

Pruebas de bondad de ajuste para distribuciones

Versión PDF

II-1123 Estadística para Ingeniería Industrial II

Steven García Goñi

steven.garciagoni@ucr.ac.cr

26 de febrero de 2026



Agenda

- Preguntas generadoras
- Concepto de bondad de ajuste
- Rutinas gráficas
- Test de χ^2
- Pruebas de normalidad
- Otras rutinas de identificación de distribuciones



Preguntas generadoras

- ¿Cómo se selecciona la distribución que mejor se ajusta a los datos?
- ¿Por qué es necesario conocer qué distribución se ajusta a los datos?



¿Qué es bondad de ajuste?

- Es el grado de acoplamiento que existe entre los datos originales y los valores teóricos.
- Describe que tan bien se ajusta un conjunto de observaciones. Tales medidas, por lo general, describen o resumen la discrepancia entre los valores observados y los valores esperados en el modelo de estudio.
- Para ello se emplean contrastes de hipótesis.
- La bondad de ajuste es un concepto estadístico que mide qué tan bien un conjunto de **datos observados** se ajusta a una distribución de **probabilidad teórica** específica.
- En otras palabras, evalúa si los datos podrían razonablemente provenir de una distribución determinada (por ejemplo, binomial, Poisson, normal, exponencial, etc.).



Rutinas gráficas

- Se usan como herramienta para la identificación de distribuciones.
- Son **procedimientos visuales** que se usan para comparar los datos observados en una distribución teórica, con el fin de evaluar de **forma preliminar** si el modelo propuesto parece adecuado.
- No sustituyen bajo ningún concepto a las pruebas formales que van a ser estudiadas, pero permiten detectar patrones de desviación, asimetrías o valores atípicos antes de aplicar dichas pruebas.
- Su ventaja principal es que la **vista humana es muy sensible para reconocer discrepancias** que podrían pasar desapercibidas en un solo estadístico numérico.



Recuerde: Visualización

- Una buena visualización mostrará cosas que no se esperaban o hará surgir nuevas preguntas acerca de los datos. También puede dar pistas acerca de si se están haciendo las preguntas equivocadas o si necesitas recolectar datos diferentes.
- **“Un simple gráfico ha brindado más información a la mente del analista de datos que cualquier otro dispositivo” - John Tukey**



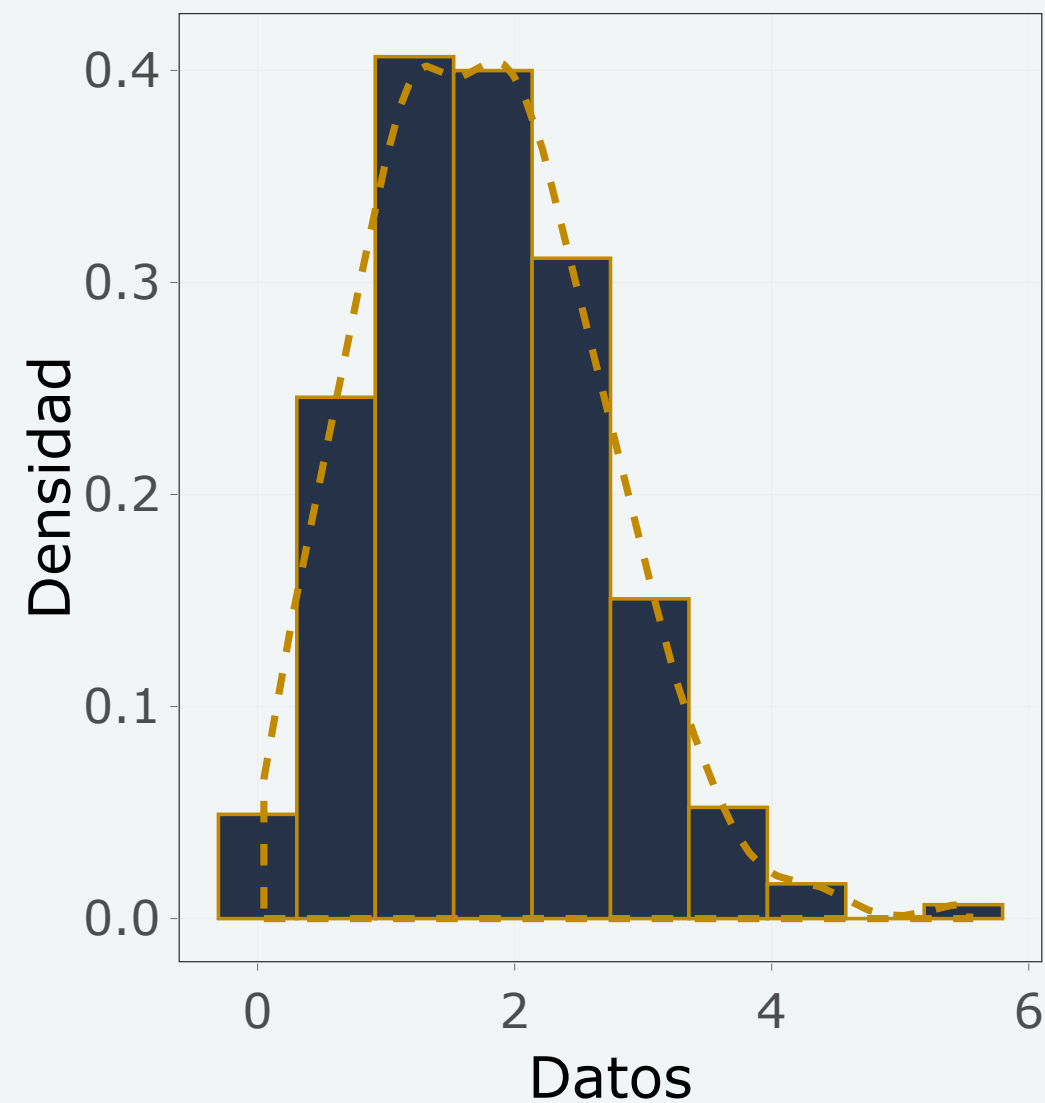
Rutinas gráficas: Formalmente

- Una rutina gráfica es un **conjunto de pasos estandarizados** para elaborar y analizar gráficamente la relación entre los datos observados y una distribución teórica, de manera que se pueda evaluar visualmente la concordancia antes de realizar pruebas formales.



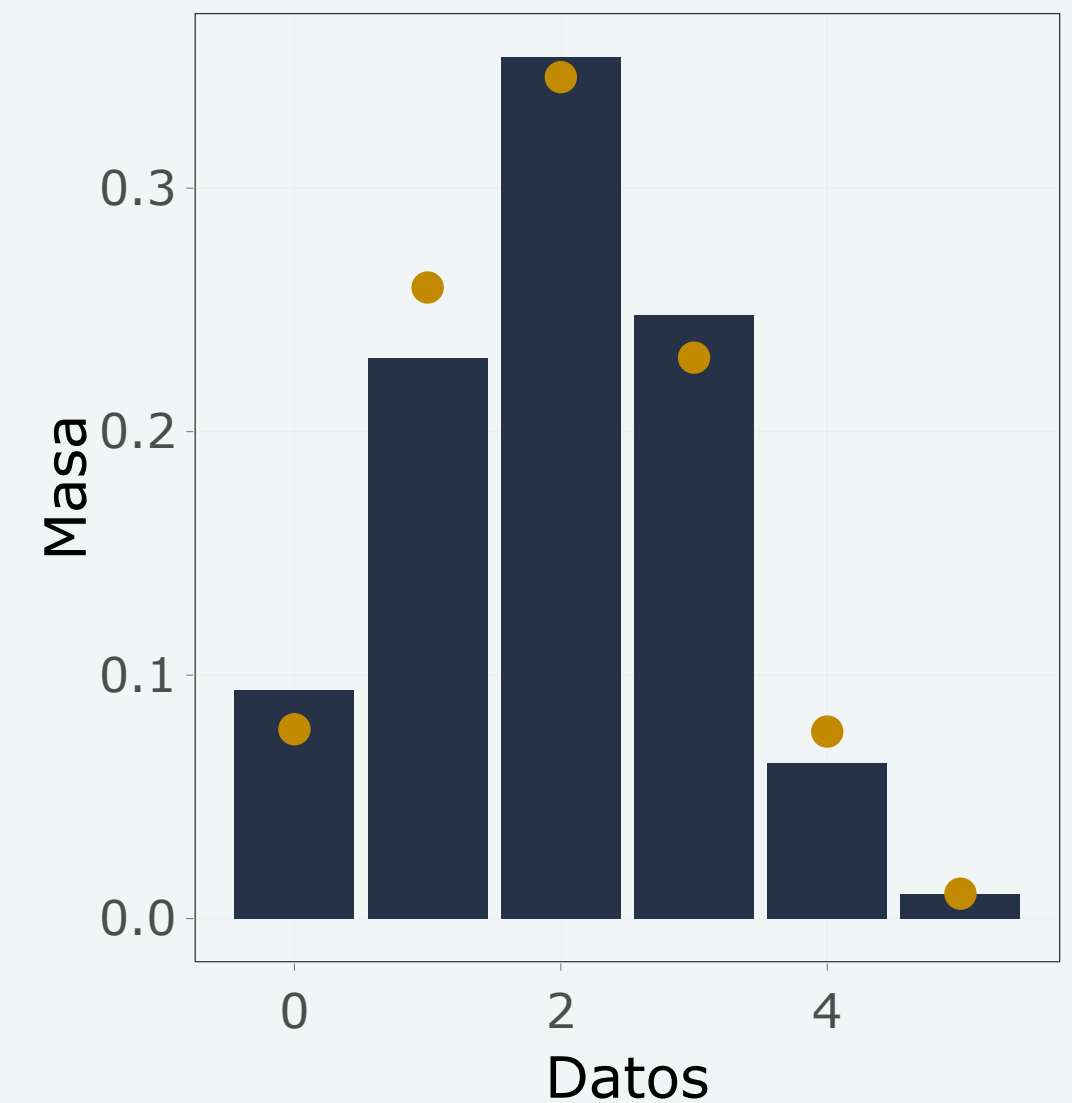
¿Cuáles gráficos usar?

Variable continua



- Histograma (variables continuas)
 - Puede agregar una “curva” teórica superpuesta para evaluar el ajuste.
- Gráfico de barras (Variables discretas)
 - Puede agregar puntos teóricos que le permitan evaluar el ajuste.

Variable discreta



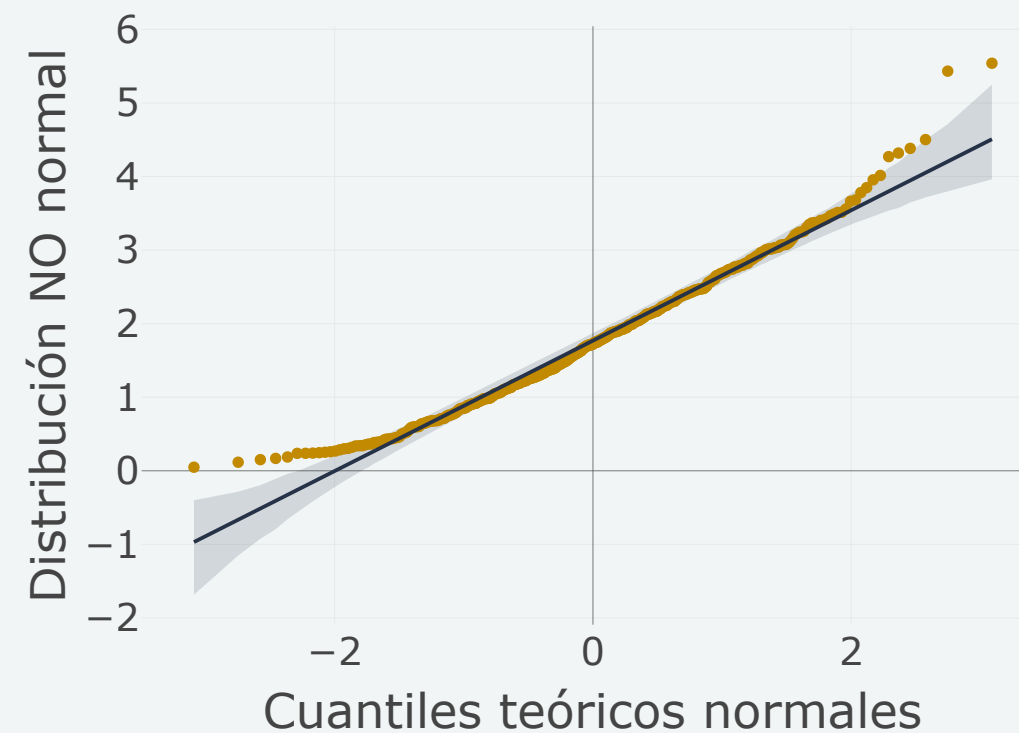
¿Cuáles gráficos usar?

- Puede usar un gráfico de probabilidad (**Probability plot**) o un gráfico cuantil-cuantil (**QQ Plot**).

— La idea que sigue este gráfico es que si los puntos se alinean a una recta (teórica), el ajuste es plausible, mientras que si hay desviaciones sistemáticas habría indicativos de mala concordancia.

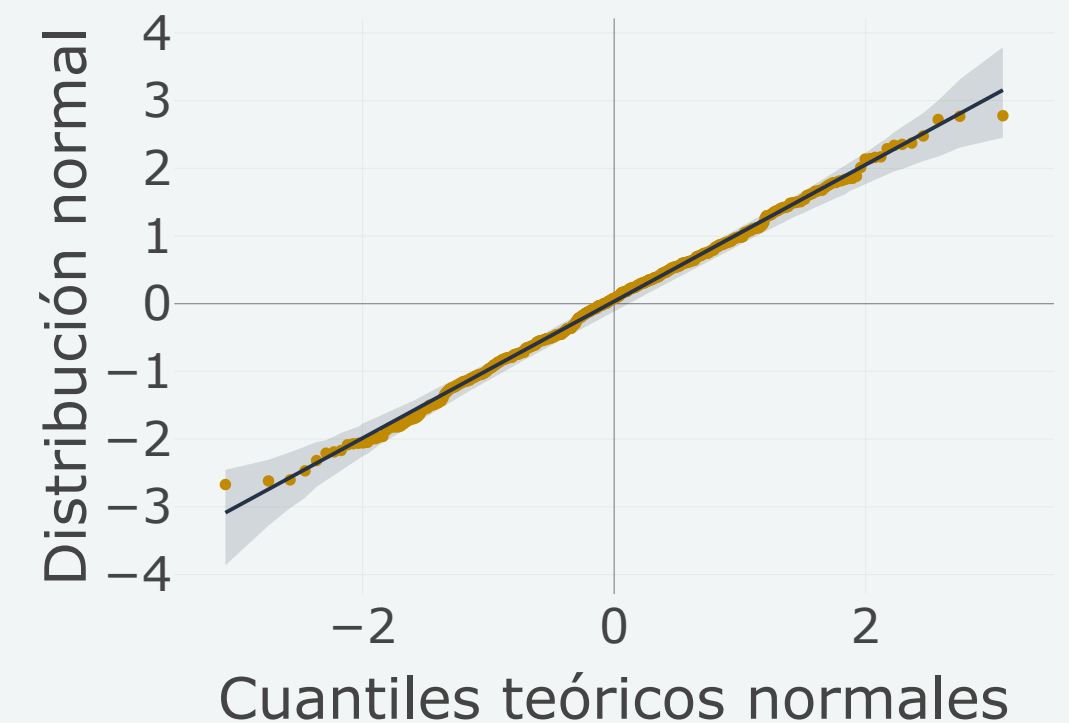
No normal

Estos datos se desvían mucho de la distribución teórica **normal**, lo que es indicativo de que no siguen la distribución normal.



Normal

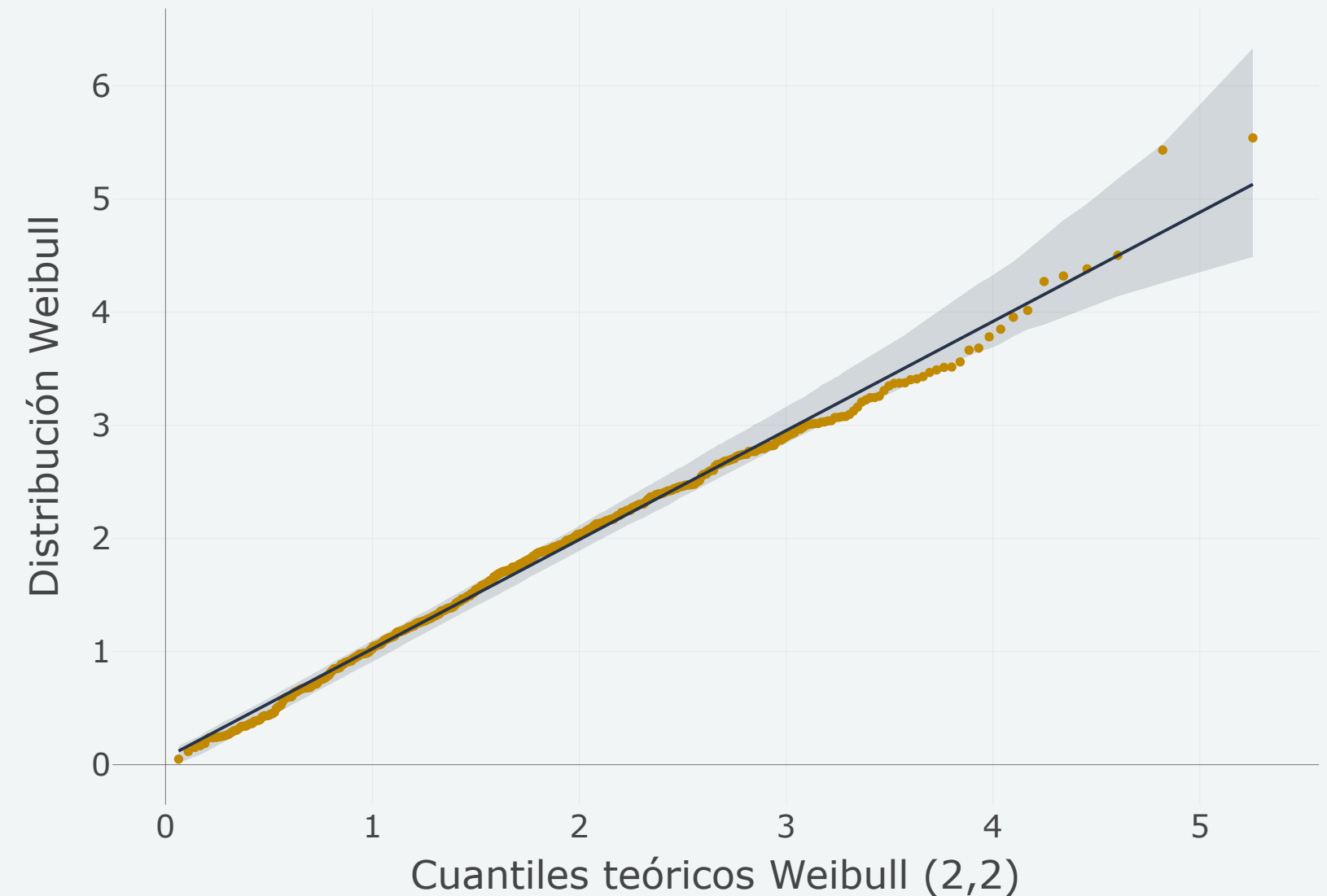
Los puntos **siguen la línea teórica**, lo que es indicativo de que en este caso se sigue la distribución normal.



¿Cuáles gráficos usar?

- El gráfico de probabilidad o el gráfico de cuantil – cuantil a menudo es “recetado” para una distribución normal, pero **se puede usar para cualquier distribución** siempre que se conozcan los cuantiles teóricos.
- Por ejemplo, en este caso, de una Weibull(2,2), que son los datos que dieron origen a la “Distribución no normal” del slide anterior, que al ser probada contra una normal, no se ajustó correctamente. Caso contrario al observado a la derecha.

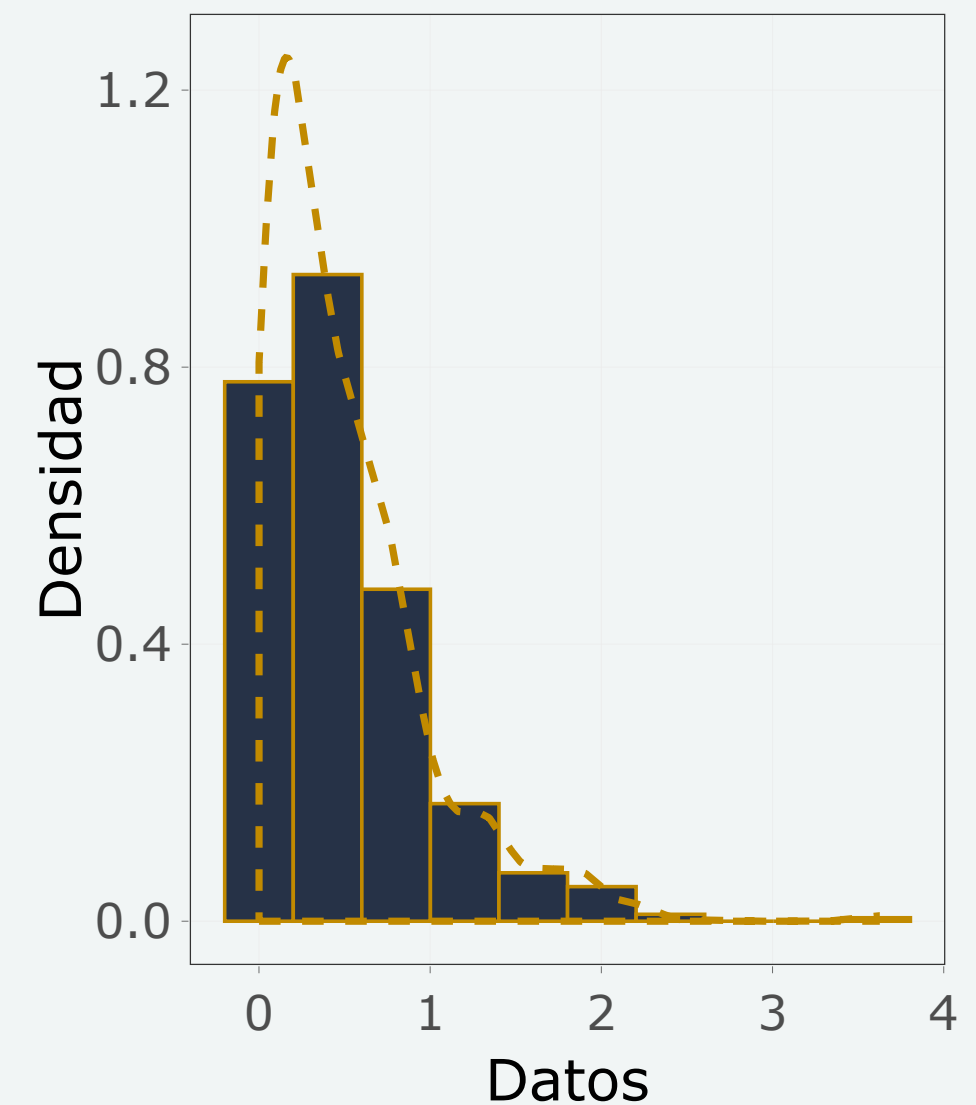
QQ-Plot (Weibull)



Importancia de la visualización

- Este tipo de gráficos **se utilizan** con frecuencia y de forma simultánea **con pruebas estadísticas formales** para contrastar la hipótesis de que un dato siga una distribución particular.
- También es relevante para saber que tipo de prueba de hipótesis realizar.
- ¿Tiene sentido aplicarle una prueba de normalidad a estos datos?
 - No ¿verdad?
- En este caso resultaría conveniente aplicar una prueba **apropiada para la forma de los datos**, por ejemplo, podemos intuir estas dos:
 - Lognormal
 - Exponencial

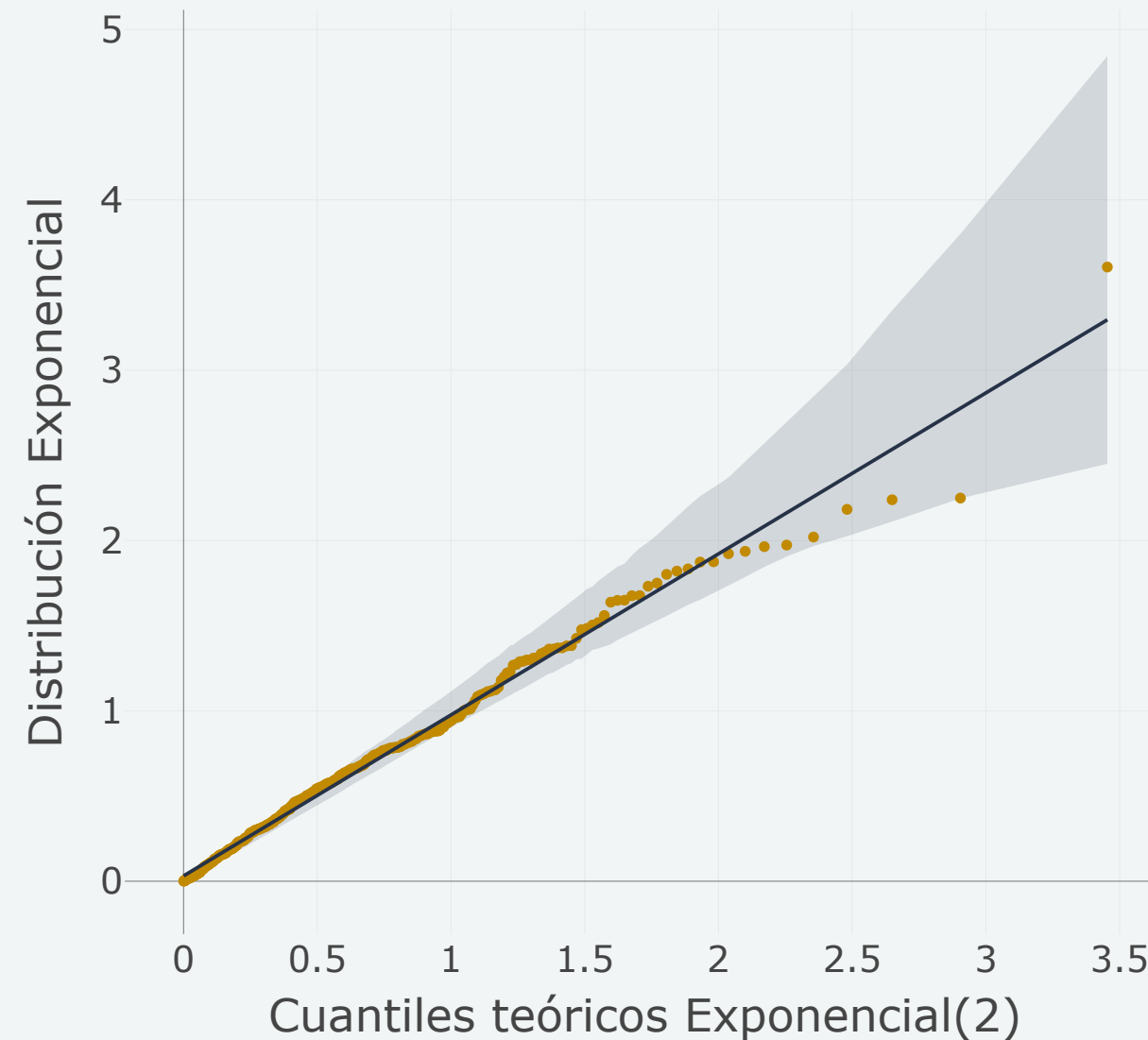
Distribución asimétrica



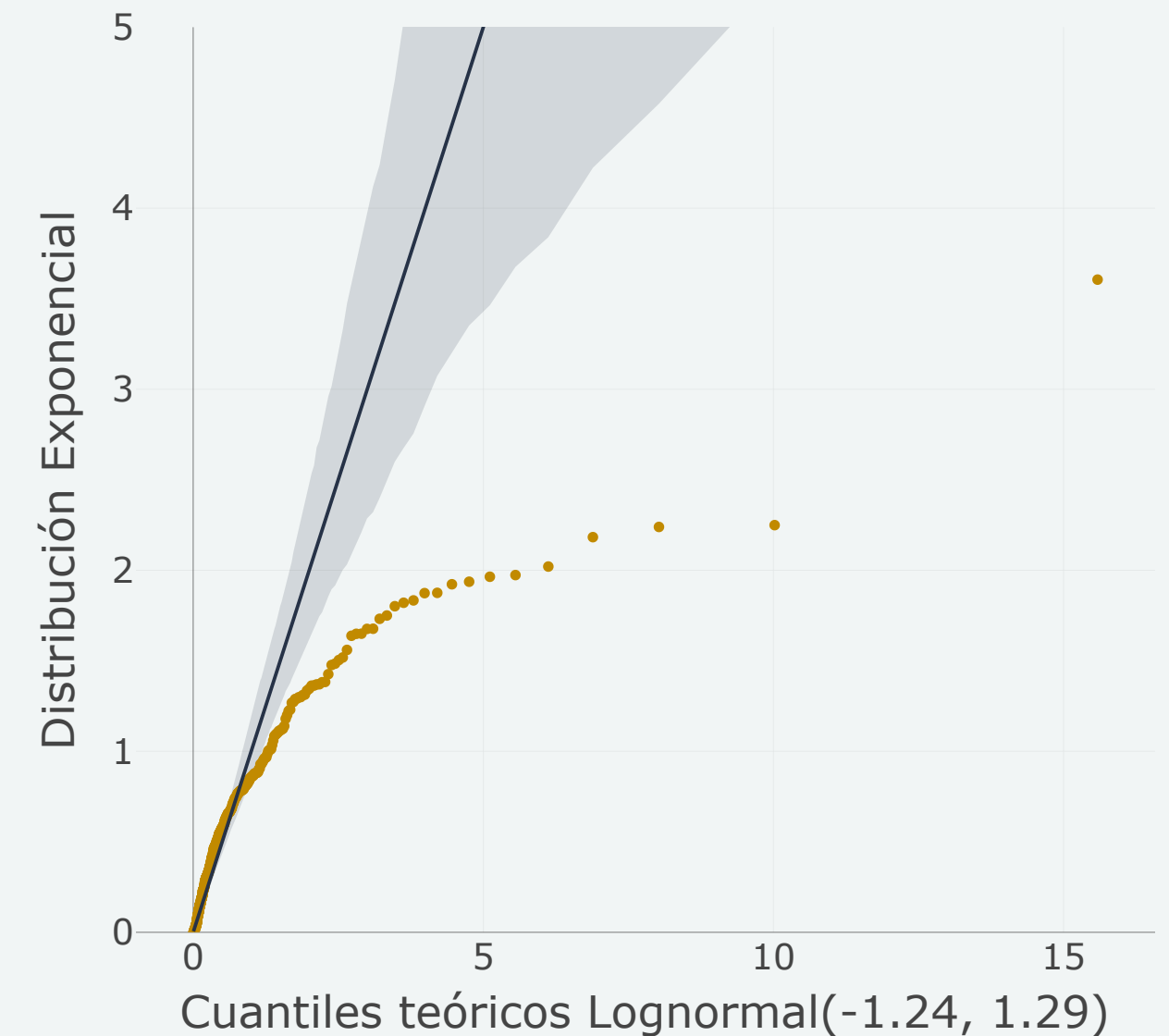
Improtancia de la visualización

- Como es obvio, la normal **no es una buena opción**, por lo que no se debe ni intentar.
- Y entre la *lognormal* y *exponencial*, esta última resulta con mejor ajuste.

QQ-plot (exponencial)



QQ-plot (lognormal)



Recordemos antes de continuar

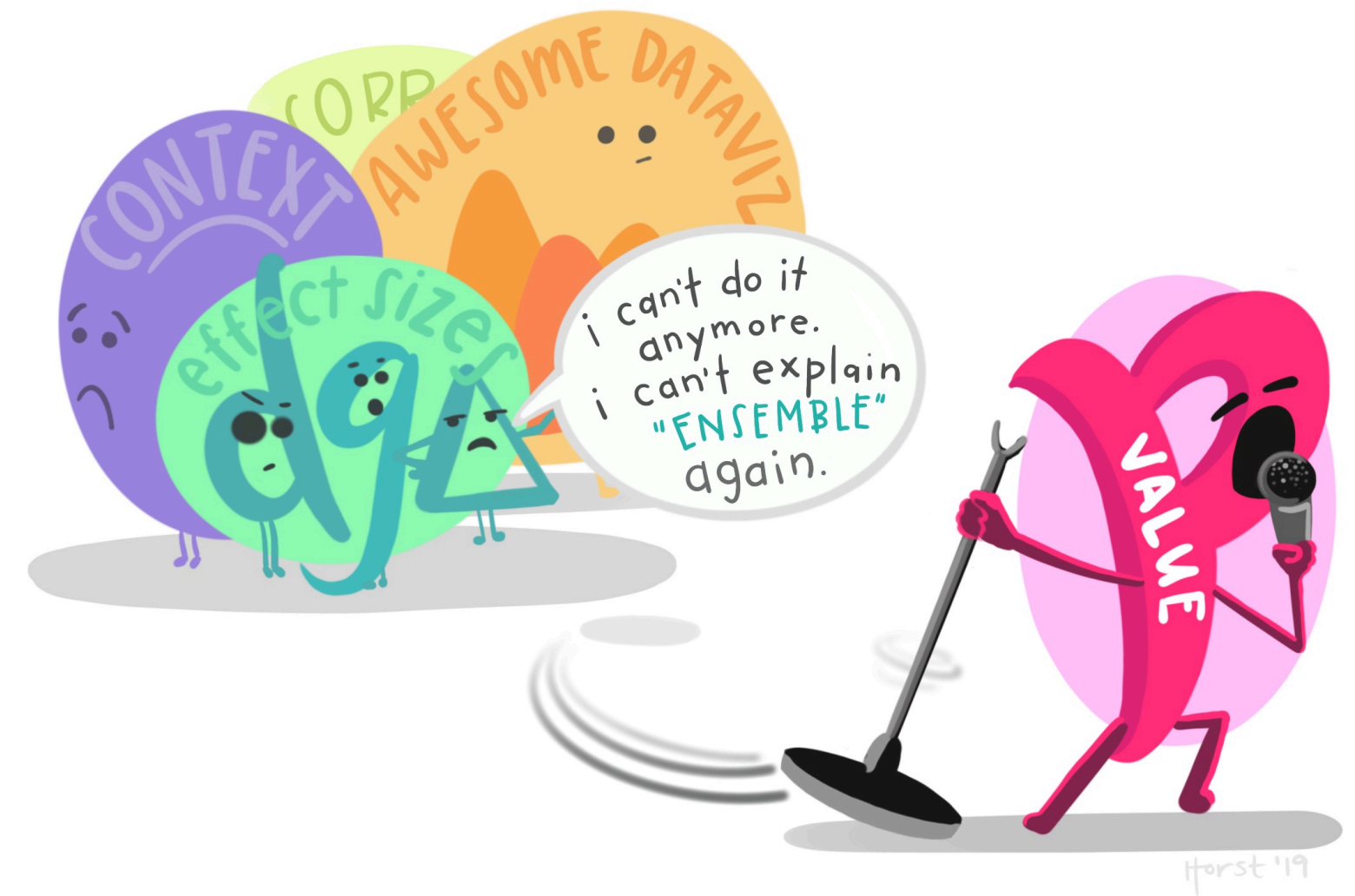
- Hacer estadística **no es solo leer valores p de forma indiscriminada.**
- La comunidad científica ha insistido en que evitemos la mala práctica de solo interpretar el valor P.
 - **No perpetuemos este error.**
- Los valores P se analizan **en conjunto** con los demás estadísticos, contexto y visualizaciones.



Valores P

- El valor P es una probabilidad que **mide la evidencia en contra de la hipótesis nula**. Entre más bajo este valor, más fuerte será la evidencia en contra.
 - Se compara contra el nivel de significancia (α).
- Formalmente: un valor P es la probabilidad bajo un modelo estadístico especificado que un resumen estadístico de los datos sería igual o más extremo que su valor observado.
- Consulte esta **infografía** para ampliar estos conceptos.

Ilustración de **Allison Horst**



Valores P

- El valor P **no es el protagonista**, es solo una medida estadística adicional, que **se interpreta en conjunto** con las demás medidas.
- En síntesis α y valor p están íntimamente relacionados, pues el primero marca el límite de decisión con base en el segundo.
- El nivel de significancia (α) se va a usar frecuentemente, este valor hay que escogerlo **cuidadosamente** en función del contexto, **nunca es una medida genérica**.
 - Como por ejemplo, usar $\alpha = 0.05$ para todo sin importar el contexto es un error. Estos valores siempre deben justificarse.



Valores P

- Por otro lado, en ocasiones nos vemos tentados a emplear terminología incorrecta como:
- Es **MUY significativo**, cuando un valor P es muy bajo.
- Es importante aclarar que la significancia estadística es un límite: **solo se puede ser o no ser significativo**.
- **No puedo ser muy o poco significativo**, pues los valores P, como bien se indica en la **infografía** no mide el tamaño de un efecto ni la importancia de un resultado.



Valores P

- Piénselo de esta manera:
- ¿Yo soy “**más tico**” por nacer en San José que alguien que nació en Los Chiles?
 - ¿Verdad que no?
- Con base en esta analogía, tampoco puedo ser más significativo con base en un valor P.
 - Así como cualquier nacimiento dentro de las fronteras costarricenses producen un “tico”, cualquier valor por debajo del nivel de significancia (α) es un resultado significativo, aún cuando ese umbral se supere ampliamente.
- Lo que puedo estar es más o menos seguro de que “soy tico” y para ello es relevante **analizar el valor P en conjunto** con:
 - **Evidencias externas** del contexto (diseño del estudio, calidad de mediciones, fiabilidad del muestreo).
 - **Evidencias internas** (otros estadísticos como la estimación del intervalo de confianza).



Test de χ^2



Introducción

- Se emplea para la bondad de ajuste de **distribuciones discretas**.
- Es un test estadístico que compara las frecuencias observadas con las frecuencias esperadas según la distribución teórica propuesta.
 - También conocida como prueba de Pearson.
- Esta se usa porque es el caso más general y fácil de entender, aunque es importante señalar que es poco sensible en tamaños de muestra pequeños.
- También es relevante dar a conocer que existen otras pruebas que pueden usarse.



Procedimiento

- Se define la prueba de hipótesis
 - H_o : la distribución se ajusta a los datos
 - H_i : la distribución *NO* se ajusta a los datos
- Se calcula el estadístico χ^2

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

- Donde O_i es lo observado y E_i es lo esperado.
 - A partir de ese estadístico χ^2 se puede obtener un valor P que **en conjunto con las rutinas gráficas**, y una elección adecuada del nivel de significancia (α) da origen a las conclusiones acerca de la bondad de ajuste.



Aclaraciones

- Esta prueba **también puede usarse con datos continuos**, pero no es "*nativa*" para este tipo de datos, por lo que dichos datos deben "*discretizarse*" en clases (como los de un histograma) y luego comparar las frecuencias observadas en cada clase.
- Pero no se recomienda usarla en estos casos por la falta de potencia, son mejores las opciones que veremos adelante.



Ejemplo 01

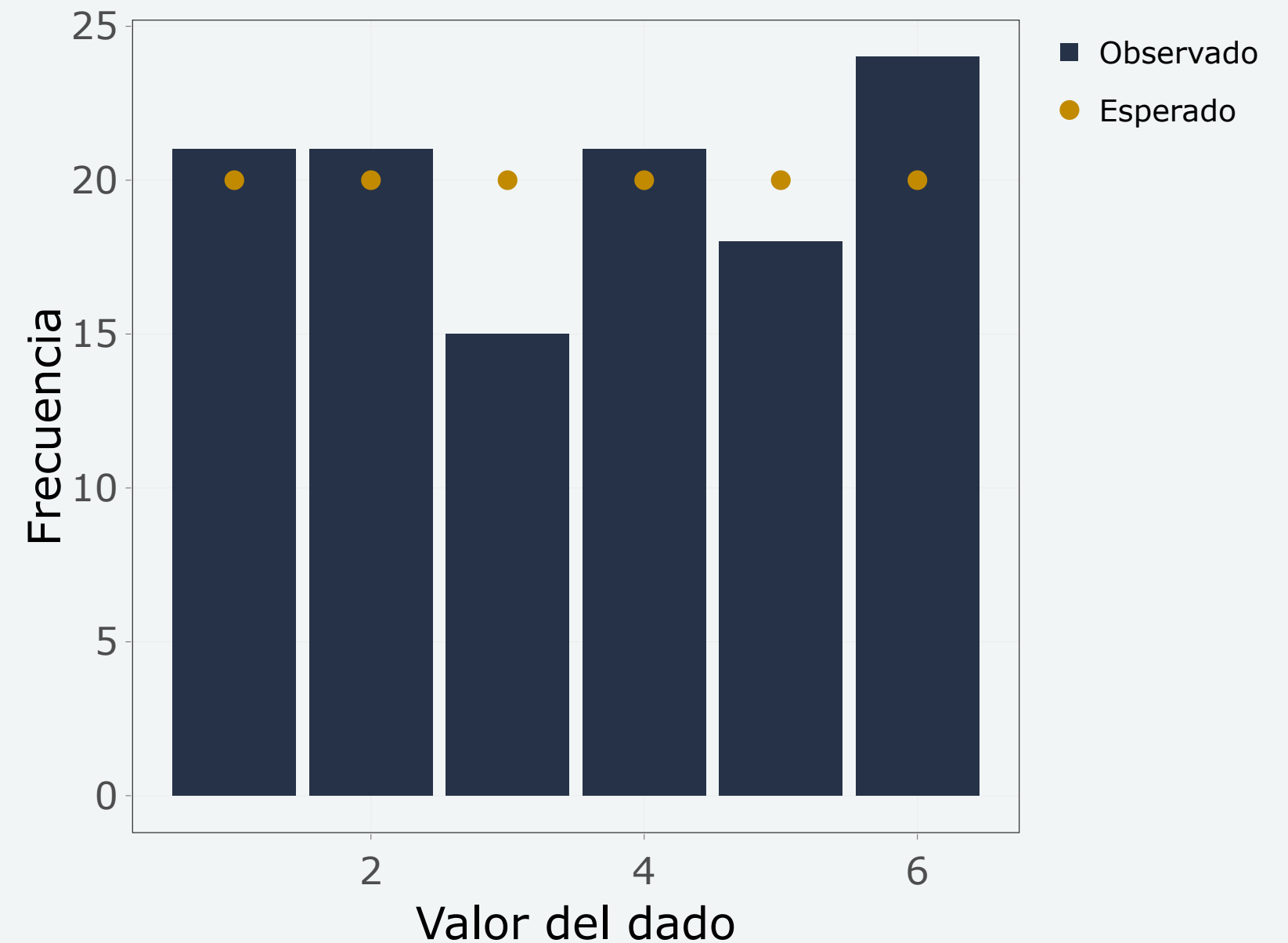
- Se realiza el lanzamiento de un dado en 120 ocasiones. Se anotan los resultados.
- ¿Los datos están balanceados? Es decir, ¿se sigue una distribución uniforme discreta?
 - Responda con $\alpha = 0.05$
- En este **Excel** puede encontrar los datos sin procesar, así como los ejemplos resueltos.

Lanzamiento	Conteo (Observado)
1	21
2	21
3	15
4	21
5	18
6	24



Ejemplo 01

- El primer paso, antes de buscar obtener un valor P a como de lugar, es realizar las rutinas gráficas.
 - Como este conjunto de datos es discreto, lo que conviene es realizar un gráfico de barras.
- El gráfico mostrado a la derecha, ¿parece ser una uniforme discreta?
 - Aunque pueda generarle dudas, razonablemente parece seguir una uniforme. Esta visualización es importante, porque inmediatamente le hace ver que el valor de 3 difícilmente alcanza el valor esperado y que el de 6 lo supera con creces.
 - No obstante, no se puede ser más estricto de la cuenta.



Ejemplo 01

Dado	Observado (O_i)	P(Esperada)	Esperado (E_i)	$\frac{(O_i - E_i)^2}{E_i}$
1	21	0.167	20	0.05
2	21	0.167	20	0.05
3	15	0.167	20	1.25
4	21	0.167	20	0.05
5	18	0.167	20	0.20
6	24	0.167	20	0.80
Total	120	1.000	120	2.40



Ejemplo 01

- Obtenga el valor P a partir del estadístico χ^2
- En Excel hay dos formas de hacerlo (fórmulas en *inglés*)
 - Forma corta: CHISQ.TEST(O_i , E_i)
 - Forma larga: 1-CHISQ.DIST(χ^2 , $n - 1$, TRUE)
- En R `chisq.test(x = c(21, 21, 15, 21, 18, 24), p = rep(1/6, 6))`
- $\chi^2 = 2.4$
- Valor P (forma corta) = 0.7915
- Valor P (forma larga) = 0.7915
- Significancia = 0.05
- No hay evidencia suficiente para rechazar la hipótesis nula de que el dado se comporta como una distribución uniforme discreta de probabilidad $\frac{1}{6}$, dado que $p > \alpha$, aunado a lo mostrado en el gráfico anterior.



Ejercicio 01

- Se ha recolectado 400 observaciones sobre la cantidad de personas que se reciben cada minuto en una estación de tren de la capital.
- Por la naturaleza de este proceso, se sospecha que estos datos son una variable aleatoria que sigue la **distribución de Poisson**.
- En **Excel** se le hace entrega de los datos y del ejercicio resuelto, procure realizar este ejercicio por su cuenta y utilice el ejercicio resuelto solo en caso de dudas.

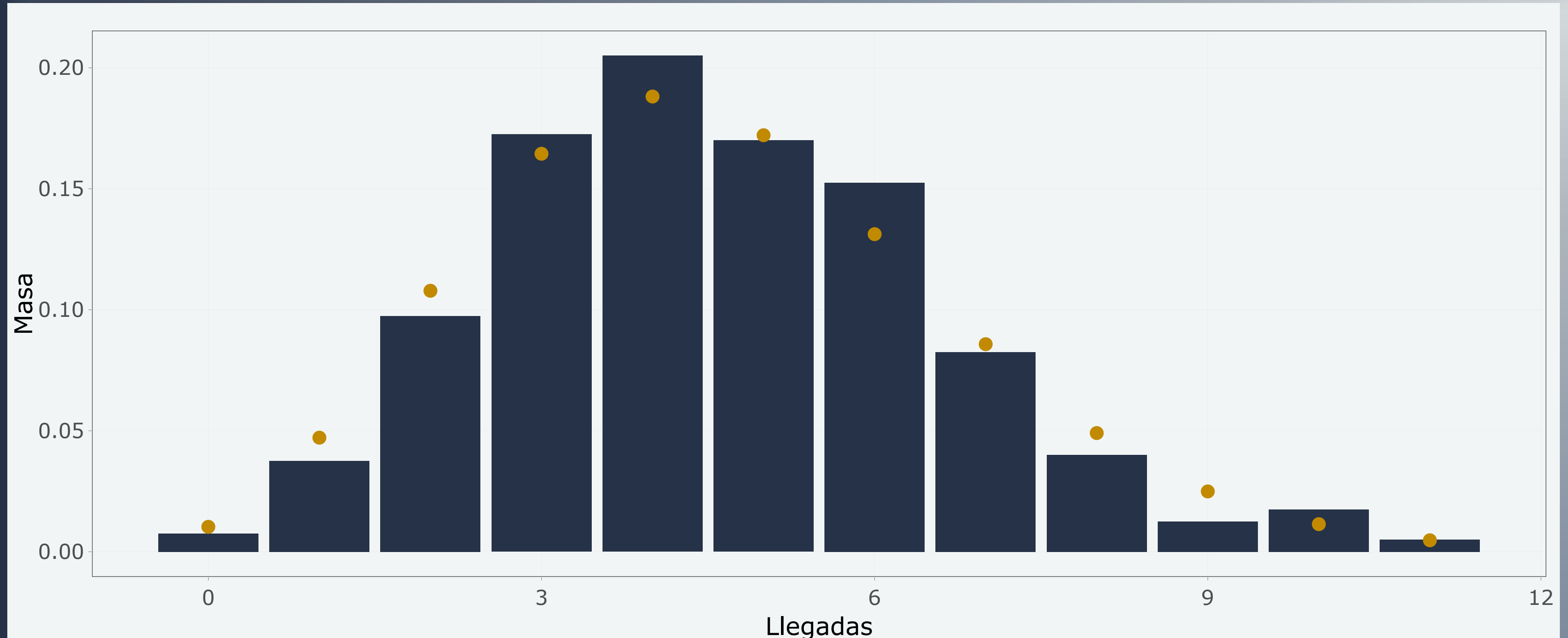


Ejercicio 01

- Determine el parámetro de la distribución. Como el enunciado deja en claro, esta es una Poisson, por lo tanto, el parámetro es λ .
 - Como es una muestra, el parámetro se estima. En una Poisson se sabe que $E(x) = \lambda$.
 - $\lambda = 4.575$
- Realice una rutina gráfica.
 - Recomendación: E_i debe sumar 400 (es decir, las probabilidades siempre deben suma 1), por lo que quizá deba agregar datos en cero, por ejemplo:
 - 15 personas observadas 0 veces, esperadas 0.03.
 - Salvo el cálculo de λ , resuelva con dos decimales.



Ejercicio 01



Ejercicio 01

- Por ejemplo, en este caso las probabilidades no suman 1, ni los valores esperados 400, por lo que debemos agregar más valores con 0 observaciones.
- No obstante, **Cochran** y otros autores mencionan que se debe cumplir una regla básica:
 - Todos los $E_i \geq 3$ o $E_i \geq 5$
- Por lo que estos valores deben agregarse. Para efectos de este curso usaremos $E_i \geq 5$ como una simplificación, la regla global es 1) Ningún $E_i < 1$ y al menos el 80% de los $E_i \geq 5$.

Llegadas	O_i	$P(E_i)$	E_i	$\frac{(O_i - E_i)^2}{E_i}$
0	3	0.010	4.123	0.306
1	15	0.047	18.861	0.790
2	39	0.108	43.143	0.398
3	69	0.164	65.794	0.156
4	82	0.188	75.252	0.605
5	68	0.172	68.855	0.011
6	61	0.131	52.502	1.375
7	33	0.086	34.314	0.050
8	16	0.049	19.623	0.669
9	5	0.025	9.975	2.481
10	7	0.011	4.564	1.301
11	2	0.005	1.898	0.005
Total	400	0.997	398.903	8.148



Ejercicio 01

Llegadas	O_i	$P(E_i)$	E_i	$\frac{(O_i - E_i)^2}{E_i}$
0	3	0.0103	4.1225	0.3057
1	15	0.0472	18.8605	0.7902
2	39	0.1079	43.1435	0.3979
3	69	0.1645	65.7938	0.1562
4	82	0.1881	75.2516	0.6052
5	68	0.1721	68.8552	0.0106
6	61	0.1313	52.5021	1.3754
7	33	0.0858	34.3139	0.0503
8	16	0.0491	19.6233	0.6690
9	5	0.0249	9.9752	2.4814
10	7	0.0114	4.5636	1.3007
11	2	0.0047	1.8981	0.0055
12	0	0.0018	0.7236	0.7236
13	0	0.0006	0.2547	0.2547
14	0	0.0002	0.0832	0.0832
15	0	0.0001	0.0254	0.0254
16	0	0.0000	0.0073	0.0073
Total	400	1.0000	399.9974	9.2423

En la primera tabla se obtiene un valor de $\chi^2 = 9.24$ con un valor $P = 0.9031$. No obstante, no se cumple la regla de que $E_i \geq 5$.

Por tanto, procedemos a agrupar valores que incumplen la regla con el más cercano posible, por ejemplo $E_0 + E_1 = 3 + 15 = 18$. Obteniendo un valor $P = 0.6264$.

Llegadas	O_i	$P(E_i)$	E_i	$\frac{(O_i - E_i)^2}{E_i}$
0-1	18	0.0575	22.9832	1.0805
2	39	0.1079	43.1437	0.3980
3	69	0.1645	65.7942	0.1562
4	82	0.1881	75.2521	0.6051
5	68	0.1721	68.8557	0.0106
6	61	0.1313	52.5025	1.3753
7	33	0.0858	34.3141	0.0503
8	16	0.0491	19.6234	0.6690
9	5	0.0249	9.9752	2.4814
≥ 10	9	0.0189	7.5559	0.2760
Total	400	1.0000	400.0000	7.1025



Sensibilidad de χ^2

- El cálculo estadístico de χ^2 y su comparación con un valor crítico de la distribución χ^2 permite al investigador evaluar si los recuentos de celdas observados son significativamente diferentes de los recuentos de celdas esperados.
- Debido a la forma en que se calcula el valor de χ^2 , es **extremadamente sensible al tamaño de la muestra**:
 - Cuando el tamaño de la muestra es demasiado grande (~500), casi cualquier pequeña diferencia parecerá estadísticamente significativa.
 - χ^2 es sensible porque las diferencias pequeñas multiplicadas por n generan E_i grandes, y el estadístico acumula discrepancias sistemáticas.
- Por tanto, debe analizarse con más cuidado, utilizando otras evidencias además del valor P .



Pruebas de normalidad



Introducción

- Nos vamos a adentrar en el universo de las pruebas de normalidad, pero la idea es hacerlo estando preparados.
- Por lo general, desarrollar pruebas de normalidad de forma manual (con apoyo de software como Excel), puede ser muy laborioso.
- Por tanto, este abordaje será basado en el uso de software estadístico ([R](#), [Minitab](#), entre otros).



Introducción

- La idea básica que se quiere desarrollar es que las pruebas de normalidad NO SON INTERCAMBIABLES sin más:
 - No tienen la misma **sensibilidad**
 - Su **potencia** varía con el tamaño de muestra
 - Los **contextos** en los que se aplican son diversos
- Pero todas parten de la misma hipótesis:
 - H_o : los datos siguen una distribución normal



Introducción

- En esta sesión de trabajo vamos a aprender qué características tienen las pruebas de normalidad, para que pueda escogerlas más adelante.
- Por *tradición* se suele usar [Minitab](#) para hacerlo, pero se nombraran otras que no están presentes en este software, pues no es el único que existe ni el que se usa en ambientes fabriles.



Características

- Un conjunto de datos puede **no seguir la distribución normal** por al menos dos grandes motivos:
 - La distribución es asimétrica
 - Las colas son pesadas
- Por lo que la selección de una prueba **debe ser concordante con el tipo de desviación que nos preocupe**. Ya las pruebas, por lo general, suelen darle más importancia a uno de los dos motivos anteriores

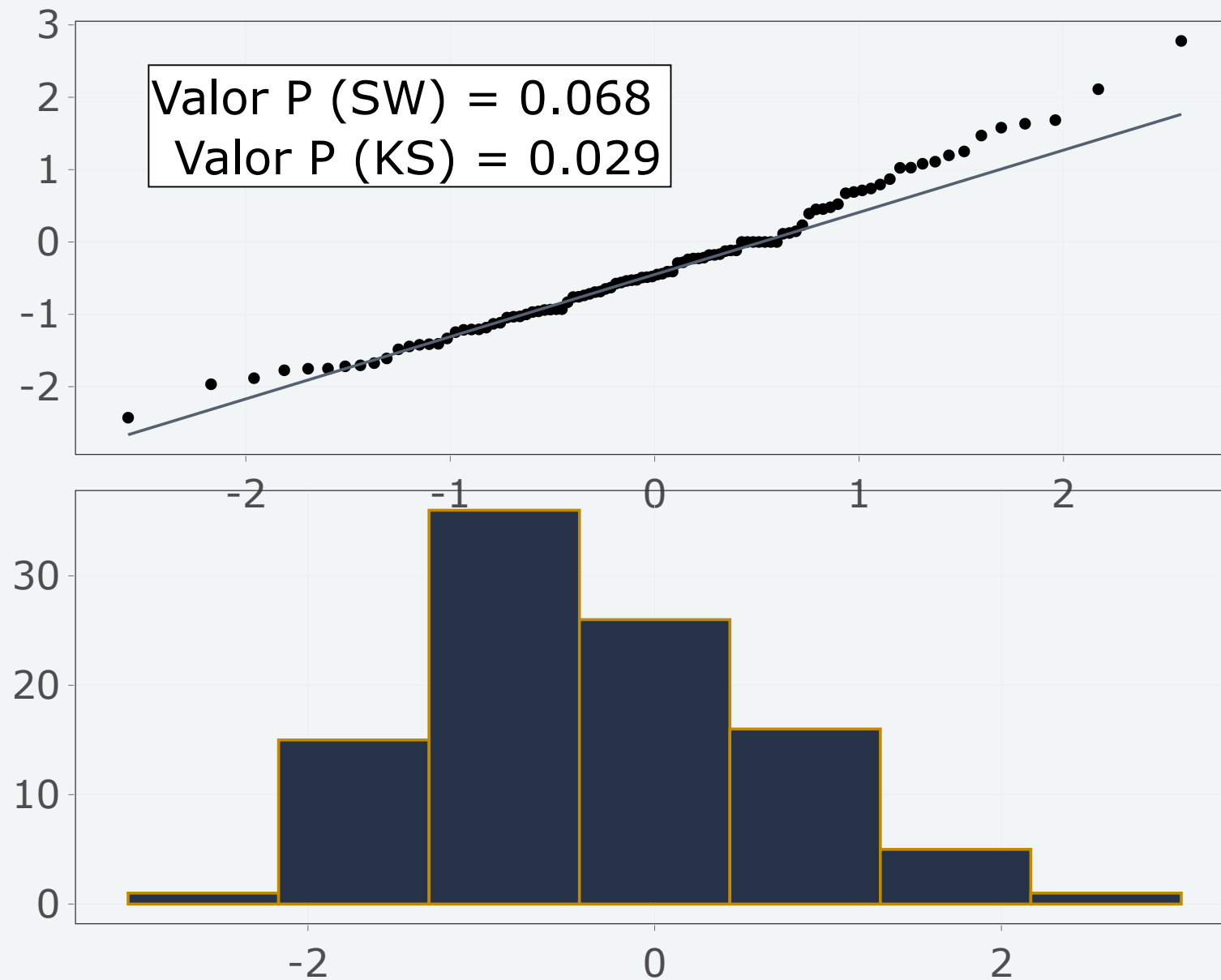


Por ejemplo

- Un análisis de regresión (*tema que se atiende luego en el curso*) es robusto a desviaciones ligeras de la normalidad, por lo que conviene utilizar una prueba que no penalice en exceso dichas desviaciones.
- Por otro lado, los análisis de capacidad (*tema de ingeniería de calidad*) son muy sensibles a desviaciones en el supuesto de normalidad por las colas, por lo que conviene utilizar una prueba de normalidad que sea apropiada para este contexto.
 - Desviaciones en la normalidad a causa de las colas pesadas incrementa la probabilidad de valores extremos y aumenta la estimación de la desviación estándar
- En resumen, **es el contexto** el que dicta la adecuación de las pruebas de normalidad usadas.



Pruebas de normalidad



- ¿Por qué en este caso la prueba de Shapiro – Wilk (SW), con un 95 % de confianza, no rechaza la hipótesis nula, mientras que Kolmogórov-Smirnov (KS) si lo hace?
- ¿Cuál prueba escojo?
 - Discuta con compañer@s y persona docente.



Pruebas de normalidad

- ¿Qué sucede en la práctica? Y que no necesariamente es una buena práctica.
- Las personas realizan varias pruebas de normalidad y se reportan la que les da bien según sus objetivos (rechazar o no rechazar).
- Esto no es **éticamente correcto** y por ello se les insta a comprender cuando funciona mejor cada prueba de normalidad.



Anderson - Darling (AD)

- Tiene una sensibilidad especial en las colas de la distribución y es justo por este motivo que es la primera opción en Minitab y otros software, pues desviaciones en las colas genera problemas en la estimación de índices de calidad (tema de ingeniería de calidad abordado en otros cursos).
- Su principal limitación deviene en que, cuando n es grande tiende a sobre – rechazar la hipótesis nula.



Shapiro - Wilk (SW)

- Si bien no está presente en software como [Minitab](#) de forma explícita, es necesario explicarla.
- Es una prueba que se centra en las diferencias de forma global (asimetría y colas).
- Su limitación se basa en que con n muy grande puede rechazar H_0 por desviaciones irrelevantes.
- Esta suele ser más potente que KS en muestras pequeñas.



Ryan - Joiner (RJ)

- Es similar a SW, pues se centra en desviaciones generales. Es por este motivo que suele preferirse en muestras “pequeñas”.
- Posee las mismas limitaciones que SW.
- La lógica del estadístico está más centrada en la alineación global, donde indefectiblemente la mayor contribución proviene del centro de la distribución.
- Es decir, las colas si influyen, pero no tanto como en AD, porque su efecto está diluido.



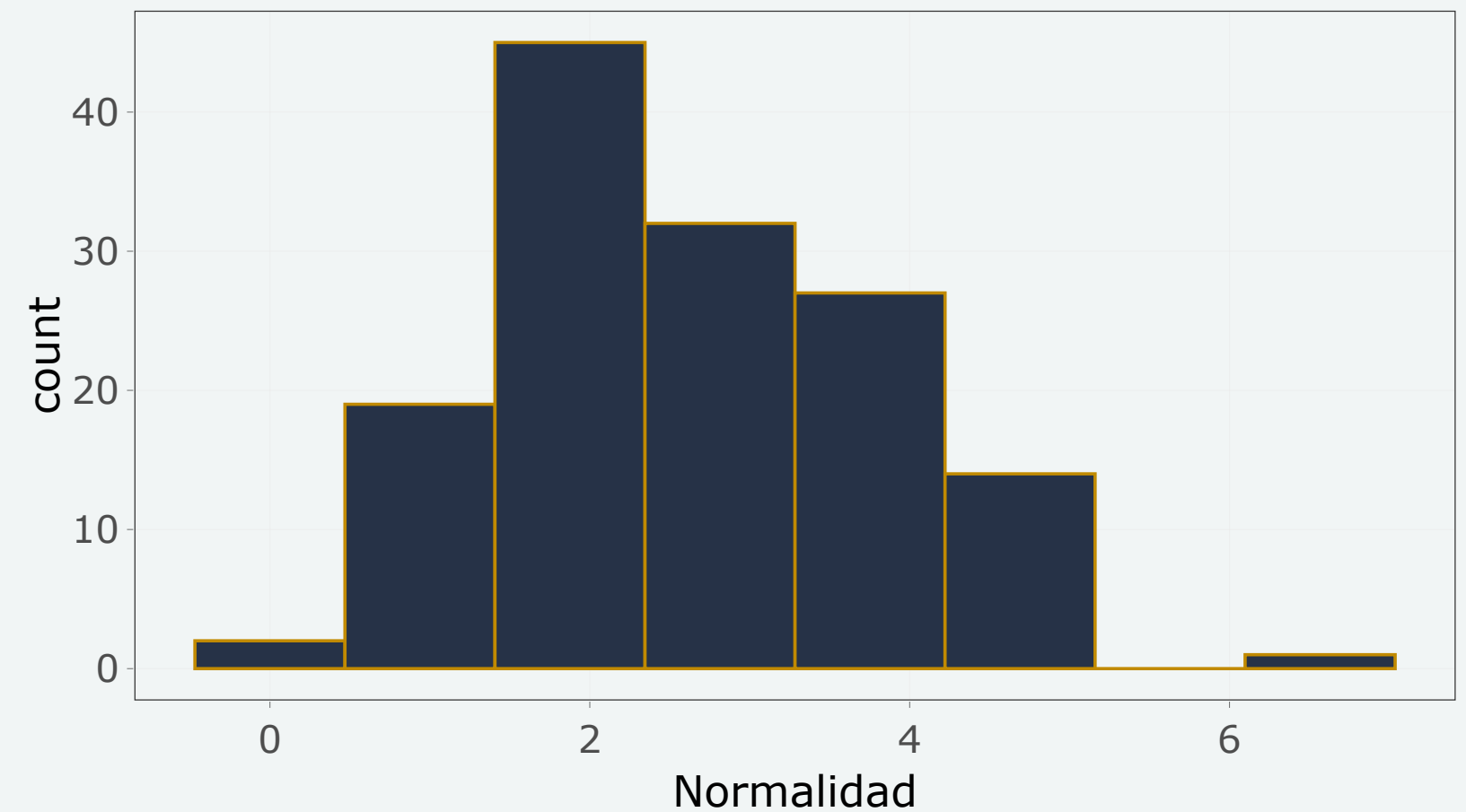
Kolmogórov - Smirnov (KS)

- Está centrado en la desviación general en la forma de la distribución. Requiere de parámetros conocidos, por lo que la versión de Minitab es menos potente, pues trabaja con estimaciones ($\bar{x}$ en lugar de μ). Al usar estimaciones se realiza la prueba de Lilliefors.
- Es más sensible a las colas que RJ y SW, pero menos que AD.
- KS no es por lo general la primera opción pues en colas pesadas gana AD y en el centrado/asimetría de la distribución gana RJ y SW.
 - Rara vez es preferido cuando parámetros son estimados



Ejemplo 02

En **Excel**, en la pestaña con este mismo nombre, puede encontrar un conjunto de datos. Realice un histograma de los datos. Con un 99 % de confianza, realice al menos tres pruebas de normalidad o con las que cuente el software que esté usando.



Ejemplo 02

- Con base en el análisis gráfico y los valores p obtenidos.
- Prepare un esquema en el que responda al ¿por qué con una prueba se rechaza y con las otras no?

Prueba	Valor.p
AD	0.009
KS	0.013
RJ	0.022



Conclusiones y recomendaciones

- Si el tamaño de muestra es pequeño, no conviene penalizar fuertemente las colas, pues un solo valor, que puede ser “mala suerte” puede tener mucho peso en la decisión sobre la normalidad.
- Si n es muy grande, conviene interpretar los valores P con cautela, pues son valores del tipo “caja negra”. Es decir, preste muchas más atención a las rutinas gráficas.
- Conviene siempre realizar el análisis gráfico (rutinas gráficas) primero, antes de aplicar una prueba formal de normalidad.
- Las conclusiones finales sobre la normalidad de un conjunto de datos **no pueden** estar basadas solo en un valor P .



Otras rutinas de identificación de distribuciones



Identificación de distribuciones

- Muchos softwares estadísticos (libres o comerciales) tienen incluidas rutinas de identificación de distribuciones que funcionan de forma que prueban muchas distribuciones a la vez.
- Esto es importante, para decidir sobre el tipo de modelo y análisis estadístico a emplear.
- Nótese que las distribuciones tienen sus propias características, no se deben usar indiscriminadamente. Las pruebas estadísticas no pueden supeditar la teoría y el sentido común.
- Tome en cuenta, también, que probar muchas distribuciones simultáneamente aumenta el chance de falso positivo. Probar muchas hasta que una dé $p > \alpha$ no es una evidencia fuerte.



Identificación de distribuciones

- Es decir, que un identificador de distribuciones puede decir que mis datos siguen alguna distribución y esto no tener sentido alguno en la práctica.
- Estudie si la distribución que le tocó es adecuada o si hay otra, no tan buena, pero que tiene más sentido teórico práctico.



Ejemplo 03

- Utilice los datos del ejemplo 02 y encuentre mediante un identificador de distribuciones (puede ser en R, Python, Minitab, etc), qué distribución se ajusta mejor a los datos. Use $\alpha = 0.01$.
- Recuerde que:
 - Las rutinas gráficas ya las hizo, por eso no se repiten, pero siempre deben hacerse.
 - ¿Va a probar todas las distribuciones? Si no conoce alguna, puede que sea conveniente que la estudie de forma individual.



Ejemplo 03

- El concepto de transformación no es abarcado aún. Por lo que de obtenerlos, puede ignorarlos.
- En este y otros ejercicios **puede** obtener valores diferentes a los aquí mostrados, como consecuencia de los algoritmos empleados en los cálculos.
 - Para este ejercicio, se está usando la prueba de KS, Minitab suele usar AD.
- Nótese como este conjunto de datos se ajusta a 4 distribuciones distintas. ¿Cuál escojo? Dependerá del contexto y conveniencia.

Resultados de un identificador de distribuciones

Distribución	Valor.p
Weibull	0.072
Lognormal	0.017
Exponencial	0.000
Gamma	0.094
Uniforme	0.000
Logística	0.008
Normal	0.013



Reflexión final

- Un modelo puede no ser perfecto y aún así ser útil.
- | All models are **wrong**, but some are **useful** - George Box



Bibliografía

- Las fuentes de esta sesión provienen de diferentes libros de texto, cuya mención no es necesaria porque ahondan en exceso en conceptos que no son pertinentes al alcance de este curso.



Pruebas de bondad de ajuste para distribuciones

II-1123 Estadística para Ingeniería Industrial II

Gracias por su atención

Steven García Goñi

steven.garciagoni@ucr.ac.cr

Dudas o correcciones requeridas pueden solicitarse al correo

