema operativo, las máquinas virtuales, pero no te preocupas de todos los componentes físicos que hay por debajo.

Si pides que te traigan una pizza a casa, ya solo te preocupas de comerla y limpiar los platos. Esto se conoce como PAAS (Platform as a service), la empresa es la que se encarga de la parte física pero también de la parte de infraestructura, de gestionarte si necesitas bases de datos, de la seguridad, tú solo utilizas el servicio en el que programas, pero no te encargas de gestionar lo que necesita ese servicio para funcionar.

Por último, si directamente vas a un restaurante a comer la pizza, ya sólo te preocupas de comerla, de consumir el servicio. Esto es lo que se conoce como SAAS (Software as a

service), ya no te preocupas por nada, son servicios que utilizas directamente para ti, por ejemplo, gmail.

En este trabajo se ha utilizado Watson Assistant y node-red (ahora se va a explicar como funcionan), los dos son de la parte de PAAS.

3.2 Watson Assistant

Watson es la plataforma de inteligencia artificial de IBM. Tiene muchos servicios como un buscador inteligente, un convertidor de voz a texto y de texto a voz, machine learning...

Watson Assistant: es un agente de inteligencia artificial que sirve para construir interfaces de conversación, nos permite crear asistentes conversacionales como Rick y tú, es un sistema inteligente que aprende de su experiencia y puede ser utilizado en cualquier campo (Atención al cliente, recursos humanos, medicina...) y se basa en:

- Intenciones (Figura 3): Como ya dice la propia palabra, su objetivo es capturar la intención del usuario, para entenderlo se suelen relacionar con verbos, si hablamos de una pizzería, una intención sería que el usuario quiere comprar un pizza, otra que quiere reservar una mesa. Se representan con #.

The screenshot shows the Watson Assistant interface with the workspace titled 'WatsonPizzeria' and version 'Development'. On the left, there's a sidebar with navigation links: Entities, Dialog, Options, Analytics, Versions, and Content Catalog. The main area is titled 'Intents' and contains a table with the following data:

	Intents (7) ↑	Description	Modified ↑	Conflicts ↑	Examples ↑
<input type="checkbox"/>	#ayudar		5 months ago	6	
<input type="checkbox"/>	#comprar_pizza		5 months ago	5	
<input type="checkbox"/>	#costar		a few seconds ...	7	
<input type="checkbox"/>	#dar_propina		a month ago	2	
<input type="checkbox"/>	#empezar		a month ago	8	
<input type="checkbox"/>	#recomendar		5 months ago	5	

At the bottom of the table, it says 'Showing 1–7 of 7 intents'. To the right of the table, there are buttons for '1', '1 of 1 pages', and a blue circular button with a white icon. The top right of the interface has buttons for 'Save new version', 'Try it', and more options.

Figura 3: Intenciones de Watson Assistant

En cada intención ponemos ejemplos de distintas formas en las que un usuario puede expresar esa intención como se puede ver en la figura 4. Se necesitan como mínimo cinco ejemplos de intenciones para que Watson pueda entrenar y asociar ejemplos similares a las frases que hemos puesto con dicha intención. En el apartado 4.2 se puede ver como se han utilizado para el asistente conversacional *Rick y tú*.

The screenshot shows the Watson Assistant interface. At the top, there's a search bar with the text '#comprar_pizza'. Below it, a section titled 'User example' with the sub-instruction 'Add unique examples of what the user might say. (Pro tip: Add at least 5 unique examples to help Watson understand)'. A text input field contains the placeholder 'Type a user example here, e.g. I want to pay my credit card bill.' To the right, there are icons for download, copy, and search, along with a 'Try it' button. Below this, a table lists five user examples under the heading 'User examples (5) ↑'. The columns are 'Added ↑' and 'Conflicts (0) ↑'. The examples are:

	Added ↑	Conflicts (0) ↑
<input type="checkbox"/> ¿Podrías servirme una pizza, por favor?	5 months ago	
<input type="checkbox"/> ¿Puedo comer una pizza?	5 months ago	
<input type="checkbox"/> Me gustaría comer una pizza	5 months ago	
<input type="checkbox"/> Quiero comprar una pizza	5 months ago	

At the bottom left, it says 'Showing 1–5 of 5 examples'. On the right, there are navigation icons for page 1 of 1.

Figura 4: Ejemplos de una intención en Watson Assistant

- Entidades (Figura 5): Son objetos relevantes para la conversación, por ejemplo, los tipos de pizza, los ingredientes o las ciudades donde se encuentra un establecimiento. Se representan con @.

The screenshot shows the Watson Assistant interface for the 'WatsonPizzeria' intent. On the left, a sidebar has 'Entities' selected, with 'My entities' highlighted. The main area shows a table of entities with columns 'Entity (6) ↑', 'Values', and 'Modified ↑'. The entities listed are:

Entity (6) ↑	Values	Modified ↑
<input type="checkbox"/> @confirmar_extra	no, sí	5 months ago
<input type="checkbox"/> @pizza_ingredientes	verduras, carne, piña, queso	5 months ago
<input type="checkbox"/> @pizza_lugar	restaurante, casa	5 months ago
<input type="checkbox"/> @pizza_sitios	New York, Londres	4 months ago
<input type="checkbox"/> @pizza_tamaño	mediana, pequeña, grande	5 months ago
<input type="checkbox"/> @pizza_tipo	vegetariana, hawaiana, cuatro quesos, boloñesa, queso	5 months ago

At the bottom left, it says 'Showing 1–6 of 6 entities'. On the right, there are navigation icons for page 1 of 1.

Figura 5: Entidades de Watson Assistant

Existen entidades del sistema, que están creadas por IBM directamente ya que son ampliamente utilizadas, como fecha, que detecta todas las fechas en múltiples formatos o la de números. Podemos activarlas o no, según queramos utilizarlas. En el asistente creado en este trabajo no se ha utilizado ninguna de ellas.

- Diálogo: Para gestionar la conversación entre Watson y el usuario, se van creando distintos nodos de diálogos, como podemos ver en la figura 6, para contestar a esas intenciones o entidades que hemos creado.

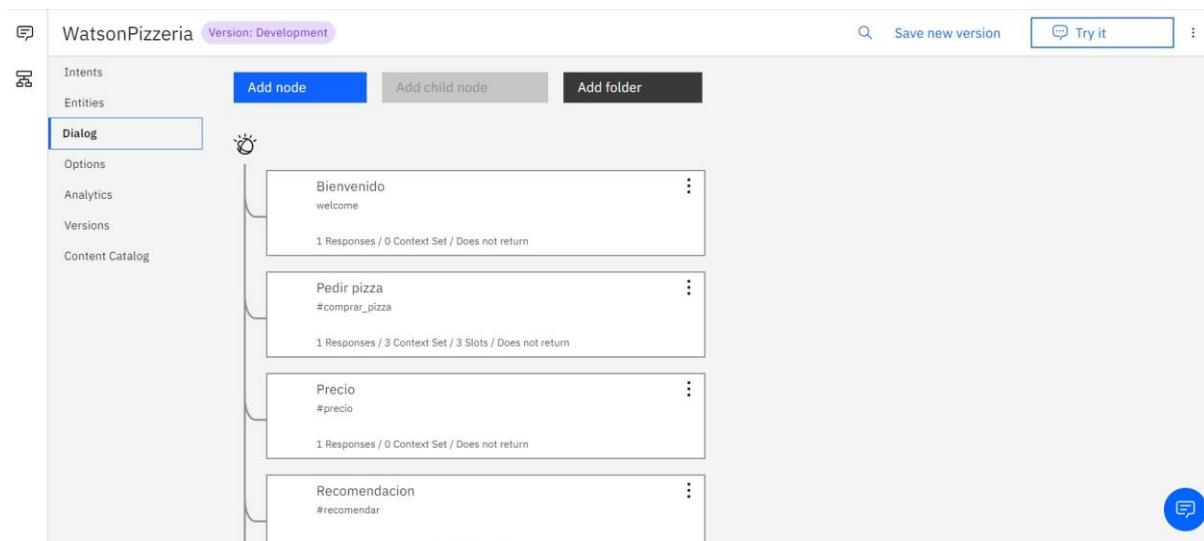


Figura 6: Nodos de diálogo en Watson Assistant

Estos nodos pueden tener (en el plan gratuito) cuatro tipos de respuesta: texto, imagen, opción y pausa. En la figura 7 se ven dos tipos más porque es una estancia de un plan superior al gratuito, pero en el asistente de este trabajo solo se puede elegir entre los cuatro dichos y solo se han utilizado texto, imagen y pausa.

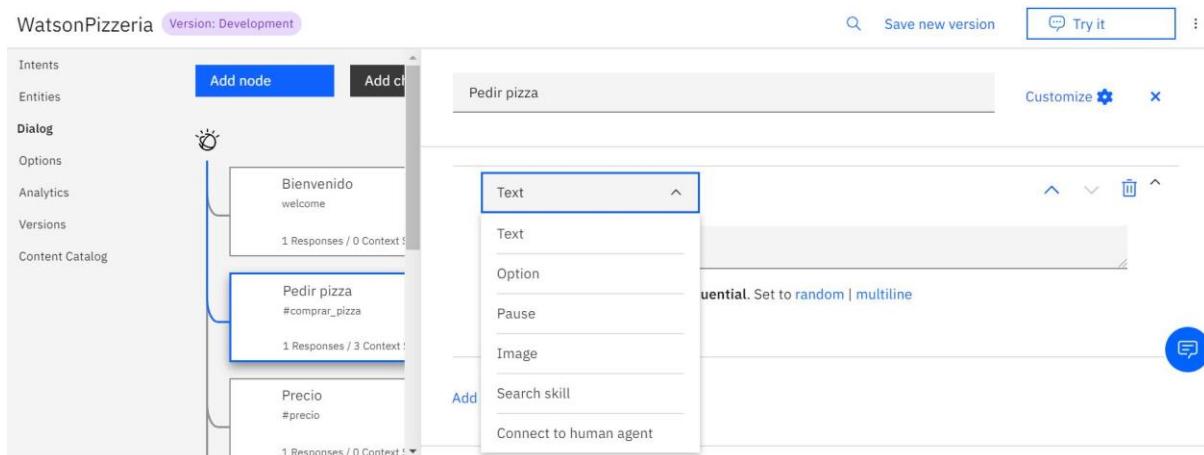


Figura 7: Tipos de respuesta en Watson Assistant

A parte de esto, Watson Assistant tiene muchas funcionalidades para mejorar su funcionamiento, vamos a explicar algunas de ellas:

- **Slots (Figura 8):** Si el asistente para responder a una determinada consulta con el usuario necesita información y no se la han dado, pregunta lo que necesita hasta que consigue toda esa información.

Watson

Usuario

Quiero una pizza
¿De qué tamaño?

Pequeña

¿De qué tipo?

Vegetariana

Gracias por pedir una pizza vegetariana pequeña

Si el usuario hubiera dicho directamente *quiero una pizza pequeña vegetariana*, Watson directamente ya contestaría *gracias por pedir una pizza vegetariana pequeña*, porque ya tiene la información que necesita para poder contestar.

The screenshot shows the Watson Assistant interface with a search bar containing "Pedir pizza". Below it, there are three rows of slot definitions:

	Check for	Save it as	If not present, ask	Type	Customize	X
1	@pizza_tama	\$pizza_tam	¿De qué tama	Required		
2	@pizza_tipo	\$pizza_tipo	¿De qué tipo	Required		
3	@pizza_lugar	\$pizza_lugar	¿Dondequier	Required		

A blue circular button with a white speech bubble icon is located at the bottom right of the slots area.

Figura 8: Slots en Watson Assistant

Para activar los slots hay que dar en *Customize* o *Ajustes*, dependiendo el idioma en el que tengamos el navegador, y activar los slots. Nos saldrá lo que podemos ver en la figura 8. En *Check for* ponemos la entidad que necesitamos detectar para contestar a la intención del usuario, en *Save it as* cómo queremos llamar a la variable de contexto (se representan con \$ y sirven para guardar información durante toda la conversación sobre datos que va dando el usuario y necesitamos utilizar en otras intenciones) en la que se guardará la información que da el usuario, y en *If not present, ask* pondremos la pregunta que queremos que se la haga

al usuario si no detectamos la entidad que necesitamos para contestarle. Estos slots pueden ser *Required* (necesarios) que quiere decir que siempre le preguntaremos al usuario por esa información si no nos la da, u *Optional* (opcionales) si queremos guardar la información en caso de que nos la dé, pero si no lo hace, no es completamente necesaria, así que no le preguntaremos por ella.

- Digresiones (Figura 9): permiten irse de una conversación para contestar una duda concreta y poder volver donde se quedó la conversación. Esta necesidad surge por la manera en la que pensamos los humanos: mientras estamos en una conversación nos surgen otras cosas aunque luego volvamos a la conversación inicial.

Watson

Usuario

Quiero una pizza pequeña
¿De qué tipo?

Quiero una pizza pequeña

¿Cuánto cuesta?

La pizza cuesta 8€

Vegetariana

Gracias por pedir una pizza pequeña
vegetariana

También dando a *Customize* dentro de un nodo podremos activar las digresiones o desactivarlas. Se puede activar la digresión de que se pueda ir a otro nodo aunque todavía el usuario no haya dado toda la información necesaria en un slot y también la digresión que nos permite llegar al nodo en el que estamos desde otro distinto. En este caso, como estamos en el nodo de pedir pizza tenemos los dos activados, ya que si el usuario pregunta cuánto cuesta una pizza queremos que salga del slot y vaya al nodo que responde el precio de las pizzas, pero luego vuelve a este nodo para seguir con el proceso de la compra de la pizza.

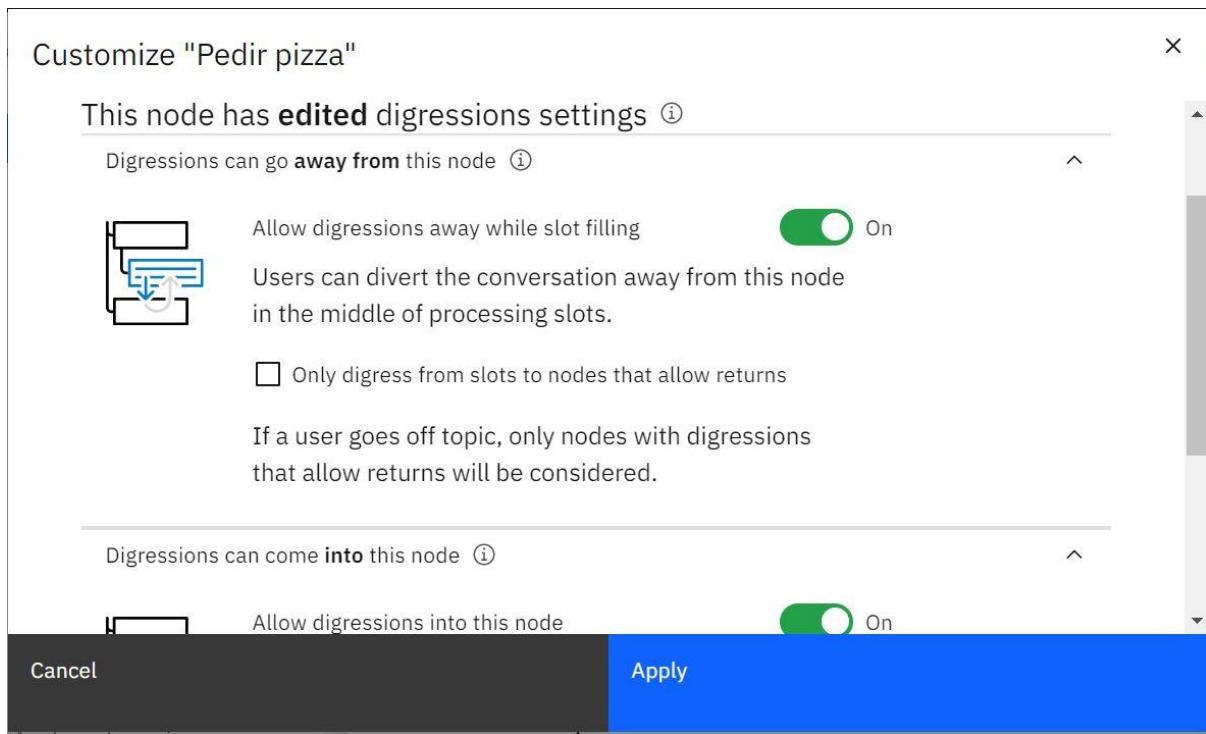


Figura 9: Digresiones en Watson Assistant

- Respuestas condicionadas (Figura 10): es capaz de dar distintas respuestas dependiendo de la condición que se detecta. En nuestro ejemplo, se puede recomendar una pizza determinada a nuestro consumidor según cuál sea su ingrediente favorito.

CONTESTA DISTINTAS RESPUESTAS SEGÚN LA CONDICIÓN

Carne → Te recomendamos una pizza boloñesa

Piña → Te recomendamos una pizza hawaiana

Queso → Te recomendamos una pizza cuatro quesos

Se activa la opción de tener *Multiple conditioned responses* (respuestas condicionadas múltiples) en *Customize*, en *If assistant recognizes* pondremos el valor de la entidad y en *Respond with* la respuesta que queremos dar cuando se detecte dicho valor.

The screenshot shows the Watson Assistant interface with a blue header bar containing the text "Recomendacion". On the right side of the header are "Customize" and a close button (X). Below the header, the text "Assistant responds" is displayed. A table lists three conditional responses:

If assistant recognizes	Respond with
1 @pizza_ingredientes:carne	Te recomendamos una boloñesa!
2 @pizza_ingredientes:piña	Te recomendamos una pizza hawaiana!
3 @pizza_ingredientes:queso	Te recomendamos una pizza clasica!

Figura 10: Respuestas condicionadas en Watson Assistant

3.3 Node-red

Node-RED App es un motor de flujos con enfoque IoT (Internet de las cosas), que permite definir gráficamente flujos de servicios, a través de protocolos estándares como REST, MQTT, Websocket, AMQP... además de ofrecer integración con apis de terceros, tales como Twitter, Facebook, Telegram...

Está programada en NodeJS y tiene una sencilla interfaz en HTML que es accesible desde cualquier navegador, en la que arrastrando y conectando nodos entre sí, es posible definir un flujo que ofrezca un servicio. Puede ejecutarse en dispositivos tan limitados como una Raspberry, hasta en plataformas complejas como IBM Cloud o Azure IoT.

Tiene muchos nodos distintos que se van conectando creando un flujo que lleva a cabo la acción que queremos realizar. Con él se pueden conectar fácilmente servicios. En el apartado 4.2, explicaremos los nodos utilizados para este trabajo y cómo se ha realizado el flujo para conectar Watson Assistant y node-red. En la figura 11 se puede ver cómo es Node-red.

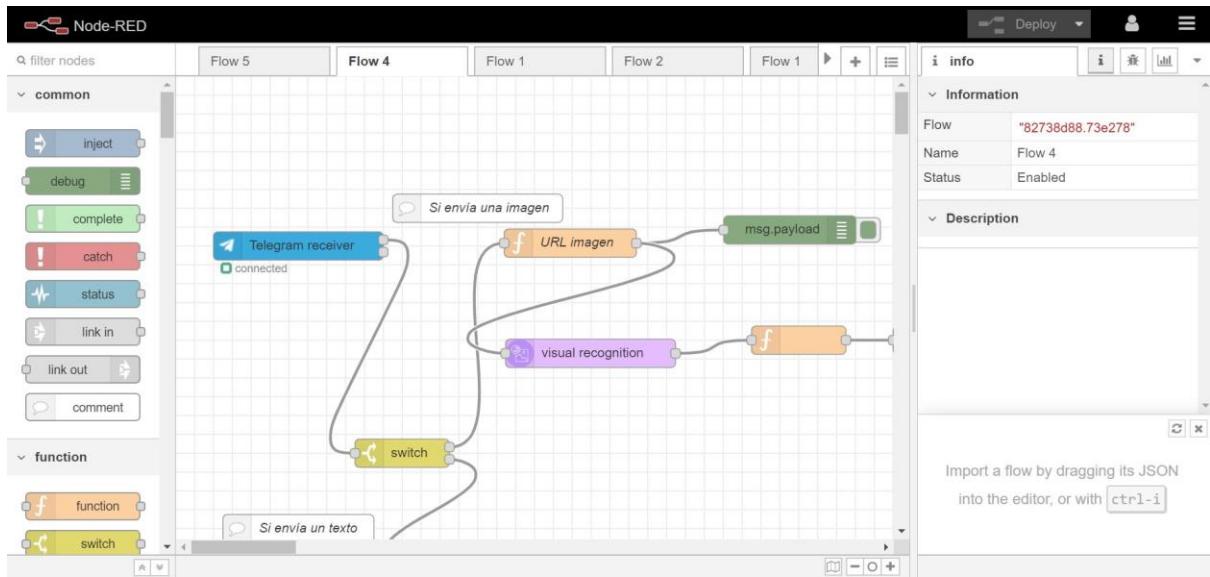


Figura 11: Node-red

3.4 Telegram

Telegram es una aplicación de mensajería instantánea al igual que Whatsapp. Con telegram tenemos la capacidad de crear bots de manera muy sencilla que realizan lo que nosotros queramos hacer con ellos.

Para crear un nuevo bot, simplemente tenemos que buscar (como buscaríamos a un contacto cualquiera de Telegram) a BotFather, poner /newbot, darle nombre y elegir un usuario. Es obligatorio que el usuario acabe en bot para que la gente sepa que es un bot. Todo el proceso de creación de Rick_y_tu_bot se puede ver en la figura 12 como ejemplo de creación de un bot con Telegram.

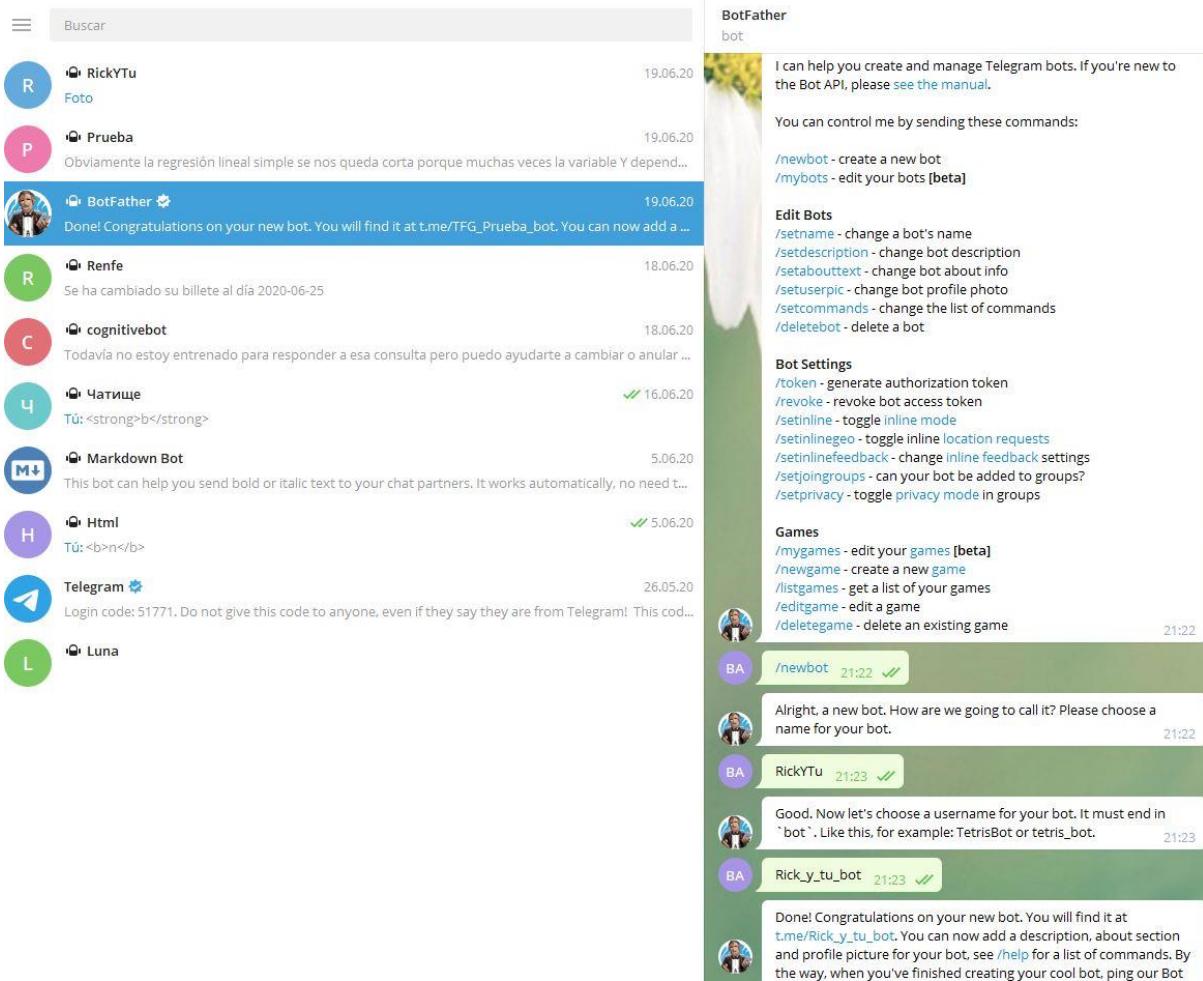


Figura 12: Crear un bot en Telegram

4. ASISTENTE CONVERSACIONAL

Rick y tú es el asistente conversacional que se ha creado para que personas que se quieran iniciar en el mundo de la estadística y tengan una base matemática, puedan hacerlo a través de él.

4.1 Icono, nombre e historia

A la hora de decidir el nombre, se quería conseguir que fuera atractivo para los usuarios y les llamará la atención para querer utilizarlo. Como se pensó hacer una historia en la que Rick (científico de la serie *Rick y Morty*) y el usuario del asistente acaban en un concurso y tienen que ir pasando misiones, se decidió llamar al asistente *Rick y tú*, ya que *Rick y Morty* es una serie muy conocida entre los jóvenes (ya que en un principio este asistente estaba enfocado a personas que fueran estudiantes de enseñanzas superiores, que suelen ser jóvenes entre 18 y 22 años) que ha tenido mucho éxito y a la mayoría de ellos les gusta, y el ver que el título de algo es Rick y ellos ya llama la atención por lo menos para mirar a ver de qué se trata.

Para el icono se quería que apareciera Rick, el usuario del asistente, algo relacionado sobre estadística y el nombre.

Para representar a Rick y el usuario del asistente, se utiliza una imagen de Rick abriendo los ojos a Morty ya que parecía perfecta para representar cómo al comenzar toda esta aventura están viendo algo que no se pueden creer. Por lo que se pone en negro a Morty y una interrogación encima, para representar a ese usuario del asistente el cual todavía no conocemos.

Para representar la estadística se elige un fondo bonito de la misma para representar el lugar al que llegan Rick y el usuario, en el cual están ensimismados viendo todo lo que hay a su alrededor sin poder creérselo. En la parte superior de la imagen se colocó el nombre del asistente con el tipo de letra *Shlop* y de color verde para estar a conjunto con el fondo.



Figura 13: Icono del asistente

En la figura 13 se puede ver el icono del asistente que se utiliza como imagen nada más iniciar el asistente y como icono del bot en telegram. Es lo primero que ve el usuario nada más iniciar el asistente, ya que es la imagen de cuando Rick y el usuario del asistente acaban de llegar al concurso y se ven envueltos en toda esta aventura y todavía Rick no conoce a este usuario, con el cual, a medida que avanza la aventura va cogiendo confianza y acaban siendo buenos amigos. ¿Qué joven no quiere ser ahora mismo amigo de Rick?

Para crear la historia se ha visto la serie desde HBO y se han consultado las siguientes fuentes (respuestas.me, 2020), (Páez, 2020), (elgranporque, 2017), (Marcos, 2017), (lascosasquenoshacenfelices, 2020) y (korosenai, 2018).

Para la historia, se tenían varias ideas, al final se acabó eligiendo la opción del *Concurso Intergaláctico* ya que:

1. Como en esta situación Rick y el usuario del asistente no se conocen, era muy fácil preguntarle el nombre al usuario sin que quede claro y sea al principio.
2. Era la situación en la que se cree que el usuario se puede llegar a meter más en el papel y se puede sentir más identificado.
3. Le da un aire competitivo al tener que demostrar que son los mejores del espacio, a la vez que se ve compañerismo ya que tienen que ir superando misiones junto con Rick.

4. Al ser un concurso, se aprovecha la excusa de dar las normas del mismo para meter las instrucciones de uso del asistente sin que parezca un rollo.
5. Es muy fácil meter el concepto de misiones.

Para las misiones se ha decidido poner situaciones que Rick y otros personajes de la serie han vivido durante la misma o podrían serles útiles pero, también, situaciones que puede vivir un usuario del asistente para hacerle sentir también partícipe de la historia y pueda ver que la estadística le puede ayudar en situaciones cotidianas y vea directamente una aplicación práctica de la misma.

Durante toda la exposición de la teoría de estadística, Rick y el usuario del asistente hacen comentarios graciosos que se les viene a la cabeza cuando van descubriendo el contenido de estadística, esta relación está inspirada en mi amiga Marina y yo cuando estudiábamos juntas la carrera.

Entre medias de la historia hay un código secreto que se puede encontrar en varios apartados durante toda la historia y, al final, se dice la respuesta del mismo, consiste en escribir las palabras con la letra que se encuentra a la derecha de la que queremos escribir en el teclado del ordenador español. Esto se hace para meter un reto más de forma lúdica y picar al usuario para que quiera seguir viendo todo el asistente para descubrir cuál será el código.

Al final se decide que el premio es que con las vivencias de cada uno de los personajes se va a construir un asistente para poder ayudar a las personas de cada uno de los planetas participantes y que todos ellos irán juntos al mundo de la estadística. Tiene el objetivo de dejar claro que muchas veces hay cosas que hacemos que no tienen por qué tener una recompensa sino, que el proceso en sí ya lo es. Aún así se da la misma recompensa a todos los participantes para que puedan seguir aprendiendo y pasen un buen rato y la satisfacción de que lo que han hecho va a poder ayudar a otras personas.

4.2 Proceso de creación y problemas surgidos

Lo primero que se creó fue el asistente conversacional con Watson Assistant. Antes de empezar, decidimos plantearnos cómo iba a ser la estructura del mismo ya que iba a contener mucha materia, para simplificar a la hora de manejarnos por él.

Antes de explicar la estructura del asistente se va a hacer un pequeño inciso necesario para entender la misma. En el apartado 3.2, cuando se ha explicado Watson Assistant, se ha dicho que el diálogo se estructura a partir de unos nodos de diálogo, estos nodos de diálogo pueden tener nodos hijos a los que sólo se puede acceder si ya se ha pasado por el nodo padre.

Entonces, inicialmente, se decidió definir un nodo padre para cada parte de la estadística que se iba a explicar y dentro de ella los nodos hijos necesarios (y los nodos hijos de esos nodos hijos necesarios). La estructura era la misma que la que se puede ver en el índice del contenido del chatbot, es decir, por ejemplo, estadística descriptiva tiene cinco nodos hijos (conceptos surclave, medidas de posición, medidas de dispersión, medidas de forma y representaciones gráficas), el primer nodo hijo tiene a su vez tres nodos hijos (Población y

muestra, variable estadística y tabla de frecuencias) y así. En definitiva se estructuran como el índice: los puntos principales (los que tienen un sólo número) son los nodos padre, los del segundo nivel (los que tienen dos números: 1.1, 2.3, 5.1...) son los nodos hijos de esos nodos padre y así.

Cada nodo tenía su intención y sólo se llegaba a él si esa intención se detectaba, por ejemplo, el nodo *Estadística descriptiva* respondía cuando se detectaba la intención *#Estadística_descriptiva*, que tenía como ejemplos maneras en las que la gente puede querer preguntar algo relacionado con la estadística descriptiva.

Primer problema con el que nos encontramos: en el trabajo siempre trabajamos con planes superiores al gratuito por lo que no sabía que había límites de uso (sí pone cuando te haces una instancia lite, que es la gratuita, que hay límite de 100 nodos de diálogo, pero probé a crear más de 150 y sin problema), así que cuando quise crear la intención 101 no me dejaba y ya me di cuenta que el límite estaba en 100 intenciones. Por lo que necesitaba un cambio de estructura si quería continuar. También comprobé que el límite de entidades era 25, para no volver a tener imprevistos en esta nueva estructura.

Al necesitar un cambio de estructura completo y tener que rehacer parte del trabajo, me di cuenta que no quería la estructura actual, ya que yo quería que una vez se haya pasado la historia y se diga el temario disponible, se pudiera preguntar cualquier cosa y se respondiera directamente esa cosa, y no lo que pasaba con la estructura antigua, que si alguien pregunta por una media, se le conteste con estadística descriptiva y de ahí tenga que saber que es una medida de posición y ya ir a la media.

Por lo que se sigue manteniendo el orden que se puede ver en el índice del contenido a la hora de guiar al usuario por el asistente (si sigue el orden), pero puede preguntar en cualquier momento cualquier concepto que esté en el asistente y se le contestará con él. Para ello, todos los nodos menos los del último nivel (sin contar misiones y sus soluciones), pasan a ser nodos padre, a estar en el primer nivel y siguen funcionando como hasta ahora, si se detecta su intención se muestra su contenido. Como los nodos del último nivel siguen siendo nodos hijos de algunos de esos nodos que ahora están en el primer nivel, se crea una entidad por cada nodo del primer nivel que sigue teniendo nodos hijos y se introducen como ejemplos de esa entidad cada uno de los nodos hijos del último nivel que dependen de este nodo. Para que esos nodos hijos puedan ser encontrados aunque no estén en el primer nivel, el nodo penúltimo solo se reconoce cuando la intención de ese penúltimo nodo (que tiene como ejemplos de intención frases relacionadas tanto con ese nodo como con sus nodos hijo) es detectada y, luego tiene un slot optional, que si reconoce la entidad de ese nodo, contesta con el ejemplo de la entidad que detecta, y si no, contesta con el mensaje de ese penúltimo nodo. Como es un poco difícil de entender sin verlo, en la figura 14 podemos ver cómo todos los nodos menos los del último nivel han pasado al primer nivel, en la figura 15, podemos ver cómo es el nodo del penúltimo nivel, y en la figura 16 los nodos hijo que responden cuando se detecta tanto la intención como la entidad y el último nodo hijo es el que responde si solo se detecta la intención y tiene el mensaje de ese penúltimo nodo.

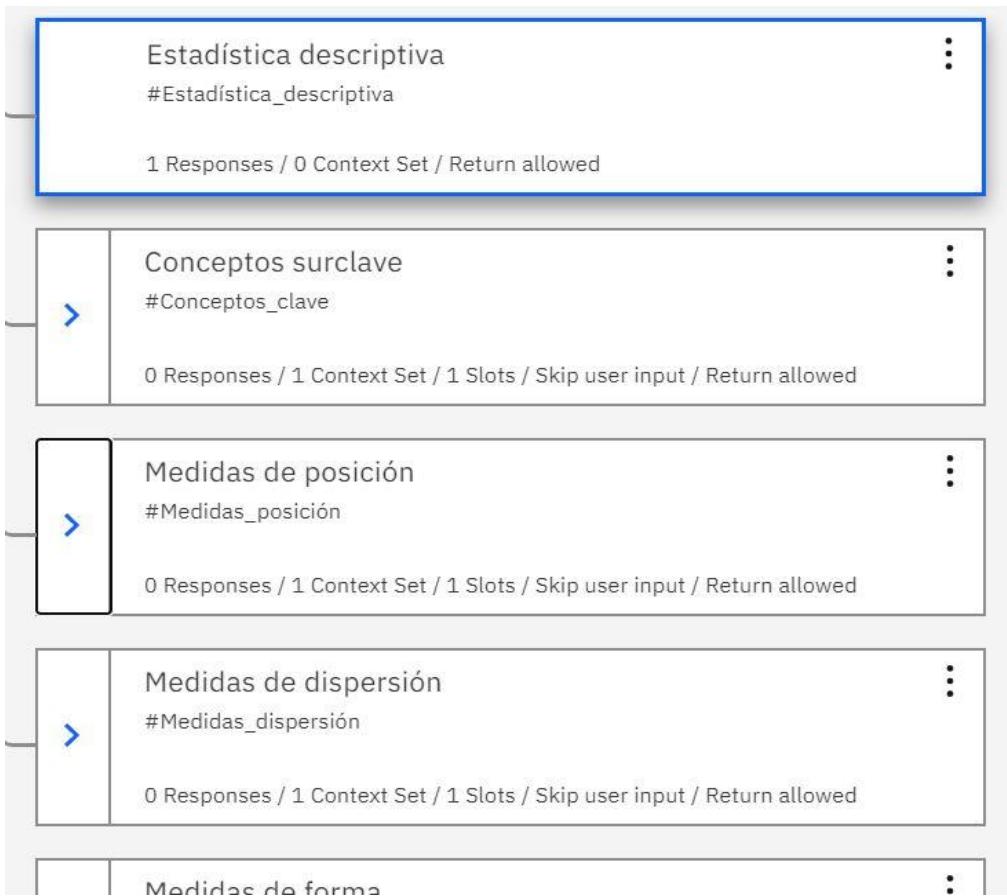


Figura 14: Nueva estructura del asistente

Medidas de forma Customize ⚙ X

Node name will be shown to customers for disambiguation so use something descriptive. Settings

If assistant recognizes

#Medidas_forma untitled +

Then check for 0 Manage handlers

Check for	Save it as	If not present, ask	Type
1 @MedidasFor	\$MedidasForr	Enter prompt	Optional ⚙️ untitled

Figura 15: Nodo del penúltimo nivel del asistente

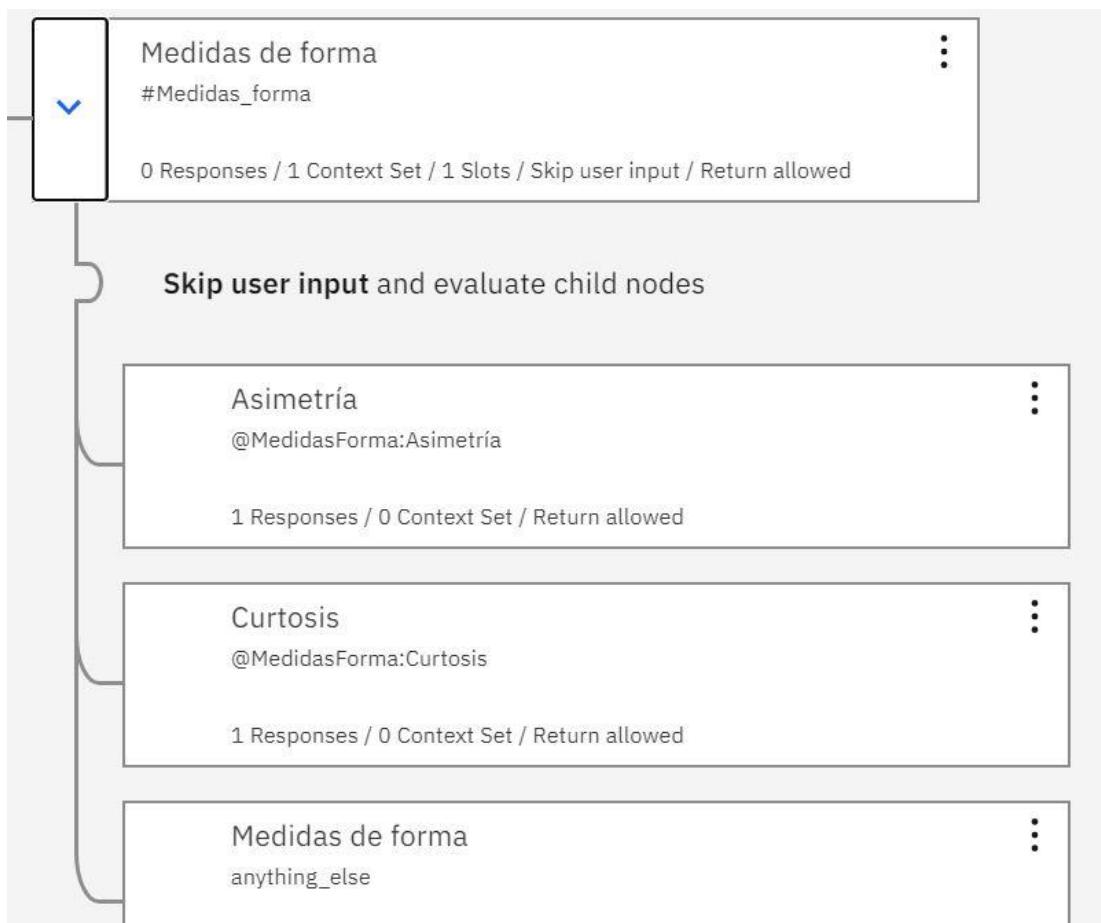


Figura 16: Nodos del último nivel del asistente

Para las misiones se crea un nodo misión (Figura 17) que funciona igual que los nodos de penúltimo nivel: hay una intención misión que se detecta cuando alguien dice algo sobre misiones y una entidad misión que tiene como ejemplo desde la misión 1 hasta la misión 29 (número de misiones que hay) y responderá con la misión que sea según el número de misión que quieras ver, cada una de esas misiones tiene dos nodos hijo, solución correcta y solución incorrecta, estos son los únicos nodos a los que sólo se puede acceder si estás en la misión correspondiente para que no haya problemas. Esto se debe a que si se hiciera lo mismo con las soluciones que con las misiones, si alguien mete un número de manera aleatoria y es una solución contestaría con la solución cuando no viene a cuento, por lo que son a los únicos nodos que no se puede responder desde cualquier lugar. También porque solo queremos que puedan ver la solución de una misión si han visto el enunciado de la misma, si no, no tiene sentido.

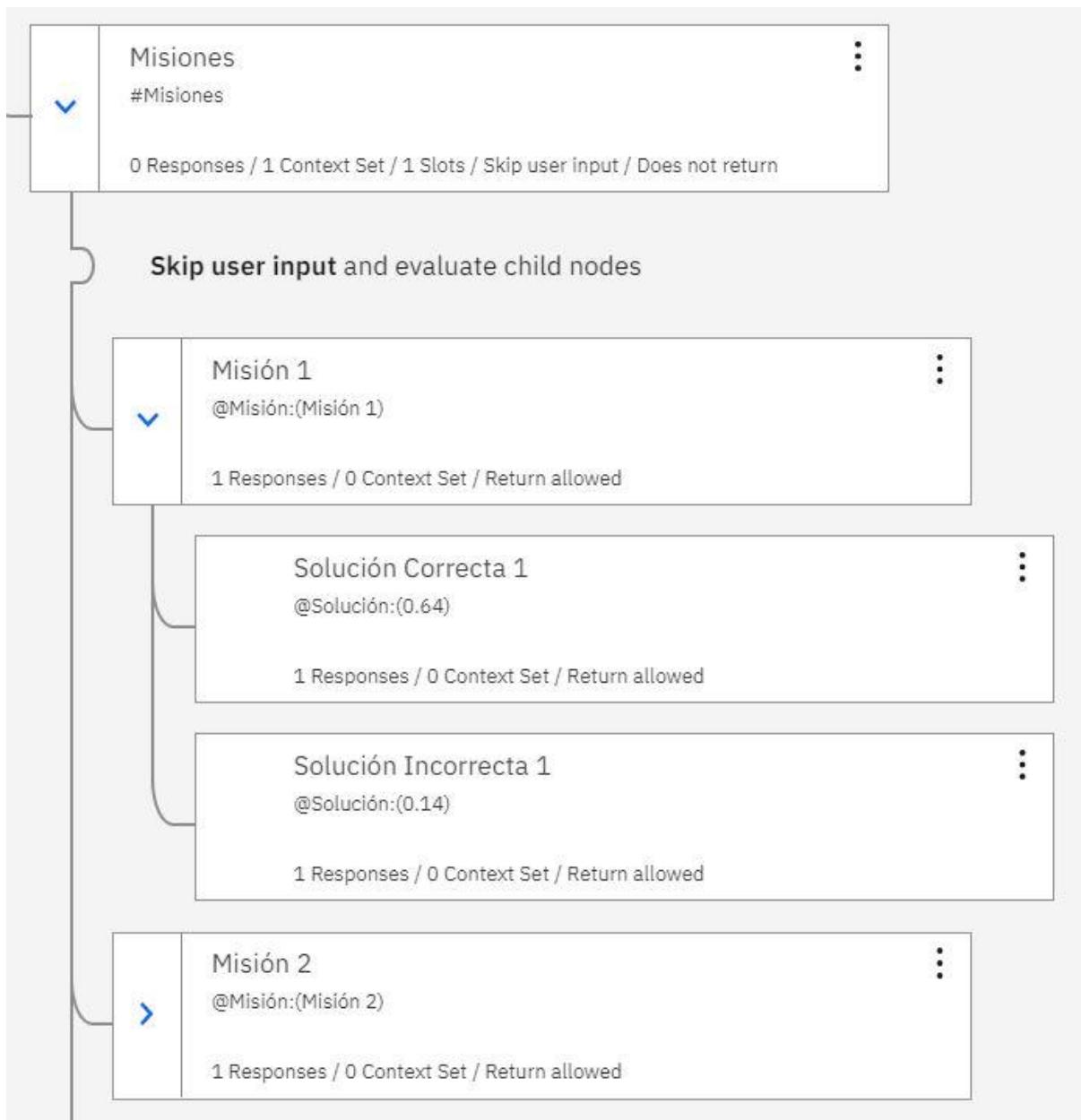


Figura 17: Nodo Misiones del asistente

Otro problema que nos surge, es que cuando se creó este asistente, existía la entidad del sistema nombre, que tenía muchos nombres utilizados por todo el mundo, (que se utilizaba para recoger el nombre del usuario), pero hace dos meses (aproximadamente) esa entidad del sistema fue eliminada, por lo que teníamos que buscar una alternativa. Hablando con un compañero de trabajo que había hecho un asistente para enseñar programación, me dijo que él había creado una entidad que tenía como ejemplos cada uno de los nombres españoles que eran utilizados por más de diez personas en España; los había conseguido a partir del INE. Por lo que me pasó esta entidad. Si el nombre que pone el usuario está en esta entidad, se guarda en la variable de contexto \$nombre (ya que se utiliza a lo largo de todo el proceso) y ya empieza la historia. Si no se detecta, se le dice al usuario que introduzca de nuevo su nombre y ya se guarda en la variable de contexto \$nombre lo que escribe el usuario con la

primera letra en mayúscula, así si hay personas que quieren que se las llame por su mote u otra cosa, pues se puede realizar y no se queda en un bucle.

Al principio, el nodo nombre estaba como nodo padre (nodo del primer nivel), pero al probar el chatbot, al poner Chi-cuadrado, detectó que Chi (está como ejemplo de nombre en la entidad @Nombre) es un nombre, lo guardó en la variable de contexto \$nombre y contestó con el principio de la historia, por lo que se decidió ponerle como nodo hijo del nodo Bienvenido (Figura 18) para no tener este problema y que sólo se llegue al nodo nombre si alguien está al principio o si quiere cambiar el nombre, volver a empezar o volver a ver la historia, es decir, volver a empezar de nuevo.



Figura 18: Nodo Bienvenido del asistente

Por defecto, Watson Assistant te da un link previo para poder ver el asistente y utilizarlo, pero como en el trabajo había aprendido a conectarlo a Telegram a través de node-red, decidí que también era muy atractivo que los usuarios pudieran acceder desde Telegram, ya que es una aplicación que se puede tener tanto en el móvil como en el ordenador y, como es una conversación más, se puede dejar en cualquier momento y retomarlo donde se deja.

El flujo de Node-Red que se utiliza para conectar Telegram con Watson Assistant se puede ver en la figura 19.

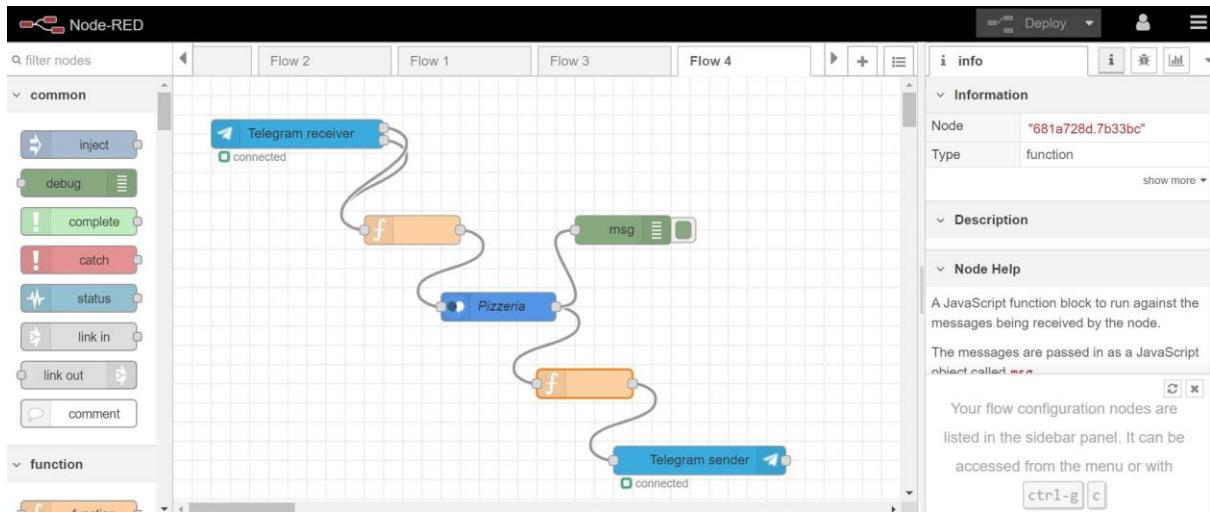


Figura 19: Flujo inicial de Node-red

Los nodos azules claritos (*Telegram receiver* y *Telegram sender*) son los nodos de Telegram que hay que configurar para que se conecten con el bot adecuado, y, el nodo que es de un tono de azul un poco más oscuro (*Pizzería*) es el nodo del servicio de Watson Assistant, que también hay que configurar para que esté conectado con la instancia del asistente que se quiere utilizar.

Los nodos naranjas, son unos nodos llamados *function*, en los que se puede meter código en javascript y, en este caso, el que está conectado a *Telegram receiver* y *Pizzería* se utiliza para adecuar el mensaje como le manda Telegram a como le entiende Watson Assistant, y, el que está conectado a *Pizzería* y *Telegram sender*, se utiliza para adecuar el mensaje de como lo manda Watson Assistant a cómo Telegram necesita recibirllo para poder mostrarlo como se quiere.

Por lo que este flujo funciona de la siguiente manera: el usuario del asistente escribe su mensaje en Telegram y le envía, este mensaje le recibe el nodo *telegram receiver*; que se le manda al nodo al que está conectado, el primer nodo *function* (naranja), que adecúa el mensaje a la forma en la que el nodo *asistente* lo puede entender para mandarle ese mensaje a la instancia de Watson Assistant con la que le hemos conectado, Watson Assistant responde y este mensaje pasa al segundo nodo *function* (naranja) que adecua la respuesta al formato de Telegram y se lo manda al nodo *Telegram Sender* que ya se lo manda a Telegram y aparece como cuando un amigo nos manda un nuevo mensaje.

El nodo que se llama *msg* (el nodo verde) que está conectado a la salida de Watson Assistant, es un nodo *debug* que se utiliza para poder ver en la pestaña *debug*, lo que envía ese nodo al siguiente al que está conectado y sirve para saber cómo se envían las cosas y poder adecuar nuestro código a lo que necesitemos, es un nodo que se ha utilizado mucho para resolver los problemas que vamos a comentar a continuación.

El primer problema con el que topamos, es que siempre que había usado esto, el único tipo de respuesta que se había utilizado en Watson Assistant era texto, entonces me di cuenta de que ni las imágenes ni opciones salían en telegram y me daba error. Preguntando a

compañeros, ninguno sabía cómo hacerlo, así que me puse a buscar en Internet, porque me parecía imposible que nadie se hubiese encontrado con este mismo problema. Cuando ya casi había perdido la esperanza, encuentro el blog de una persona que estuvo en IBM de becario, en el que explica cómo se puede adecuar a los tres tipos de respuesta que necesitaba saber. Él lo hacía por separado ya que en cada nodo del diálogo sólo tenía un tipo de respuesta, generalizando y adecuando este código (se puede encontrar en el Anexo I y es el código que se utiliza en el segundo nodo *function* del flujo para adecuar el mensaje de como nos le da Watson Assistant a como Telegram necesita recibirla) y probándolo, acabe consiguiendo que me detectara texto, imagen y opciones.

Después, como en Watson Assistant se utiliza HTML para dar formato al texto, se necesitaba una manera de que Telegram detectará HTML para que diera formato en vez de poner literal el código, buscando en el foro de Node-red y en la documentación de Telegram, se consiguió que Telegram reconociera HTML añadiendo el código que se puede encontrar a continuación tanto en el tipo de respuesta texto como en el de opciones.

```
msg.payload.options = {  
    parse_mode:'HTML'  
}
```

El último problema que se encontró fue que, como el asistente tiene nodos que en su respuesta tienen tanto imagen como texto y opciones, se confundía con el tipo mensaje e imagen, por lo que no sacaba bien la información y, como el mensaje venía todo junto, la única opción que se encontró fue añadir nodos *switch* para cada elemento del objeto mensaje. El nodo *switch* es un nodo que permite poner condiciones y conectar distintos nodos dependiendo de la condición con la que se encuentre, así que, si detectaba que el mensaje era de tipo imagen, lleva a un nodo *function* en el que está el código para ese tipo, y si era tipo texto, a otro nodo *function* distinto en el que está el código tanto para opciones como para texto. El flujo final se puede ver en la figura 20.

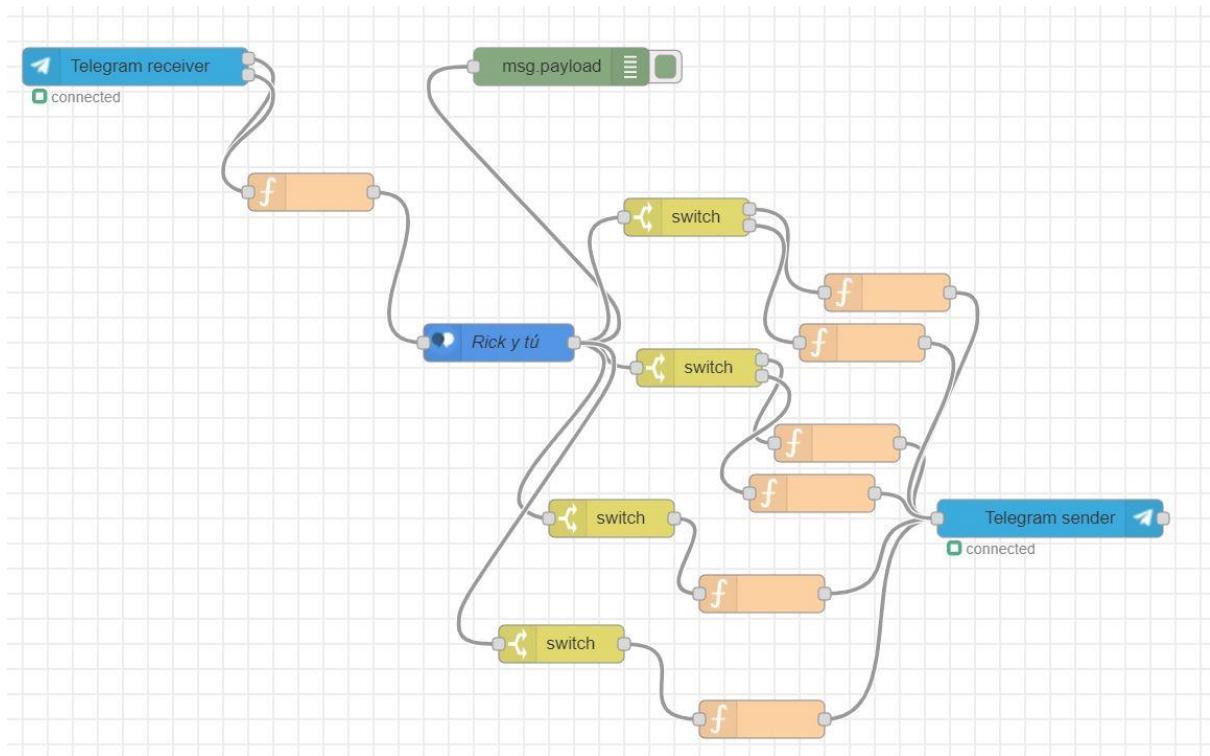


Figura 20: Flujo final de Node-red

4.3 Contenido

Para elaborar este apartado se han consultado (González, 2014), (Menéndez, 2015), (Rodríguez, Probabilidad, 2015), (Rodríguez, Inferencia, 2016), (Barba, 2016), (Quintela, El estadístico accidental, 2014), (Quintela, Página de Alejandro Quintela, 2020), (Wikipedia, 2020), (Derivando, 2020), (Emojipedia, 2020), (La investigación, 2020), (Línea de código, 2020), (webusable, 2020), (unam, 2020), (Wikipedia, 2020), (indifferentlanguages, 2020).

En este punto se puede encontrar el contenido del asistente conversacional, nada más empezar el asistente sale la figura 21 y la frase de debajo y ya todo está como en el asistente.

HISTORIA



Figura 21: Imagen de Rick y tú

Rick: ¿Qué narices? Sí que me pasé ayer... oye, y tú, ¿quién eres?
(Introduce tu nombre)

Si el usuario no introduce un nombre guardado:

Rick: ¿Eso es un nombre? Cada día ponen nombres más raros a la gente, ¿en serio quieres que te llame así?
(Vuelve a introducir tu nombre)

Se guarda el nombre de la persona en la variable de contexto \$nombre.

Y ya, para los dos casos (que tengamos el nombre en la entidad @Nombre o que no y guardamos lo que escribe el usuario):

Voz: Bienvenidos al *Concurso Intergaláctico*, desde el planeta Tierra os presento a Rick y \$nombre que han venido hasta aquí para demostrar que su planeta es el mejor en estadística.

Rick: Eeee para el carro, ¿qué gano yo con eso?

Voz: Aprender estadística.

Rick: Ja, ja, ja, en serio pensáis que me vais a enseñar algo que yo no sepa.

Voz: A ver, ¿cómo se hace una media?

Rick: Mmmm, ¿qué narices?

Voz: Hemos borrado todos los recuerdos de los participantes sobre la estadística para que sea un concurso justo.

Rick: ¡Diantres! ¡Vamos a por todas! ¿Tú que opinas \$nombre?

Si elige la opción: ¡Vamos a por todas!

Rick: ¡Así se habla \$nombre! Venga, que, ¿cuándo empieza esto?

Le saldrá la opción: Eso, ¿cuándo empezamos? ¡Que tengo unas ganas!

Si elige la opción: Puff... paso bastante, pero que remedio...

Rick: Vaya ánimos que das, más te vale cambiar esa actitud o no ganaremos y a mi nadie me gana. Bueno, ¿cuándo empieza esto?

Le saldrá la opción: Eso, ¿cuándo empezamos? Que cuanto antes nos lo quitemos de encima mejor...

En los dos casos le lleva a:

Voz: Ya estáis en el concurso, estas son las normas:

1. Se os pone al alcance todo lo necesario para poder completar las misiones que irán surgiendo, las mejor realizadas y originales ganarán el concurso.
2. Desde cualquier sitio podréis volver aquí escribiendo **INICIO**.
3. Se aconseja seguir el orden (cuando hay varias opciones la que irá primera será la que aparece en primer lugar y así) preestablecido, aun así podéis pasar del orden y preguntar en cada momento lo que os interese saber (¡allá vosotros!).
4. Se aconseja clicar en los botones (desplegar las opciones y elegir una) en vez de escribir manualmente y enviar el texto, aun así podéis escribir el texto de las opciones/botones manualmente y enviarlo, menos escribir literalmente **SIGUIENTE**, ya que si clicais al botón siguiente nos escribirá lo que viene después de eso pero si escribís literalmente siguiente os recordaré esto.
5. Para que no haya trampas, una vez dada la respuesta a una misión (aunque sea incorrecta), no se puede escribir la otra solución, tiene que volver a ponerse la misión para poder dar la solución correcta.

Esto es todo, ¡Feliz Concurso Intergaláctico! Y que la suerte esté siempre de vuestra parte.

1. Estadística descriptiva

La **rdysfodyovs descriptiva** es un conjunto de técnicas y procedimientos que ayudan a describir, mostrar y resumir la información de un conjunto de datos.

\$nombre: ¿Qué significa rdysfodyovs?

Rick: Me da que se les ha olvidado traducir esa palabra

\$nombre: A lo mejor quieren que aprendamos su idioma... ¡A ver si averiguamos cómo funciona!

Rick: Eso sí que te motiva, ¿eh?

\$nombre: La verdad es que sí y, oye, no se andan con bobadas, parece que empezamos al lío directamente.

Rick: Como deber ser, lo bueno, si breve, dos veces bueno.

1.1 Conceptos clave

Pasamos a conocer una **serie de conceptos** que nos van a ser útiles para comprender más todo este mundo.

\$nombre: ¿Qué será eso del sur? 🤔

Rick: a saber, a lo mejor piensan que está de moda o es una gracia del traductor.

1.1.1 Población y Muestra

La **población** es el conjunto de individuos (cada uno de los elementos que la forman) sobre los que se va a realizar el estudio estadístico.

Por ejemplo: habitantes de Rodilana, herretes fabricados en 2019...

\$nombre: anda mira, ¡habrán visto Phineas y Ferb!

Rick: Las maravillas de la televisión interdimensional, oye, pero ¿dónde narices está Rodilana?

\$nombre: 🧑.

La **muestra** es un conjunto REPRESENTATIVO de la población sobre el que se trabaja.

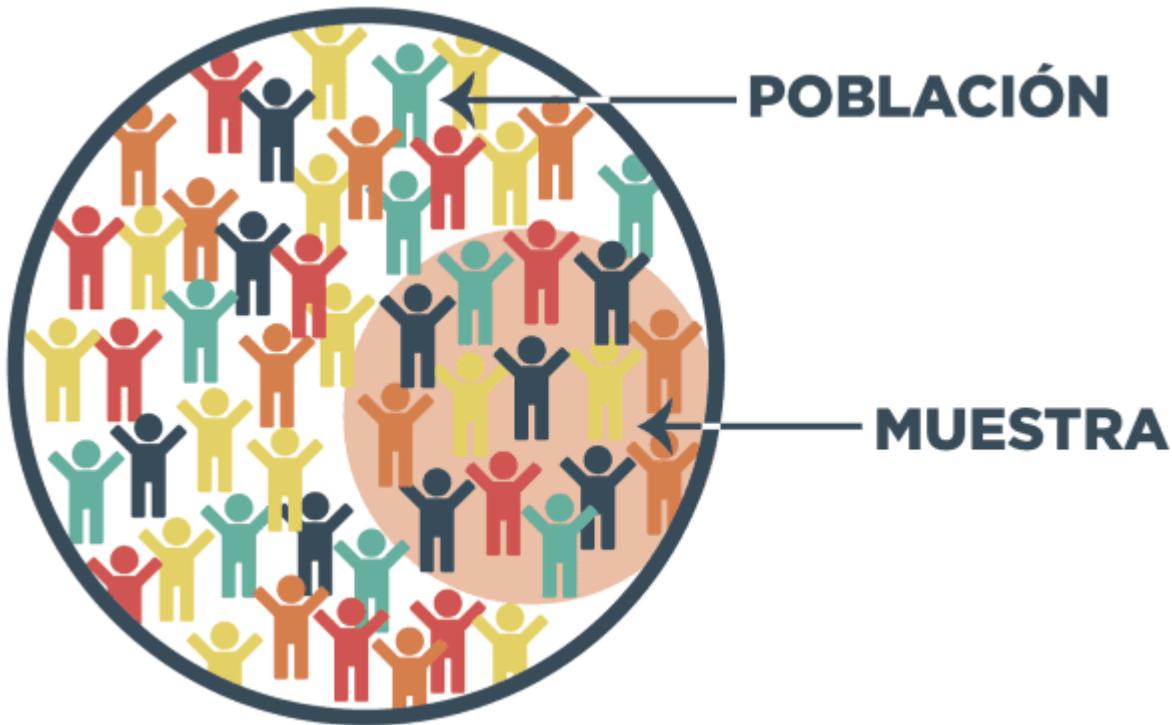


Figura 22: Población y muestra

1.1.2 Variable

Una **variable** es una característica cambiante de los individuos de la población que nos interesa estudiar.

Hay dos tipos de variables:

1.1.2.1 Variable cualitativa

Una **variable cualitativa** (atributo) es una característica que no puede ser medida con números. Hay dos tipos:

- **Nominal:** no admiten un criterio de orden. Por ejemplo, el color de pelo: rubio, moreno o pelirrojo.

\$nombre: La rubia es...

Rick: ...

\$nombre: Tienes que repetir... ¿no te la sabes?

Rick: Estos españoles tenéis unas cosas...

- **Ordinal:** admiten un criterio de orden. Por ejemplo, el grado de satisfacción: insatisfecho, poco satisfecho, muy satisfecho.

Una variable cualitativa se puede convertir en cuantitativa si a cada una de sus clases se le asigna un número (1=rubio, 2=moreno, 3= pelirrojo).

1.1.2.2 Variable cuantitativa

Una **variable cuantitativa** es una característica medida a través de los números. Hay dos tipos:

- **Discreta**: cuando no puede tomar valores entre dos consecutivos. Por ejemplo, número de ornitorrincos por persona (una persona puede tener uno, dos o tres ornitorrincos pero nunca podrá tener un ornitorrinco y medio).

\$nombre: Vale, ¡definitivamente han visto Phineas y Ferb! 😊

- **Continua**: cuando puede tomar cualquier valor. Por ejemplo, cl consumidos de vodka por persona/día (pueden ser 33.333 cl).

Rick: ¡esto ya si que es lo mío! 😎

1.1.3 Tabla de frecuencias

Una **tabla de frecuencias** es una forma de organizar los datos. Cada una de las filas está formada por:

- **Datos** (x_i): cada uno de los valores.
- **Frecuencia absoluta** (n_i): número de veces que aparece el valor.
- **Frecuencia absoluta acumulada** (N_i): es el resultado de ir sumando las frecuencias absolutas.
- **Frecuencia relativa** (f_i o h_i): es la frecuencia absoluta del valor dividido por el número de valores total.
- **Frecuencia relativa acumulada** (F_i o H_i): es el resultado de ir sumando las frecuencias relativas.

Si tenemos muchos valores, los podemos agrupar por intervalos e indicar el número de observaciones que hay dentro de cada intervalo (por ejemplo, si nuestros datos son 2,3,5,6,7,5 en el [2,5] hay tres observaciones y en el (5,8] hay dos).

1.1.3.1 Misión 1

\$nombre: Nuestra primera misión, ¡qué ilusión!

Rick: ¡Waba laba dab-dab!

Water-T está realizando un estudio estadístico sobre su planeta *Alphabetrium*, para ello recoge las edades de una muestra de sus habitantes:

1,1,3,5,6,6,7,12,14,16,16,18,19,24,24,24,26,27,28,29,29,31,31,33,35,37,37,46,46,47,47,47,52,54,55,57,62,62,68,69,75,80. Pero como ni idea de estadística tiene, necesita vuestra ayuda, necesita una tabla de frecuencias agrupando las edades de diez en diez (1-10,11-20,21-30....).

¿Cuál es la frecuencia relativa acumulada de las edades de 31-40 años?

Solución correcta 1:

¡Aleluya!

Solución incorrecta 1:

Rick: Empezamos con buen pie...

\$nombre: Pues si... Nos hemos liado con la frecuencia relativa y su acumulada 😥.

Solución:

Intervalo (i)	Edades	Frecuencia Absoluta (fi)	Frecuencia Acumulada (Fi)	Frecuencia Relativa (hi)	Frecuencia Relativa Acumulada (Hi)
1	1 - 10	7	7	7 / 42 = 0,17	7 / 42 = 0,17
2	11 - 20	6	7+6= 13	6 / 42 = 0,14	13 / 42 = 0,31
3	21 - 30	8	13+8= 21	8 / 42 = 0,19	21 / 42 = 0,5
4	31 - 40	6	21+6= 27	6 / 42 = 0,14	27 / 42 = 0,64
5	41 - 50	5	27+5= 32	5 / 42 = 0,12	32 / 42 = 0,76
6	51 - 60	4	32+4= 36	4 / 42 = 0,1	36 / 42 = 0,86
7	61 - 70	4	36+4= 40	4 / 42 = 0,1	40 / 42 = 0,95
8	71 - 80	2	40+2= 42	2 / 42 = 0,05	42 / 42 = 1
	Amplitud 9	N: 42		hi= fi / N	Hi= Fi / N

Figura 23: Misión 1

1.2 Medidas de posición

Procedemos a dar una **serie de medidas** que resuman la información existente en los datos que queramos analizar y representen a la distribución en su conjunto y ayuden a la comparación entre distintas variables.

\$nombre: 🤔 ¿En qué posición estoy?

Rick: 🙄

1.2.1 Media

La **media** es el valor promedio de todas las observaciones. Cuando hablamos de media, nos referimos a la media aritmética. Para calcularla sumamos todos los valores y dividimos esa suma por cuántos son, es decir, utilizamos la siguiente fórmula:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Ecuación 1: Media

Sus características son:

1. Si a cada observación de una variable X se le suma una constante b , se tiene una nueva variable $Y = X + b$ con media igual a la de X más la constante b .
2. Si se multiplica una variable X por una constante a , la variable resultante $Y = aX$ tendrá media igual a a por la media de X .
3. La media aritmética se ve muy alterada por valores extremos de la variable.

Rick: Por lo que si a X se le suma y multiplica una constante, $Y = aX + b$.

\$nombre: Te crees muy listo, ¿no?

1.2.1.1 Misión 2

Rick: Oye, ¿y si calculamos la media de las veces que me tengo que llenar la petaca a la hora?

\$nombre: Mmmm, vale, si te hace ilusión... ¿Cuántas veces la has llenado las últimas 7 horas?

Rick: Pues mirando en mi memoria... La primera hora 8, la segunda 9, 10, 11, 16, 17 y la última 6.

\$nombre: Te va a venir esto bien y todo, mira qué poco has bebido desde que llegamos.

Rick: ¡Calla! A ver, ¿cuánto es la media?

Solución correcta 2:

¡αλληλούια!

Solución incorrecta 2:

Rick: Ni una media sabemos hacer... vamos a ser el ridículo del espacio...

Solución:

$$\bar{X} = \frac{8 + 9 + 10 + 11 + 16 + 17 + 6}{7} = \frac{77}{7} = 11$$

Figura 24: Misión 2

1.2.2 Mediana

La **mediana** parte la muestra ordenada en dos, dejando el 50% de los datos a cada lado. Para calcularla hay que ordenar los datos. Si son un número impar, la mediana es el valor que se encuentra en el medio, es decir, $x_{n/2}$, y si son uno par será la media de los dos elementos centrales:

$$\frac{x_{n/2} + x_{n/2+1}}{2}$$

Ecuación 2: Mediana

\$nombre: Anda claro, por eso si una familia tiene tres hijos, ¡la segunda se llama mediana!
Rick: definitivamente me ha tocado el gracioso del grupo 

1.2.2.1 Misión 3

Morty quiere saber la mediana de las horas que le dura la superinteligencia pasajera después de la disolución de las megasemillas para saber cuándo tiene que introducírselas para aprobar los exámenes. Para ello ha probado seis veces y estos son los resultados: 8, 10, 9, 7, 12, 11. ¿Cuál es la mediana?

Solución correcta 3:

¡Hallelujah!

Solución incorrecta 3:

\$nombre: Yo creo que en el fondo le estamos haciendo un favor... 

Rick: Si tú lo dices... 

Solución:

Cálculo de la mediana

1 Ordenamos los datos de menor a mayor.

6 5 2 4 3 4 5 6 5

2 Si la serie tiene un número impar de medidas la mediana es la puntuación central de la misma.

2, 3, 4, 4, 5, 5, 5, 6, 6 $Me= 5$

3 Si la serie tiene un número par de puntuaciones la mediana es la media entre las dos puntuaciones centrales.

7, 8, 9, 10, 11, 12 $Me= 9.5$

Figura 25: Misión 3

1.2.3 Moda

La **moda (absoluta)** será el valor que más veces se repita y serán **modas relativas** los valores que se repitan más veces que los valores adyacentes. Por ejemplo, para 1,3,3,4,5,5,5,7,8 la moda absoluta será cinco y la relativa 3 ya que se repite más veces que el 1 y el 4.

Rick: ¿Quién va a la moda? ¿eh? \$nombre ¿Quién va a la moda? - dice mientras se pasea con una boa al cuello que a saber de dónde la ha sacado.

\$nombre: ¡Y luego dices de mí! 

1.2.3.1 Misión 4

Summer se piensa que esto de calcular la moda la ayudará a ir a la última y conseguir más likes en sus publicaciones de Instagram, así que quiere saber la moda absoluta y las relativas de los likes que le han dado en sus últimas once publicaciones: 10, 20, 15, 20, 25, 30, 35, 20, 20, 30, 15.

Solución correcta 4:

¡哈利路亞!

Solución incorrecta 4:

\$nombre: Tenemos que relativizar nuestros errores, Rick.

Rick: No en este caso 

\$nombre: ¡Malísimo! 

Solución:

La moda absoluta es **20** ya que es el que aparece más veces (4) y la moda relativa es **30**, ya que aparece dos veces mientras que sus adyacentes, 25 y 35, solo una.

1.2.4 Cuantiles

Los **cuantiles** parten la muestra ordenada en dos, dejando a la izquierda una proporción p de los datos y a la derecha una $1 - p$. Para realizar su cálculo hay que ordenar la muestra y:

- Si np no es entero, es la observación que se encuentra en el lugar $[np] + 1$ (se le quita la parte decimal a np y se le suma 1).
- Si np es entero, es el promedio de las observaciones que se encuentran en $[np]$ y $[np] + 1$.

Cuantiles que destacan:

- **Primer cuartil:** $Q_1 = 0.25\text{-cuantil}$.
- **Segundo cuartil:** $Q_2 = 0.5\text{-cuantil}$.
- **Tercer cuartil:** $Q_3 = 0.75\text{-cuantil}$.
- **Percentil 100 x p = p-cuantil** donde $p = 0.01, \dots, 0.99$.

\$nombre: ¡Por lo que el segundo cuartil será la mediana! 🤪

Rick: Ha llegado un trozo de seso hasta aquí 🤪

1.2.4.1 Misión 5

Jerry está haciendo un estudio sobre los Jerrys de la guardería de Jerrys, para ello necesita el tercer cuartil de los años que los Jerry's calculan que llevan en la guardería: 10, 72, 18, 45, 32, 56, 64, 27, 60 y el primer cuartil de los años que se piensan que tienen: 29, 29.5, 30, 30.5, 30.8, 30.8, 30.8, 31, 31.5, 31.5, 32, 32, 32, 32.5, 33, 33.

Rick: es palurdo hasta para esto... ni siquiera ha preguntado a todos los Jerrys sobre las dos cosas 🙄

Jerry: ¡Joye!

\$nombre: ¿Nos estáis escuchando?

Jerry: Pues claro, no veis que estáis en un concurso y tenemos una tele interdimensional en casa... si es que... luego el tonto soy yo 🤦 Bueno qué, ¿cuánto da el tercer cuartil de los años que piensan que llevan en la guardería?

Solución correcta 5:

¡Alleluia!

Solución incorrecta 5:

Rick: ¡Que se vaya a freír espárragos! Que esto no sirve para nada.

Jerry: ¿Por qué eres siempre tan malo conmigo? 😭

Rick: Por qué se casaría mi hija contigo, ¡esa es la pregunta!

Solución:

Lo primero que tenemos que hacer es ordenar los datos (si están desordenados).

- Para el primer caso, como hay nueve datos y nos mandan calcular el tercer cuartil, $np = 9 \times 0.75 = 6.75$, como no es entero, quitamos su parte decimal y le sumamos 1 que da 7, ya que hay nueve datos y nos mandan calcular el tercer cuartil, por lo que la respuesta es **60** ya que es el valor que está en el séptimo lugar.

- Para el segundo caso, como hay 16 datos y nos han mandado calcular el primer cuartil, $np = 16 \times 0.25 = 4$, como es un número entero, la solución es **30.65**, que es la media de los valores que están en las posiciones 4 y 5.

1.3 Medidas de dispersión

Son una **serie de medidas** que indican el nivel de concentración de los datos e informan sobre la bondad de los promedios calculados como representativos del conjunto de datos.

\$nombre: ¡Concentración es lo que necesitamos!

Rick: Vamos estando en la misma onda 😊😊

1.3.1 Varianza, cuasivarianza y desviación típica

La **varianza** representa la variabilidad de una serie de datos respecto a su media. Se calcula sumando la resta al cuadrado de cada dato menos la media y dividiendo el resultado por el número de datos. Esta es la fórmula:

$$Var(X) = \frac{\sum_1^n (x_i - \bar{X})^2}{n}$$

Ecuación 3: Varianza

Las características de la varianza son:

1. Si se le suma una constante a una variable, la varianza de la nueva variable no cambia (porque la distancia de los datos a la media sigue siendo la misma).
2. Si se multiplica una variable por una constante, la varianza de la nueva variable es igual a la de la antigua multiplicada por la constante al cuadrado (al multiplicar los datos, multiplicamos la dispersión. Pensemos en $X=1,2,3$, e $Y = 2X=2,4,6$. Hemos multiplicado por 2 la distancia entre los datos).

Rick y \$nombre: ¡Así que si multiplicamos y sumamos una constante a $X, Y = a^2 \cdot X$ simplemente!

La **cuasivarianza** se calcula igual que la varianza pero cambiando el denominador por $n - 1$ y la desviación típica es la raíz cuadrada con signo positivo de la varianza.

1.3.1.1 Misión 6

Rick: ¡\$nombre! ¡Apunta el número de ojos de esta especie de moscas que hay por aquí!

\$nombre: ¿Para qué? ¡Loco la vía!

Rick: Vamos a calcular su varianza, cuasivarianza y desviación típica.

\$nombre: ¡Sí que te está afectando esto! jajaja a ver, a ver... mmm 3, 2, 3, 3, 2, 6 y 3.

Rick: Quiero que se den cuenta que somos los mejores y la originalidad fijo que suma, a ver, ¿cuál es la cuasivarianza?

Solución correcta 6:

اسبحوا الراب

Solución incorrecta 6:

Rick: Lo que sí está variando es nuestra posibilidad de ser los mejores.

\$nombre: Pues si... nos tenemos que esforzar más....

Solución:

$$\bar{x} = \frac{3 + 2 + 3 + 3 + 2 + 6 + 3}{7} = \frac{22}{7} = 3.14$$

$$\begin{aligned}\sigma^2 &= \frac{(3 - 3.14)^2 + (2 - 3.14)^2 + (3 - 3.14)^2}{7} + \\&+ \frac{(3 - 3.14)^2 + (2 - 3.14)^2 + (6 - 3.14)^2 + (3 - 3.14)^2}{7} = \\&= \frac{0.02 + 1.3 + 0.02 + 0.02 + 1.3 + 8.18 + 0.02}{7} = \\&= \frac{10.86}{7} = 1.55\end{aligned}$$

Figura 26: Misión 6

La cuasivarianza es $10.86/6 = 1.81$.

La desviación típica es **1.245** (la raíz cuadrada de la varianza).

1.3.2 Recorrido, rango intercuartílico y coeficiente de variación

El **recorrido** (también llamado rango o amplitud) es la diferencia entre el mayor y el menor de los valores. $R = \max - \min$.

Rick: ¡Volando voy! ¡Volando vengo, vengo! ¡Por el recorrido, yo me entretengo!

\$nombre: Voy preparándome...

El **rango intercuartílico** es la diferencia entre el tercer y primer cuartil. $RIQ = Q3 - Q1$.

El **coeficiente de variación** es la división de la desviación típica entre el valor absoluto de la media (por lo que es adimensional, es decir, no tiene unidades) y permite comparar la dispersión de varias distribuciones.

1.3.2.1 Misión 7

Rick: Estoy pensando ahora en mi atracción *Piratas del Páncreas* y me gustaría mejorarla.

\$nombre: Oye, y, ¿por qué no la probamos y apuntamos los minutos que estamos en ella?

Rick: Venga va, y así, podemos calcular el recorrido, el rango intercuartílico y el coeficiente de variación y vemos qué podemos hacer.

... *Tres horas más tarde* ...

\$nombre: ¡Que subidón! ¡Ha sido una pasada!

Rick: Me cuesta reconocerlo 🤔 pero estoy mayor para estas cosas 😱

\$nombre: ¡Puaj! ¡Qué asco, Rick! Anda trae la hoja que ya me encargo yo: 17, 16, 21, 18, 13, 16, 12 y 11, mmm ¿cuál será el recorrido?

Solución correcta 7:

¡Alleluja!

Solución incorrecta 7:

Rick: ¡No se te puede dejar solo!

Solución:

El recorrido es $21 - 11 = 10$.

El rango intercuartílico es $18 - 12.5 = 5.5$.

El coeficiente de variación es $3.122 / 15.5 = 0.2014$.

1.3.3 Tchebychev

La **desigualdad de Tchebychev** (fue formulada por un matemático que se apellidaba Chebyshov pero al final todo el mundo lo llama de todo tipo de formas menos así) proporciona un intervalo de confianza que nos permite relacionar la media y la varianza.

\$nombre: Pobre Chebyshov, ya la podría llamar todo el mundo así y no liarnos.

Rick: o haber elegido un nombre más fácil para llamarla 🤪

$$fr(|x_i - \bar{X}| \leq kS) \geq \left(1 - \frac{1}{k^2}\right), \quad k > 0$$

$$fr(|x_i - \bar{X}| > kS) \leq \left(\frac{1}{k^2}\right), \quad k < 0$$

Sé que os habéis quedado igual, pero la desigualdad de Tchebychev es más fácil de lo que parece, vais a verlo.

Lo que nos viene a decir es que en un intervalo de centro la media y de radio k veces la desviación típica, es decir, $(\bar{X} - kS, \bar{X} + kS)$ se encuentran al menos el $\left(1 - \frac{1}{k^2}\right) \times 100\%$ de los datos y por fuera están a lo sumo el $\left(\frac{1}{k^2}\right) \times 100\%$ de los datos, por ejemplo, si $k = 2$, el radio será 2 veces la desviación típica y el intervalo contendrá, al menos, el 75% de los datos por lo que por fuera del intervalo quedará, a lo sumo, el 25% de los datos.

Figura 27: Tchebychev

1.3.3.1 Misión 8

Rick: En aquella época remota en la que todavía no tenía un mundo en mi batería si quise hacer pruebas para ver qué necesitaría para que funcionará.

\$nombre: ¿Ah, sí? ¿Cómo qué?

Rick: Pues estuve investigando muchas cosas, pero ahora se me ocurre una más; vale, imaginemos que tuviera 100 componentes electrónicos que les pongo a trabajar hasta que fallen, veo que la duración media fue de 103.45 h y la varianza de 9973.45 h.

\$nombre: Anda, pues por ejemplo, podríamos mirar si es posible que algún componente durará más de 1000 horas.

Rick: mmm, venga va, me gusta, ¿cuál sería el porcentaje que queda por fuera?

Solución correcta 8:

¡Sñrñius!

\$nombre: Mira Rick, parece que han vuelto a usar su idioma, Aleluya es así, entonces, ¿cómo funcionará?

Rick: Ya te digo que su idioma no creo que sea, será un código o un error del traductor o lo que sea, pero a saber, todo es posible.

Solución incorrecta 8:

Rick: Esto nos pasa por fiarnos de la intuición en vez de ponernos en serio.

\$nombre: Ya te digo Rick, tendremos que espabilar...

Solución:

Como nos están diciendo que dure más de 1000 horas, están hablando del extremo derecho del intervalo:

$$\bar{X} + kS = 1000$$

Por lo que $k = 8.97742$ y $1/k^2 = 0.012$. Así que **1.24%** de los componentes pueden haber superado las mil horas así que es posible que uno lo haya superado.

1.4 Medidas de forma

Son unas medidas que se fijan en la forma de la distribución.

\$nombre: Se han herniado, ¿eh?

1.4.1 Asimetría

Una distribución o variable es simétrica si al levantar gráficamente un eje sobre la media (o mediana) el dibujo a ambos lados es idéntico, si no, es asimétrica.

Rick: ¡Eso es una verdad de perogrullo!

Los **coeficientes de asimetría** son valores numéricos que indican si la distribución es simétrica y, en caso de no serlo, el signo de su asimetría. Un coeficiente de asimetría muy utilizado es el de Fisher (*CA* o *As*):

$$As = \frac{\sum_{i=1}^n \frac{(x_i - \bar{x})^3}{n}}{S^3}$$

Ecuación 4: Coeficiente de asimetría

Hay dos tipos de asimetría:

- **Positiva:** cola de la distribución más larga a la derecha y $CA > 0$.
- **Negativa:** cola de la distribución más larga a la izquierda y $CA < 0$.

Si es **simétrica**, $CA = 0$.

1.4.1.1 Misión 9

Rick: ¿Sabes, \$nombre? Me gustaría saber la forma del número de cristales *kalaxianos* que he utilizado en mis últimas fiestas.

\$nombre: ¿Cuántos has utilizado en las últimas?

Rick: A ver, a ver, pues 6, 15, 9, 12, 9, 12, 12 y 17. ¿Cuánto da el coeficiente de asimetría?

Solución correcta 9:

¡Alléluia!

Solución incorrecta 9:

\$nombre: Han ido a pillar... ¡no vale!
Rick: ¡Bienvenido al espacio!

Solución:

El numerador se ha calculado restando a cada dato la media y elevándolo al cubo $(6 - 11.5)^3 + \dots + (17 - 11.5)^3$. El denominador es simplemente multiplicar el número de datos por el cubo de la desviación típica. Por lo que $CA = 12/(8 \times 3.279^3) = 0.035$. Por lo que tiene un poco de asimetría positiva.

1.4.2 Curtosis

Las **medidas de curtosis** (o apuntamiento) tratan de estudiar la distribución de frecuencias en la zona central de la distribución. La mayor o menor concentración de frecuencias alrededor de la media dará lugar a una distribución más o menos apuntada. Ese grado de apuntamiento se calcula a través del **coeficiente de curtosis**:

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^4}{nS^4} - 3$$

Ecuación 5: Coeficiente de curtosis

Si vale **0**, querrá decir que tiene el mismo apuntamiento que una $N(0,1)$ y se dice que la distribución es mesocúrtica (que tiene el **mismo apuntamiento que la normal**).

Si es **menor que 0** se dice que es platicúrtica (la variable es **menos apuntada que la normal**).

Si es **mayor que 0** es leptocúrtica (es **más apuntada que la normal**).

1.4.2.1 Misión 10

Rick: Sabes lo que toca ahora, ¿no?

\$nombre: Sí, sí, calcular el coeficiente de curtosis para el número de cristales kalaxianos de tus últimas fiestas, que eran, 6, 15, 9, 12, 9, 12, 12 y 17.

Rick: Estoy pensando que a mi próxima fiesta vienes sí o sí. Estoy divagando un montón, ¡que ganas! Va a ser la mejor fiesta que hayas visto nunca, nada parecido por lo menos, pero ahora vamos a centrarnos, que se me va, venga, ¿cuánto da?

Solución correcta 10:

¡aleluya!

Solución incorrecta 10:

\$nombre: Están tocanarices últimamente...

Rick: Pues si, pero así son las reglas del espacio, hay que ser perfecto si quieres ser el mejor.

Solución:

El numerador se ha calculado restando a cada dato la media y elevándolo a la cuarta $(6 - 11.5)^4 + \dots + (17 - 11.5)^4$. El denominador es simplemente multiplicar el número de datos por la desviación típica a la cuarta. Por lo que $CA = 2058.5/(8 \times 3.279^4) = 2.23$. A esta cantidad la tenemos que restar 3, por lo que el coeficiente de curtosis da **-0.77**. Al ser menor que 0 esta variable es menos apuntada que la normal.

1.5 Representaciones gráficas

Vamos a ver diferentes formas de representar los datos.

1.5.1 Diagrama de sectores

El **diagrama de sectores** representa los valores de la variable en sectores en un círculo proporcionales a su frecuencia. Sobre todo se usa con variables cualitativas.



Figura 28: Diagrama de sectores

- Si lo queremos hacer a mano, se va calculando el ángulo de cada sector, $(360^\circ \times n_i)/n$, y se dibuja.

- Si lo queremos hacer con R, definimos las frecuencias relativas o absolutas en un vector, por ejemplo: `x=c(0.3,0.3,0.4)`, asociamos los nombres, `names(x)=c("guatebuena", "guatemala", "guatepeor")` y realizamos el gráfico, `pie(x)`, si ejecutamos `help(pie)` podemos ver todas las opciones que tenemos (título, colores...).

\$nombre: Anda, mira, mi padre siempre me decía, te has ido de guatemala a guatepeor.

Rick: Tenéis un humor tan raro... 😢 Me gusta 👍

1.5.2 Diagrama de barras

El **diagrama de barras** casi siempre se utiliza para representar variables cualitativas. En el eje x se ponen los diferentes valores de la variable y sobre ellos se levantan las barras con altura el valor de su frecuencia relativa o absoluta. Si lo queremos hacer con R, representaríamos las frecuencias y sus nombres como en el diagrama de sectores y luego utilizaríamos `barplot(x)`.

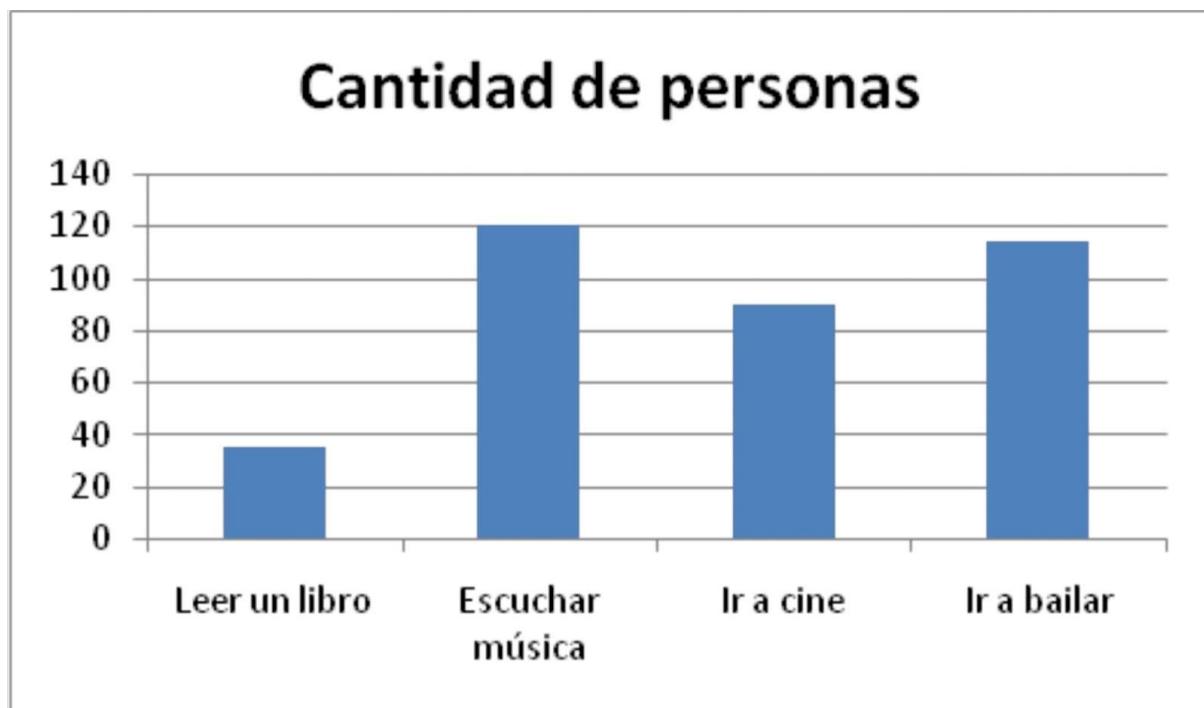


Figura 29: Diagrama de barras

Si tenemos más de una variable se utiliza un **diagrama de barras múltiple** que nos sirve para comparar los datos.

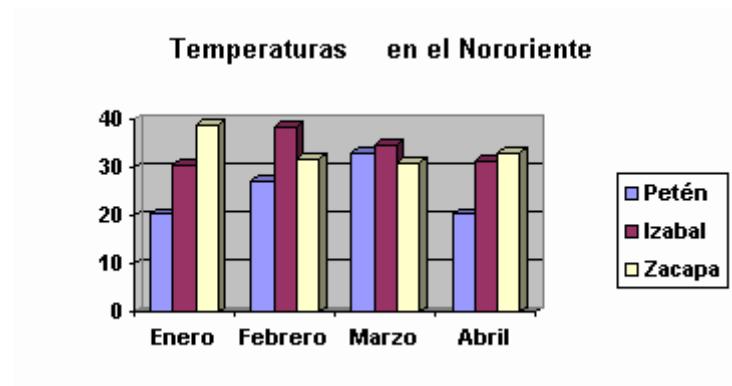


Figura 30: Diagrama de barras múltiple

1.5.3 Diagrama de Pareto

El **diagrama de Pareto** es como un diagrama de barras pero ordenando las frecuencias de mayor a menor y una línea que va representando la frecuencia acumulada. Solo se utiliza para variables cualitativas nominales.

\$nombre: 😂 😂

Rick: ¿Qué te pasa ahora?

\$nombre: Lo siento, pero no puedo evitarlo, me ha sonado a diagrama de Paleo 😂 😂

Rick: 🙄

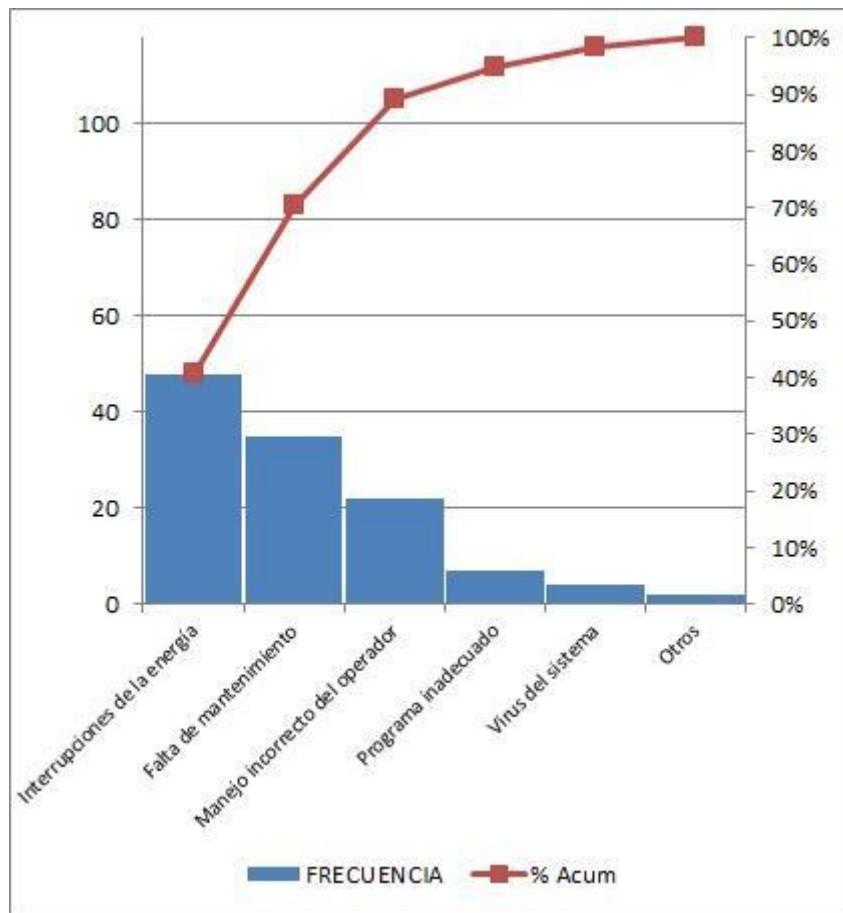


Figura 31: Diagrama de Pareto

Para hacerlo en R tenemos que instalar el paquete *qcc* ejecutando *install.packages("qcc")* y *library(qcc)*, luego es poner las frecuencias en un vector *x* y ejecutar el comando *pareto.chart(x)*.

1.5.4 Histograma

El **histograma** se utiliza para las variables cuantitativas continuas.

Rick: Si que es histórico esto 😂😂

\$nombre: 🧑

Título: Histograma Calificación obtenida en la prueba

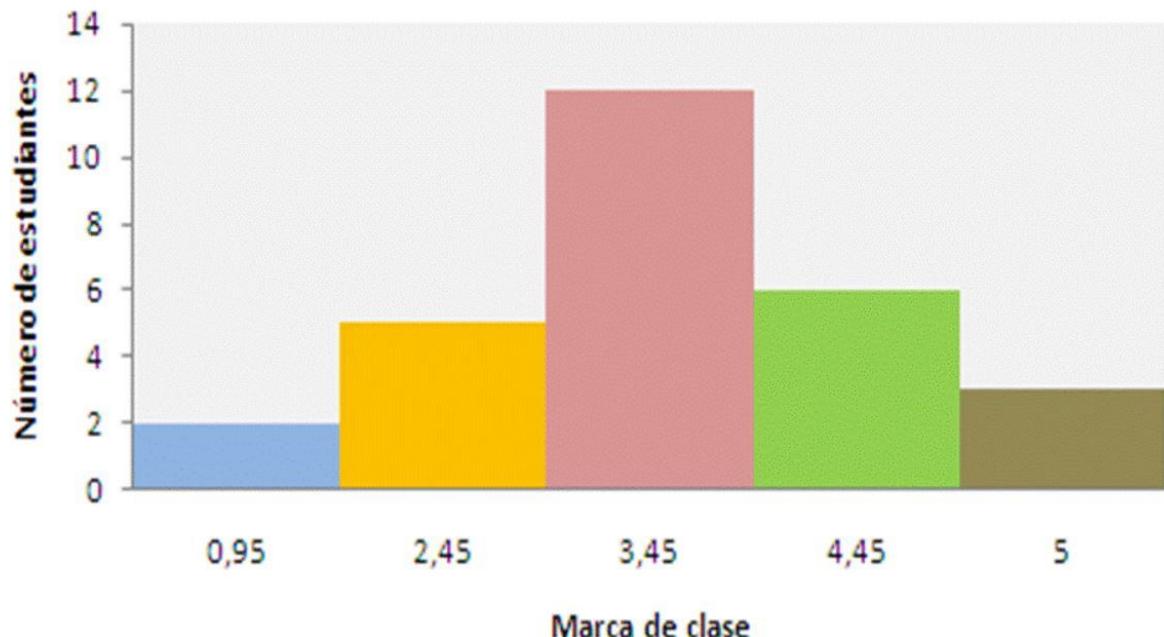


Figura 32: Histograma

Se realiza clasificando los datos en intervalos y, levantando sobre cada uno de ellos un rectángulo de altura la frecuencia absoluta o relativa. Si la longitud de todos los intervalos es la misma, el área de cada rectángulo también será igual a la frecuencia. También es importante la elección del número de intervalos, porque puede cambiar mucho el aspecto de un histograma y dar la información de manera distinta, por ejemplo, R utiliza $\log_2(n) + 1$. Para hacerlo en R, metemos los datos en un vector (esta vez sin nombres) y ejecutamos `hist(x)`.

1.5.5 Polígono de frecuencias

Se crea el **polígono de frecuencias** empezando en el eje X a una distancia igual a la mitad de un intervalo del histograma, se van uniendo los puntos medios de los rectángulos superiores del mismo hasta acabar otra vez a la misma distancia del histograma en el eje X . Esto se debe a que si sumamos el área creado debajo del polígono da uno si hablamos de frecuencias relativas, y n (n^o de datos) si hablamos de absolutas.

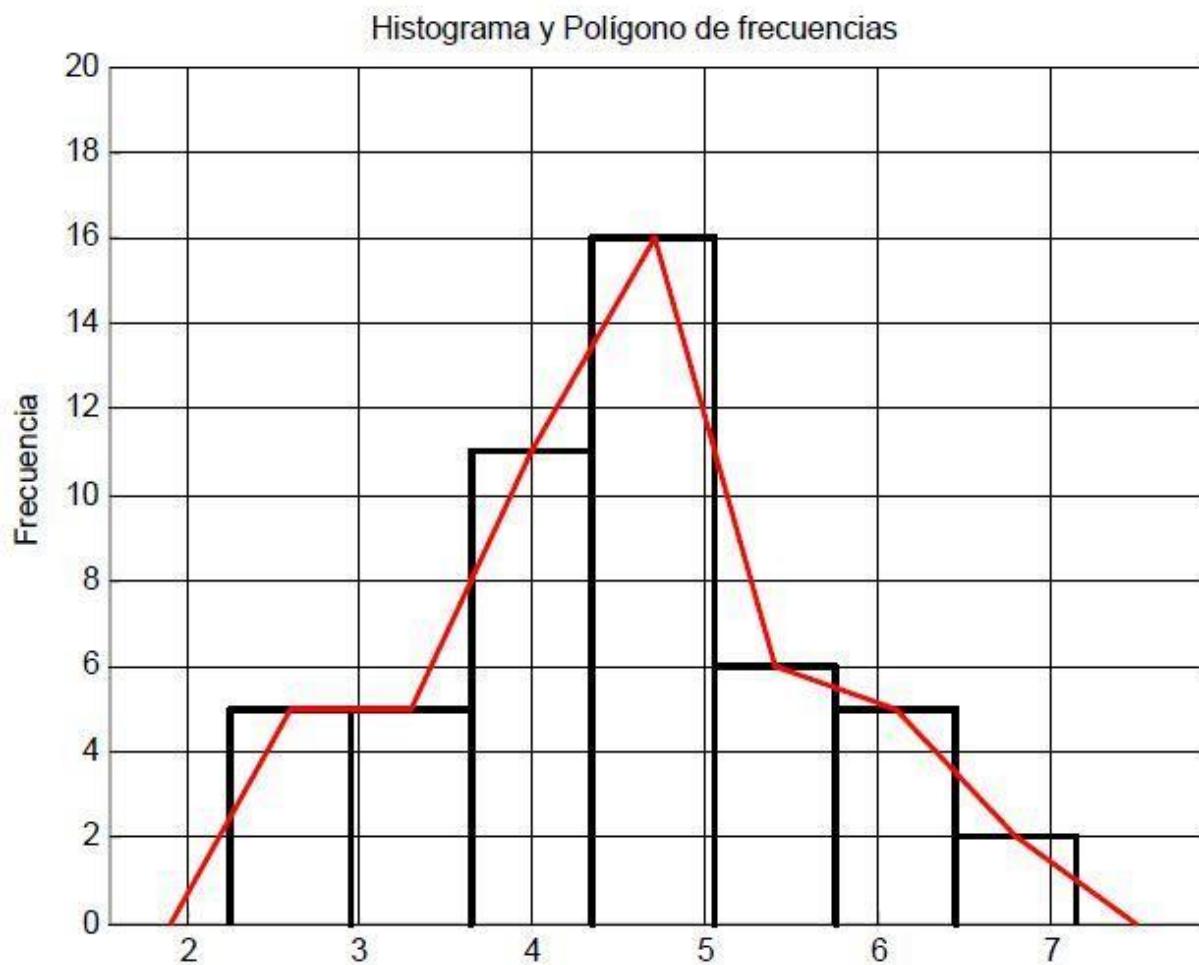


Figura 33: Polígono de frecuencias

1.5.6 Diagrama de tallo y hojas

Es un tipo de representación que ya casi está en desuso pero que fue muy utilizado en los años 80. Se utiliza para variables cuantitativas cuando tenemos pocos datos y es muy fácil de hacer a mano. Nos sirve para ver la distribución, el rango y la simetría de los datos. Se construye separando el último dígito de cada valor, que será la hoja, de los demás, que formarán el tallo. Dibujamos dos columnas, en la primera ponemos los tallos en orden ascendente y en la segunda las hojas que pertenecen a cada tallo de menor a mayor. Para hacerlo con R, solamente hay que ejecutar `stem(x)` siendo x el vector. Vamos a verlo gráficamente para que quede más claro.

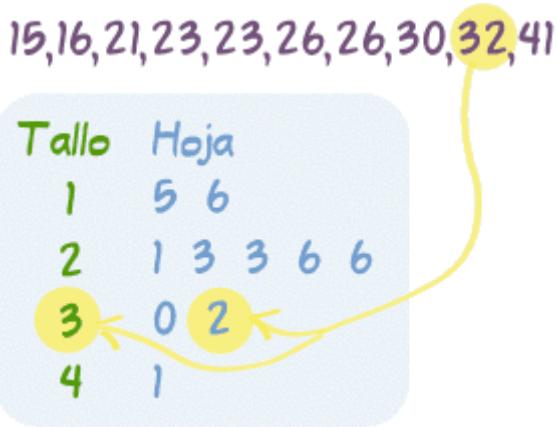


Figura 34: Diagrama de tallo y hojas

Rick: 🧑 ¿Así caeré mejor al diagrama de tallo y hojas? ¿Tú crees?

\$nombre: 💡💡 no puedo contigo de verdad, ¡estás fatal! 💡💡

1.5.7 Diagrama de caja

El **diagrama de caja** se utiliza para estudiar tanto la dispersión como la forma de una variable y compararla con otras. Para construirlo, dibujamos un eje X y de Q_1 a Q_3 dibujamos una caja, que recogerá el 50% de los datos. Sin atravesar la caja, unimos a través de una línea a límite inferior, $LI = Q_1 - 1.5 \times RIQ$, (recordemos que $RIQ = Q_3 - Q_1$) con el límite superior, $LS = Q_3 + 1.5 \times RIQ$, todos los datos que queden fuera de estos límites serán atípicos. En R ponemos `boxplot(x)` siendo x el vector.

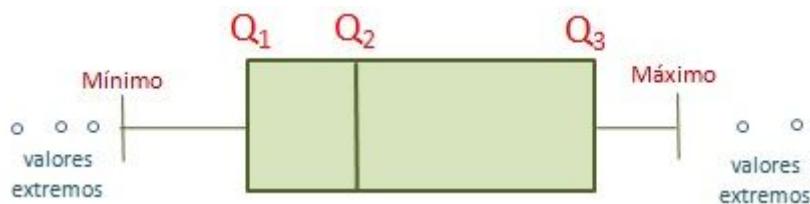


Figura 35: Diagrama de caja

Si tenemos más de una variable utilizamos un **diagrama de caja múltiple**.

Calidad del aire

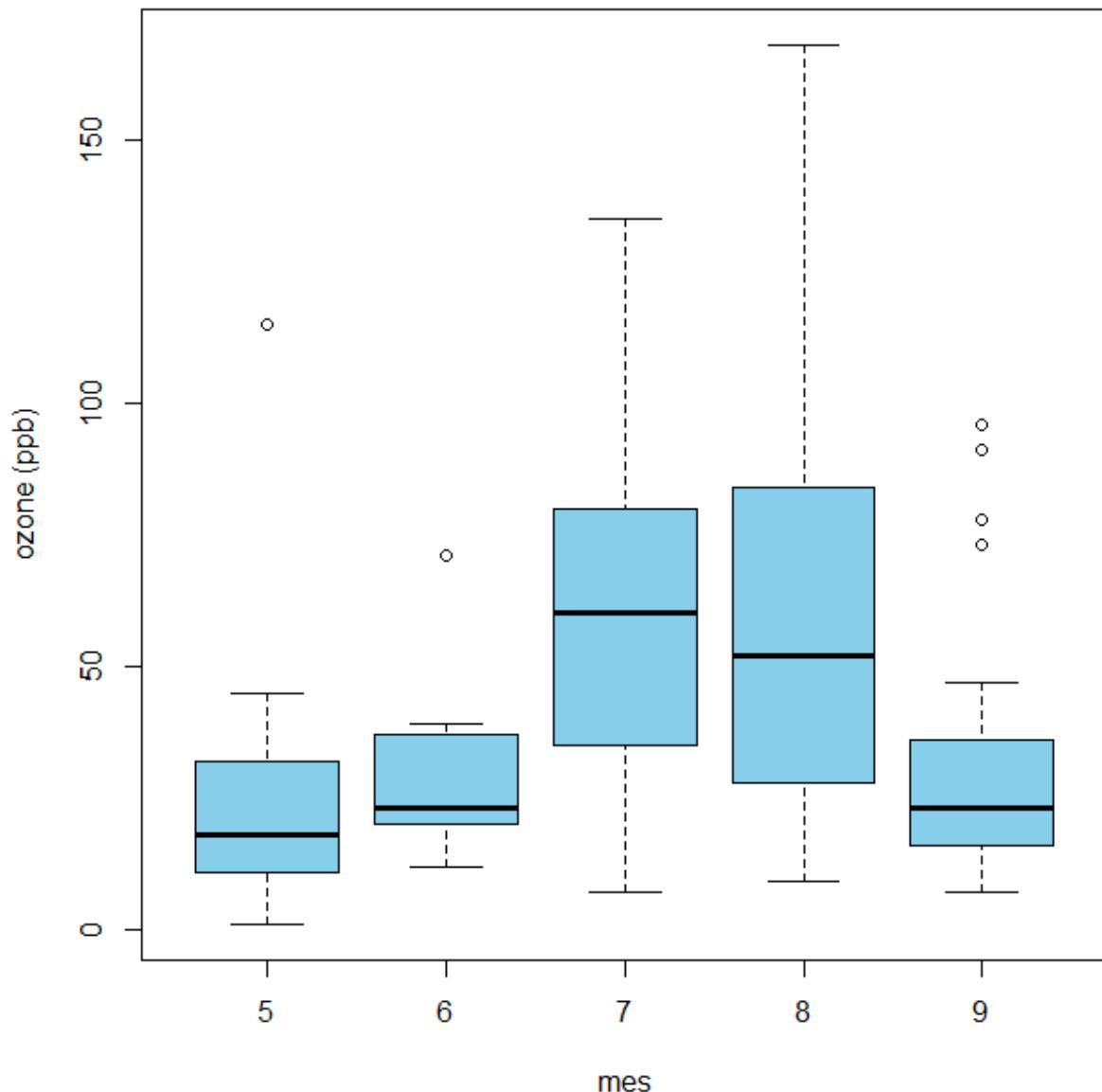


Figura 36: Diagrama de caja múltiple

Rick: Oye, \$nombre, ¡nos merecemos un trago!

\$nombre: ¡Pero si no has parado!

Rick: Lo que me queda por enseñarte...

2. Probabilidad

La **probabilidad** es la parte de la estadística que estudia la posibilidad de que ocurra un evento.

Rick: ¡Esto pinta  y interesante!

\$nombre: Si tú lo dices...

Rick: ¡Ya lo verás!

2.1 Experimento aleatorio

Un **experimento aleatorio** es aquel que cumple las siguientes propiedades:

1. Su resultado es imposible de saber antes de realizarse.
2. Siempre se puede realizar en condiciones idénticas.
3. Se conocen cuáles son los posibles resultados que pueden ocurrir.

Los posibles resultados de un experimento aleatorio reciben el nombre de **espacio muestral**, representado por Ω .

Por ejemplo, para el experimento aleatorio: "lanzamiento de un dado" Espacio muestral $\Omega=\{1,2,3,4,5,6\}$.

Rick: ¡Después de esto vamos a ser los reyes de los casinos del espacio! 

\$nombre: 

2.2 Suceso

Suceso: Es un subconjunto del espacio muestral, $A \subset \Omega$.

Suceso elemental: cada uno de los resultados de un experimento aleatorio.

Suceso imposible: suceso que nunca puede ocurrir, se denota por \emptyset (conjunto vacío).

Suceso seguro: Es otra forma de denominar al propio espacio muestral, E , que siempre ocurre.

Sucesos incompatibles: Aquellos que no se pueden dar simultáneamente.

En el ejemplo del lanzamiento de un fsfp, un suceso sería "salir número impar" = 1,3,5, un suceso elemental, 1, un suceso imposible 8, un suceso seguro: 1,2,3,4,5,6 y dos incompatibles serían "salir número par" y "salir número impar".

\$nombre: Mira, ¡otra vez!

Rick: Si sacas alguna conclusión, ya me dices 

2.2.1 Operaciones

Recordemos las operaciones de los conjuntos, ya que son las mismas:

Unión ($A \cup B$): suceso formado por los elementos que están en A o B , es decir, con que esté en uno de los dos ya cuenta.

Intersección ($A \cap B$): formado por los elementos que están en A y B simultáneamente (tiene que estar en los dos).

Diferencia ($A - B$): formado por los elementos que está en A pero no en B .

Complementario (\bar{A}): todos los elementos del espacio muestral que no están en A .

Rick: ¡Vamos a prepararnos! 🧑

\$nombre: 🎉

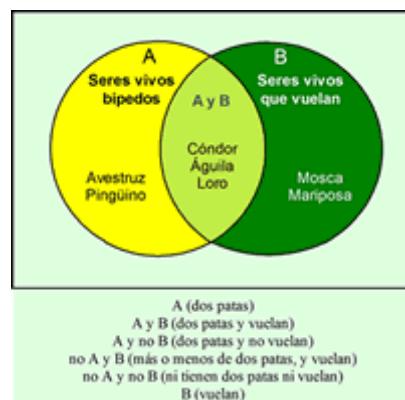


Figura 37: Operaciones de sucesos

2.2.1.1 Misión 11

Rick: Para ir practicando un poco para cuando vayamos al casino, vamos a tirar un dado...

\$nombre: ... y contemplamos el suceso $A = "salir un número par"$, es decir, $A = \{2, 4, 6\}$ y el suceso $B = "salir un número mayor que 3"$, es decir, $B = \{4, 5, 6\}...$

Rick: ... y calculamos su unión, intersección y diferencia (tanto $A - B$ como $B - A$). ¿Cuál es $A - B$?

Solución correcta 11:

¡Al-leluia!

Solución incorrecta 11:

Rick: ¡Nos vamos a arruinar!

\$nombre: Tenemos que practicar más....

Solución:

$$A \cup B = "El resultado es par o mayor que 3" = \{2, 4, 5, 6\}$$

$$A \cap B = "El resultado es par y mayor que 3" = \{4, 6\}$$

$$A - B = "El resultado es par pero no mayor que 3" = \{2\}$$

$$B - A = "El resultado es mayor que tres pero no par" = \{5\}$$

2.2.2 Sistema completo de sucesos

Un **sistema completo de sucesos** es un conjunto de sucesos que verifican:

1. Son incompatibles entre sí.
2. El total es la unión de todos ellos.
3. La probabilidad de todos es distinta de cero.

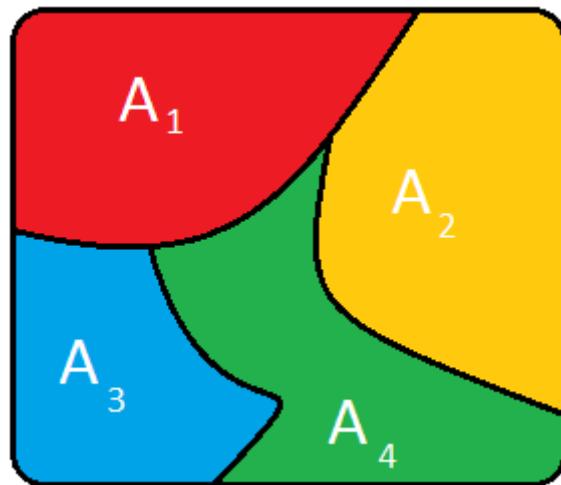


Figura 38: Sistema completo de sucesos

En cristiano, cogemos un espacio muestral y lo dividimos en n sucesos como si fueran piezas de un puzzle.

2.3 Variables aleatorias

Es una **variable** que asocia a cada resultado de un experimento aleatorio un número. Definimos **rango o soporte** de una variable aleatoria como todos sus posibles valores. Según su rango hay dos tipos de variable aleatoria.

2.3.1 Variable aleatoria discreta

Una bstosnñr aleatoria es **discreta** si su rango es un conjunto finito o numerable infinito de números. Un conjunto numerable infinito es el que está formado por números que no tienen valores entre ellos, como los naturales: 1,2,3.... hasta infinito.

\$nombre: ¡Hala, otra! ¡Qué intriga!

Rick: ¡Qué seráá, serááááá!

2.3.1.1 Función de probabilidad

Una variable aleatoria discreta queda definida por su **función de probabilidad** (también conocida como ley o distribución de probabilidad), por sus valores y las probabilidades de estos valores (suman 1):

$$X \quad P(X = x_i)$$

$$x_1 \quad p_1$$

...

$$x_k \quad p_k$$

donde $p_1 + \dots + p_k$.

Gráfica de función de probabilidad f(x)

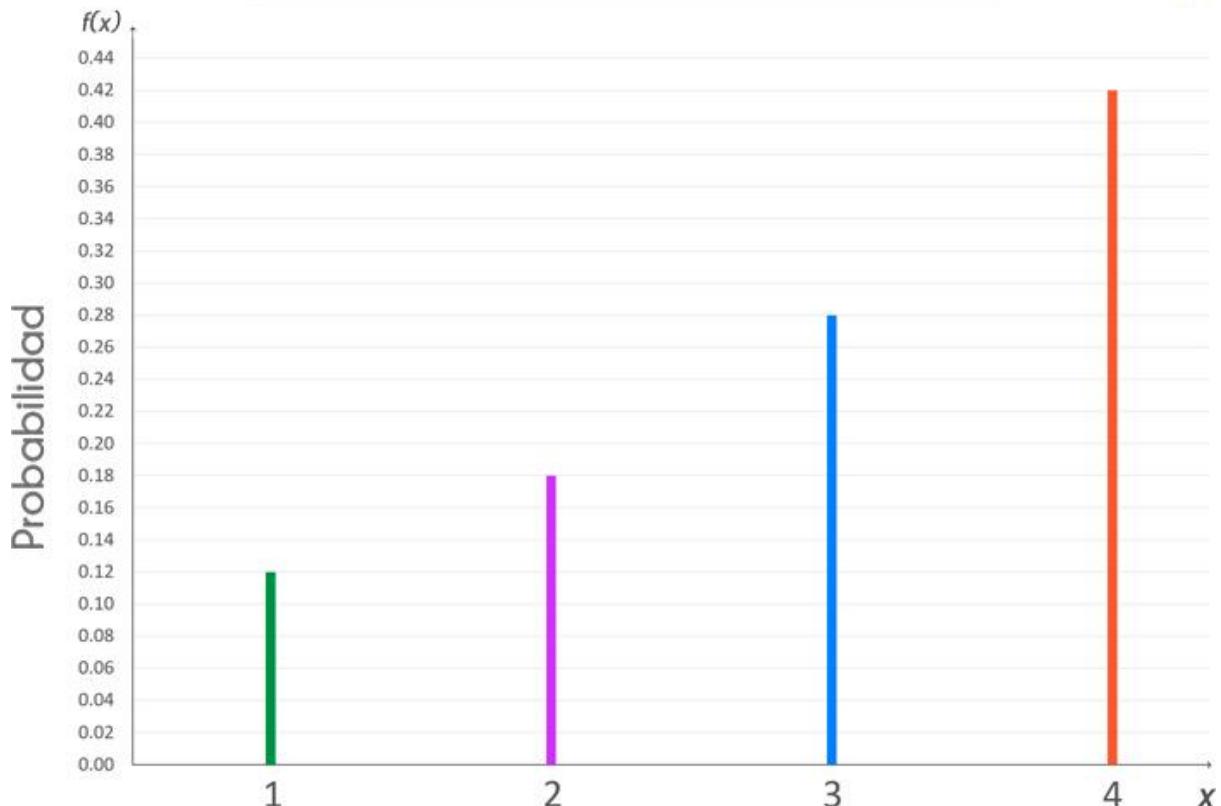


Figura 39: Función de probabilidad

2.3.1.2 Función de distribución discreta

La **función de distribución discreta** es la función que asocia a cada punto a la probabilidad acumulada hasta él. Por lo tanto, es una función no decreciente que varía entre 0 y 1.

$F(a) = P(X \leq a) = P(X = x_1) + \dots + P(X = x_i)$, donde todos los $x_i \leq a$.

Por lo que, matemáticamente se escribe:

$$\begin{aligned}
 0 & \quad \text{si } x < x_1 \\
 p_1 & \quad \text{si } x \in [x_1, x_2) \\
 p_1 + p_2 & \quad \text{si } x \in [x_2, x_3) \\
 \dots & \\
 p_1 + \dots + p_i & \quad \text{si } x \in [x_i, x_{i+1}) \\
 \dots & \\
 1 & \quad \text{si. } x \geq x_n
 \end{aligned}$$

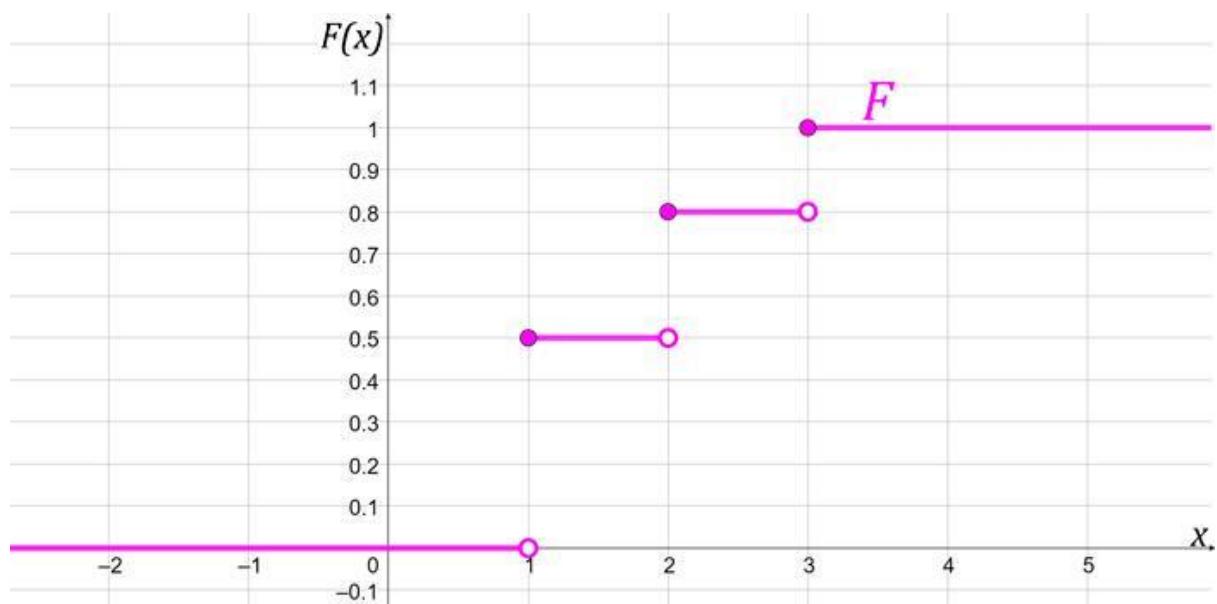


Figura 40: Función de distribución discreta

\$nombre: Ya empezamos a funcionar 😊

Rick: 🙋

2.3.2 Variable aleatoria continua

Una variable aleatoria es **continua** si su rango es un intervalo de números reales, es decir, puede tomar cualquier valor entre 2 consecutivos.

Rick: ¡Continúa la fiesta!

\$nombre: 🙋

2.3.2.1 Función de densidad

La **función de densidad** surge de la generalización del polígono de frecuencias y es una generalización de la **función de probabilidad**. Sus propiedades son:

1. En todo momento es mayor a 0 → $f(x) \geq 0$

$$f(x) \geq 0$$

Ecuación 6: La función es mayor o igual que 0

2. El área bajo la curva es 1.

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Ecuación 7: El área bajo la curva es 0

3. El área entre dos puntos es su probabilidad.

$$\Pr(a \leq X \leq b) = \int_a^b f(x) dx = F(b) - F(a)$$

Ecuación 8: El área entre dos puntos es su probabilidad

4. La probabilidad de un punto es 0.

Por lo que $P(a < X < b) = P(a \leq X < b) = P(a \leq X \leq b) = P(a < X \leq b)$.

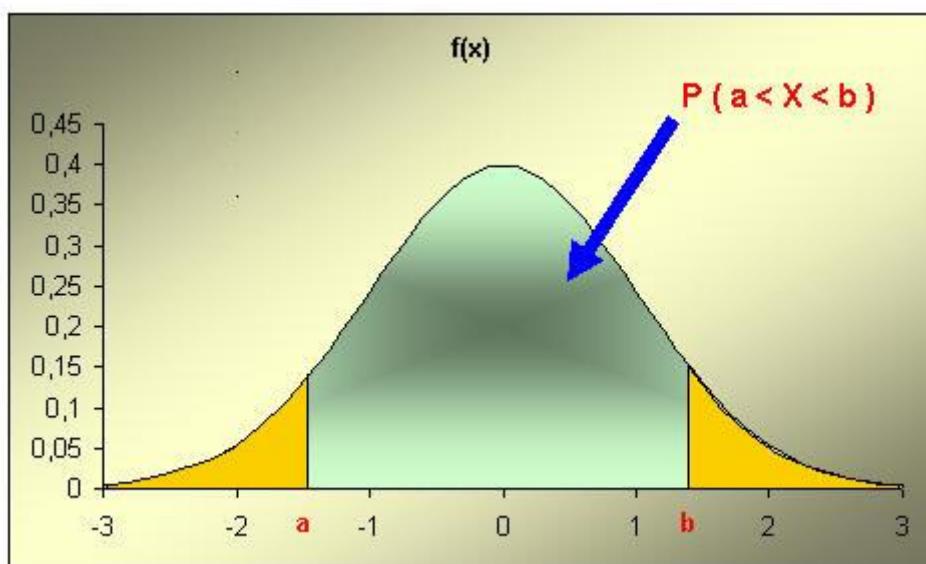


Figura 41: Función de densidad

2.3.2.2 Función de distribución continua

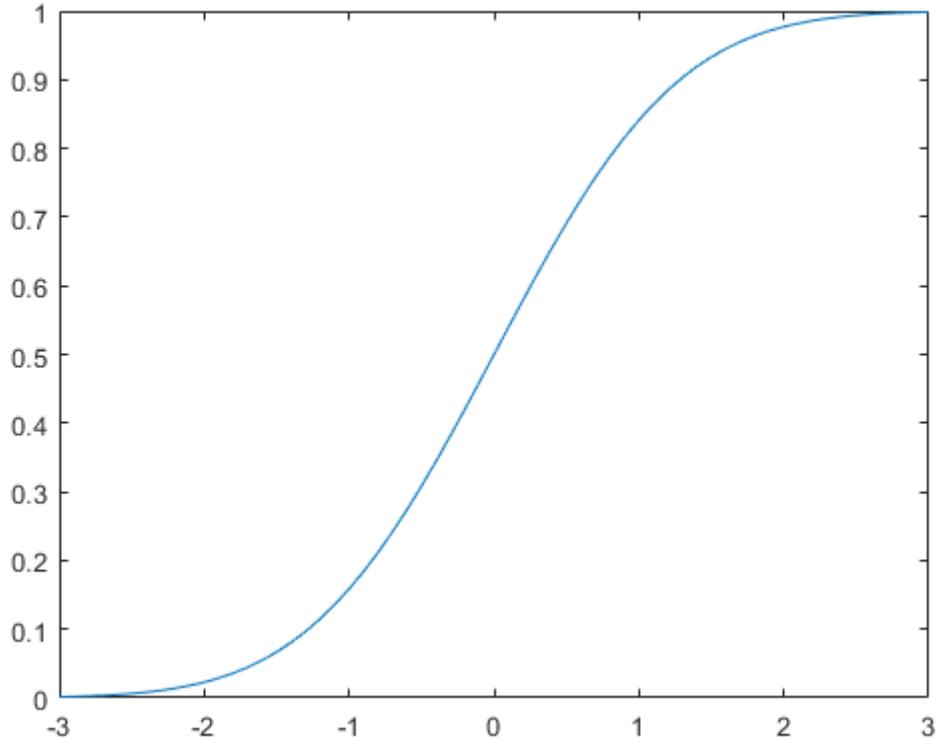


Figura 42: Función de distribución continua

Tiene el mismo significado que en el caso de una variable aleatoria discreta. En este caso será la integral, ya que la probabilidad acumulada es el área debajo de la curva hasta ese punto a .

$$P(X \leq a) = P(X < a) = F(a) = \int_{-\infty}^a f(x) dx$$

Ecuación 9: Probabilidad de que X sea menor o igual a a

Se calcula como la **integral de la función de densidad**. Por eso mismo su derivada es igual a la función de densidad.

También, sacamos la siguiente conclusión:

$$P(X \geq a) = P(X > a) = 1 - F(a) = \int_a^{+\infty} f(x) dx$$

Ecuación 10: Probabilidad de que X sea mayor o igual a a

2.3.3 Medidas de una variable aleatoria

Una variable aleatoria tiene medidas de posición, dispersión y forma al igual que las variables estadísticas y sus significados coinciden. Veremos los más utilizados.

2.3.3.1 Esperanza

La **esperanza** es la generalización de la media aritmética y se presenta con μ , por lo que tiene sus mismas propiedades. Por resumirlo:

$$Y = aX + b \rightarrow E(Y) = aE(X) + b.$$

Con variables independientes: $E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n)$.

Si X es una variable aleatoria discreta será la suma de sus valores por sus probabilidades (con la media hacíamos lo mismo, pero suponíamos que todas tenían la misma probabilidad, $1/n$).

\$nombre: La esperanza es lo último que se pierde.

Rick: 😊

$$\mathbb{E}[X] = x_1 p(X = x_1) + \dots + x_n p(X = x_n) = \sum_{i=1}^n x_i p(x_i)$$

Ecuación 11: Esperanza de una variable aleatoria discreta

Si X es una variable aleatoria continua, la esperanza será la integral de todos los valores y la función de densidad.

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx$$

Ecuación 12: Esperanza de una variable aleatoria continua

2.3.3.2 Varianza

La **varianza** mide la dispersión de la variable, se representa por σ^2 y su raíz cuadrada positiva es la **desviación típica**, al igual que con las variables estadísticas, por ello, también tiene sus mismas propiedades. Por resumirlas:

$$Y = aX + b \rightarrow \text{Var}(Y) = a^2 \cdot \text{Var}(X).$$

Con variables independientes: $\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)$.

Rick: Seguimos desvariando.

\$nombre: 😊

Si X es una variable aleatoria discreta, su varianza será:

$\sigma^2 = (x_1^2 \cdot p_1 + \dots + x_n^2 \cdot p_n) - \mu^2$. Lo que es igual a:

$$\sigma_X^2 = \left(\sum_{i=1}^n p_i \cdot (x_i - \mu)^2 \right)$$

Ecuación 13: Varianza de una variable aleatoria discreta

Si X es una variable aleatoria continua, $\sigma^2 = \int x^2 f(x) dx - \mu^2$. Lo que es igual a:

$$\sigma_X^2 = \int (x - \mu)^2 f(x) dx$$

Ecuación 14: Varianza de una variable aleatoria continua

2.3.3.3 Mediana, cuantil y moda

Son lo mismo que para las variables estadísticas, así que solo refrescaremos un poco la memoria.

La **mediana** es el valor que deja a cada lado el 50% de la población.

La generalización de la mediana son los **cuantiles**, que dado un valor p (toma valores entre 0 y 1), el cuantil x_p es el punto que deja una probabilidad p a su izquierda. Recordemos a los cuartiles, deciles y percentiles.

La **moda** es el valor más probable, por lo que, será el valor x_i con mayor p_i en el caso discreto, o el que maximiza la función de densidad en el continuo. Si solo tiene una moda (absoluta) será unimodal, si tiene varias (además tiene modas relativas) será multimodal.

Rick: ¿No se te ocurre ninguna chorrada que decir?

\$nombre: No, ¿y a ti?

Rick: Tampoco, ¡qué desaboríos estamos por dió!

2.4 Propiedades

A continuación vamos a ver las propiedades de la probabilidad.

2.4.1 Axiomas de Kolmogorov

Los **axiomas de Kolmogorov** son tres propiedades que verifica la probabilidad, función que asigna a cada suceso un número que conoceremos como probabilidad de A , $P(A)$:

1. $0 \leq P(A) \leq 1$, la probabilidad de A siempre está entre 0 y 1.
2. $P(\Omega) = 1$, la probabilidad del espacio muestral es 1.
3. Si tenemos n sucesos incompatibles, la probabilidad de sus uniones es la suma de sus probabilidades. $P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n)$

\$nombre: No podrían haberle puesto un nombre más difícil, ¿no?

Rick: ¡Ya te digo!

2.4.2 Resultados

De los axiomas de Kolmogorov sacamos los siguientes resultados:

P1) Probabilidad del complementario: $P(\bar{A}) = 1 - P(A)$.

P2) Probabilidad del suceso imposible: $P(\emptyset) = 0$.

P3) Si $A \subset B$, entonces $P(A) \leq P(B)$.

P4) Probabilidad de la unión de sucesos: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Si son incompatibles simplemente es $P(A \cup B) = P(A) + P(B)$. Si generalizamos $P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n) - P(A_1 \cap A_2) - \dots - P(A_{n-1} \cap A_n) + P(A_1 \cap A_2 \cap A_3) + \dots + P(A_{n-2} \cap A_{n-1} \cap A_n) - \dots + (-1)^{n+1} P(A_1 \cap \dots \cap A_n)$.

P5) $P(A - B) = P(A) - P(A \cap B)$.

- Además, al cumplirse las **leyes de Morgan**, que dicen:

a) el complementario de la unión es la intersección de los complementarios:

$$\overline{A_1 \cup \dots \cup A_n} = \bar{A}_1 \cap \dots \cap \bar{A}_n,$$

b) el complementario de la intersección es la unión de los complementarios:

$$\overline{A_1 \cap A_2 \cap \dots \cap A_n} = \bar{A}_1 \cup \dots \cup \bar{A}_n,$$

podremos utilizar también que

P6) $P(\bar{A}_1 \cap \dots \cap \bar{A}_n) = 1 - P(A_1 \cup A_2 \cup \dots \cup A_n)$ y $P(\bar{A}_1 \cup \dots \cup \bar{A}_n) = 1 - P(A_1 \cap A_2 \cap \dots \cap A_n)$.

Figura 43: Leyes de Morgan

2.5 Regla de Laplace

Cuando todos los elementos de un espacio muestral tienen la misma probabilidad de ocurrir se utiliza la **regla de Laplace** para calcular la probabilidad de que ocurra un suceso A , que consiste en dividir el número de resultados del suceso A (se conocen como casos favorables) entre el total de resultados.

$$P(A) = \frac{\text{casos favorables}}{\text{casos posibles}}$$

Ecuación 15: Regla de Laplace

2.5.1 Misión 12

En la boda de persona pájaro y Tammy se sortea un viaje al Roscacosca entre los 120 asistentes a la misma. Teniendo en cuenta que 65 de ellos tiene más de 2 ojos, 80 tienen más de 3 piernas y 45 tienen más de dos ojos y de tres piernas. ¿Cuál es la probabilidad de que le toque el viaje a un ser con 2 o menos ojos y con 3 o menos piernas?

Solución correcta 12:

אַלְעָלוּאִיא

Solución	incorrecta	12:	
Rick: Vamos a tener que prestar más atención...	XD		
\$nombre: Pues sí, porque no quiero seguir haciendo el ridículo...	WTF		
Solución:			
Con los datos que nos han dado y rellenando los que faltan hacemos una tabla:			
	Piernas > 3	Piernas ≤ 3	Total
Ojos ≤ 2	35	20	55
Ojos > 2	45	20	65
Total	80	40	120

Figura 44: Misión 12

Por la regla de Laplace sabemos que la probabilidad de un suceso es el número de veces que ocurre el mismo entre el total, por lo que, $20/120 = 1/6 = 0.167$.

2.6 Probabilidad condicionada

La probabilidad de A condicionada a B , es decir, la **probabilidad de A sabiendo que B ha ocurrido**, se calcula a partir de:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Ecuación 16: Probabilidad condicionada

Vamos a pensar un momento, en realidad, esta fórmula tiene mucho sentido porque es lo mismo que la regla de Laplace pero incorporando la información de que B ha ocurrido.

El denominador es B , que ahora mismo es el espacio muestral (ya que ha ocurrido) y el numerador es $A \cap B$ porque son los casos favorables para el suceso A ya que todo lo que no coincide con B será un suceso imposible.

La probabilidad condicionada tiene las mismas propiedades que la probabilidad, por ejemplo, $P(\bar{A}/B) = 1 - P(A/B)$.

Rick: Todo en esta vida está condicionado por algo.

\$nombre: Todo no Rick... todo no...

2.6.1 Misión 13

\$nombre: Antes de todo esto no pensaba que pudiera ser real lo que estoy viviendo.

Rick: En realidad, casi nadie de la población cree en este tipo de cosas. Una vez leí que el 12% de los habitantes de nuestro planeta creen en el mundo de la fantasía (que existen las brujas, los vampiros, la magia...) y solamente un 25% de ellos creen en la existencia de los dragones.

\$nombre: Oye, pues, ¿qué porcentaje de habitantes de nuestro planeta cree en el mundo de la fantasía pero no en la existencia de dragones?

Solución

correcta

13:

¡Apliuya!

Solución incorrecta 13:

Rick: Todo es posible en el mundo del espacio...

\$nombre: Como equivocarnos en esto...

Solución:

Consideramos los sucesos $F = "creer en el mundo de la fantasía"$ y $D = "creer en los dragones"$. Tenemos que $P(F) = 0.12$ y $P(D/F) = 0.25$. Lo que nos piden es:

$$P(F \cap \bar{D}) = P(F)P(\bar{D}/F) = P(F)(1 - P(D/F)) = 0.12(1 - 0.25) = 0.09$$

Por lo que el **9%** de los habitantes de nuestro planeta cree en el mundo de la fantasía pero no en los dragones.

2.7 Regla del producto

De la probabilidad condicionada sacamos la **probabilidad de la intersección de dos sucesos A y B**, que es la probabilidad de que A y B ocurran simultáneamente:

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A).$$

Si la generalizamos, obtenemos la **regla del producto** que nos dice que la intersección de n sucesos es igual a la multiplicación de las probabilidades de todos los sucesos condicionados a que ocurran los sucesos inmediatamente anteriores:

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) \times \dots \times P(A_n/A_1 \cap \dots \cap A_{n-1})$$

Rick: Como decía mi abuela, mejor multiplicar y sumar que dividir y restar.

\$nombre: Bueno... en este caso es peor multiplicar que dividir 😕

2.7.1 Misión 14

\$nombre: Oye Rick, ¿Sabes que en mi clase había dos personas que cumplían años el mismo día?

Rick: Anda tú, ¡mira qué coincidencia!

\$nombre: Estoy pensando que no es una coincidencia... ¿Por qué no calculamos la probabilidad de que eso pase?

Rick: Venga va, ¿cuántos erais en clase?

\$nombre: Éramos 23 y vamos a suponer que nacimos en un año no bisiesto por hacerlo más general, ¿cuál es la probabilidad?

Solución correcta 14:

¡aplliluyá!

\$nombre: Guau, pensaba que era mucho más pequeña...

Solución incorrecta 14:

\$nombre: Es mucho más creíble la que yo creía que la que es...

Solución:

En vez de calcular el suceso "*al menos dos personas en un grupo de 23 cumplen años el mismo día*" calcularemos el suceso opuesto "*las 23 personas cumplen años en días diferentes*", que es más fácil de calcular y se lo restaremos a uno, consiguiendo la probabilidad que buscamos. En este caso A_i es el suceso, "*el número sacado en el intento i es distinto a los anteriores*". La primera vez que sacamos un número, la probabilidad de que sea distinto a los anteriores es, 365 números que podemos sacar entre 365 números posibles, así que es 1, ya que es el primero que sacamos, ya tenemos $P(A_1)$. Para que el siguiente número que saquemos sea distinto al anterior tenemos 364 posibilidades (ya que una de ellas ya ha sido cogida por el primero) entre 365 números posibles, que es $P(A_2/A_1)$. La probabilidad de sacar el siguiente número sin que sea igual a ninguno de los anteriores es

363/365, que es $P(A_3 | A_1 \cap A_2)$ ya que es la probabilidad de que el número sacado en el intento 3 sea distinto al 2 y al 1 teniendo en cuenta/condicionado a que esos no son iguales a los anteriores a ellos. Para calcular cada una de estas probabilidades hemos utilizado la regla de Laplace, casos favorables entre casos totales. Si seguimos así hasta el suceso "el número sacado en el intento 23 es distinto a los anteriores" tendremos: $P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2/A_1)P(A_3/A_2 \cap A_1) \times \dots \times P(A_n/A_1 \cap \dots \cap A_{n-1}) = 365/365 \times 364/365 \times 363/365 \times \dots \times 343/365 = 0.493$ que es la probabilidad de que los 23 números sean diferentes, por lo que, $1 - 0.493 = 0.507$ será la probabilidad de que en un grupo de 23 personas al menos dos de ellas cumplan años el mismo día. Si lo hicieramos en un grupo de 50 personas la probabilidad sería 0.97 y a partir de 60 la probabilidad es mayor a 0.99, esto se conoce como la paradoja del cumpleaños ya que sorprende a mucha gente porque no te esperas que vaya a haber una probabilidad tan alta, pero, ¡cuidado! esto no es lo mismo a que en un grupo de 23 personas haya una persona con tu mismo cumpleaños, esa probabilidad es $1 - (364/365)^{23} = 0.061$, que si es baja.

2.8 Independencia de sucesos

Dos sucesos A y B son **independientes** cuando la ocurrencia de uno de ellos no influye en la ocurrencia del otro, por lo que se verifica que

$P(A/B) = P(A)$ y $P(B/A) = P(B)$ y que $P(A_1 \cap \dots \cap A_n) = P(A_1) \times \dots \times P(A_n)$, lo que quiere decir que la intersección de n sucesos independientes es igual a la multiplicación de sus probabilidades.

Rick: ¡Bendita independencia!

2.8.1 Misión 15

Morty: Rick, \$nombre, necesito vuestra ayuda, tengo un examen de estadística en 10 minutos y necesito aprobarlo.

Rick: Podríamos volver a meterle la idea en un sueño a Goldenfold, pero sueña cosas tan raras que paso.

\$nombre: Otra posibilidad es que Morty estudie su examen en un sueño, ya que el tiempo va más lento y le podría dar tiempo.

Morty: Tened en cuenta que al estar soñando, tengo una probabilidad 3/4 de pensar en Jessica y distraerme.

Rick: Y que las pastillas que tengo para tener sueños lúcidos fallan el 1% de las veces.

\$nombre: Y que hay un 1/5 de probabilidad de que no encuentres el libro de estadística...

Morty: Vale, entonces, ¿cuál es la probabilidad de que tenga un sueño lúcido, encuentre el libro y me concentre?

Solución correcta 15:

¡aleluya!

Morty: Jo tío...

Solución

incorrecta

15:

\$nombre: Ya nos hemos confundido...

Morty: Prefiero creer esta probabilidad que la real...

Solución:

Consideramos los sucesos F = "las pastillas funcionan", L = "Morty encuentra el libro" y C = "Morty se concentra".

La probabilidad de que las pastillas funcionen es $P(F) = 1 - 0.01 = 0.99$.

La probabilidad de que Morty encuentre el libro es $P(L) = 1 - 1/5 = 4/5$.

La probabilidad de que Morty esté concentrado es $P(C) = 1 - 3/4 = 1/4$.

Como estos tres sucesos son independientes (ya que la efectividad de una pastilla no depende de la concentración de Morty ni de que encuentre el libro y así) la probabilidad de que Morty tenga un sueño lúcido, encuentre el libro y se concentre es: $P(F \cap L \cap C) = P(F) \times P(L) \times P(C) = 0.99 \times 4/5 \times 1/4 = 0.198$.

2.9 Teoremas

Veremos dos teoremas de la probabilidad.

2.9.1 Teorema de las probabilidades totales

Considerando un sistema completo de sucesos y otro suceso cualquiera B , su probabilidad será:

Teorema de la Probabilidad Total

$$P(B) = P(A_1).P(B|A_1) + P(A_2).P(B|A_2) + \dots + P(A_n).P(B|A_n) = \sum_{i=1}^n P(A_i).P(B|A_i)$$

Ecuación 17: Teorema de las probabilidades totales

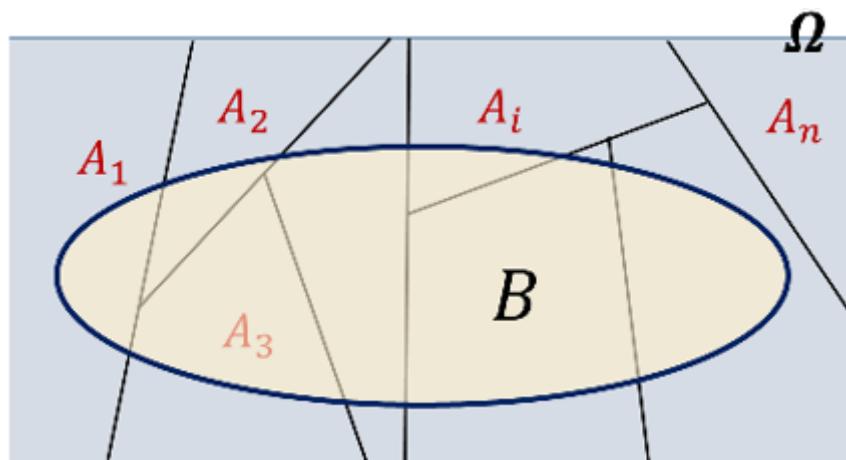


Figura 45: Teorema de las probabilidades totales

Para entender esta fórmula, nos fijamos en la imagen y vemos que B es la unión de las intersecciones de B con cada uno de los sucesos A_i , por lo que su probabilidad será la suma de todos los $P(B \cap A_i)$ y recordamos su fórmula $P(A_i)P(B/A_i)$.

2.9.1.1 Misión 16

Summer susurrando: \$nombre, despiértate, \$nombre, des-piér-ta-tee...
\$nombre balbuceando: ¿Qué pasa?

Summer susurrando: Shhh... Sepárate de Rick... Sí, ahí debería bastar.
\$nombre: ¿Se puede saber qué pasa?

Summer: Pues es que necesito que me ayudes con una cosa y Rick no se puede enterar, por lo que he aprovechado cuando os he visto dormidos. A ver, te cuento, creo que estoy embarazada...

\$nombre interrumpiendo a Summer: ¿qué, qué?

Summer: ¡Déjame hablar! A ver, la cuestión es que en el último mes he tenido encuentros con varios chicos, el 30% de las veces han sido con Abradolf Lincler, el 60% de las veces con Ethan y el 10% restante con Toby y tengo un 90% de probabilidades de quedarme embarazada de Abradolf Lincler, un 75% de Ethan y un 50% de Toby. Así que, ¿cuál es la probabilidad de que esté embarazada?

Solución correcta 16:

¡Areruia!

\$nombre: La que has liao pollito 🐔

Summer: Di que sí, tú encima con cachondeo 😞

Solución incorrecta 16:

Summer: Encima dándome falsas esperanzas... 😞

\$nombre: Perdóname 🙏

Solución:

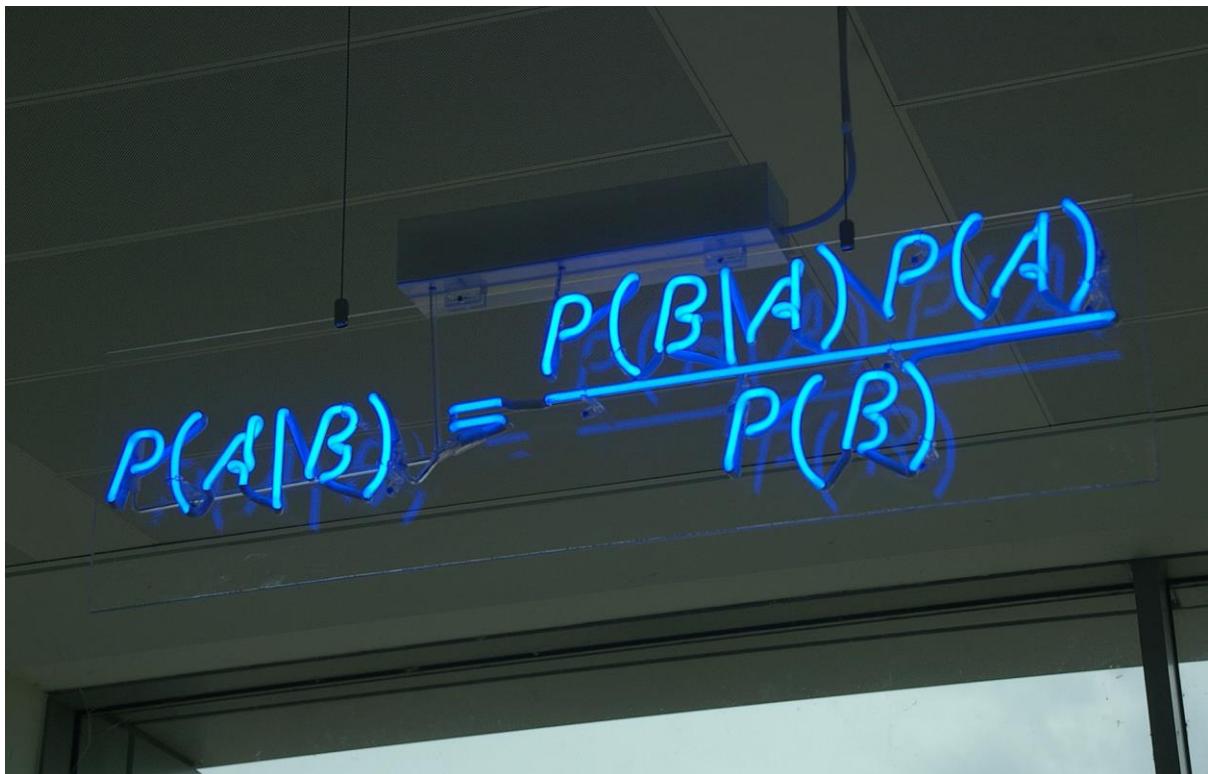
Primero definimos los sucesos $A_1 = "encuentros\ con\ Abradolf\ Lincler"$, $A_2 = "encuentros\ con\ Ethan"$, $A_3 = "encuentros\ con\ Toby"$ y $B = "quedarse\ embarazada"$. Por el enunciado sabemos que $P(A_1) = 0.3$, $P(A_2) = 0.6$ y $P(A_3) = 0.1$, como sus probabilidades suman 1 y son sucesos incompatibles forman un sistema completo de sucesos. También sabemos que $P(B/A_1) = 0.9$, $P(B/A_2) = 0.75$ y $P(B/A_3) = 0.5$.

Queremos saber cuál es la probabilidad de que Summer esté embarazada, por lo que:

$P(B) = P(B/A_1)P(A_1) + P(B/A_2)P(A_2) + P(B/A_3)P(A_3) = 0.9 \times 0.3 + 0.75 \times 0.6 + 0.5 \times 0.1 = 0.77$ es la probabilidad de que Summer esté embarazada.

2.9.2 Teorema de Bayes

De la fórmula de la probabilidad condicionada, si sustituimos el numerador $P(A \cap B)$ por su fórmula $P(B/A)P(A)$ tenemos la **regla de Bayes**:



Ecuación 18: Regla de Bayes

\$nombre: Parece como el cartel fluorescente de alguna tienda... 😐

Rick: Cuenta la leyenda que hay un planeta de la estadística donde sus tiendas tienen letreros así.

\$nombre: ¡En cuanto ganemos el concurso vamos!

Rick: ¡Hecho!

La probabilidad $P(A)$ se conoce como *probabilidad "a priori"* y la probabilidad $P(A/B)$ se llama *probabilidad "a posteriori"*.

Si ahora el denominador, $P(B)$, lo sustituimos por la fórmula de la probabilidad total, tenemos el **teorema de Bayes** (que en realidad fue descubierto por Laplace) siendo $A_1 \dots A_n$ un sistema completo de sucesos y B un suceso cualquiera. Las probabilidades $P(A_i)$ son conocidas como *probabilidades "a priori"* y las probabilidades $P(A_i/B)$ como *probabilidades "a posteriori"*.

$$P[A_n/B] = \frac{P[B/A_n] \cdot P[A_n]}{\sum P[B/A_i] \cdot P[A_i]}$$

Ecuación 19: Teorema de Bayes

2.9.2.1 Misión 17

Summer susurrando: \$nombre, despiértate, \$nombre, des-piér-ta-tee...
\$nombre balbuceando: ¿Otra vez tú?

Summer susurrando: Shhh... Sepárate de Rick... Sí, ahí debería bastar.

\$nombre: ¿Se puede saber qué pasa ahora Summer?

Summer: Te acuerdas que te dije que es probable que esté embarazada (recordamos los datos de la misión 16: Summer cree que está embarazada y en el último mes ha tenido varios encuentros, el 30% de las veces con Abradolf Lincler, el 60% de las veces con Ethan y el 10% restante con Toby y tiene un 90% de probabilidades de quedarse embarazada de Abradolf Lincler, un 75% de Ethan y un 50% de Toby), pues ahora me gustaría saber cuál es la probabilidad de que sea de Abradolf Lincler.

\$nombre: ¡Ah, claro! Como Rick no puede ni ver a Abradolf Lincler, quieres saber cuál es la probabilidad de que el futuro bebé sea de él, porque suficiente disgusto ha tenido con que tu madre esté con Jerry, como ahora su nieta esté embarazada de su fallido experimento...

Summer: ¡Exacto! Por lo que necesito tu ayuda urgente, por favor, dime que es una probabilidad baja 🤦

Solución correcta 17:

¡byen fò!

Summer: Uff... Bueno es mejor de lo que pensaba, gracias, \$nombre. ¡No digas nada a Rick!

\$nombre: Tranqui Summer, soy una tumba 😊

Solución incorrecta 17:

Summer: Menos mal que esta no es la realidad, espero que seas mejor guardando secretos que calculando probabilidades.

\$nombre: Encima de que te intento ayudar 😞

Solución:

(Aunque ya teníamos definidos los sucesos y el denominador de la anterior misión, vamos a repetir todo para refrescarlo) Primero definimos los sucesos $A_1 = \text{"encuentros con Abradolf Lincler"}$, $A_2 = \text{"encuentros con Ethan"}$, $A_3 = \text{"encuentros con Toby"}$ y $B = \text{"quedarse embarazada"}$. Por el enunciado sabemos que $P(A_1) = 0.3$, $P(A_2) = 0.6$ y $P(A_3) = 0.1$, como sus probabilidades suman 1 y son sucesos incompatibles forman un sistema completo de sucesos. También sabemos que $P(B/A_1) = 0.9$, $P(B/A_2) = 0.75$ y $P(B/A_3) = 0.5$.

Queremos saber cuál es la probabilidad de que esté embarazada de Abradolf Lincler, por lo que:

$$P\left(\frac{A_1}{B}\right) = \frac{P\left(\frac{B}{A_1}\right)P(A_1)}{P\left(\frac{B}{A_1}\right)P(A_1) + P\left(\frac{B}{A_2}\right)P(A_2) + P\left(\frac{B}{A_3}\right)P(A_3)} = \frac{0.9 \times 0.3}{0.9 \times 0.3 + 0.75 \times 0.6 + 0.5 \times 0.1} = 0.35$$

Por lo que hay una probabilidad de **0.35** que de Summer esté embarazada de Abradolf Lincler.

3. Modelos probabilísticos

En todo este tiempo, hemos podido observar que los procesos se rigen por una serie de patrones, estos patrones siguen unos modelos que son los llamados **modelos probabilísticos**. Estos modelos probabilísticos sirven para explicar la variabilidad de los distintos comportamientos de estos procesos, preverlos y poder calcular fácilmente sus probabilidades y medidas.

Voz: ¡Atención a todos los participante! Para esta parte del concurso tenemos a unas invitadas muy importantes: las DISTRIBUCIONES, han venido aquí desde el mundo de la estadística para explicarnos qué son, esperamos que sea de ayuda, y recordad, ¡que la probabilidad esté siempre de vuestra parte!

3.1 Modelos discretos

Los **modelos discretos**, son modelos probabilísticos de variables aleatorias discretas.

3.1.1 Bernoulli

Bernoulli: Una variable X sigue un modelo de **Bernoulli** cuando solo puede tomar dos posibles resultados, éxito o fracaso, que tendrán una probabilidad p y q (que es igual a $1 - p$), respectivamente. Por ejemplo, la variable toma valor 1 si sale cara al lanzar una moneda, y 0 si sale cruz.

Se representa como $X \rightarrow B(p)$, X puede tomar los valores 0 ó 1 y su función de probabilidad es:

$$\begin{array}{ll} X & P(X = x_i) \\ 0 & q \\ 1 & p \end{array}$$

Ecuación 20: Bernoulli

Puesto de otra forma, $P(X = k) = p^k q^{1-k}$, donde x puede tomar los valores 0 ó 1.

Fácilmente podemos obtener su media y varianza:

$$E(X) = q \cdot 0 + p \cdot 1 = p.$$

$$Var(X) = E(X^2) - \mu^2 = (q \cdot 0^2 + p \cdot 1^2) - p^2 = p - p^2 = p \cdot (1 - p) = pq.$$

\$nombre: Que guay, Rick, que las propias distribuciones nos puedan explicar qué son.

Rick: La verdad es que sí, lo que me hace gracia, es que por lo menos esta, parece que no habla de sí misma, ¡habla todo el rato en tercera persona!

3.1.2. Binomial

Binomial: Si realizamos n experimentos de Bernoulli independientes entre sí, la variable X , número de éxitos en estos n experimentos, seguirá un modelo **Binomial**. $X = X_1 + \dots + X_n$, siendo X_i una variable de Bernoulli. Por ejemplo, la variable número de caras obtenidas al lanzar una moneda 10 veces.

Se representa con $X \rightarrow Bi(n, p)$, X puede tomar los valores $0 \dots n$ (son el número de éxitos que puede haber) y sus probabilidades vienen dadas por su función de probabilidad:

$$p(x = k) = \binom{n}{k} p^k \cdot q^{n-k}$$
$$\binom{n}{k} = \frac{n!}{k!(n - k)!}$$

Ecuación 21: Binomial

Tiene sentido, ya que queremos saber cuál es la probabilidad de que haya k éxitos, eso quiere decir que habrá $n - k$ fracasos y, como son independientes, su probabilidad es la multiplicación: $p^k q^{(n-k)}$, pero claro, tenemos que multiplicarla por todas las posibles formas de combinar k éxitos y $n - k$ fracasos.

Como en variables independientes la media de una suma es la suma de sus medias y la varianza es la suma de varianzas, la media y la varianza de una variable Binomial son:

$$E(X) = np.$$

$$Var(X) = npq.$$

Esta variable tiene la propiedad aditiva por lo que es reproductiva, pero sólo respecto al parámetro n . Esto quiere decir que si tenemos dos variables $X \rightarrow B(n_1, p)$ e $Y \rightarrow B(n_2, p)$, su suma $X + Y \rightarrow B(n_1 + n_2, p)$. Esto se puede aplicar a la suma de n variables.

\$nombre: Madre mía, qué compenetración han tenido los procesos de Bernoulli para hablar todos a la vez sobre lo que es su suma.

Rick: ¡Lo que te queda por aprender!

3.1.2.1 Misión 18

\$nombre: ¿Sabes qué? Siempre he pensado que algún profesor mío para aprobar y suspender a los alumnos hacía lo típico de tirar los exámenes a algún sitio y si quedaban dentro, aprobaba al examen, y si estaba fuera, le suspendía.

Rick: Ya sabes lo que opino del sistema de educación... pero, ya entiendo por dónde van los tiros. Te gustaría calcular la probabilidad de aprobar, si eso es así ¿no?

\$nombre: Exactamente, me gustaría saber la probabilidad de que apruebe la mitad de una clase de 20 alumnos, sabiendo que la probabilidad de que caiga en ese sitio es 0.3.

Rick: venga, va, pues a ello, ¿cuánto es?

Solución correcta 18:

¡Halleeluuya!

Solución incorrecta 18:

\$nombre: ¡Incrédulo que soy!

Solución:

La variable X = "número de exámenes lanzados" sigue una distribución $B(20,0.3)$. La probabilidad de que al menos aprueben la mitad de alumnos es:

$$P(X \geq 10) = 1 - P(X < 10) = 1 - P(X \leq 9) = \\ = 1 - \sum_{k=0}^9 \binom{20}{k} \cdot 0.3^k \cdot (1 - 0.3)^{20-k} = 1 - 0.952 = 0.048$$

Para calcular la probabilidad hemos utilizado el comando `sum(dbinom(0:9, 20, 0.3))` de R. Por lo que la probabilidad de que apruebe más de la mitad de la clase es **0.048**.

\$nombre: Ahora me explico muchas cosas... 🤪

3.1.3 Geométrica

Geométrica: Si ahora repetimos un experimento de Bernoulli hasta que obtengamos el primer éxito, la variable X , número de la repetición en la que conseguimos el primer éxito, seguirá una distribución **geométrica**. Por ejemplo, número de intento en el que sacamos cara lanzando una moneda. Se presenta por $X \rightarrow G(p)$ y su función de probabilidad para $k = 1, 2, \dots$ (ya que el primer éxito puede ocurrir en la primera prueba, en la segunda...):

$$P(X = k) = q^{k-1}p.$$

Ecuación 22: Geométrica

Esto es así ya que para que ocurra un éxito en el k intento, tiene que haber $k - 1$ fracasos (con probabilidad q) y un éxito (con probabilidad p).

Su media y varianza son:

$$E(X) = 1/p.$$

$$Var(X) = q/p^2.$$

En ocasiones se puede hablar de la variable geométrica Y como número de fracasos antes del primer éxito, entonces sería $Y = X - 1$, su función de probabilidad $P(Y = k) = q^k p$, su $E(Y) = q/p$ y su $Var(Y) = q/p^2$. En la función de probabilidad aparece p porque si no, solo estaríamos mirando la probabilidad de que haya k fracasos, no sabríamos si el siguiente intento sería fracaso o éxito.

Rick: Qué maja la geométrica.

\$nombre: Sí, me ha caído "to" bien.

3.1.3.1 Misión 19

Rick: No sé si alguna vez te he contado que tengo una amante llamada Unidad, bueno el caso es que es una mente colmena, entonces, pues, como cuando estoy con ella es totalmente aleatorio con qué "cuerpo" voy estando, si la primera vez que hablo con ella y, si consideráramos que en ese momento en la sala que estamos están 40 "cuerpos", que no es así pero por simplificarlo un poco, vamos, pues me gustaría saber la probabilidad de que tenga que estar con más de 8 "cuerpos" para volver a coincidir con el primero.

\$nombre: Guau, rick, cada día me asombras más... mmm pues la probabilidad sería....

Solución correcta 19:

!הלויה!

Solución incorrecta 19:

\$nombre: ¡Una y no más, santo Tomás!

Solución:

Como la probabilidad de que esté en cada uno de los cuerpos es la misma, pues $p = 1/40$. La variable X = "número de vez en la que vuelve a hablar con el primer cuerpo" sigue una distribución $G(p)$. La probabilidad que queremos saber es:

$$P(X > 8) = 1 - P(x \leq 8) = 1 - \sum_{k=1}^{8} (1 - 0.025)^{k-1} \cdot 0.025 = 1 - 0.183 = 0.817$$

Para calcular esta probabilidad con R tenemos que tener en cuenta que define a la geométrica como "número de fracasos antes del primer éxito". Esto quiere decir que calcularíamos $P(Y > 7) = 1 - P(Y \leq 7) = 1 - F(7)$. Por lo que el comando que utilizamos es `pgeom(7,0.025)` que nos da 0.183, por lo que hay una probabilidad de **0.817** de que Rick tenga que estar con más de 8 "cuerpos" para volver a estar con el que estuvo la primera vez.

3.1.4 Binomial negativa

Binomial negativa: La generalización de una variable geométrica es una variable binomial negativa, es decir, la variable X es el número de intentos hasta el éxito número r . Por ejemplo, número de intento en el que hemos sacado cinco caras lanzando una moneda. Su diferencia con la binomial, es que aquí permanece fijo el número de éxitos y en la binomial el número de intentos.

Se presenta con $X \rightarrow BN(r,p)$ y su función de probabilidad para $k = r, r + 1, r + 2, \dots$ (ya que como mínimo tendrá que haber r intentos):

$$f(k) = p(X=k) = \binom{k-1}{r-1} p^r \cdot (1-p)^{k-r}$$

Ecuación 23: Binomial negativa

Sabemos que va a haber r éxitos (con probabilidad p), por lo que habrá $k - r$ fracasos (con probabilidad $1 - p$). Como son independientes, su probabilidad es la multiplicación: $p^r(1 - p)^{k-r}$, multiplicamos por el número de combinaciones posibles y así hemos obtenido la función de probabilidad.

Su media y varianza serán:

$$E(X) = r/p.$$

$$Var(X) = r \cdot q/p^2.$$

¡Cuidado! En algunos sistemas informáticos (como R o statgraphics), en vez de considerar el número de intentos, definen la variable binomial negativa Y como número de fracasos antes del éxito r , por lo que tenemos que restarle a la variable los éxitos $Y = X - r$ y $P(X = k) = P(Y = k - r)$.

\$nombre: Lo estoy pensando con más detenimiento y me parece muy fuerte que las distribuciones existan como tal.

Rick: ¡Todo es posible en el universo!

3.1.4.1 Misión 20

Rick: Venga va, ahora que sabemos más cosas, vamos a liar más lo de Unidad para poder calcular algo que ya tenga más chicha.

\$nombre: ¡Va! ¿Qué te gustaría saber?

Rick: Pues, teniendo en cuenta otra vez, que estoy en una sala con 40 "cuerpos" controlados por unidad, me gustaría saber la probabilidad de estar 5 veces con el "cuerpo" con el que hablé la primera vez cuando esté en el punto de que Unidad haya cambiado más de 90 veces el "cuerpo" con el que me habla.

\$nombre: Cada día que paso en este concurso, más pienso que es un sueño...

Rick: ¡Céntrate! La probabilidad es...

Solución correcta 20:

إسبحان الله!

Solución incorrecta 20:

\$nombre: Pues esta parecía más razonable...

Rick: La estadística es impredecible, no puedes actuar por intuición, siempre hay que calcular las cosas porque nos pueden sorprender, ¡acuérdate de la paradoja del cumpleaños!

\$nombre: Ya te digo Rick...

Solución:

La variable X = "número de fracasos hasta conseguir el quinto éxito" sigue una distribución $BN(5,0.025)$. Tenemos que calcular $P(X > 85)$ ya que es el número de fracasos que hay en 90 intentos si 5 son éxitos:

$$P(X > 85) = 1 - P(X \leq 85) = 1 - \sum_{k=0}^{85} \binom{5+k-1}{k} \cdot (1-0.025)^k \cdot 0.025^5 = 1 - 0.0755 = 0.9245$$

Para calcular la probabilidad $P(X \leq 85)$ hemos ejecutado en R el comando `pbinom(85,5,0.025)` que da 0.0755 y lo que nos piden es $1 - 0.0755$. Por lo que hay una probabilidad del **92.45%** de que le toque estar con más de 90 cuerpos para estar cinco veces con el que estuvo primero.

3.1.5 Hipergeométrica

Hipergeométrica: Tenemos una población de N individuos en la que K de ellos tienen una característica y los demás no. La variable X , número de individuos con esa característica en una muestra de tamaño n sigue un modelo **hipergeométrico**, se representa por $X \rightarrow H(N, n, K)$ y siendo $k = \max(0, n - N + K) \dots \min(n, K)$, su función de probabilidad es:

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

Ecuación 24: Hipergeométrica

Si llamamos éxito a que el individuo tenga esa característica, ¿no es lo mismo que la variable binomial? No, pero ¿por qué? Vamos a verlo con un ejemplo que es como puede quedar más claro. Si cogemos 5 cartas de una baraja, todas a la vez, sin reemplazamiento y queremos saber cuántas de ellas son copas ($N=40$ y $K=10$), la variable X , número de oros en cinco cartas seguirá una $H(40,5,10)$, en cambio si cogemos una carta, comprobamos qué es, la volvemos a devolver a la baraja y lo hacemos 5 veces, es decir, con reemplazamiento, X seguirá una $B(5,10/40)$.

Pero cuando N es muy grande respecto a n (exactamente n/N es menor a 0.1) no hay casi diferencia en coger la muestra con o sin reemplazamiento por lo que la variable hipergeométrica se puede aproximar por la binomial.

Por eso, su esperanza y varianza son muy parecidas (en el caso de la esperanza iguales) a la de la binomial (recordemos que $p = K/N$ y $q = 1 - p$):

$$E(X) = n \cdot K/N = np.$$

$$Var(X) = n \cdot K/N \cdot (1 - K/N) \cdot (N - n)/(N - 1) = npq \cdot (N - n)/(N - 1).$$

\$nombre: Esta es complicada...

Rick: Todo es asentar bien lo que nos ha dicho.

3.1.5.1 Misión 21

\$nombre: Tres amigos y yo fuimos el otro día a un bar, invitó una ronda la casa...

Rick: ¿Cómo no me llamaste? ¡Esas cosas se avisan!

\$nombre: ¡Claro! Como te conocía en aquella época... Bueno, déjame seguir, el caso, que pusieron en la barra 10 cervezas negras y 15 cervezas rubias, cuando las personas que estábamos en el bar habíamos cogido una cerveza cada uno, nos avisaron que quienes habían cogido una cerveza negra, tenían premio, ¿cuál es la probabilidad de que nos tocará al menos a uno de nosotros?

Solución correcta 21:

¡bài hát ca ngợi!

\$nombre: ¡Pero si no nos tocó! Flipante...

Rick: La vida es dura...

Solución incorrecta 21:

\$nombre: Esta ya me parecía alta... ¡No nos tocó a ninguno!

Rick: Que te he dicho de elegir las cosas sin pensar, ¡tenemos que pensar más!

Solución:

La variable X = "número de cervezas negras en una muestra de 4 personas" sigue una distribución $H(25,4,10)$. La probabilidad de que les toque una cerveza negra al menos a uno de ellos es:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{10}{0} \cdot \binom{15}{4}}{\binom{25}{4}} = 0.89$$

Por lo que hay un **89%** de probabilidades de que le toque a uno de ellos.

3.1.6 Poisson

Poisson: Queremos ver los sucesos ocurridos en un soporte continuo (tiempo, longitud...), por ejemplo, personas que van a la relojería Alonso en una hora. Por lo que la variable X , número de sucesos que ocurren en un intervalo sigue un modelo de **Poisson** de parámetro λ , se representa por $X \rightarrow Poi(\lambda)$ y siendo $k = 0,1,2\dots$ su función de probabilidad es:

$$P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

Ecuación 25: Poisson

Su media y varianza es la misma:

$$E(X) = Var(X) = \lambda.$$

La distribución de Poisson también tiene la propiedad aditiva, por lo que es reproductiva respecto a su parámetro, así que la suma de dos variables independientes de Poisson será $X + Y \rightarrow Poi(\lambda_1 + \lambda_2)$.

Si tenemos una muestra grande, n es mayor que 30, y una probabilidad baja de éxito, p es menor que 0.1, una variable binomial puede aproximarse por una de Poisson (ya que es una generalización de ella) de parámetro $\lambda = np$.

\$nombre: ¿Nos envenenaremos estudiando esto?

Rick: 

3.1.6.1 Misión 22

Rick: Vale, vamos a suponer que el número de frases coherentes dichas por Jerry cuando le hago una pregunta es, en media, de dos, y ya es mucho suponer...

Jerry: Ya estás metiéndote otra vez conmigo... ¡Qué habré hecho yo para merecer esto!

Rick: Esto de que dejen hablar a más gente por subir la audiencia me pone negro...

\$nombre: Bueno, ¡sigamos!

Rick: ¡Así me gusta, \$nombre! Si le hago cinco preguntas a Jerry, ¿cuál es la probabilidad de que conteste con 10 frases coherentes?

Solución correcta 22:

¡Aló!

Rick: Y demasiado me parece...

Jerry: A veces digo cosas coherentes, jo 

Solución incorrecta 22:

Rick: Ya decía yo que esto no era posible...

Solución:

La variable X = "número de frases coherentes dichas en una respuesta" siguen una distribución $Poi(2)$; como es reproductiva podemos considerar la variable Y = "número de frases coherentes dichas en 5 respuestas" que sigue una $Poi(10)$. Por lo que la probabilidad que nos piden es:

$$P(Y = 10) = e^{-10} \cdot \frac{10^{10}}{10!} = 0.125$$

Por lo que hay una probabilidad del **12.5%** de que Jerry diga 10 frases coherentes en 5 respuestas.

3.2 Modelos continuos

Los **modelos continuos**, son modelos probabilísticos de variables aleatorias vpmymisid.

\$nombre: ¡Vale! ¡Ya he pillado lo que pasa!

Rick: ¿Qué es lo que pasa?

\$nombre: Piensa un poquito, al final si no lo pillas, te lo digo 😊

Rick: Es a ti al que le interesa esto, yo paso...

\$nombre: Ya, ya, eso dices pero en el fondo te pica la curiosidad, que lo sé yo 😊

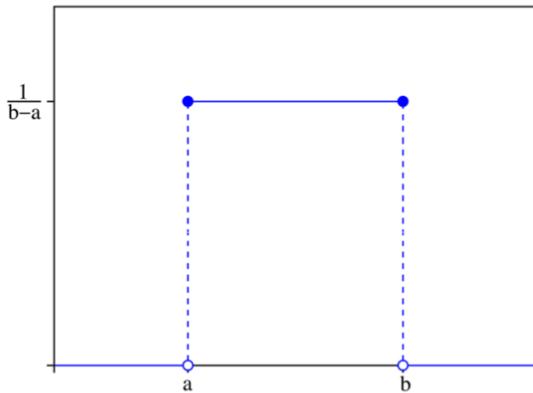
Rick: Lo que tú digas... 😓

3.2.1 Uniforme

Uniforme: Una variable X sigue un modelo **Uniforme** entre dos valores a y b y se representa como $X \sim U(a, b)$ si su función de densidad y distribución, son, respectivamente:

Función de densidad:

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{en otro caso} \end{cases}$$



Función de distribución:

$$F(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a < x < b \\ 1 & \text{si } x \geq b \end{cases}$$

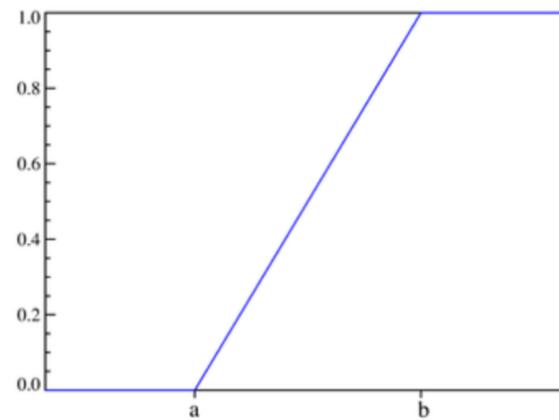


Figura 46: Uniforme

Es la generalización de la variable discreta uniforme que proporciona la misma probabilidad a todos los puntos, así que los intervalos que tienen una misma longitud, también tienen la misma probabilidad. La media y la varianza de una uniforme son:

Rick: ¿Me queda bien? 

\$nombre: 

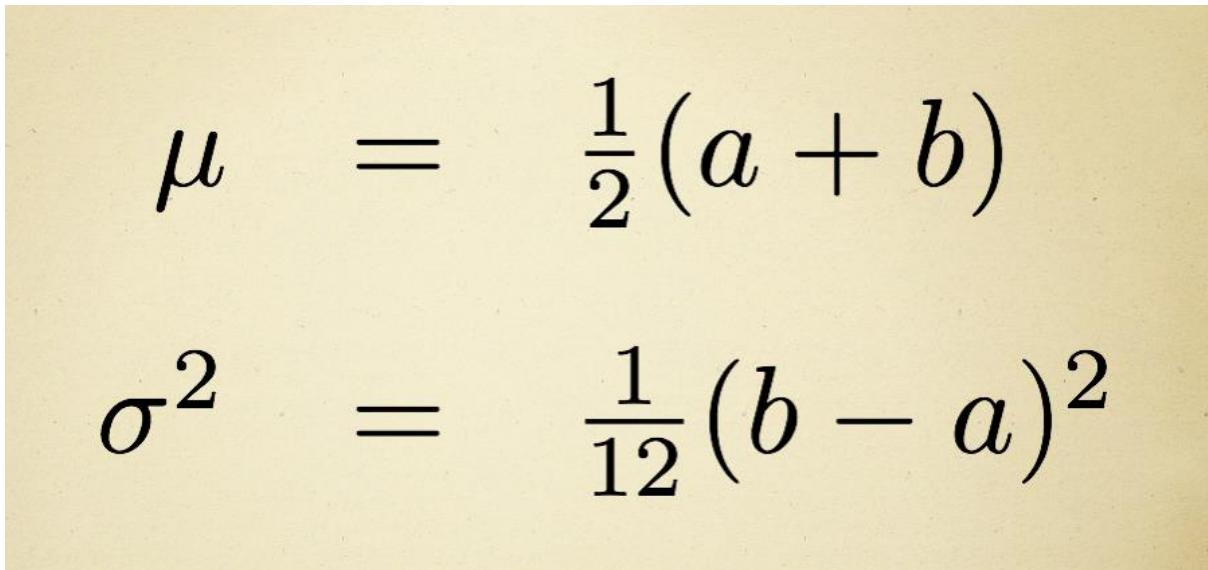

$$\mu = \frac{1}{2}(a + b)$$
$$\sigma^2 = \frac{1}{12}(b - a)^2$$

Figura 47: Esperanza y varianza de una uniforme

3.2.1.1 Misión 23

\$nombre: Tengo un amigo que llega entre media hora y una hora más tarde de la hora en la que hemos quedado inicialmente.

Rick: ¿Y tú eres igual?

\$nombre: No, yo no aguento la impuntualidad, así que como ya sé esto, siempre llego media hora más tarde y le doy 10 minutos de cortesía, si tarda más me voy.

Rick: Entonces, ¿cuál es la probabilidad de que te vayas?

Solución correcta 23:

¡அல்லேவுடா!

Solución incorrecta 23:

\$nombre: Pensaba que era un poco más paciente, la verdad... 

Rick: Conmigo no vas a tener estos problemas, porque si te necesito, iré directamente sin quedar contigo jajaja

Solución:

El tiempo que tarda es una variable $X \rightarrow U(30,60)$ expresado en minutos. Así que la probabilidad de que tarde más de 40 minutos es:

$$P(X > 40) = 1 - P(X \leq 40) = 1 - F(40) = 1 - \frac{40 - 30}{60 - 30} = 0.67$$

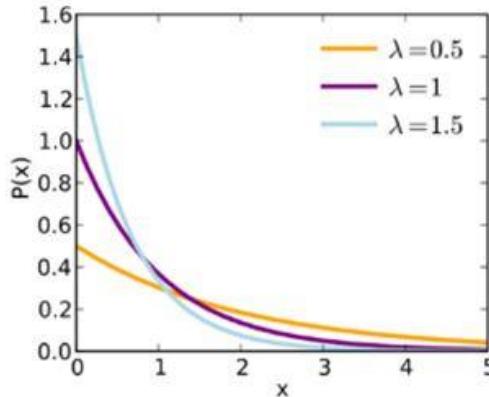
Por lo que hay una probabilidad del **67%** de que \$nombre se vaya antes de que llegue su amigo.

3.2.2 Exponencial

Exponencial: La variable X que suele representar la duración de un sistema biológico o mecánico, aunque también se adecúa a más casos sigue un modelo **exponencial** que se representa con $X \rightarrow Exp(\lambda)$ siendo $\lambda > 0$ y su función de densidad y distribución son:

Función de densidad

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{para } x \geq 0 \\ 0 & \text{en caso contrario} \end{cases}$$



Función de distribución

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{para } x < 0 \\ 1 - e^{-\lambda x} & \text{para } x \geq 0 \end{cases}$$

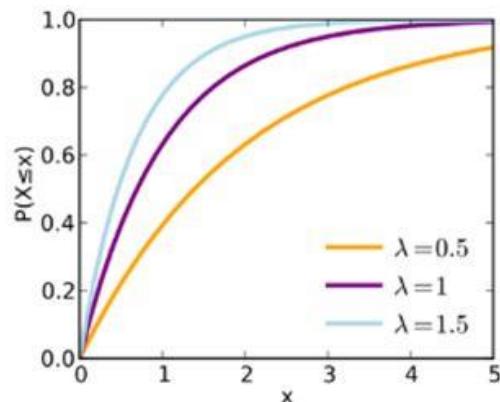


Figura 48: Exponencial

Su media es $1/\lambda$ y su varianza es $1/\lambda^2$.

3.2.2.1 Misión 24

Beth: Hola chicos, ando de camino a casa que me ha llamado Jerry porque tiene una urgencia.
Rick: ¿Se le ha roto una uña otra vez? ¡Ay, pobre!

Beth: Tampoco se ha explayado mucho pero me da miedo llegar cuando sea demasiado tarde y la haya liado, pero teniendo en cuenta que acabo de salir del trabajo y suelo tardar una media de 30 minutos en llegar dependiendo del tráfico, ¿cuál es la probabilidad de que tarde menos de veinte minutos y me dé tiempo a llegar antes de que la haya liado?

Solución correcta 24:

¡P�ะเจ້າ!

Beth: Bueno, no está todo perdido, ¡la esperanza es lo último que se pierde!

Solución incorrecta 24:

Beth: No me deis falsas esperanzas 😞

Solución:

X = "tiempo de espera en minutos" sigue una exponencial con una media de 30, por lo que α será 1/30:

$$P(X \leq 20) = F(20) = 1 - e^{-\frac{20}{30}} = 0.49$$

Por lo que hay una probabilidad de **49%** de que Beth llegue a tiempo.

3.2.3 Normal

Normal: La variable X que suele medir caracteres como el peso, la puntuación de un examen, errores en medidas... sigue un modelo **mpt,sñ**, es la distribución continua más importante por la frecuencia con la que aparece y sus propiedades. Se representa con $X \sim N(\mu, \sigma)$ y su función de densidad (ya que nadie ha conseguido calcular su función de distribución):

\$nombre: ¡Rick! ¡Mira cómo escriben normal! ¡Aparece una coma! No hay sólo letras... eso tiene que ser una señal...

Función de densidad

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

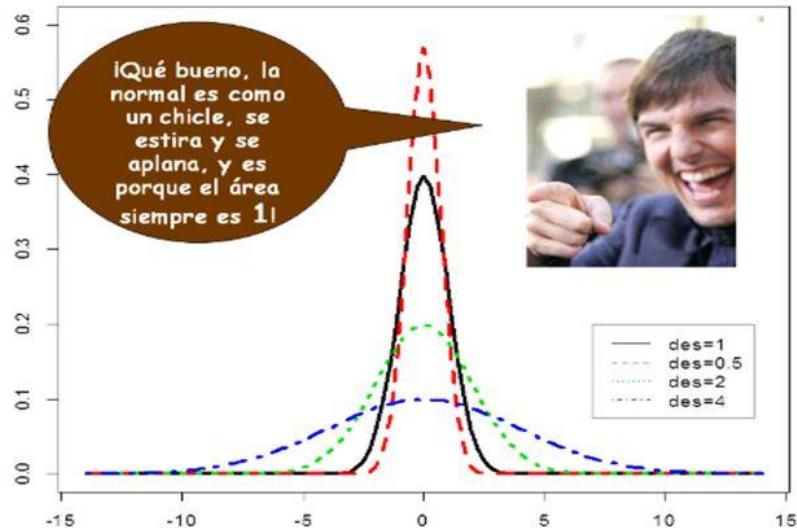
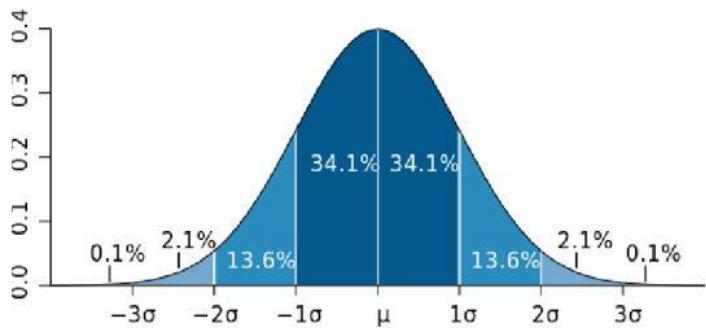


Figura 49: Normal

Como podemos ver, tiene forma de campana (que es conocida como campana de Gauss), es simétrica respecto a la media y podemos ver, como en la área que va desde $\mu - k\sigma$ hasta $\mu + k\sigma$, siendo $k = 2$, se concentran el 95.4% de los datos.

Rick: Ahora me explico por qué la gente me dice que no soy normal...

\$nombre: Ya te digo, tron, jajaja

Su media es μ y su varianza es σ^2 .

La variable normal también tiene la propiedad aditiva, por lo que es reproductiva; la suma de variables aleatorias normales es otra variable aleatoria normal de media la suma de las medias y varianza la suma de varianzas.

Para comparar poblaciones solemos utilizar la normal estándar o tipificada, que es una normal de media 0 y varianza 1. Para tipificar una normal, simplemente le restaremos su media y le dividiremos su desviación, y ya tendremos una normal tipificada. También nos sirve para calcular sus probabilidades:

$$P(a < X < b) = P\left(\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{b - \mu}{\sigma}\right) = P\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right)$$

Ecuación 26: Cálculo de probabilidades de una Normal

3.2.3.1 Misión 25

Rick: No voy a contarte detalles de por qué o qué es lo que vamos a calcular ahora porque es un tema que es mejor no sacar.

\$nombre: ¡Pero no me digas eso! Que me dejas con toda la intriga...

Rick: Te voy conociendo y si no, ibas a hacer muchas preguntas, el caso si sabemos que las mantis siguen una $N(80,20)$ y los cronembergs siguen una $N(140,50)$, si una mantis pesa 120 kg y un cronemberg pesa 200 kg, ¿cuál de los dos sería el que más destaca? ¿Del ejemplar que destaque más, cuál sería el porcentaje de ejemplares que podría llegar a mover si consigo mover sin problemas hasta 100 kg?

\$nombre: no sé de qué narices estás hablando ni para qué quieres saber eso, pero el porcentaje de ejemplares que podrías mover sería...

Solución correcta 25:

¡Theta!

Rick: ¡El más fuerte yo! 🤘

\$nombre: Baja modesto que sube Rick...

Solución incorrecta 25:

Rick: Ya me estoy infravalorando, si es que...

\$nombre: Si, como te infravaloras mucho tú... 😅

Solución:

Lo que hacemos es tipificar ambos valores para saber cuál de los dos destaca más. $(120 - 80) / 20 = 2$ y $(200 - 140) / 50 = 1.2$. Por lo que destaca más la mantis. Así que ahora definimos X = "peso de las mantis" y calculamos:

$$P(X < 100) = P\left(\frac{X - 80}{20} < \frac{100 - 80}{20}\right) = P(Z_{N(0,1)} < 1) = 0.8413$$

Hemos buscado en la tabla de la normal, en el valor $z = 1$. Para calcularlo con R sólo tendríamos que ejecutar el comando `pnorm(100,80,20)`, si lo que quisieramos fuera calcular 100 a partir del 0.8413, poniendo `qnorm(0.8413,80,20)` lo conseguiríamos. Por lo que Rick puede mover al **84.13%** de las mantis

3.2.4 Chi-cuadrado

Chi-cuadrado: Cuando se suman k variables aleatorias elevadas al cuadrado que siguen una distribución $N(0,1)$, da como resultado una variable que sigue una distribución **Chi-cuadrado**. Se representa por χ^2_k ya que depende de un número entero positivo k llamado *grados de libertad*.

\$nombre: Chi, chi.

Rick: 🤘

$$\chi_k^2 = X_1^2 + X_2^2 + \dots + X_k^2, \quad \text{con } X_i \in N(0,1)$$

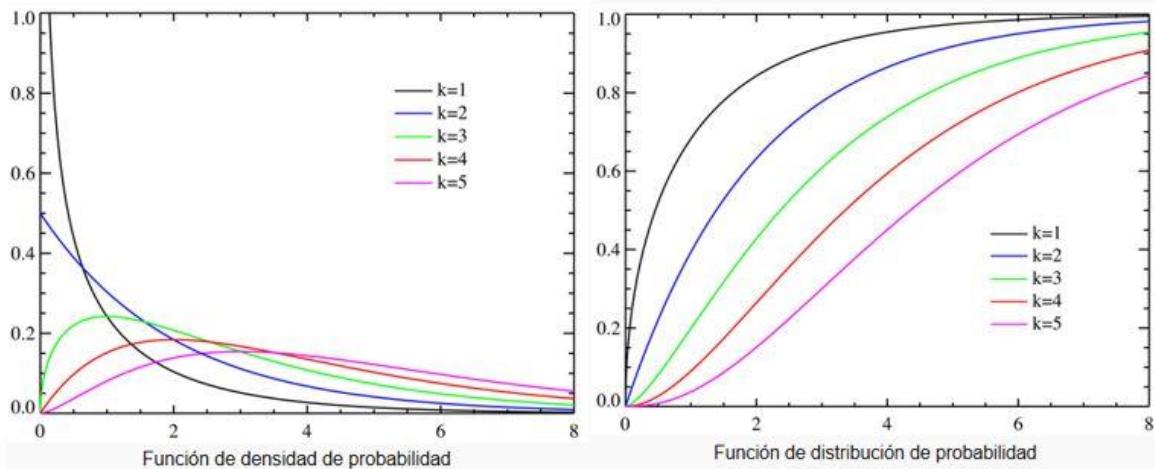


Figura 50: Chi-cuadrado

3.2.5 T de Student

T de Student: Si una variable Z que sigue una distribución $N(0,1)$ la dividimos por una variable Y independiente a ella que sigue una distribución Chi-cuadrado con n grados de libertad conseguimos una variable t_n con n grados de libertad. Esta variable tiene forma de campana de Gauss como la Normal pero su media es siempre cero y tiene colas más pesadas que ésta. Su función de densidad fue formulada por Gosset mientras trabajaba en una fábrica de cervezas Guinness en Dublín.

Rick: Esta me gusta mucho, que ganas de echar un trago...

\$nombre: ¡Cómo que no lo hicieras ya de por sí! 🍺🍺

$$t_n = \frac{Z}{\sqrt{Y/n}}$$

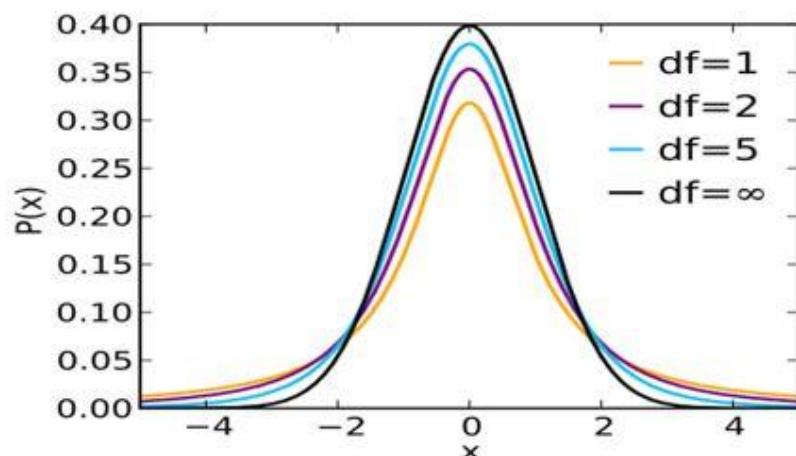


Figura 51: T de Student

3.2.6 Fisher-Snedecor

Si tenemos X e Y , dos variables independientes que siguen una distribución chi-cuadrado y dividimos una por la otra resulta la variable W que sigue una distribución **F de Fisher-Snedecor** con n y m grados de libertad, los de las chi-cuadrado.

\$nombre: La verdad... No podían tener nombres más raros jajaja siempre haciendo difícil la tarea de recordarlos.

Rick: Pero nosotros al haber tenido este contacto directo con ellas, yo creo que no se nos van a olvidar.

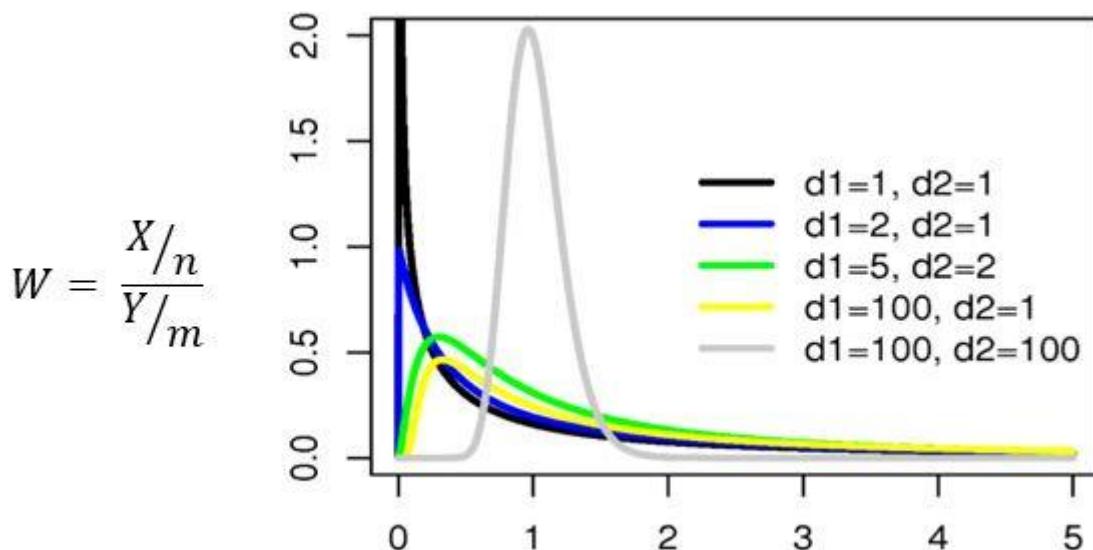


Figura 52: Fisher-Snedecor

3.3 Teorema central del límite

El **teorema central del límite** nos recalca la importancia de la distribución normal, y nos dice que cuando sumamos un gran número (a partir de n mayor que 30 ya da buenos resultados) de variables aleatorias el resultado es una variable aleatoria que sigue aproximadamente una normal de media la suma de las medias, y varianza la suma de varianzas.

Rick: Esto es lo malo, que la distribución normal es muy importante igual que en la sociedad lo es ser normal...

\$nombre: Te lo tomas todo muy demasiado a pecho jajaja

$$Y = X_1 + X_2 + \dots + X_n \approx N\left(\sum_{i=1}^n \mu_i, \sqrt{\sum_{i=1}^n \sigma_i^2}\right)$$

Figura 53: Teorema central del límite

4. Inferencia estadística

Ahora el problema que nos surge es que normalmente no sabemos el modelo que sigue una población, pero sí podemos sacar una muestra de ella y estudiarla. Para esto está la Inferencia Estadística que, a partir del cálculo de probabilidades y la información que le da la muestra, valida un posible modelo para la población y estima sus parámetros. Veremos estimación puntual con la que sacaremos el valor de parámetros, estimación por intervalos en el que diremos entre qué valores puede estar un parámetro con un nivel determinado de confianza y contraste de hipótesis paramétrico para poder decidir si un parámetro está más cerca de un valor que de otro.

4.1 Muestreo aleatorio simple

El muestreo consiste en elegir una muestra de la población (ya que muchas veces no es posible estudiarla entera) que sea representativa a la misma, con el fin de estudiar sus propiedades estadísticas. Hay muchas formas distintas de muestreo (estratificado, por conglomerados, sistemático...) pero nosotros solo vamos a ver el muestreo aleatorio simple ya que es el que vamos a utilizar.

En el **muestreo aleatorio simple** (muchas veces diremos **m.a.s.** por abreviar) todos los individuos tienen la misma probabilidad de ser elegidos. Cogemos un individuo, miramos la característica que queremos estudiar y lo devolvemos a la población y ya cogemos al siguiente y repetimos el proceso (si cada vez que cogemos un individuo no lo devolvemos a la población sería muestreo aleatorio sin reemplazamiento, otro tipo de muestreo). Cómo la población es muy grande, la probabilidad de coger dos veces al mismo individuo es casi nula.

\$nombre: ¡Me he quedado con curiosidad sobre los otros tipos de muestreo!

Rick: Sí, tendremos que investigar sobre ellos, ¡a ver de qué van!

4.2 Estimación puntual

La **estimación de tipo puntual** tiene el objetivo de obtener estimaciones de los parámetros (μ , σ^2) de las variables aleatorias (binomial, normal...), que nos permiten medir

características (peso, altura...) de una población (mayores de 18 años, españoles...) a partir de una muestra de la misma.

4.2.1 Definiciones

Estadístico: Una función de variables aleatorias $g(X_1, \dots, X_n)$ donde X_i es una variable aleatoria).

Estimador: Estadístico que se usa para estimar un parámetro desconocido. Se representa con la letra del parámetro que estima con un acento circunflejo (sombro para los amigos) encima.

Estimación: Valor concreto del estimador para una muestra en particular, por lo que es una predicción del valor teórico desconocido (el de la población).

Cualquier estadístico utilizado para estimar un parámetro, es un estimador. Cuando ese estimador se cuantifica (cuando ya tenemos los datos de la muestra y lo valoramos), tenemos una estimación.

4.2.2 Estimación de la media

Para obtener una aproximación de la media de una población realizamos una estimación de la misma. Uno de los estimadores posibles es la **media muestral**. Por ejemplo, queremos estimar el peso medio de los niños de 10 a 12 años españoles. Para ello cogemos una muestra aleatoria simple de n niños españoles y medimos n veces la variable X = "peso" por lo que obtenemos

X_1, \dots, X_n , n variables aleatorias. $g(X_1, \dots, X_n) = X_1, \dots, X_n / n$ es un estadístico (media muestral) y un estimador (de la media poblacional), ya que aparte de ser una función de variables aleatorias, estima el valor de un parámetro, en este caso, la media. Sólo para ver cómo se calcularía, si cogemos 5 niños con pesos de 32, 33, 35, 37 y 38:

$$\hat{\mu} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{32 + 33 + 35 + 37 + 38}{5} = 35$$

35 sería la estimación de la media.

Podríamos considerar otros estadísticos aparte de la media muestral como estimadores de la media poblacional y utilizar métodos matemáticos para elegir entre distintos estimadores.

Rick: A mí esto de estimar mira que no me va... Me gusta que sea exacto...

\$nombre: ¡Mejor esto que nada!

4.2.3 Estimación de la varianza

Para obtener una aproximación de la bstosmxs de una población realizamos una estimación de la misma. Uno de los estimadores posibles es la **varianza muestral**.

Si en vez de dividir por n se divide por $n - 1$, se tiene otro estimador que se llama **cuasi-varianza muestral**.

Su relación con la varianza muestral es muy sencilla.

Rick: ¡Ya lo he pillado!

\$nombre: ¡A ver!

Rick: ¡Dilo tú!

\$nombre: Bua, eso es que no tienes ni idea... jajaja y quieres que te lo diga

Rick: Lo que tú digas 😊

Varianza muestral Cuasi-varianza muestral

$$\hat{s}_n^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})}{n} \quad \hat{s}_n^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})}{n-1}$$

Ecuación 27: Varianza y cuasivarianza muestral

Cuando n es grande, los valores son similares pero si lo miramos desde un punto de vista teórico se puede demostrar que la cuasi-varianza muestral tiene mejores propiedades y, por eso, en muchos programas estadísticos como R, se utiliza directamente la cuasi-varianza muestral.

\$nombre: ¡Desvariando voy! ¡Desvariando vengo vengo!

Rick: Normal jajaja ya llevamos un buen tiempo por aquí jajaja

4.2.4 Estimación de una proporción

Si queremos saber la proporción p de individuos que tienen una característica (por ejemplo, número de personas con sobrepeso) se elige una m.a.s. de tamaño n e iremos apuntando 1 si el individuo tiene la característica y 0 si no (cada vez que elegimos un individuo observamos una variable Bernoulli de parámetro p , que es lo que queremos estimar) y estimamos p a través de su estimador llamado **proporción muestral**:

$$\hat{p} = \frac{\text{número de unos}}{n}$$

Ecuación 28: Proporción muestral

4.2.5 Distribución muestral de un estimador

Como un estimador es una variable aleatoria seguirá una distribución que la llamamos distribución muestral o distribución en el muestreo.

Rick: Cómo nos gusta lo que nos encanta, oyes jajaja

\$nombre: ¡A qué narices vendrá eso ahora! 🤪

Supongamos que se quiere saber la altura media de los donantes de sangre en Valladolid, para ello se ha tomado la altura de las primeras 30 personas que han ido a donar en cada uno de los puestos, que son 100. Haremos la media de la altura de cada puesto. Si hicieramos un histograma de la tercera columna de la tabla, podríamos ver cómo tiene forma de campana de Gauss, esto ocurre ya que la media muestral es la suma de n variables divididas por n , y, como nos dice el Teorema Central del Límite, seguirá una **distribución normal**.

Muestra	Valores del peso	Estimación
1	(64.37, 56, ..., 94.5)	$\bar{x}_1 = 67.03$
2	(65, 91.2, ..., 84.9)	$\bar{x}_2 = 71.78$
...
300	(87.3, 49.2, ..., 80)	$\bar{x}_3 = 79.02$

Figura 54: Distribución muestral de un estimador

4.2.6 Calidad de un estimador

Ya habíamos dicho que se puede mirar por métodos matemáticos si un estimador es mejor que otro. Ahora veremos el sesgo y la consistencia de los estimadores que son dos de las propiedades que nos permiten estudiarlo.

\$nombre: Calidad como calidad 😊

4.2.6.1 Sesgo de un estimador

El **sesgo** es la distancia entre la media del estimador y el parámetro que estamos estimando, cuanto menor es el sesgo, mejor es el estimador y si es 0, se dice que el estimador es insesgado.

$$\text{Sesgo}(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

Ecuación 29: Sesgo de un estimador

Puede demostrarse matemáticamente que la media muestral y proporción muestral son estimadores insesgados de la media y proporción, respectivamente. La cuasivarianza muestral es un estimador insesgado de la varianza pero la varianza muestral no, este es uno de los motivos por los que muchos programas estadísticos cuando les decimos que nos calculen la varianza, directamente nos calculan la cuasivarianza.

Rick: Estoy rodeado de sesgadez.

\$nombre: Y yo entonces... ni teuento ya...

4.2.6.2 Consistencia de un estimador

Cuanto mayor sea la muestra que tomamos mejor será la estimación.

Un estimador se dice *consistente* si $\lim_{n \rightarrow \infty} \hat{\theta} = \theta$.

Ecuación 30: Consistencia

Los estimadores que hemos visto (media muestral, varianza muestral, cuasivarianza muestral y proporción muestral) son **consistentes**.

\$nombre: Y rodeados de muchas cosas inconsistentes.

Rick: ¡Ya te digo!

4.3 Estimación por intervalos

Ahora vamos a ver cómo dar un intervalo en el que está el parámetro que queremos estimar con una probabilidad alta.

Dicho de manera formal, un intervalo de confianza de nivel α (número comprendido entre 0 y 1 conocido como nivel de significación) para estimar un parámetro θ es un intervalo (a, b) que contiene al parámetro con probabilidad $1 - \alpha$ (que es el nivel de confianza).

¿Esto qué quiere decir? Que si escogemos un $\alpha = 0.05$, la estimación del parámetro del 95% de las muestras que hayamos tomado estará dentro del intervalo y el 5% no.

Generalmente construimos el intervalo de confianza a través de un estadístico llamado pivote, con distribución conocida (normal, Chi-cuadrado...) al que le restamos y sumamos (para conseguir el lado izquierdo y derecho del intervalo, respectivamente) una cantidad que depende del tamaño de la muestra y de α . Cuanto mayor sea el nivel de confianza, mayor longitud tendrá el intervalo, en cambio, cuanto mayor sea el tamaño de la muestra menor será la longitud del mismo (lo cual tiene mucho sentido si lo pensamos).

Rick: Uff, ahora ver intervalos de confianza, con lo que me gusta confiar.

\$nombre: Te ayudará a hacerlo, ¡ya lo verás!

4.3.1 Intervalo de confianza para la media

Sea X una variable aleatoria normal $X \sim N(\mu, \sigma)$ y partiendo de una m.a.s.:

- Si **conocemos σ** (desviación típica) el intervalo es:

$$\left(\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right),$$

donde \bar{x} ,

$$\bar{x} = \frac{x_1 + \dots + x_n}{n}$$

Figura 55: Intervalo de confianza para la media si conocemos σ

es la media muestral, y $Z_{\alpha/2}$ el valor de una distribución $N(0,1)$ (buscaremos ese valor en la tabla), es decir, el cuantil que deja a la izquierda una probabilidad $1 - \alpha/2$.

Rick: Mira otra cosa que es que no... con lo que me cuesta a mí confiar.

\$nombre: Pero confiar en la estadística es distinto hombre.

- Si **no conocemos σ** la estimaremos con la cuasi-desviación típica muestral y, en este caso, utilizaremos la distribución t de Student siendo $t_{n-1,\alpha/2}$ su valor (lo buscaremos en su tabla) con $n - 1$ grados de libertad que es el cuantil que deja a la izquierda un área de $1 - \alpha/2$.

$\frac{\hat{s}_{n-1}}{\sqrt{n}}$ suele recibir el nombre de error típico o error estándar de la media. El intervalo de la media en este caso es:

$$\left(\bar{x} - t_{n-1,\alpha/2} \frac{\hat{s}_{n-1}}{\sqrt{n}}, \bar{x} + t_{n-1,\alpha/2} \frac{\hat{s}_{n-1}}{\sqrt{n}} \right)$$

Figura 56: Intervalo de confianza para la media si no conocemos σ

4.3.1.1 Misión 26

\$nombre: El otro día me compro una bolsa de patatas y ponía que venían 80g, como no me fío mucho yo de eso me compré 6 paquetes de patatas y las estuve pesando en casa a ver si era verdad.

Rick: Si tienes tiempo libre si...

\$nombre: El caso es que no sabía cómo mirar si me estaban mintiendo o no pero ahora podemos verlo. Si los resultados que obtuve fueron: 77.8g, 82.2g, 77.8g, 78g, 81g y 79.5g, ¿cuál sería el extremo derecho del intervalo si tomamos una confianza del 95%?

Solución correcta 26:

¡ハレレヤ!

Solución incorrecta 26:

Rick: No damos ni una...

Solución:

Como $\bar{x} = 79.38$, $\hat{s}_{n-1} = 1.87$, $\alpha = 0.05$, $t_{6-1,0.05/2} = 2.5706$ (lo hemos buscado en la tabla) nos sale un intervalo de confianza para la media de $(77.42, 81.45)$. Lo podríamos haber hecho directamente en R, ejecutando el comando `z = c(77.8, 77.8, 82.2, 78, 81, 79.5)` y el `t.test(z, conf.level=0.95)`

Por lo que el extremo derecho del intervalo de confianza para la media es **81.45**.

4.3.2 Intervalo de confianza para la desviación típica

Nota: para la varianza solo hay que elevar los resultados al cuadrado.

- Si **conocemos μ** (la media) el intervalo es:

$$\left(\sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{\chi_{n,\alpha/2}^2}}, \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{\chi_{n,1-\alpha/2}^2}} \right)$$

Figura 57: Intervalo de confianza para la desviación típica si conocemos la μ

donde $\chi_{n,\alpha/2}^2$ es el valor de una Chi-cuadrado con n grados de libertad que es el cuantil que deja a su derecha un área de $1 - \alpha/2$.

- Si **no conocemos μ** la Chi-cuadrado tendrá $n - 1$ grados de libertad y el intervalo es:

$$\left(\sqrt{\frac{(n-1)\hat{S}_{n-1}^2}{\chi_{n-1,\alpha/2}^2}}, \sqrt{\frac{(n-1)\hat{S}_{n-1}^2}{\chi_{n-1,1-\alpha/2}^2}} \right)$$

Figura 58: Intervalo de confianza para la desviación típica si conocemos la μ

4.3.2.1 Misión 27

Rick: \$nombre me ha dado cuenta que cuando terminemos esto quiero volver a hacer una fiesta con todos mis amigos.

Summer: Sí, por favor, ¡repítámoslo!

Morty: No, por favor, otra vez, ¡no!

Rick: ¿Alguien os ha pedido opinión? Por dónde iba..., para que no nos falte de nada, me gustaría mirar cuánto se consumió la última vez; me acuerdo del número de mejunjes que se bebieron 9 personas: 2, 8, 9, 12, 13, 15, 17, 18, 32. Así que podemos calcular el intervalo de

la media y el de la desviación típica. ¿Cuánto da: extremo derecho del intervalo de la media + 2 x extremo derecho del intervalo de la desviación típica?

Solución correcta 27:

¡d'eseñor! ¡d'eseñor!

Rick: A mí no me importa que sobre, que lo voy a aprovechar, ¡así que tiremos la casa por la ventana!

\$nombre: Así puedo invitar yo a unos amigos también.

Rick: Claro, claro, ese es el quid de la cuestión.

Solución incorrecta 27:

Rick: En esto sí que es una cosa en la que no podemos permitirnos fallar.

\$nombre: Ya te digo, nunca hay que quedarse corto...

Solución:

Para calcular el intervalo de confianza para la media hemos ejecutado los comandos de R:

$w = c(2,8,9,12,13,15,17,18,32)$

$t.test(w,conf.level=0.95)$

Que nos da (7.57, 20.43).

Para calcular el intervalo de confianza para la desviación típica, como no conocemos la media de la población ya que Rick no se acuerda de todos los mejunjes que se bebieron todos los invitados, pues utilizamos el intervalo en el que no conocemos la media. Entonces, $n - 1 = 8$, la cuasivarianza muestral es igual a 70 y con los comandos de R $qchisq(0.025,8)$ y $qchisq(0.975,8)$ tenemos los valores 2.18 y 17.53 que son los valores de la Chi-cuadrado que necesitamos, el intervalo de confianza nos queda: (5.65, 16.03).

Por lo que extremo derecho de la media + 2 x extremo derecho de la desviación típica es igual a **52.49** que son los sucesos considerados como raros.

4.3.3 Intervalo de confianza para una proporción

Si queremos saber la proporción p de individuos de una población que tienen una determinada característica podemos hacer un **intervalo de confianza de nivel α para p** :

$$\left(\hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

Figura 59: Intervalo de confianza para una proporción

4.3.3.1 Misión 28

Rick: Tengo un váter al que voy muy a menudo a hacer mis necesidades y me gustaría saber la proporción de veces que lo utilizo.

\$nombre: Todo el mundo tiene un váter Rick, no me parece nada del otro mundo.

Rick: Bueno, el mío es muy especial, pero no tengo porque darte explicaciones, la cuestión es que de cada 60 veces que necesito ir al váter, 42 voy a ese. ¿Cuál es el extremo izquierdo del intervalo de la proporción p veces que voy a ese baño si tomamos un nivel de significación del 1%?

Solución correcta 28:

শরের যশোকীর্তনমূলক সঙ্গীতবিশেষ!

Solución incorrecta 28:

Rick: Yo pensaba que este mínimo sería el extremo izquierdo...

\$nombre: Aplícate el consejo que das de más calcular y menos elucubrar.

Solución:

Nivel de significación del 1 por ciento: $\alpha = 0.01 \Rightarrow 1 - \alpha = 0.99 \Rightarrow Z_{\alpha/2} = 2.64$. Por otro lado, nos dicen que $\hat{p} = \frac{42}{60}$ y que $n = 60$. El intervalo sale:

$$\left(\frac{42}{60} - 2.64 \cdot \sqrt{\frac{\frac{42}{60} \cdot (1 - \frac{42}{60})}{60}}, \frac{42}{60} + 2.64 \cdot \sqrt{\frac{\frac{42}{60} \cdot (1 - \frac{42}{60})}{60}} \right) = (0.54, 0.85),$$

Figura 60: Misión 28

Por lo que el extremo izquierdo del intervalo es **0.54** y la proporción de veces que Rick va a su váter es entre el 54% y el 58% de las veces.

4.3.3 Intervalo de confianza para la diferencia de medias

Sean X e Y variables aleatorias normales $X \rightarrow N(\mu_X, \sigma_X)$ e $Y \rightarrow N(\mu_Y, \sigma_Y)$ y partiendo de sus respectivas m.a.s.:

- Si **conocemos** σ_x y σ_y

$$\left((\bar{x} - \bar{y}) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} \right)$$

- Si **no conocemos las desviaciones** pero suponemos:

1. Que son iguales (las desviaciones)

$$\left((\bar{x} - \bar{y}) \pm t_{n+m-2, \alpha/2} \sqrt{\frac{(n-1)\hat{S}_{n-1}^2 + (m-1)\hat{S}_{m-1}^2}{n+m-2}} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}} \right)$$

2. Que el tamaño de muestra es grande ($n, m \geq 30$)

$$\left((\bar{x} - \bar{y}) \pm Z_{\alpha/2} \sqrt{\frac{\hat{S}_{n-1}^2}{n} + \frac{\hat{S}_{m-1}^2}{m}} \right)$$

3. Que el tamaño de muestra es pequeño ($n, m < 30$)

$$\left((\bar{x} - \bar{y}) \pm t_{n+m-2-\Delta, \alpha/2} \sqrt{\frac{\hat{S}_{n-1}^2}{n} + \frac{\hat{S}_{m-1}^2}{m}} \right)$$

siendo Δ el entero más próximo a

$$\frac{\left((m-1) \frac{\hat{S}_{n-1}^2}{n} - (n-1) \frac{\hat{S}_{m-1}^2}{m} \right)^2}{(m-1) \left(\frac{\hat{S}_{n-1}^2}{n} \right)^2 + (n-1) \left(\frac{\hat{S}_{m-1}^2}{m} \right)^2}$$

Figura 61: Intervalo de confianza para la diferencia de medias

4.3.4 Intervalo de confianza para la razón de varianzas

El **intervalo de la razón o cociente de dos varianzas** utiliza la distribución F de Snedecor con $n - 1$, $m - 1$:

$$\left(F_{n-1, m-1, 1-\alpha/2} \frac{\hat{S}_{m-1}^2}{\hat{S}_{n-1}^2}, \quad F_{n-1, m-1, \alpha/2} \frac{\hat{S}_{m-1}^2}{\hat{S}_{n-1}^2} \right)$$

Figura 62: Intervalo de confianza para la razón de varianzas

4.3.5 Intervalo de confianza para la diferencia de proporciones

Para calcular la **diferencia entre dos proporciones** p_1 y p_2 , se toman dos muestras de tamaño n_1 y n_2 de dos poblaciones de las que se quiere estudiar la misma característica.

$$\left((\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1 \cdot (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 \cdot (1 - \hat{p}_2)}{n_2}} \right)$$

Figura 63: Intervalo de confianza para la diferencia de proporciones

4.4 Contraste de hipótesis paramétricas

En muchas ocasiones, más que una estimación, lo que buscamos es tomar una decisión sobre si una opción es mejor que otra y ver qué riesgo conlleva equivocarnos. Esto lo conseguimos a través del **contraste o test de hipótesis**, que es el procedimiento por el cual aceptamos o rebatimos una hipótesis estadística, que es una afirmación que se hace sobre una característica de la población.

Si esta hipótesis se refiere a los parámetros (media, desviación...) de la población, se conoce como hipótesis paramétrica y se realizará un contraste de hipótesis paramétrico (que son los que vamos a ver), y si se refiere a otras cosas que no sean los parámetros (como puede ser qué distribución sigue la variable), se conoce como hipótesis no paramétrica y se realizará un contraste de hipótesis no paramétrico.

En este capítulo supondremos que todas las muestras se cogen a través de muestreo aleatorio simple y con un tamaño muy grande. Es muy importante coger una "buena" muestra que sea representativa a la población que queremos estudiar.

Rick: Esto ya me va gustando más, un ayudante de tomar decisiones, que luego haré lo que me dé la gana, pero me va gustando más.

4.4.1 Conceptos esenciales

A continuación veremos una serie de conceptos esenciales del contraste de hipótesis. Todos estos conceptos los entenderemos cuando veamos el ejemplo en el apartado de *cómo realizar un contraste de hipótesis paramétrico* pero primero necesitamos saber qué son, para entender ese ejemplo.

\$nombre: Otra vez lo del sur 😕

Rick: ¡Yo ya he perdido el norte!

4.4.1.1 Hipótesis nula y alternativa

Cuando realizamos un contraste tenemos dos hipótesis:

- **Hipótesis nula (H_0)**: se asume como correcta, como que es la que cumple la población, es la que queremos contrastar, los datos pueden refutarla y no debe rechazarse sin una buena razón por lo que es cierta mientras los datos no prueben con gran certidumbre lo contrario.

- **Hipótesis alternativa (H_1)**: pretendemos contrastarla como opuesta a la hipótesis nula, quiere demostrar lo contrario, los datos pueden ser favorables a ella y no debe aceptarse sin grandes evidencias a su favor.

Si pensamos en un acusado en un juicio:

H_0 : inocente. H_1 : culpable.

Una persona es inocente hasta que se demuestre lo contrario.

\$nombre: mira, Rick, como tú ⚡.

Rick: ¿qué has desayunado hoy? ¿payaso? 😅.

\$nombre: ⚡⚡.

Por lo que se dice que rechazar H_0 pasando a aceptar H_1 es una **decisión fuerte** ya que supone que se ha hallado gran evidencia en favor de H_1 . Mientras que aceptar H_0 es una **decisión débil** ya que sólo supone que, al no haber suficiente evidencia en su contra, no se la rechaza.

4.4.1.2 Error tipo I y error tipo II

		Realidad	
		H_0 es cierta	H_0 es falsa
Decisión que tomamos	Rechazamos H_0	Error de tipo I (α) Falso Positivo	Decisión correcta 😊
	No rechazamos H_0	Decisión correcta 😊	Error de tipo II (β) Falso negativo

Figura 64: Error de tipo I y error de tipo II

A $1 - \beta$ le llamamos **potencia del contraste** y mide la probabilidad de acertar cuando rechazamos H_0 .

Si pensamos en un veredicto a un paciente en un hospital:

H_0 : enfermo. H_1 : sano.

Aquí $1 - \beta$ es la probabilidad de rechazar que el paciente está enfermo cuando es cierto que no lo está, α es la probabilidad de rechazar que el paciente está enfermo cuando lo está y β es la probabilidad de aceptar que está enfermo cuando no lo está. Como podemos ver el **error de tipo I (α)** es más grave que el **error de tipo II (β)** ya que es peor no tratar a un paciente enfermo, por eso siempre intentaremos minimizar el error de tipo I. Lo idóneo sería poder bajar la probabilidad de cometer los dos errores pero si bajamos la probabilidad de cometer un error subimos la de cometer el otro (si ingresáramos a todos los pacientes aumentaría la probabilidad de ingresar a un paciente enfermo pero disminuiría la de mandar a casa a uno sano).

Rick: Pues a mí me parece más importante reducir el error de tipo II.

\$nombre: Rick, especialista en llevar la contraria al mundo, próximamente en los mejores cines.

4.4.1.3 Contrastes unilaterales y contrastes bilaterales

Los contrastes que se refieren a parámetros pueden ser de dos estilos (cuando ponemos θ nos referimos a un parámetro teórico y cuando ponemos θ_0 a un número que puede tomar ese parámetro):

- **Contrastes bilaterales**, un contraste bilateral tiene la forma:

$$H_0: \theta = \theta_0$$

$$H_1: \theta \neq \theta_0$$

- **Contrastes unilaterales**, un contraste unilateral puede tomar las formas:

$$H_0: \theta = \theta_0 \text{ ó } H_0: \theta \geq \theta_0$$

$$H_1: \theta < \theta_0$$

ó

$$H_0: \theta = \theta_0 \text{ ó } H_0: \theta \leq \theta_0$$

$$H_1: \theta > \theta_0$$

4.4.1.4 Región crítica de un contraste

Primero tenemos que elegir un **estadístico de contraste** (luego veremos los que se suelen utilizar) del que sabremos su distribución bajo H_0 .

Para construir la **región crítica o de rechazo** tendremos que buscar el valor de la distribución que sigue el estadístico de contraste (lo buscaremos en la tabla de esa distribución) que deja a la derecha o a la izquierda un área o probabilidad α en el caso unilateral y un área $\alpha/2$ a cada lado en el caso bilateral. Esa área que dejamos a la derecha, izquierda o ambos lados es lo que conocemos como región crítica o de rechazo.

Por lo que si el valor numérico del estadístico de contraste está en la región crítica o de rechazo, rechazaremos H_0 y si no, no.

\$nombre: ¡Por fin ponen un nombre a algo que ayuda a entender su significado!

Rick: .

4.4.1.5 P-valor

El **p-valor** es la probabilidad de tener un valor del estadístico (o discrepancia) igual o mayor que el observado (es decir, más cerca de H_1) cuando H_0 es cierta. Al ser una probabilidad toma valores entre 0 y 1.

Como ya dijimos el error de tipo I (α) es el más importante y nosotros decidimos cómo de grande queremos que sea (cuánto nos queremos arriesgar a que ese error se cometa). Por eso, si:

- $\alpha \geq p\text{-valor} \rightarrow \text{Rechazamos } H_0$
- $\alpha < p\text{-valor} \rightarrow \text{No rechazamos } H_0$

Vamos a dar los valores que normalmente se usan para entenderlo mejor:

- Si el p-valor < 0.01 ó 0.05 (dependiendo del riesgo que estemos dispuestos a asumir) rechazamos H_0 .
- Si el p-valor > 0.05 no rechazamos H_0 .

H_0 no se demuestra, simplemente se mantiene a no ser que haya evidencias fuertes en su contra que entonces la rechazamos y, por eso, siempre decimos que no rechazamos H_0 en vez de aceptamos.

4.4.2 Estadísticos utilizados para hacer contrastes paramétricos

Vamos a ver los **estadísticos** que se suelen utilizar cuando queremos hacer un contraste paramétrico y la distribución que siguen cuando la hipótesis nula es cierta.

4.4.2.1 Estadístico para la media

Siendo $H_0: \mu = \mu_0$

Si **se conoce la desviación típica σ** , el estadístico que utilizaremos será:

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \in N(0,1)$$

Figura 65: Estadístico para la media si se conoce σ

Si **no se conoce la desviación típica σ** , utilizaremos:

$$\frac{\bar{x} - \mu_0}{\hat{S}_{n-1}/\sqrt{n}} \in t_{n-1}$$

Figura 66: Estadístico para la media si no se conoce σ

4.4.2.2 Estadístico para la varianza

Siendo $H_0: \sigma^2 = \sigma_0^2$

Si **se conoce la media μ** , el estadístico que utilizaremos será:

$$\frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} \in \chi_n^2$$

Figura 67: Estadístico para la varianza si se conoce μ

Si **no se conoce la μ** , utilizaremos:

$$\frac{n\hat{S}_n^2}{\sigma_0^2} = \frac{(n-1)\hat{S}_{n-1}^2}{\sigma_0^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} \in \chi_{n-1}^2$$

Figura 68: Estadístico para la varianza si no se conoce μ

4.4.2.3 Estadístico para la diferencia de medias

Sean X e Y variables aleatorias normales $X \rightarrow N(\mu_X, \sigma_X)$ e $Y \rightarrow N(\mu_Y, \sigma_Y)$, partiendo de sus respectivas m.a.s. y siendo $H_0: \mu_X = \mu_Y$ ó $H_0: \mu_X - \mu_Y = 0$, los estadísticos que utilizamos son (Nota: el símbolo \approx indica que la distribución, en lugar de ser exacta, es una aproximación):

- **Si conocemos σ_X y σ_Y**

$$\frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \in N(0,1)$$

- **Si no conocemos la desviaciones pero suponemos:**

1. Que son iguales (las desviaciones)

$$\frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{(n-1) \cdot \hat{S}_{n-1}^2 + (m-1) \cdot \hat{S}_{m-1}^2}{n+m-2} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}}} \in t_{n+m-2}$$

2. Que el tamaño de muestra es grande ($n, m \geq 30$)

$$\frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\hat{S}_{n-1}^2}{n} + \frac{\hat{S}_{m-1}^2}{m}}} \approx N(0,1)$$

3. Que el tamaño de muestra es pequeño ($n, m < 30$)

$$w = \frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\hat{S}_{n-1}^2}{n} + \frac{\hat{S}_{m-1}^2}{m}}} \approx t_{n+m-2-\Delta},$$

siendo Δ el entero más próximo a

$$\frac{\left((m-1) \cdot \frac{\hat{S}_{n-1}^2}{n} - (n-1) \cdot \frac{\hat{S}_{m-1}^2}{m} \right)^2}{(m-1) \left(\frac{\hat{S}_{n-1}^2}{n} \right)^2 + (n-1) \left(\frac{\hat{S}_{m-1}^2}{m} \right)^2}.$$

Figura 69: Estadístico para la diferencia de medias

4.4.2.4 Estadístico para la razón de varianzas

Siendo $H_0: \sigma_X/\sigma_Y = cte$, el estadístico a utilizar es:

$$\frac{\hat{S}_{n-1}^2 \sigma_Y^2}{\hat{S}_{m-1}^2 \sigma_X^2} \in F_{n-1, m-1}$$

Figura 70: Estadístico para la razón de varianzas

Este test nos sirve para saber si podemos considerar que las **varianzas de dos variables son iguales o no, para contrastar después la igualdad de las medias**. El contraste quedaría con la razón de H_0 igualada a uno y la de H_1 distinta de 1.

4.4.2.5 Estadístico para una proporción y su diferencia

Si $H_0: p = p_0$, el **estadístico para una proporción** será:

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \approx N(0,1)$$

Figura 71: Estadístico para una proporción

Si $H_0: p_1 = p_2$ ó $H_0: p_1 - p_2 = 0$, utilizaremos el **estadístico para la diferencia de proporciones**:

$$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \approx N(0,1)$$

Figura 72: Estadístico para la diferencia de proporciones

4.4.3 Cómo realizar un contraste de hipótesis paramétrico

En la vida real lo que haremos es tomar una m.a.s. de la población y estimar el parámetro. Para ver si el parámetro es muy distinto a lo que hemos estimado necesitaremos alguna medida de discrepancia o estadístico de contraste. También plantearemos un contraste de hipótesis con una hipótesis nula y una alternativa. Para ayudarnos a tomar esta decisión utilizaremos la región crítica, si está en ella rechazaremos H_0 y si no, no, también utilizaremos el p-valor, que si es menor a α rechazaremos la hipótesis nula y si no, no.

\$nombre: Me hace gracia cuando ponen cosas como en la vida real, ¿en cuál estamos ahora?

4.4.3.1 Misión 29

Morty: Oye, me han dicho que un hechizo de enamoramiento funciona en más de 70 ocasiones de las 100 que se utiliza, pero Squanchy me ha dicho que a él le funciona 68 de cada 100 veces.

\$nombre: Mmm, vale, pues podemos hacer un contraste a ver si es cierto.

Rick: ¿Cuál será el p-valor?

Solución correcta 29:

¡UO. QC!

Solución incorrecta 29:

Rick y \$nombre: ¡Somos un verdadero desastre!

Solución:

Definimos p = "proporción de veces que funciona el hechizo". Por lo que nuestro contraste de hipótesis será:

$$\begin{aligned} H_0: p &\geq 0.7 \\ H_1: p &< 0.7 \end{aligned}$$

Vamos a utilizar el estadístico de contraste para la proporción que ya hemos visto que sigue una $N(0,1)$ si H_0 es cierta y que da -0.44, por lo que el p-valor será **0.33**, por lo que no rechazaremos la hipótesis nula.

5. Regresión Lineal Simple

En el apartado de estadística descriptiva vimos el estudio de una sola variable, esto muchas veces se nos queda corto ya que lo que nos puede interesar es la relación entre distintas variables. Por eso ahora introducimos unos conceptos para tener una base de cómo ver relaciones entre variables.

Existen muchos tipos de regresiones pero nosotros nos centraremos sólo en la **regresión lineal simple** (es lineal y la variable independiente Y solo depende de una variable X) ya que este tema puede alargarse hasta el infinito y solo queremos dar unas pinceladas a esto.

\$nombre: ¡Hasta el infinito y más allá!

Rick: .

5.1 Covarianza

La **covarianza** (S_{xy}) mide la relación lineal entre las variables y su sentido. Es simétrica ya que S_{xy} es lo mismo que S_{yx} . Lo malo de la covarianza es que tiene unidades, por eso veremos la correlación lineal que no las tiene.

\$nombre: Todo tiene siempre una parte mala.

Rick: Todo no, \$nombre, todo no...

$$Cov(X, Y) = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Ecuación 31: Covarianza

La relación es positiva si $S_{xy} > 0$, es decir, si una variable crece, la otra también.

La relación es negativa si $S_{xy} < 0$, por lo que, si una variable crece, la otra decrece.

No existe relación si $S_{xy} = 0$.

Para calcularlo en R, utilizamos la orden `cov(x,y)`.

5.2 Correlación lineal

El coeficiente de **correlación lineal o de Pearson**, al igual que la covarianza, mide la relación lineal entre las variables y tiene su mismo signo pero es una variable adimensional y sus valores están entre -1 y 1.

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}}$$

Ecuación 32: Correlación lineal

Si el coeficiente es mayor que 0, entonces la correlación es positiva, si es menor, es negativa, y si es 0, no existe relación lineal, pero, ¡cuidado!, sí puede existir otro tipo de relación entre las variables.

Al cuadrado de este coeficiente se le conoce como coeficiente de determinación (R^2). Toma valores entre 0 y 1, y cuanto mayor es, mayor es la relación lineal entre las variables.

Rick: Lógicamente.

5.3 Diagrama de dispersión o nube de puntos

El **diagrama de dispersión o minr de puntos** se utiliza para estudiar y medir la relación entre dos variables cuantitativas. Para construirlo colocamos una en el eje X y la otra en el Y y se representa el valor (x_i, y_i) con un punto, por eso también es conocido como nube de puntos. El punto que viene dado por la media de X y la de Y se conoce como punto de gravedad. Para hacerlo con R ponemos $plot(x,y)$ donde x e y son dos vectores numéricos.

Rick: Minr... Nube.... Vale, sí, me pillaste, antes dije que lo sabía pero era mentira...

\$nombre: ¡Lo sabía!

Rick: El código es que están escribiendo cada letra con la letra de la derecha del teclado español del ordenador, ¡por eso me costaba tanto verlo!

\$nombre: Excusas, excusitas jajaja

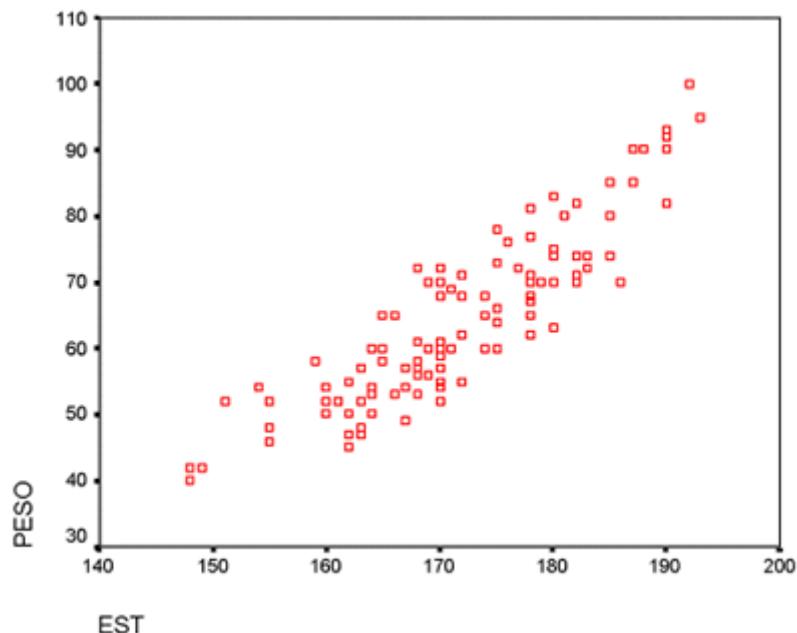


Figura 73: Nube de puntos

5.4 Recta de regresión

Sean $(x_1, y_1), \dots, (x_n, y_n)$ los valores observados, los representamos con una nube de puntos. Sea $y = a + bx$ la recta de ajuste de los valores de Y en función de los de X , llamada **recta de regresión** nos servirá para poder relacionar dos variables de manera lineal y poder predecir un nuevo valor de Y para uno nuevo de X .

Rick: ¡Reeeeeeeeeeeeec - tos!

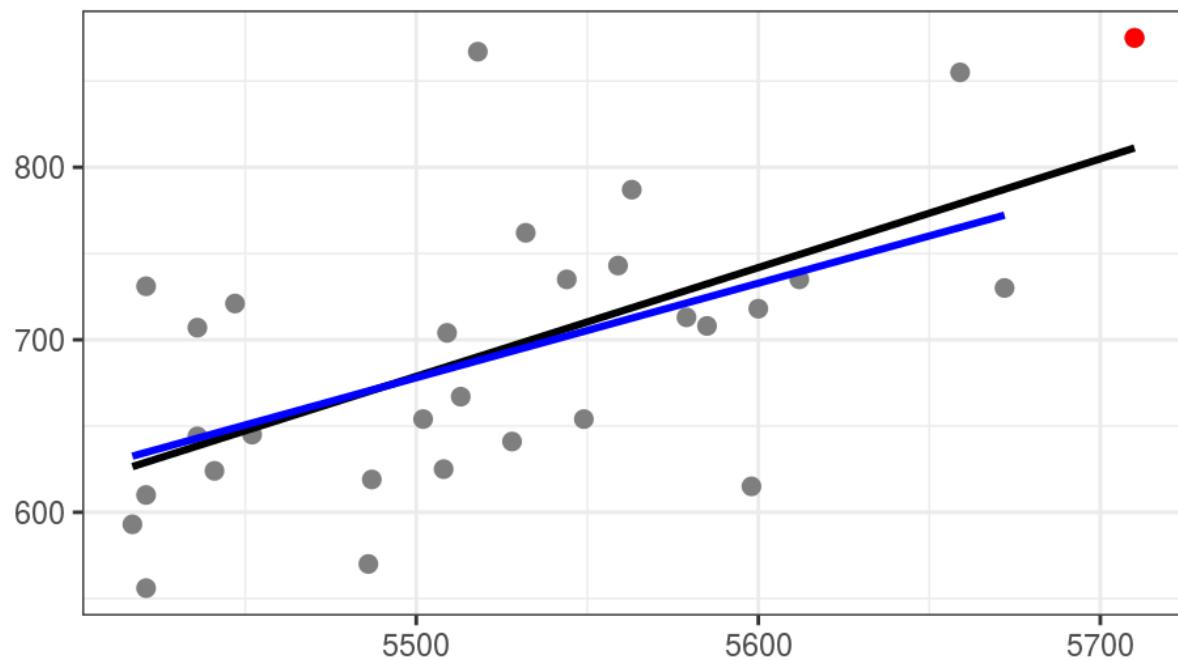


Figura 74: Recta de regresión

5.5 Mínimos cuadrados

Pero claro, podría haber muchas rectas de regresión posibles, una de las técnicas utilizadas para encontrar la mejor recta de regresión es **mínimos cuadrados** que suma todos los residuos (distancia entre el verdadero punto y su valor predicho en la recta, son las líneas rojas en la figura 75) elevados al cuadrado para que no se compensen los positivos con los negativos.

$$F(a, b) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - (a + bx_i))^2$$

Recta de regresión y residuos

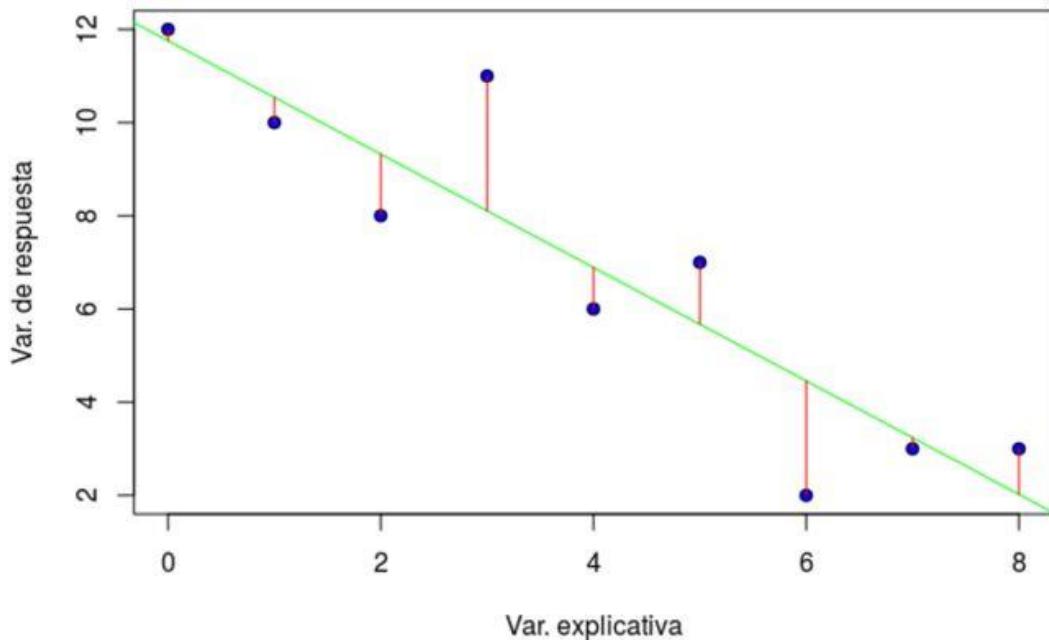


Figura 75: Mínimos cuadrados

Los parámetros a y b son los valores que minimizan la función.

$$a = \bar{y} - b\bar{x}, \quad b = \frac{S_{xy}}{S_x^2}$$

Por lo que la recta de regresión quedará:

$$y - \bar{y} = \frac{S_{xy}}{S_x^2} (x - \bar{x})$$

5.6 Problemas de la regresión lineal simple

Obviamente la regresión lineal simple se nos queda corta porque muchas veces la variable Y dependerá de más de una variable y muchas veces no será ni lineal. También da problemas porque para poder aproximar un modelo por una recta de regresión los residuos deben ser independientes; para cada valor de X , la varianza de los residuos debe ser la misma, y tiene que seguir una distribución normal de media 0. Hay maneras de mirar si estas condiciones se cumplen, y de realizar transformaciones (box-cox, logarítmica...) si no se cumplen. También se podría elegir otro tipo de regresión que funcionara mejor con nuestros datos (regresión logística, no lineal...).

Rick: Ya decía yo que era demasiado fácil para quedarse ahí la cosa jajaja

\$nombre: ¡Qué ganas de saber más!

6. Otros

6.1 Final

Voz: ¡Enhorabuena a todos los participantes por su gran esfuerzo! Que sepáis que habéis ayudado mucho a que la estadística llegue a los seres de vuestros planetas, ya que con vuestra experiencia se ha creado un chatbot (como Alexa o Siri) para poder ayudar a que la gente aprenda de vuestra experiencia. Además, todos los participantes tenéis un viaje ahora al mundo de la estadística para descubrirle y seguir aprendiendo sobre este mundo tan extenso y en el que no habéis hecho nada más que empezar. ¡Muchas gracias a todos! ¡Ha sido una experiencia muy divertida para todo el universo seguiros!

\$nombre: ¡Guau, ha sido súper emocionante! ¡No me he dado ni cuenta de que estaba aprendiendo, me lo he pasado genial!

Rick: Y, ¡vamos a poder descubrir el mundo de la estadística! Eso sí, nada más volvamos a casa hago la fiesta que todos estamos deseando con nuestros amigos, ¡nos lo merecemos!

6.2 Siguiente

Voz: Mira que os dije que si escribíais literalmente siguiente en vez de clicar en el botón siguiente pasaría esto, pero claro, os sentís rebeldes al no seguir las normas ehhh y mira que son pocas; como veo que no las tenéis claras os las voy a recordar.

6.3 Ayuda

Voz: Rick, \$nombre, recordad que estoy aquí para ayudaros en todo lo que necesitéis pero como tengo que ser imparcial y no puedo ayudar más a un planeta que a otro, lo único que puedo deciros es lo que ya sabéis que estáis en el *Concurso Intergaláctico* para demostrar

que vuestro planeta es el mejor aprendiendo estadística, clicando el botón *Normas* o escribiendo *Normas* en cualquier momento podréis recordar las normas del concurso y ver el contenido del mismo (estadística descriptiva, probabilidad, modelos probabilísticos, inferencia estadística y regresión lineal simple) que es lo que pueden ver todos los participantes y hasta donde os puedo ayudar para ser justa con todos.

6.4 En otras cosas

Voz: Nos os distraigais, recordad que estáis en el *Concurso Intergaláctico* de la televisión interdimensional y os está viendo todo el mundo, aquí estamos a lo que estamos, ¡la estadística! 😢 (Si en realidad estáis hablando de estadística perdonadme, simplemente eso que queréis saber no entra dentro de este concurso, lo siento, aunque lo tendré en cuenta para futuras ediciones 😊).

5. ENCUESTA DE OPINIÓN Y RESULTADOS

En este capítulo se va a hablar sobre los resultados de la encuesta (se puede ver la encuesta completa en el Anexo III) realizada tanto a profesores como alumnos de la carrera de estadística o de otras donde enseñan o estudian estadística.

Esta encuesta tiene el objetivo de conocer la opinión tanto de personas que conocen ampliamente la estadística (estudiantes, titulados o profesores de estadística) como de personas que han estudiado o están estudiando una carrera de la rama de ingenierías, económicas o salud, donde dan la asignatura de estadística, para saber si este asistente tiene un buen contenido de estadística, si es fácil de usar, si la historia y las misiones son divertidas y si creen que es útil y cumple su objetivo de que con él, las personas con una base matemática puedan iniciarse en el mundo de la estadística. También tiene el objetivo de conocer cómo mejorarían ellos cada uno de los aspectos de este asistente para conseguir dicho fin.

Primero se pregunta al encuestado si es enseñante de estadística o de qué rama es su carrera, luego si está familiarizado con el uso de asistentes.

Para conocer si el asistente ha resultado fácil de usar y su contenido es adecuado, se han diseñado unas respuestas cerradas que dan cuatro opciones, una a favor, otra a favor pero un poco menos, otra en contra y otra en contra pero menos, buscando que el encuestado elija la opción con la que se siente más identificado. No se ha puesto una quinta opción intermedia para saber la opinión de las personas, porque con cuatro opciones, aunque no te sientas completamente identificado con una de ellas por lo menos habrá una que represente más o menos tu idea.

A continuación de cada una de estas preguntas se ha dejado una pregunta abierta sobre las anteriores para saber cómo el encuestado pueda dar su opinión sobre esa pregunta y decir cómo mejoraría este aspecto.

Se han representado las preguntas con opciones a través de representaciones gráficas para poder ver mejor el resumen de las opiniones y se han sacado unas propuestas de mejora a partir de las ideas que los encuestados han ido dejando en las preguntas abiertas.

Al final del todo se ha dejado una pregunta abierta para que los encuestados puedan decir qué opinan de manera general sobre el asistente y puedan expresar alguna idea que no se ha preguntado explícitamente en la encuesta.

PREGUNTA 1: De las 108 personas que contestaron a la encuesta, en la figura 76, podemos ver el porcentaje de personas que eran profesores de estadística, alumnos de la carrera de estadística o alumnos de otras carreras que tienen la asignatura de estadística.

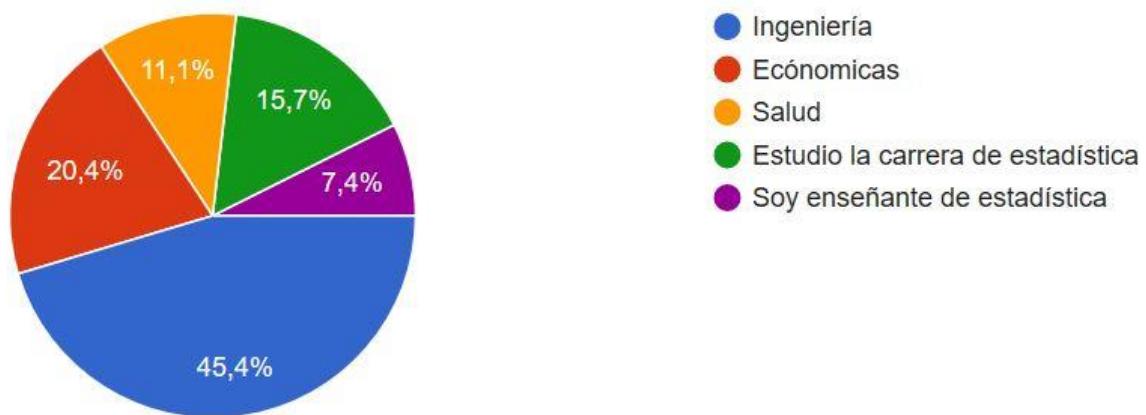


Figura 76: Pregunta 1 de la encuesta

PREGUNTA 2: Más del 30% de los encuestados nunca utilizan asistentes conversacionales pero todos saben lo que son, más de la mitad de ellos no lo usan nunca, sólo el 5.6% de ellos los usan bastante, y ese mismo porcentaje los usa a diario, como podemos ver en la figura 77.

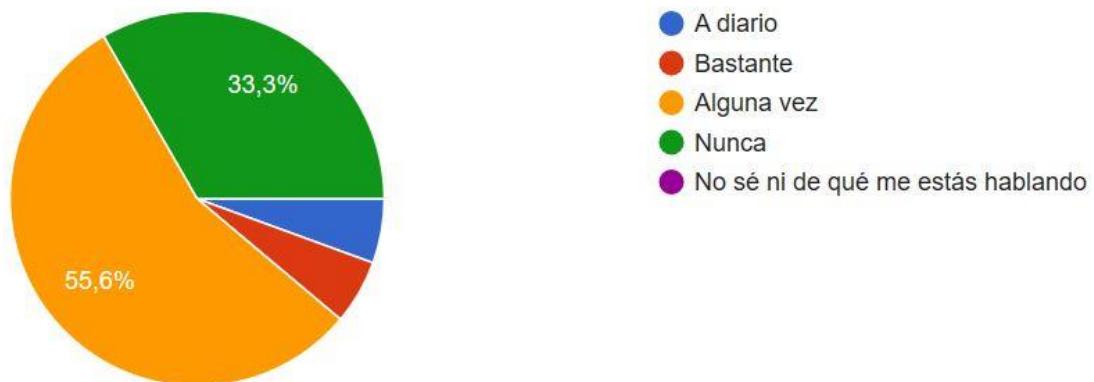


Figura 77: Pregunta 2 de la encuesta

PREGUNTA 3: En la figura 78 se ve que las normas del Concurso Intergaláctico (instrucciones sobre como utilizar el asistente) han ayudado mucho al 53.7% de los encuestados, también han ayudado al 39.8% aunque creen que se podrían mejorar. Tan solo el 6.5% opina que ayudan bastante poco, y es de destacar que no les han parecido inútiles a nadie.

Propuestas de mejora recogidas: Los comentarios sobre esta pregunta es que las normas deberían ser más cortas y concisas y que deberían simplificarse porque si no, la gente no las lee.

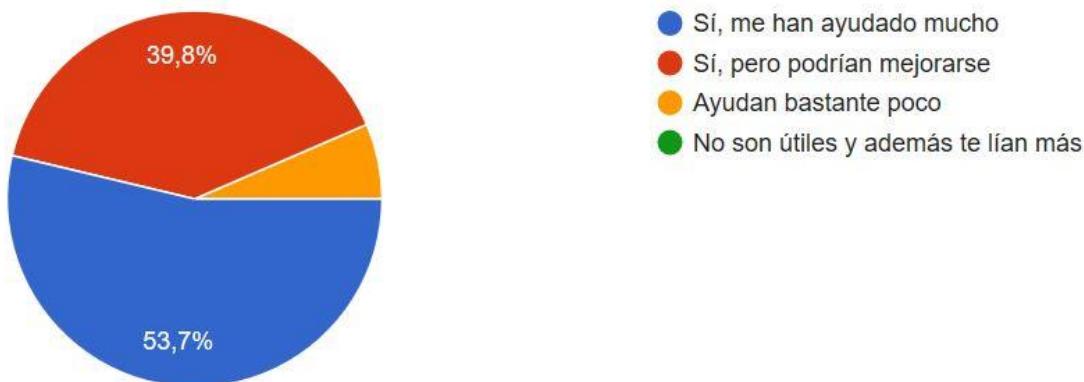


Figura 78: Pregunta 3 de la encuesta

PREGUNTA 4: Como podemos ver en la figura 79, el 79.6% de los encuestados piensan que el contenido del asistente conversacional es adecuado. El 16.7% piensa que no está mal pero cambiarían cosas sobre ese contenido. Y tan solo el 3.7% de los encuestados piensan que debería enseñarse otro contenido distinto a estadística descriptiva, probabilidad, inferencia y regresión a una persona que quiere iniciarse en el mundo de la estadística.

Propuestas de mejora recogidas: Los comentarios escuchados es que habría que dar algo más de contenido en algún apartado, que se podrían explicar más ciertos conceptos y soluciones a misiones y que se deberían poner links con más información sobre lo que se explica para que la gente pueda ampliar el contenido del asistente sin hacérselo pesado.

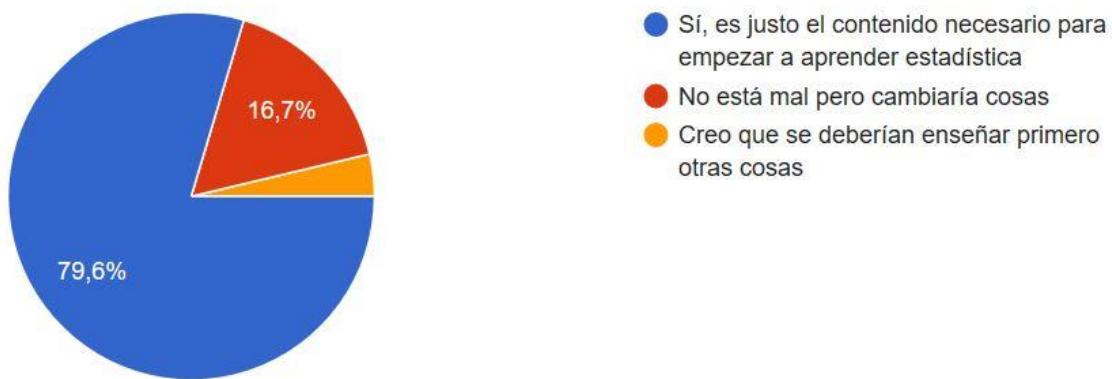


Figura 79: Pregunta 4 de la encuesta

PREGUNTA 5 (Figura 80): Respecto a la pregunta sobre si la historia y las misiones ayudan a entender mejor los conceptos, un 59.3% de los encuestados piensa que sí. Un 35.2% piensa que hace más entretenida la tarea de aprender aunque no ayude mucho con la teoría. Tan solo un 2.8% de los encuestados opina que la historia y las misiones no ayuda a aprender los conceptos de estadística.

Propuestas de mejora recogidas: Los comentarios dados es que el texto se hace demasiado largo y la historia lo alarga aún más, que la gamificación se conseguiría hacer más llamativa con otros sistemas más interactivos y que habría que integrar más la historia con la teoría.



Figura 80: Pregunta 5 de la encuesta

PREGUNTA 6 (Figura 81): Respecto a la pregunta sobre la facilidad de uso del asistente conversacional, un 77.8 % piensa que es muy fácil de usar, el 19.4% que se puede aprender a usar, tan solo el 2.8% piensa que se puede mejorar bastante y no hay nadie que piense que el asistente es muy difícil de usar.

Propuestas de mejora recogidas: Por los comentarios parece que muchos usuarios no están acostumbrados a utilizar asistentes, una persona dice que al principio no sabía si él estaba poniendo preguntas o respuestas, otro que es lioso y otro que preferiría ver el contenido del asistente como un diagrama de árbol. Hay una persona que sí parece que está acostumbrada porque pone que funciona muy bien con todo menos con errores gramaticales.

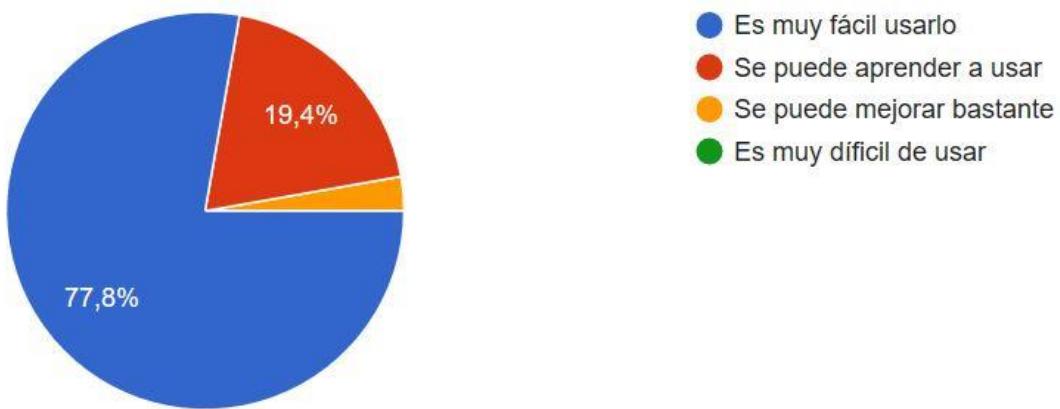


Figura 81: Pregunta 6 de la encuesta

PREGUNTA 7: A la pregunta sobre si les gusta que se les dé opciones para guiarles en la conversación y dar un posible orden a seguir, como se puede ver en la figura 82, el 56.5% dice que si no, habrían estado muy perdidos. El 38.9% opina que no está mal. El 3.7% piensan que hay maneras mejores de guiar al usuario y el 0.9% de los encuestados dice que preferiría que se dejara libre sin dar opciones ni un posible orden a seguir.

Propuestas de mejora recogidas: Comentan que se podría hacer una historia lineal o seguir un orden según lo que vaya sucediendo en la historia y las misiones en vez de por el contenido de estadística.



Figura 82: Pregunta 7 de la encuesta

PREGUNTA 8: A la pregunta sobre si este chatbot podría ayudar a la gente a aprender estadística o entenderla mejor, el 51.9% de los encuestados opina que podría ayudar a mucha gente, el 46.3% dice que podría ayudar a determinadas personas, y el 1.9% de los encuestados concluyen que se debería realizar el asistente de otra forma para conseguir el objetivo de que la gente pueda aprender estadística con él. Destaca que nadie opina que lia más de lo que enseña. Esto se puede ver en la figura 83.

Propuestas de mejora recogidas: Dicen que se debería dar una explicación más completa para ciertos conceptos, que se podría hacer más conciso. Una persona señala que el tema de gamificación en ciertas edades está bien pero cuando la gente es más adulta puede dar la impresión de lentitud y no gustar. Hay quienes proponen guardar el progreso, dar recompensas o puntos por cada misión acertada, poner más misiones. Alguien indica que una persona que se acaba de iniciar en estadística necesitaría más contenido pero para alguien que quiere refrescarla está bien y que esto puede ayudar a las personas que se enganchan con juegos o una buena historia pero no a todo el mundo.



Figura 83: Pregunta 8 de la encuesta

Como ya se ha dicho al principio de este apartado, al final de la encuesta se da la posibilidad de hacer cualquier tipo de comentario sobre la herramienta en pregunta abierta. Los comentarios aportados son bastante positivos, los encuestados dicen que la idea les parece buena y que esto puede convertir en entretenido algo que veían como un rollo, que sí se podrían mejorar aspectos (poner más materia, meter más misiones, explicar más las cosas o poner colores distintos según hable un personaje u otro) pero que en general es una idea buena y se ha llevado bien a cabo.

6. CONCLUSIONES Y TRABAJO FUTURO

La estadística es una materia que se ha hecho imprescindible en cualquiera de las áreas de la vida actual, por lo que tiene que ir apareciendo en los contenidos de matemáticas desde edades más tempranas de lo que se está haciendo ahora y con más peso en la misma.

Con la revolución digital que estamos viviendo actualmente es necesario que se utilicen estas nuevas herramientas digitales que tenemos al alcance de la mano para buscar nuevas formas de enseñanza que sean más atractivas para las nuevas generaciones, ya que son nativos digitales y todo lo que está hecho alrededor de este mundo, de por sí, les llama más la atención.

Como conclusión de este trabajo, podemos decir que los objetivos planteados han sido alcanzados. Se ha creado un asistente conversacional, Rick y tú, que es una herramienta digital que permite una iniciación en el mundo de la estadística, de una forma amena y divertida.

Esto está refrendado por los resultados de la encuesta que fue elaborada para obtener realimentación sobre la herramienta desarrollada, y que fue contestada por 108 personas. Hemos podido constatar que a un alto porcentaje de los encuestados les parece adecuado el contenido del asistente conversacional, que la historia y las misiones ayudan a entender los conceptos o al menos hacen más divertido el aprendizaje, que es una herramienta fácil de usar, les gusta cómo está construido el flujo de aprendizaje, y que puede ser una buena herramienta para usar en el campo de la estadística como iniciación en la misma.

Aunque al principio se creó este asistente con el objetivo de utilizarlo en enseñanzas superiores, ya que actualmente todavía no se estudia mucha estadística hasta llegar hasta este nivel, en un futuro en el que en las enseñanzas medias tengan más contenido de estadística se podría utilizar en ellas. También podría utilizarlo cualquier persona que tenga una base matemática, quiera iniciarse en el mundo de la estadística y no haya tenido esta formación académica antes, o personas que quieren recordar esta parte de la estadística.

Como líneas de trabajo futuro se plantean las siguientes, hay tanto sugerencias por parte de la autora y tutoras del trabajo sugerencias que han dado en la encuesta los encuestados:

- A nivel técnico, se podría conectar la instancia de Watson Assistant con otros servicios de Watson para añadirle más funcionalidad, como discovery (un buscador inteligente) que se podría cargar con todos los apuntes de la asignatura y para que el usuario pueda preguntar en el asistente dudas, le enseñaría el párrafo donde se contesta la misma y le diría en qué documento se encuentra. Además, se podría conectar con Speech to text y Text to speech para conseguir que el usuario pudiera hablar al asistente y él mismo le pudiera contestar por voz.
- En cuanto al contenido de estadística, se podrían seguir añadiendo más explicaciones, se podrían poner enlaces a páginas donde se pueda profundizar los conceptos que se van explicando, introducir más misiones (ejercicios) y explicar más como se resuelven y ampliar

con otras partes de la estadística más avanzadas como series temporales o análisis de datos.

- En cuanto al aspecto pedagógico y de aprendizaje mediante gamificación, se podría mejorar la parte de la historia y las misiones para que haya más contenido lúdico y hacer la herramienta más entretenida.

BIBLIOGRAFÍA

- Acuña, M. (30 de Junio de 2018). *evirtual plus*. Obtenido de <https://www.evirtualplus.com/storytelling-en-la-educacion/> (Último acceso: 27/06/2020).
- Barba, L. (2016). Regresión y anova.
- Derivando. (2020). *Youtube*. Obtenido de https://www.youtube.com/watch?v=7uzx6D_0V7M (Último acceso: 27/06/2020).
- elgranporque*. (31 de Enero de 2017). Obtenido de <http://www.elgranporque.com/por-que-rick-y-morty-se-llama-asi/> (Último acceso: 27/06/2020).
- Emojipedia*. (2020). Obtenido de <https://emojipedia.org> (Último acceso: 27/06/2020).
- González, T. (2014). Estadística descriptiva.
- IBM. (2020). *Cognitive class*. Obtenido de <https://cognitiveclass.ai/> (Último acceso: 27/06/2020).
- IBM. (2020). *IBM Cloud*. Obtenido de <https://www.ibm.com/cloud> (Último acceso: 27/06/2020).
- indifferentlanguages*. (2020). Obtenido de <https://www.indifferentlanguages.com/es/palabra/aleluya> (Último acceso: 27/06/2020).
- korosenai*. (9 de Abril de 2018). Obtenido de <https://www.korosenai.es/series/rick-y-morty/historia-rick-c-137/> (Último acceso: 27/06/2020).
- La investigación*. (2020). Obtenido de <https://lainvestigacion.com/muestreo/probabilidad/aleatorio-simple/> (Último acceso: 27/06/2020).
- lascosasquenoshacenfelices*. (Último acceso: 27/06/2020).
- (2020). Obtenido de <https://www.lascosasquenoshacenfelices.com/category/series/rick-y-morty-analisis-de-los-capitulos/> (Último acceso: 27/06/2020).
- Línea de código*. (2020). Obtenido de <http://lineadecodigo.com/html/alfabeto-griego-en-html/> (Último acceso: 27/06/2020).
- Malvido, A. (30 de Julio de 2019). *Cursos femxa*. Obtenido de <https://www.cursosfemxa.es/blog/gamificacion-estrategia-educativa> (Último acceso: 27/06/2020).
- Marcos, N. (7 de Noviembre de 2017). *El país*. Obtenido de https://elpais.com/cultura/2017/11/06/televisión/1509992020_145523.html (Último acceso: 27/06/2020).
- Martín, E. (17 de Enero de 2020). *Emprenderalia*. Obtenido de <https://www.emprenderalia.com/que-es-la-gamification/> (Último acceso: 27/06/2020).
- Menéndez, J. A. (2015). Modelos probabilísticos.
- Páez, M. Á. (31 de Enero de 2020). *Xataka*. Obtenido de <https://www.xataka.com/cine-y-tv/un-fisico-nos-explica-por-que-rick-y-morty-es-fantastica-para-ensenar-y-aprender-ciencia> (Último acceso: 27/06/2020).
- Quintela, A. (2014). *El estadístico accidental*.
- Quintela, A. (2020). *Página de Alejandro Quintela*. Obtenido de <https://www.alejandroquintela.com/> (Último acceso: 27/06/2020).
- respuestas.me*. (2020). Obtenido de <https://respuestas.me/q/que-bebida-alcoholica-hay-en-el-frasco-de-cadera-de-rick-62112148373> (Último acceso: 27/06/2020).

- Rodríguez, P. (2015). Probabilidad.
- Rodríguez, P. (2016). Inferencia.
- Socialmood. (2020). *40 de fiebre*. Obtenido de <https://www.40defiebre.com/que-es/storytelling> (Último acceso: 27/06/2020).
- Telegram. (2020). Obtenido de <https://core.telegram.org/bots/api#sendmessage> (Último acceso: 27/06/2020).
- unam. (2020). Obtenido de
https://www.ccg.unam.mx/~vinuesa/R4biosciences/docs/Tema9_regresion.html
(Último acceso: 27/06/2020).
- UVA. (Marzo de 2020). *Universidad de Valladolid*. Obtenido de
<http://www.uva.es/export/sites/uva/2.docencia/2.01.grados/2.01.02.ofertaformativagrados/2.01.02.01.alfabetica/index.html> (Último acceso: 27/06/2020).
- webusable. (2020). Obtenido de <http://webusable.com/CharsExtendedTable.htm> (Último acceso: 27/06/2020).
- Wikipedia. (2020). *Wikipedia*. Obtenido de https://es.wikipedia.org/wiki/Distribución_exponencial (Último acceso: 27/06/2020).
- Wikipedia. (2020). *Rick y Morty*. Obtenido de https://es.wikipedia.org/wiki/Rick_y_Morty (Último acceso: 27/06/2020).

ANEXO I - CÓDIGO INICIAL DEL NODO FUNCTION EN NODE-RED

```
general = msg.payload.output.generic
msg.payload.content = []
for(let i = 0; i < general.length; i++){
    if(general[i].response_type == "text"){
        msg.payload.content += general[i].text +"\n"
        msg.payload.chatId=msg.chatId
        msg.payload.type="message"
    }
    if(general[i].response_type == "image"){
        if(general[i].title){
            msg.payload.content += general[i].title +"\n"
        }
        if(general[i].description){
            msg.payload.content += general[i].description +"\n"
        }
        msg.payload.content += general[i].source +"\n"
        msg.payload.chatId=msg.chatId
        msg.payload.type="photo"
    }
    if(general[i].response_type == "option"){
        opciones = general[i].options
        keyboard_array = []
        for(let j = 0; j < opciones.length; j++){
            keyboard_array.push([{"text: general[i].options[j].label}])
        }
        msg.payload.chatId=msg.chatId
        if(general[i].title){
            msg.payload.content+= general[i].title +"\n"
        }
        if(general[i].description){
            msg.payload.content += general[i].description +"\n"
        }
        msg.payload.type="message"
        msg.payload.options={
            reply_to_message_id:msg.messageId,
            reply_markup:{
```

```
        keyboard: keyboard_array,  
        'resize_keyboard': true,  
        'one_time_keyboard': true  
    }  
}  
}  
}  
return msg;
```

ANEXO II - ACCESO AL ASISTENTE RICK Y TÚ

Para acceder al asistente se puede acceder por el link que genera automáticamente Watson Assistant o por Telegram.

Para acceder desde Telegram se pone en el buscador de Telegram el nombre del asistente: *Rick_y_tu_bot*, como pondríamos el nombre de cualquier contacto que quisiéramos buscar en él.

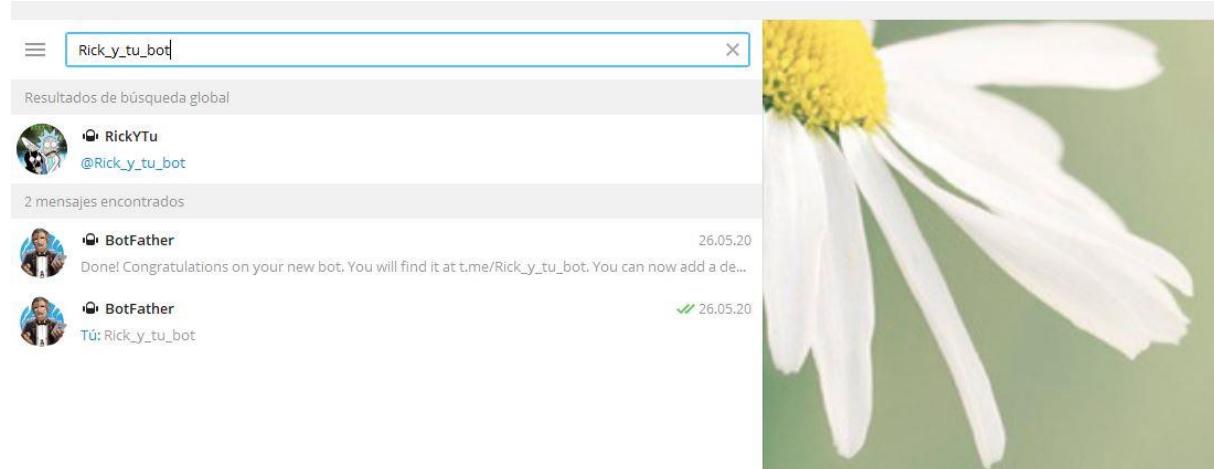


Figura 84: Acceso a *Rick y tú* en Telegram

Clicamos en él (es el primero que aparece en la figura 84) y le damos a Iniciar (el botón que se puede ver en la figura 85).

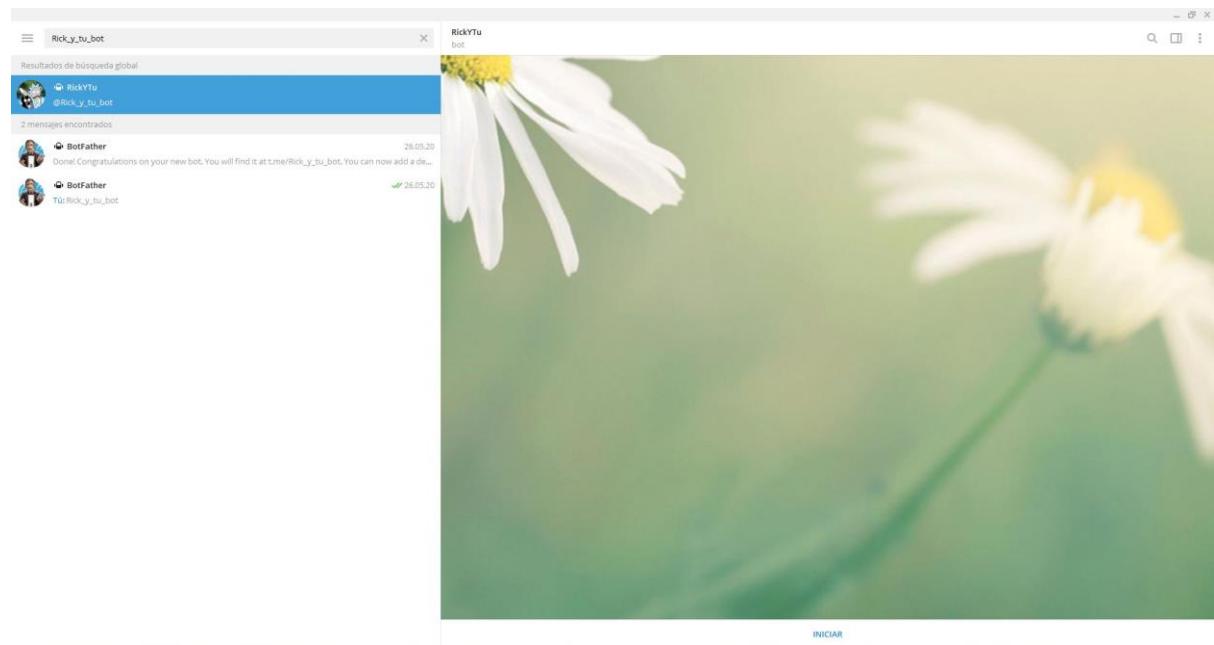


Figura 85: Iniciar *Rick y tú* en telegram

Y nos aparecerá el mensaje del nodo Bienvenida (Figura 86).

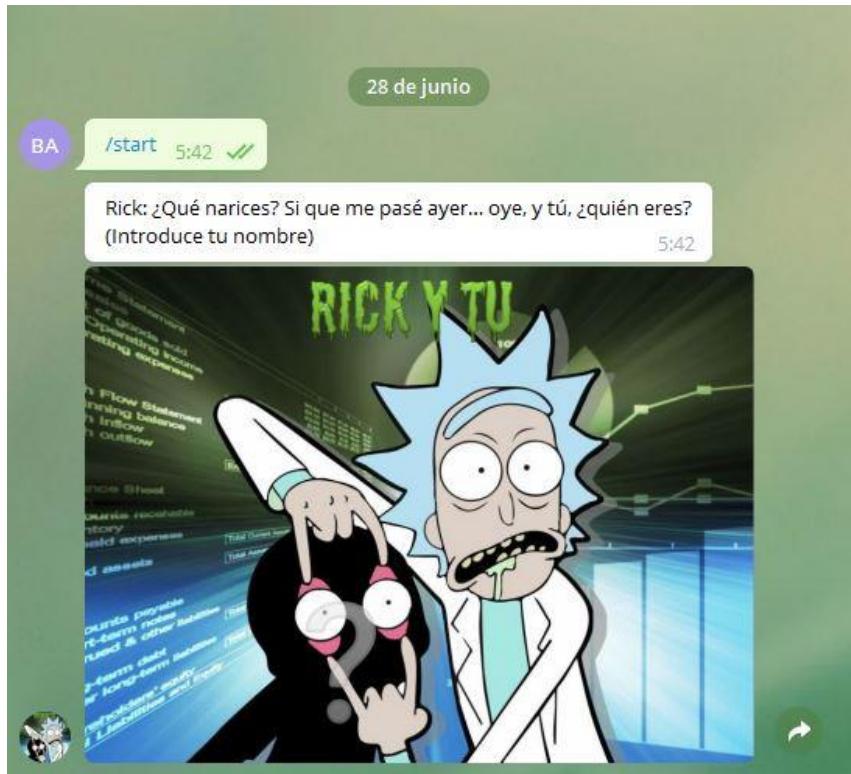


Figura 86: Mensaje de bienvenida de Rick y tú

Metemos nuestro nombre y lo guarda para poder utilizarlo durante toda la conversación, como se puede ver en la figura 87. Las opciones en Telegram salen debajo de la barra donde se escribe, solo con clicar en ese botón se escribe lo que pone en el botón para no tener que escribirlo a mano, guiar la conversación y que sea más cómodo y rápido de usar que si tienes que poner todo escribiendo.

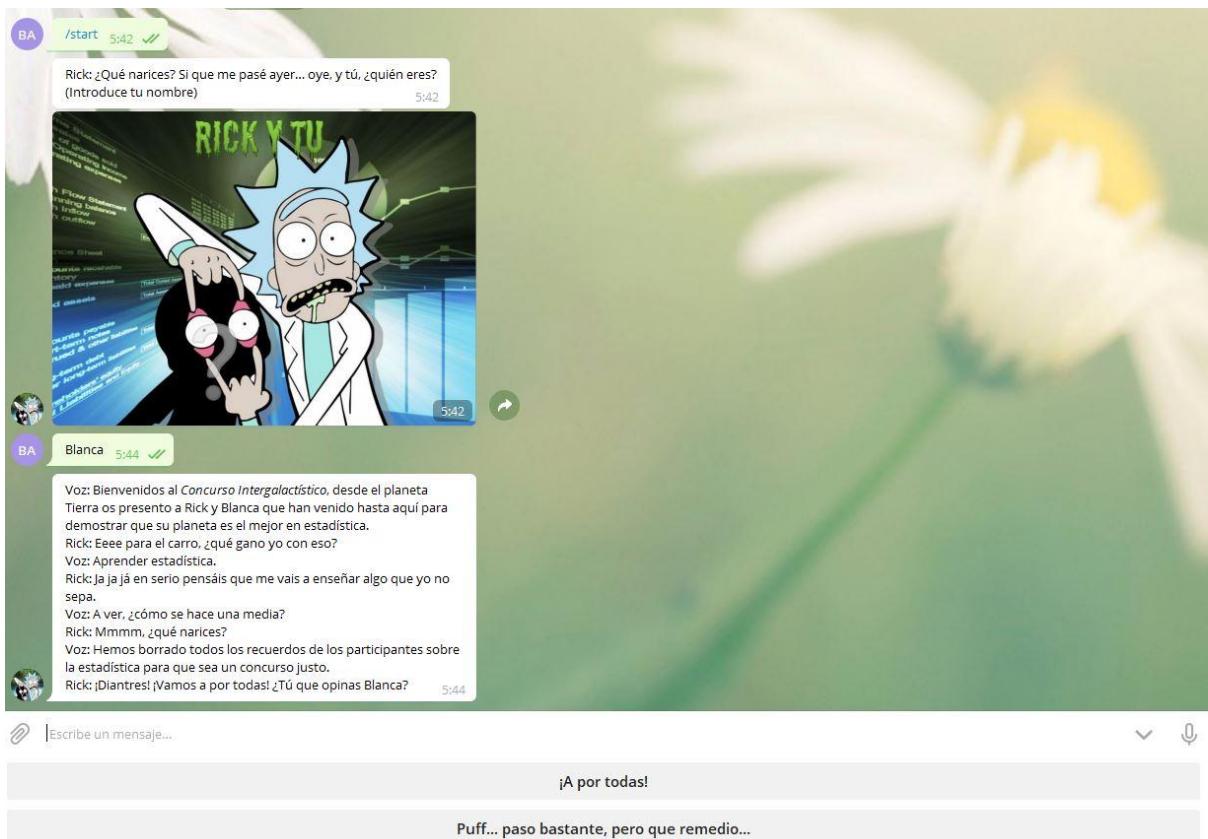


Figura 87: El asistente guarda el nombre del usuario

Clicamos en el botón *¡A por todas!* y después en el de *Eso, ¿cuándo empezamos? ¡Qué tengo unas ganas!* (Figura 88).

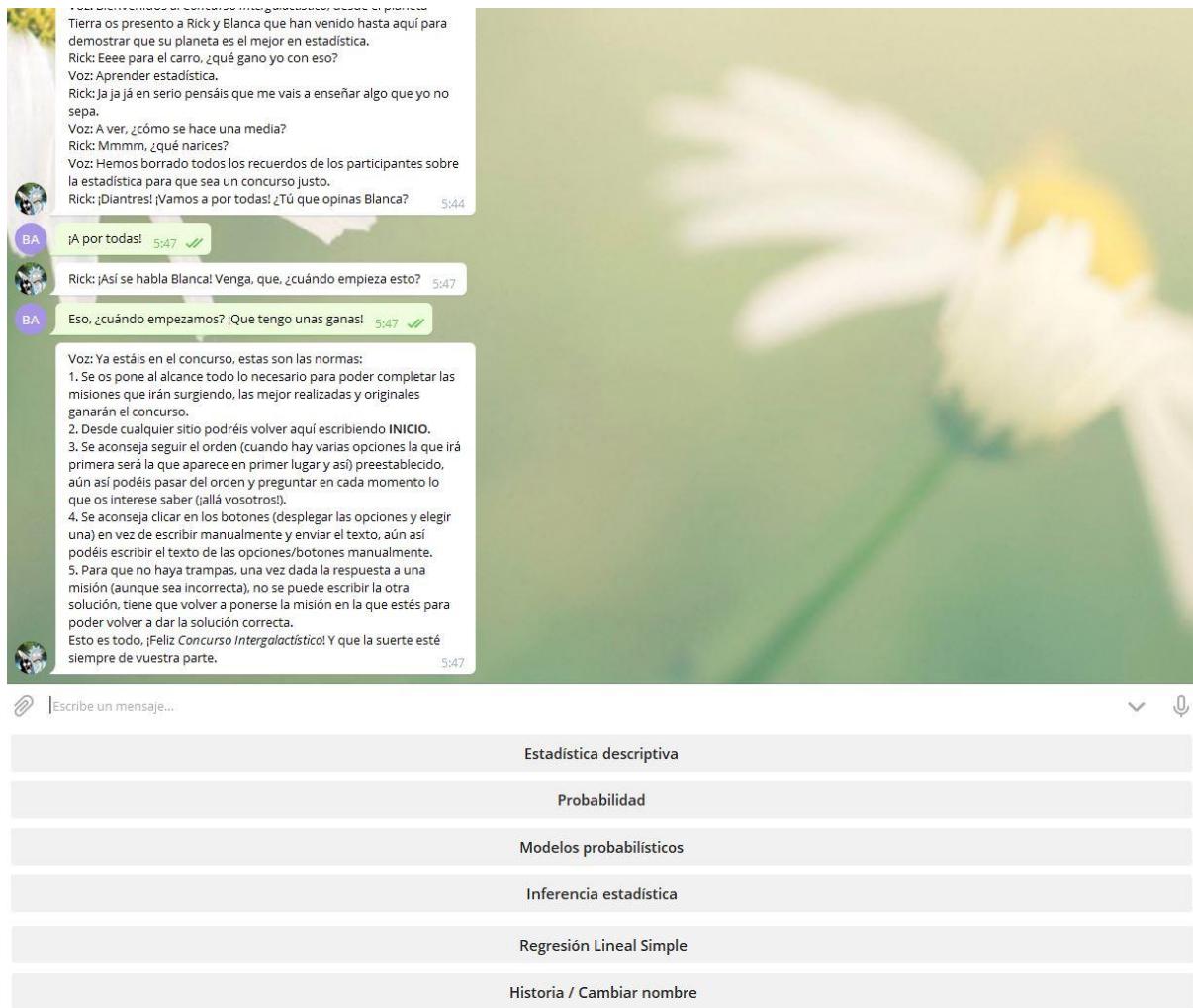


Figura 88: Normas de Rick y tú en Telegram

Ya podemos ver el mensaje con las normas del concurso, que en realidad son las instrucciones para ayudar al usuario a manejar el asistente, pero se quieren dar con un aire divertido. Nos salen las distintas partes de la estadística que trata el asistente y un botón para volver a la historia y poder cambiar el nombre.

A partir de este nodo ya se puede preguntar lo que se quiere saber y si el asistente sabe responderte a esto lo hará y si no, saltará al nodo de “cualquier otra cosa” y nos responderá un mensaje, se puede ver en la figura 89, para volvemos a guiar en la conversación.

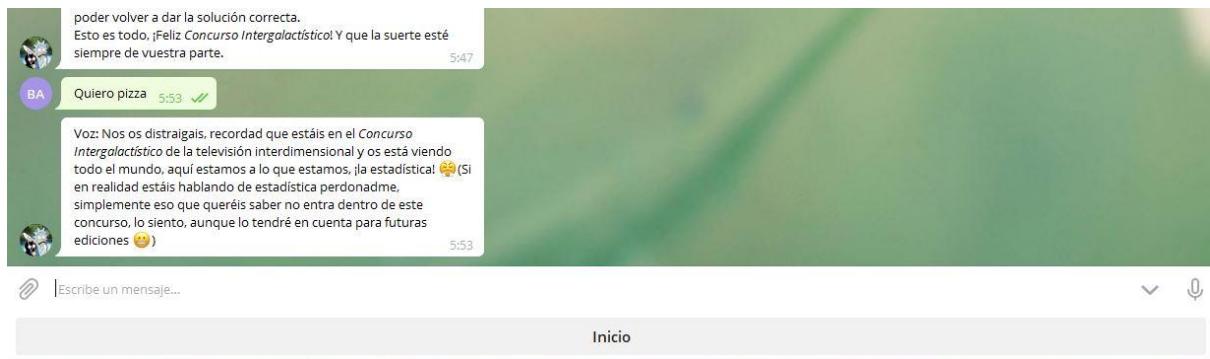


Figura 89: Mensaje “en otras cosas” de Rick y tú

Para acceder por el link tenemos que clicar en el mismo o ponerle en un navegador. Si clicas [aquí](#) puedes acceder a él, una vez que estás en él se puede ver la figura 90.

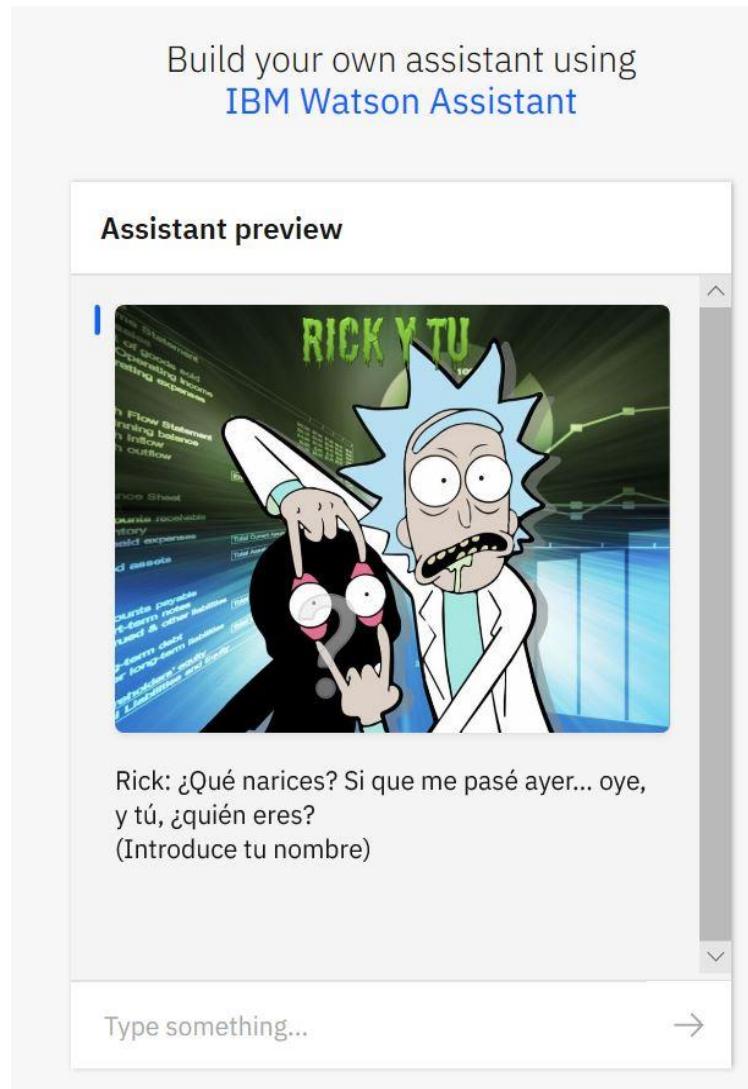


Figura 90: Link previo del asistente

Todo es igual que en Telegram excepto dos salvedades: la primera es que la norma cuatro nos da más explicación como se puede ver en la figura 91 ya que el nombre del botón, en vez de ser el nombre del apartado que va después, es siguiente, pero si clicamos en él, nos responderá con la respuesta del siguiente apartado como podemos ver en la figura 92 pero si escribimos literalmente siguiente nos aparecerá el mensaje que podemos ver en la figura 93.

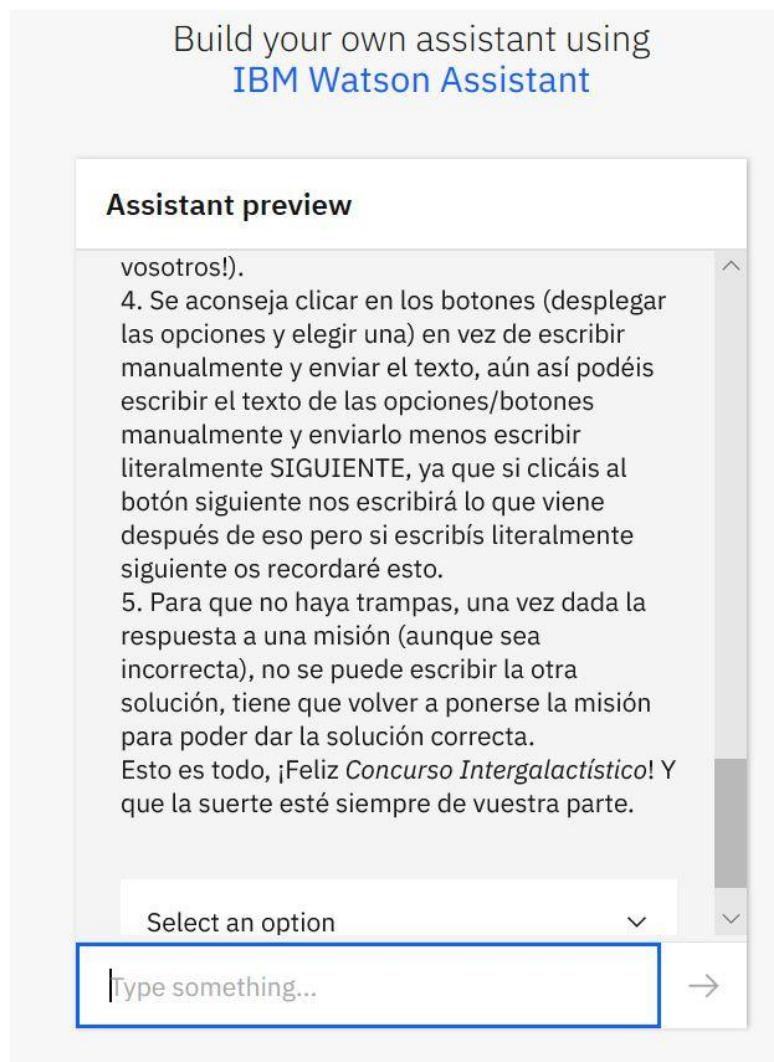


Figura 91: Norma cuatro en el link previo del asistente

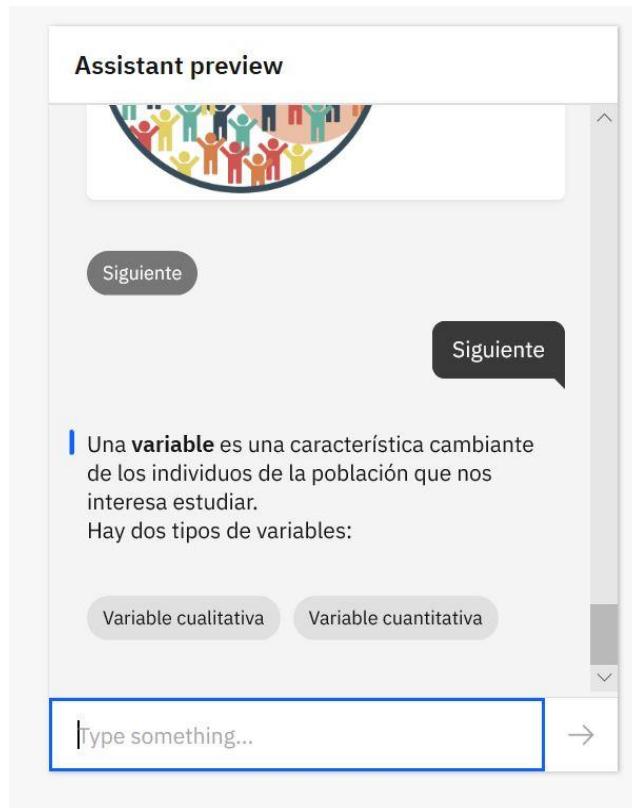


Figura 92: Clicar siguiente en el link previo del asistente

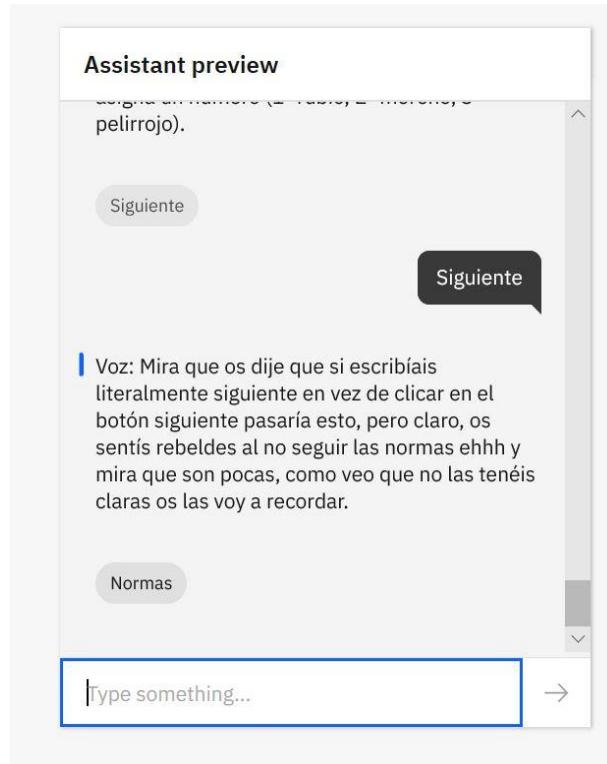


Figura 93: Escribir siguiente en el link previo del asistente

La segunda es, que cuando hay varias opciones, en vez de salir botones, sale un desplegable con el nombre de Select an option. Cuando clicamos en él se despliega con todas las opciones que podemos elegir, cómo se puede ver en la figura 94.

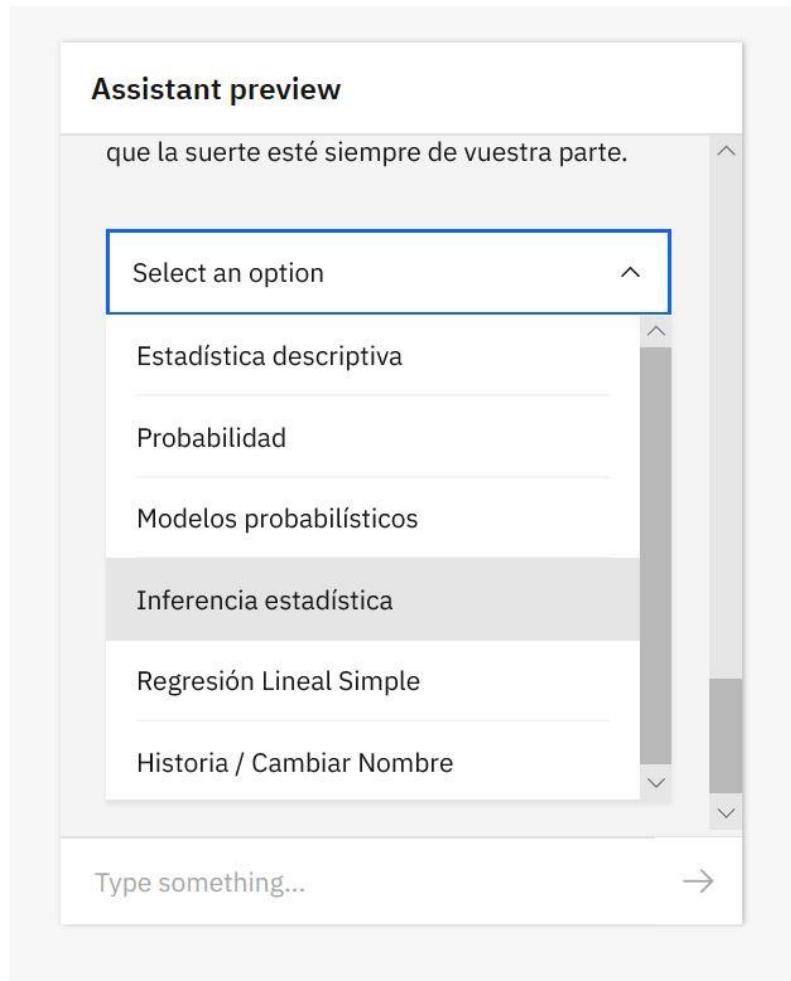


Figura 94: Seleccionar una opción en el link previo del asistente

Si el asistente detecta que el usuario necesita ayuda o no sabe a qué puede responder el asistente, le responde con el mensaje que se puede ver en la figura 95.

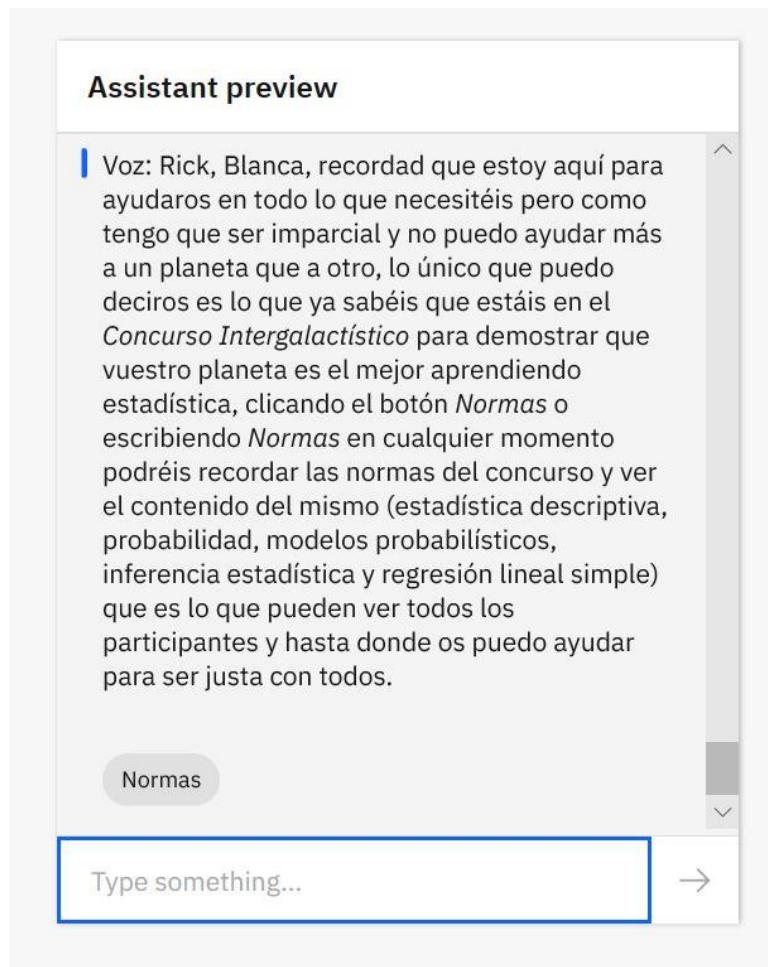


Figura 95: Mensaje de “ayuda” del asistente

Cuando se llega al apartado de *problemas de la regresión lineal simple*, saldrá el botón final que lleva al desenlace de la historia de Rick y tú en el *Concurso Intergaláctico*. ¡Qué nervios por llegar y saber qué pasa!

ANEXO III – ENCUESTA DE OPINIÓN

Si clicas [aquí](#) o en las figuras de este apartado (de la 96 a la 105), podrás ver la encuesta que se le paso tanto a profesores como alumnos de la carrera o asignatura de estadística para saber su opinión sobre el asistente para plantear futuras mejoras que se podrían hacer.



Rick y tú

Muchas gracias por dedicar 2 o 3 minutos en contestar esta encuesta que tiene la finalidad de mejorar el chatbot "Rick y tú" que, a través de una historia en la que Rick (el científico de la serie Rick y Morty) y el usuario del chatbot (tú) aprenden conceptos sobre estadística. Se puede ir leyendo la encuesta e ir probando el chatbot lo necesario para poder contestar a las preguntas de la misma. ¡Muchas gracias por dedicar unos minutillos de tu tiempo en ayudarme y más en las épocas que estamos que justo es el tiempo lo que nos falta!

*Obligatorio

Figura 96: Explicación de la encuesta

Pregunta 1: ¿De qué rama es la carrera que estudias/estudiaste? *

- Ingeniería
- Económicas
- Salud
- Estudio/terminé la carrera de estadística
- Soy enseñante de estadística

Figura 97: Elegir rama de carrera en la encuesta

Pregunta 2: ¿Con que frecuencia utilizas chatbots (Siri, Alexa, Cortana, los que hay en alguna página web....)? *

- A diario
- Bastante
- Alguna vez
- Nunca
- No sé ni de qué me estás hablando

Figura 98: Elegir frecuencia con la que se usan asistentes en la encuesta

Pregunta 3: ¿Están claras las "normas" del concurso y te han ayudado a manejarlo con el chatbot? (Estas normas aparecen en el chatbot, después de la pequeña historia inicial) *

- Sí, me han ayudado mucho
- Sí, pero podrían mejorarse
- Ayudan bastante poco
- No son útiles y además te lian más

Si la anterior respuesta es negativa, ¿por qué? ¿qué "normas" añadirías para facilitar el uso del chatbot? (O has contestado sí pero quieres decir algo al respecto de esto)

Tu respuesta

Figura 99: Pregunta sobre las normas de la encuesta

Pregunta 4: ¿Crees que el contenido del chatbot es adecuado para una persona que quiere iniciarse en el mundo de la estadística y tiene base matemática? *

- Sí, es justo el contenido necesario para empezar a aprender estadística
- No está mal pero cambiaría cosas
- Creo que se deberían enseñar primero otras cosas

Si has contestado que cambiarias cosas o que se deberian enseñar otras, ¿qué añadirías, quitarías, cambiarías...? (O has contestado sí pero quieres decir algo al respecto de esto)

Tu respuesta

Figura 100: Pregunta sobre el contenido del asistente de la encuesta

Pregunta 5: ¿Te parece que la historia y las misiones hacen más amena la tarea de aprender estadística? *

- Sí, además creo que ayuda a entender mejor los conceptos
- Sí, lo hace más entretenido aunque no ayude mucho con la estadística
- No mucho, creo que hay otras técnicas/métodos de enseñanza mejores
- No, además me parece completamente innecesaria y te distrae de lo importante

Si has elegido alguna de las opciones de no, ¿Por qué es, crees que contar historias (storytelling) o las misiones (gamification) no ayuda, no te gusta el tema, qué otras técnicas/métodos conoces que crees que son mejores...? (O has elegido una de las opciones de sí pero quieras decir algo al respecto de esto)

Tu respuesta

Figura 101: Pregunta sobre la historia y misiones de la encuesta

Pregunta 6: ¿Crees que es fácil de utilizar el chatbot? *

- Es muy fácil usarlo
- Se puede aprender a usar
- Se puede mejorar bastante
- Es muy difícil de usar

Si es difícil, ¿por qué, con qué problemas te has encontrado, sabes alguna forma de hacerlo más fácil de usar...? (O has contestado si pero quieres decir algo al respecto de esto)

Tu respuesta

Figura 102: Pregunta de usabilidad de la encuesta

Pregunta 7: ¿Te gusta que te dé opciones para guiarte en la conversación y un posible orden a seguir? (Ten en cuenta que puedes pasar del orden y preguntar en cualquier momento lo que te interese) *

- Menos mal que lo hace, si no, al principio, habría estado muy perdido
- No está mal
- La verdad es que yo creo que hay maneras mejores de hacerlo
- No me gusta que se den opciones, es mejor que sea libre

Si es no, ¿por qué, qué otras formas te parecen mejores para guiar a la persona y sepa a qué puede responder el chatbot....? (O has contestado sí pero quieres decir algo al respecto de esto)

Tu respuesta

Figura 103: Pregunta sobre las opciones de la encuesta

Pregunta 8: ¿Crees que este chatbot podría ayudar a la gente a aprender estadística o entenderla mejor? *

- Creo que podría ayudar a mucha gente
- Creo que podría ayudar a determinadas personas
- Creo que debería hacerse de otra forma
- Creo que lía más de lo que enseña

¿Qué se podría cambiar para conseguirlo, cómo se podría mejorar, cómo se podría hacer de forma más sencilla la explicación, directamente no crees en este tipo de enseñanza...?

Tu respuesta

Figura 104: Pregunta sobre la opinión general de la encuesta

Cualquier comentario que quieras hacer para ayudar a mejorar el chatbot o cualquier cosa que quieras que llegue a la persona que lo ha diseñado ¡Muchas gracias por ayudarme!

Tu respuesta

Enviar

Nunca envíes contraseñas a través de Formularios de Google.

Este contenido no ha sido creado ni aprobado por Google. [Notificar uso inadecuado](#) - [Términos del Servicio](#) - [Política de Privacidad](#)

Google Formularios

Figura 105: Pregunta final de la encuesta

ANEXO IV – CÓDIGO DEL ASISTENTE Y DEL FLUJO DE NODE-RED

Clicando [aquí](#) se verá la página de GitHub en la que se puede encontrar el código utilizado para llevar a cabo este trabajo.

Rick_y_tu_enlace_previo.json es el código utilizado para el asistente que se puede ver desde el enlace previo que genera por defecto Watson Assistant.

Rick_y_tu_Telegram.json es el código utilizado para el asistente que está conectado con el bot de Telegram *Rick_y_tu_bot*.

FlujoNode-Red.json es el código utilizado para conectar Watson Assistant con Telegram en Node-Red.