Tema 5
Contrastes de hipótesis

José R. Berrendero

Departamento de Matemáticas
Universidad Autónoma de Madrid

Estructura de este tema

- ¿Qué es un contraste de hipótesis?
- Elementos de un contraste: hipótesis, tipos de error, nivel de significación, región crítica.
- Ejemplos de contrastes para medias, varianzas y proporciones.
- Contrastes de bondad de ajuste.

¿Qué es un contraste de hipótesis?

Una **hipótesis** es una afirmación que se hace sobre la población.

La hipótesis es **paramétrica** si se refiere a los valores que toma alguno de los parámetros poblacionales.

Por ejemplo, una hipótesis paramétrica es: "la media poblacional es positiva" ($\mu > 0$).

Un **contraste de hipótesis** es una técnica estadística para juzgar si los datos aportan evidencia o no para confirmar una hipótesis.

Ejemplo

Los refrescos de cola *light* utilizan edulcorantes artificiales que pueden perder su efecto con el tiempo.

En un experimento se pidió a varias personas que probaran refrescos dietéticos y calificaran su grado de sabor dulce en una escala de 1 a 10.

Tras almacenar las bebidas durante un mes a alta temperatura (para imitar el efecto de 4 meses de almacenamiento a temperatura ambiente) las mismas personas probaron de nuevo los refrescos y calificaron de nuevo su grado de sabor dulce.

En la siguiente tabla aparecen las diferencias en las puntuaciones (a mayor diferencia, mayor caída del sabor):

$$2,\ 0.4,\ 0.7,\ 2,\ -0.4,\ 2.2,\ -1.3,\ 1.2,\ 1.1,\ 2.3$$

La mayoría de los datos son positivos. Es decir, la mayoría de las personas apreciaron pérdida en el nivel de sabor.

Pero las diferencias no son muy grandes (e incluso dos personas apreciaron un incremento).

La pregunta que trata de responder un contraste de hipótesis es:
¿Proporcionan estos datos evidencia de que el nivel medio de sabor decrece?

La media estimada a partir de los datos es $\bar{x} = 1.02$.

- ¿Refleja esta estimación un auténtico descenso en el nivel medio de sabor?
- ¿Se debe el resultado a razones puramente aleatorias?

Elementos de un contraste de hipótesis

La hipótesis para la que se desea encontrar evidencia se llama **hipótesis alternativa**. Se denota H_1.

La afirmación contraria a H_1 se llama **hipótesis nula**. Se denota H_0.

Llamamos μ al descenso medio (desconocido) del grado de sabor de los refrescos.

Como queremos confirmar si el grado medio realmente desciende, queremos contrastar

$$H_0 : \ \mu \leq 0 \ \text{ frente a } \ H_1 : \ \mu > 0$$

El **razonamiento básico** para hacer este contraste es:

- Supongamos que H_0 es cierta, es decir, $\mu \leq 0$.
- ¿Es el resultado obtenido a partir de los datos ($\bar{x} = 1.02$) extraño bajo esta hipótesis?
- Si esto es así, los datos aportan evidencia contra H_0 y a favor de H_1.

Para llevar a cabo el análisis anterior tenemos que estudiar qué valores son los que cabe esperar que tome $\bar{x}$ cuando H_0 es cierta.

Para simplificar suponemos de momento que la población es **normal** y que **la varianza es conocida** y vale $\sigma = 1$.

Supongamos que H_0 es cierta y que μ vale 0 (toma el valor en el que más difícil es distinguir entre H_0 y H_1).

Sabemos (tema 3) que

$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \equiv \mathrm{N}(0,1).$$

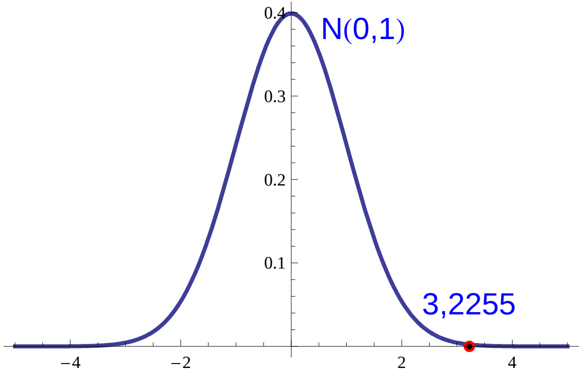
Para juzgar si el valor $\bar{x} = 1.02$ es compatible con $\mu = 0$ calculamos

$$t = \frac{1.02 - 0}{1/\sqrt{10}} = 3.2255$$

y comparamos con la distribución normal estándar.

Podemos interpretar $t = 3.2255$ como la distancia entre $\bar{x}$ y 0 *medida en desviaciones típicas*.

Mirando las tablas de la normal: $\mathrm{P}(Z > 3.2255) < 0.001$.



Si $\mu = 0$, en menos de 1 de cada 1000 muestras se obtendría un valor de t superior a 3.2255.

Si $\mu < 0$, la proporción de muestras sería todavía menor.

Como 3.2255 es un valor bastante improbable para una distribución $\mathrm{N}(0, 1)$, los datos proporcionan bastante evidencia en contra de H_0 y a favor de H_1.

Parece que la distancia entre $\bar{x}$ y 0 es “suficientemente grande” como para rechazar $H_0 : \ \mu \leq 0$.

Tipos de error

¿Qué significa "suficientemente grande"? Depende de lo seguros que queramos estar a la hora de rechazar o no la hipótesis nula. Se pueden cometer dos tipos de errores:

- **Error de tipo I**: Rechazar H_0 cuando es cierta.
- **Error de tipo II**: Aceptar H_0 cuando es falsa.

De los dos errores sólo vamos a poder controlar el error de tipo I. Por ello, se deben definir las hipótesis de forma que el error de tipo I sea el más grave (equivalentemente, H_1 debe ser la hipótesis que queremos confirmar).

Se llama **nivel de significación** α de un contraste a la mayor probabilidad de cometer un error de tipo I cuando se utiliza ese contraste.

Región crítica o de rechazo

Vamos a rechazar $H_0 : \mu \leq 0$ siempre que la distancia entre $\bar{x}$ y $\mu = 0$ sea "suficientemente grande", mayor que un valor crítico c:

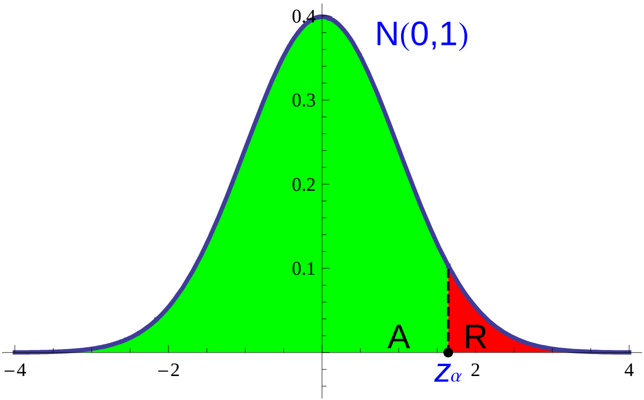
$$\frac{\bar{x} - 0}{1/\sqrt{10}} > c.$$

Para determinar c fijamos el nivel de significación α. Los valores $\alpha = 0.01$ o $\alpha = 0.05$ son los más habituales.

$$\alpha = \mathsf{P}_{H_0}(\textsc{Rechazar}) = \mathsf{P}_{H_0}\left(\frac{\bar{x} - 0}{1/\sqrt{10}} > c\right) \leq \mathsf{P}(Z > c).$$

Por lo tanto el valor $c = z_\alpha$ garantiza $\mathsf{P}_{H_0}(\textsc{Rechazar}) \leq \alpha$.

Región crítica o de rechazo



Región crítica o de rechazo

Rechazaremos $H_0 : \mu \leq 0$ a nivel α siempre que se verifique:

$$R = \left\{ \frac{\bar{x} - 0}{1/\sqrt{10}} > z_\alpha \right\}.$$

A R se le llama **región de rechazo** o **región crítica**.

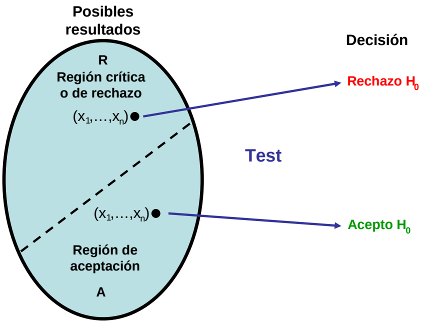
Para los datos del ejemplo recordemos que

$$\frac{\bar{x} - 0}{1/\sqrt{10}} = 3.2255$$

Para hacer el contraste a nivel $\alpha = 0.05$, buscamos en las tablas $z_{0.05} = 1.64$.

Como $3.2255 > 1.64$, estamos en la región crítica y rechazamos la hipótesis nula $\mu \leq 0$ a nivel $\alpha = 0.05$.

Región crítica o de rechazo



Para la mayoría de los contrastes la región crítica es de la forma:

$$R = \left\{ \frac{\text{Distancia entre datos y } H_0}{\text{E.T. de la distancia}} > c \ (\text{tablas}) \right\}.$$

Todos los contrastes se basan en las mismas ideas que hemos introducido hasta ahora. Vamos a ver ejemplos de aplicación de las fórmulas en distintas situaciones.

Ejemplo del edulcorante cuando σ es desconocida

Queremos contrastar $H_0 : \mu \leq 0$ frente a $H_1 : \mu > 0$ (es decir, contraste unilateral con $\mu_0 = 0$) a nivel $\alpha = 0.05$.

Suponemos ahora que σ no es conocida. La aproximamos a partir de la muestra:

$$2,\ 0.4,\ 0.7,\ 2,\ -0.4,\ 2.2,\ -1.3,\ 1.2,\ 1.1,\ 2.3$$

Para ello usamos el estimador:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = 1.196$$

Ejemplo del edulcorante cuando σ es desconocida

Calculamos el **estadístico t**:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1.02 - 0}{1.196/\sqrt{10}} = 2.697.$$

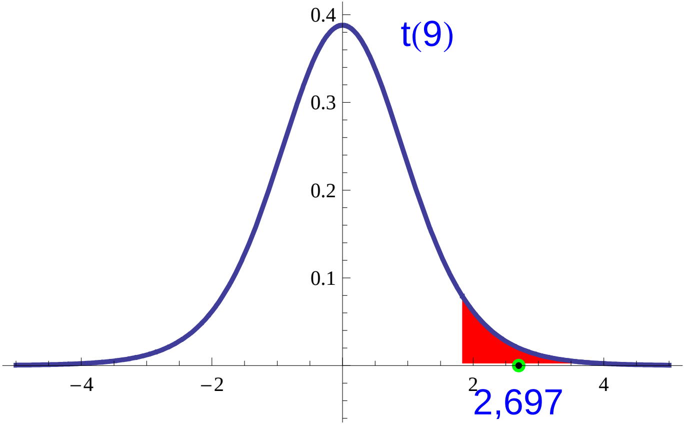
En las tablas de la t buscamos el valor:

$$t_{9,0.05} = 1.833$$

Como $2.697 > 1.833$ estamos en la región crítica y rechazamos H_0 a nivel $\alpha = 0.05$.

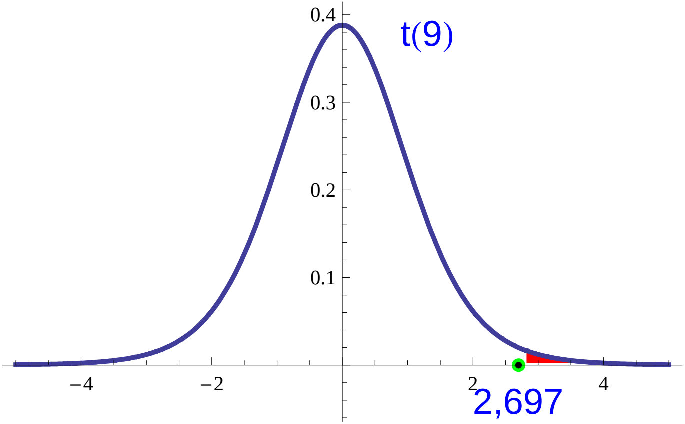
Ejemplo del edulcorante cuando σ es desconocida

Como $2.697 > 1.833$ estamos en la región crítica y rechazamos H_0 a nivel $\alpha = 0.05$.



Ejemplo del edulcorante cuando σ es desconocida

¿Cuál es la conclusión si fijamos $\alpha = 0.01$?



Contrastes para la media de una población normal (varianza desconocida)

Contrastes unilaterales:

- Hipótesis: $H_0 : \mu \leq \mu_0$ frente a $H_1 : \mu > \mu_0$
- Región crítica:

$$R = \left\{ \frac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{n-1,\alpha} \right\}.$$

- Hipótesis: $H_0 : \mu \geq \mu_0$ frente a $H_1 : \mu < \mu_0$
- Región crítica:

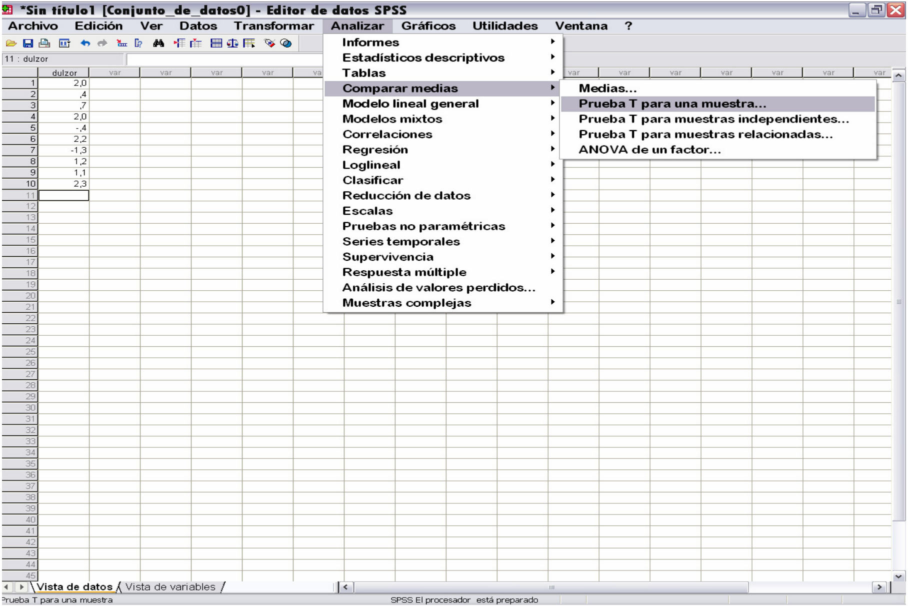
$$R = \left\{ \frac{\bar{x} - \mu_0}{s/\sqrt{n}} < -t_{n-1,\alpha} \right\}.$$

Contraste bilateral:

- Hipótesis: $H_0 : \mu = \mu_0$ frente a $H_1 : \mu \neq \mu_0$
- Región crítica:

$$R = \left\{ \frac{|\bar{x} - \mu_0|}{s/\sqrt{n}} > t_{n-1,\alpha/2} \right\}.$$

Solución con SPSS



Resultado que da SPSS

Estadísticos para una muestra

	N	Media	Desviación típ.	Error típ. de la media
dulzor	10	1,020	1,1961	,3782

Prueba para una muestra

	Valor de prueba = 0					
					95% Intervalo de confianza para la diferencia	
	t	gl	Sig. (bilateral)	Diferencia de medias	Inferior	Superior
dulzor	2,697	9	,025	1,0200	,164	1,876

- **Valor de prueba:** μ_0 (el valor de μ que separa H_0 de H_1). Por defecto $\mu_0 = 0$.
- **Sig (bilateral):** es el **p-valor** del test bilateral $H_0 : \mu = \mu_0$ frente a $H_1 : \mu \neq \mu_0$ (transparencia siguiente)

P-valor de un contraste

Dados unos datos, si el nivel de significación es pequeño resulta más difícil rechazar la hipótesis nula.

Hay un valor $\alpha = p$ a partir del cual ya no podemos rechazar H_0. Es decir si $\alpha < p$ ya no se rechaza H_0.

A p se le llama el **p-valor** del contraste. El p-valor indica el punto de división entre el rechazo y la aceptación:

- Si $\alpha < p$, no podemos rechazar H_0 a nivel α.
- Si $\alpha > p$, podemos rechazar H_0 a nivel α.

El p-valor es una medida del grado de compatibilidad de los datos con la hipótesis nula H_0.

Cuando es pequeño (usualmente menor que $\alpha = 0.05$) se considera que hay evidencia suficiente para rechazar H_0.

SPSS y p-valor

Los paquetes de software estadístico dan como resultado de un test su p-valor. `SPSS` llama `sig(bilateral)` al p-valor.

Conociendo el p-valor, el usuario puede tomar la decisión de aceptar o rechazar H_0 para cualquier α.

`SPSS` siempre calcula el p-valor para el contraste bilateral. Si queremos hacer un contraste unilateral, **usualmente tendremos que dividir entre 2 el valor calculado por SPSS**.

En el ejemplo, el p-valor del contraste es $0.025/2 = 0.0125$. Esto significa:

- Si $\alpha < 0.0125$ no podemos rechazar $\mu \leq 0$.
- Si $\alpha > 0.0125$ podemos rechazar $\mu \leq 0$.

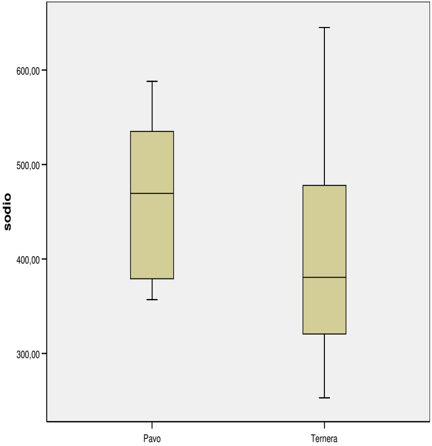
Comparación de dos medias (muestras independientes)

Se ha considerado la cantidad de calorías y de sodio en salchichas de varias marcas de dos tipos: ternera y pavo.

		Casos					
		Válidos		Perdidos		Total	
	tipo	N	Porcentaje	N	Porcentaje	N	Porcentaje
sodio	Pavo	16	100,0%	0	,0%	16	100,0%
	Ternera	20	100,0%	0	,0%	20	100,0%

Descriptivos

	tipo		Estadístico	Error típ.
sodio	Pavo	Media	460,8125	21,79435
		Mediana	469,5000	
		Varianza	7599,896	
		Desv. típ.	87,17738	
		Mínimo	357	
		Máximo	588	
		Rango	231,00	
		Amplitud intercuartil	161,50	
	Ternera	Media	401,1500	22,90510
		Mediana	380,5000	
		Varianza	10492,871	
		Desv. típ.	102,43472	
		Mínimo	253	
		Máximo	645	
		Rango	392,00	
		Amplitud intercuartil	158,75	



Parece que, en estas muestras, las salchichas de pavo tienen más sodio en media. Pero las dos muestras se solapan bastante. ¿Son las diferencias muestrales significativas?

¿Aportan evidencia estos datos para afirmar que el contenido medio de sodio de las salchichas de pavo es distinto al de las salchichas de ternera?

$X_1, \ldots, X_n$ es una muestra de $\mathrm{N}(\mu_1, \sigma)$
$Y_1, \ldots, Y_m$ es una muestra de $\mathrm{N}(\mu_2, \sigma)$

Supuestos necesarios:

- Las muestras proceden de dos poblaciones normales.
- Las varianzas son desconocidas pero iguales.
- Las dos muestras son independientes.

Hipótesis que queremos contrastar ($\alpha = 0.05$)

$$H_0: \ \mu_1 = \mu_2 \ \text{ frente a } \ H_1: \ \mu_1 \neq \mu_2$$

Región crítica (formulario)

$$R = \left\{ \frac{|\bar{x} - \bar{y}|}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}} > t_{n+m-2,\alpha/2} \right\}.$$

Estimador combinado de la varianza

$$s_p^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2}$$

Con los datos del ejemplo,

$$|\bar{x} - \bar{y}| = |460.81 - 401.15| = 59.66$$

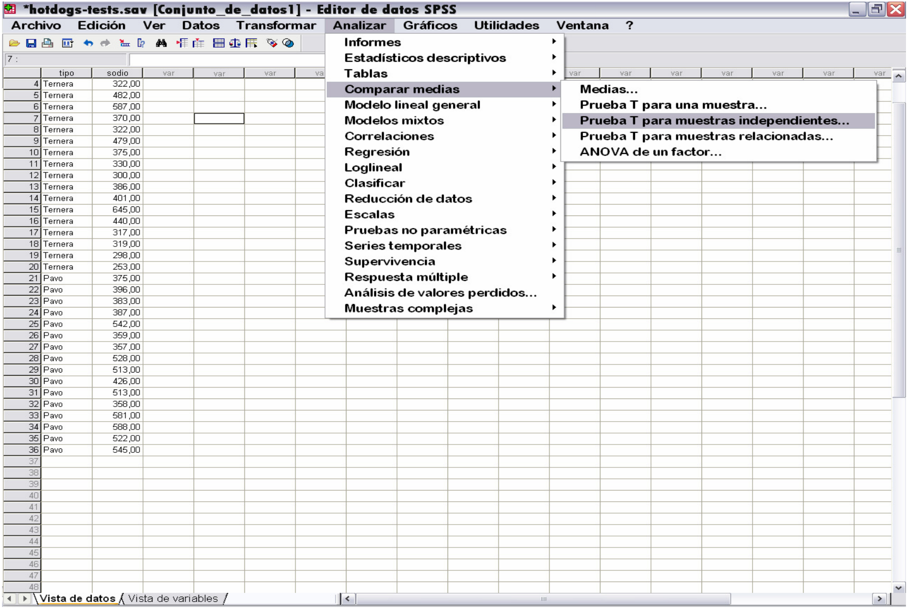
$$s_p^2 = \frac{15 \times 7599.89 + 19 \times 10492.871}{34} = 9216.556 \quad \text{y} \quad s_p = 96$$

$$t = \frac{|\bar{x} - \bar{y}|}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{59.66}{96\sqrt{1/16 + 1/20}} = \frac{59.66}{32.2} = 1.853$$

$$t_{34,0.025} \approx 2.04$$

Como $1.853 < 2.04$, no podemos rechazar H_0. Las diferencias encontradas en las cantidades medias de sodio de las dos muestras **no son significativas** al nivel $\alpha = 0.05$.

Con SPSS


*hotdogs-tests.sav [Conjunto_de_datos1] - Editor de datos SPSS
Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?
7 :
tipo sodio var var var
4 Ternera 322,00
5 Ternera 482,00
6 Ternera 587,00
7 Ternera 370,00
8 Ternera 322,00
9 Ternera 479,00
10 Ternera 375,00
11 Ternera 330,00
12 Ternera 300,00
13 Ternera 386,00
14 Ternera 401,00
15 Ternera 645,00
16 Ternera 440,00
17 Ternera 317,00
18 Ternera 319,00
19 Ternera 298,00
20 Ternera 253,00
21 Pavo 375,00
22 Pavo 396,00
23 Pavo 383,00
24 Pavo 387,00
25 Pavo 542,00
26 Pavo 359,00
27 Pavo 357,00
28 Pavo 528,00
29 Pavo 513,00
30 Pavo 426,00
31 Pavo 513,00
32 Pavo 358,00
33 Pavo 581,00
34 Pavo 588,00
35 Pavo 522,00
36 Pavo 545,00
Informes
Estadísticos descriptivos
Tablas
Comparar medias
Modelo lineal general
Modelos mixtos
Correlaciones
Regresión
Loglineal
Clasificar
Reducción de datos
Escalas
Pruebas no paramétricas
Series temporales
Supervivencia
Respuesta múltiple
Análisis de valores perdidos...
Muestras complejas
Medias...
Prueba T para una muestra...
Prueba T para muestras independientes...
Prueba T para muestras relacionadas...
ANOVA de un factor...
Vista de datos
Vista de variables

Estadísticos de grupo

	tipo	N	Media	Desviación típ.	Error típ. de la media
sodio	Pavo	16	460,8125	87,17738	21,79435
	Ternera	20	401,1500	102,43472	22,90510

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	95% Intervalo de confianza para la diferencia	
									Inferior	Superior
sodio	Se han asumido varianzas iguales	,008	,930	1,853	34	,073	59,66250	32,2003	-5,77649	125,10149
	No se han asumido varianzas iguales			1,887	33,84	,068	59,66250	31,6170	-4,60214	123,92714

- El p-valor es 0.073. Esto significa que se puede rechazar H_0 si $\alpha > 0.073$. Al nivel $\alpha = 0.05$ no podemos rechazar.
- Error típico de la diferencia: $s_p\sqrt{1/n + 1/m}$.

La hipótesis de igualdad de varianzas

$$H_0 : \sigma_1 = \sigma_2 \text{ frente a } H_1 : \sigma_1 \neq \sigma_2,$$

Idea básica

Se calcula el cociente $F = s_1^2/s_2^2$. Si H_0 fuese cierta deberíamos observar $F \approx 1$. La hipótesis nula se rechaza si F está "lejos" de la unidad.

Región crítica

Se rechaza H_0 a nivel α si

$$F < F_{m-1,n-1,1-\alpha/2} \quad \text{ó} \quad F > F_{m-1,n-1,\alpha/2},$$

donde los valores $F_{m,n,\alpha}$ se buscan en las tablas de la distribución F.

Obs. $F_{m,n,1-\alpha} = 1/F_{n,m,\alpha}$

Tablas de la distribución F

	n_1															
n_2	1	2	3	4	5	6	7	8	9	10	12	15	16	18	20	24
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	246.5	247.3	248.0	249.1
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.43	19.44	19.45	19.45
3	10.13	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786	8.745	8.703	8.692	8.675	8.660	8.639
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.912	5.858	5.844	5.821	5.803	5.774
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.678	4.619	4.604	4.579	4.558	4.527
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.000	3.938	3.922	3.896	3.874	3.841
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.575	3.511	3.494	3.467	3.445	3.410
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347	3.284	3.218	3.202	3.173	3.150	3.115
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.073	3.006	2.989	2.960	2.936	2.900
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.913	2.845	2.828	2.798	2.774	2.737
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.788	2.719	2.701	2.671	2.646	2.609
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687	2.617	2.599	2.568	2.544	2.505
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.604	2.533	2.515	2.484	2.459	2.420
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.534	2.463	2.445	2.413	2.388	2.349
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.475	2.403	2.385	2.353	2.328	2.288
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.425	2.352	2.333	2.302	2.276	2.235
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.381	2.308	2.289	2.257	2.230	2.190
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.342	2.269	2.250	2.217	2.191	2.150
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.308	2.234	2.215	2.182	2.155	2.114
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.278	2.203	2.184	2.151	2.124	2.082
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321	2.250	2.176	2.156	2.123	2.096	2.054
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297	2.226	2.151	2.131	2.098	2.071	2.028
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275	2.204	2.128	2.109	2.075	2.048	2.005
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255	2.183	2.108	2.088	2.054	2.027	1.984
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236	2.165	2.089	2.069	2.035	2.007	1.964
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220	2.148	2.072	2.052	2.018	1.990	1.946
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204	2.132	2.056	2.036	2.002	1.974	1.930
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190	2.118	2.041	2.021	1.987	1.959	1.915
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177	2.104	2.027	2.007	1.973	1.945	1.901
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165	2.092	2.015	1.995	1.960	1.932	1.887
35	4.121	3.267	2.874	2.641	2.485	2.372	2.285	2.217	2.161	2.114	2.041	1.963	1.942	1.907	1.878	1.833
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077	2.003	1.924	1.904	1.868	1.839	1.793
45	4.057	3.204	2.812	2.579	2.422	2.308	2.221	2.152	2.096	2.049	1.974	1.895	1.874	1.838	1.808	1.762
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026	1.952	1.871	1.850	1.814	1.784	1.737
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993	1.917	1.836	1.815	1.778	1.748	1.700
70	3.978	3.128	2.736	2.503	2.346	2.231	2.143	2.074	2.017	1.969	1.893	1.812	1.790	1.753	1.722	1.674
80	3.960	3.111	2.719	2.486	2.329	2.214	2.126	2.056	1.999	1.951	1.875	1.793	1.772	1.734	1.703	1.654
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927	1.850	1.768	1.746	1.708	1.676	1.627
125	3.917	3.069	2.677	2.444	2.287	2.172	2.084	2.013	1.956	1.907	1.830	1.747	1.725	1.687	1.655	1.605
150	3.904	3.056	2.665	2.432	2.274	2.160	2.071	2.001	1.943	1.894	1.817	1.734	1.711	1.673	1.641	1.590
175	3.895	3.048	2.656	2.423	2.266	2.151	2.062	1.992	1.934	1.885	1.808	1.724	1.702	1.663	1.631	1.580
200	3.888	3.041	2.650	2.417	2.259	2.144	2.056	1.985	1.927	1.878	1.801	1.717	1.694	1.656	1.623	1.572
300	3.873	3.026	2.635	2.402	2.244	2.129	2.040	1.969	1.911	1.862	1.785	1.700	1.677	1.638	1.606	1.554
400	3.865	3.018	2.627	2.394	2.237	2.121	2.032	1.962	1.903	1.854	1.776	1.691	1.669	1.630	1.597	1.545
500	3.860	3.014	2.623	2.390	2.232	2.117	2.028	1.957	1.899	1.850	1.772	1.686	1.664	1.625	1.592	1.539
750	3.854	3.008	2.617	2.384	2.226	2.111	2.022	1.951	1.892	1.843	1.765	1.680	1.657	1.618	1.585	1.532
1000	3.851	3.005	2.614	2.381	2.223	2.108	2.019	1.948	1.889	1.840	1.762	1.676	1.654	1.614	1.581	1.528

Comparación de dos medias (datos emparejados)

Se usan cinco dosis de una sustancia ferrosa para determinar si existen diferencias entre llevar a cabo un análisis químico de laboratorio o un análisis de fluorescencia por rayos X para determinar el contenido de hierro. Cada dosis se divide en dos partes iguales a las que se aplica cada uno de los dos procedimientos. Los resultados obtenidos son los siguientes:

Dosis	1	2	3	4	5
Rayos X	2.0	2.0	2.3	2.1	2.4
Análisis Químico	2.2	1.9	2.5	2.3	2.4

Se supone que las poblaciones son normales. ¿Aportan los datos evidencia suficiente a nivel $\alpha = 0.05$ para afirmar que el contenido medio de hierro detectado cuando se utiliza el análisis químico es diferente del contenido medio detectado cuando se utilizan rayos X?

Parámetros:

- μ_1 es el contenido medio detectado por rayos X
- μ_2 es el contenido medio detectado por análisis químico.

Hipótesis:

Cuando las muestras **no son independientes**, en lugar de contrastar $H_0 : \mu_1 = \mu_2$ frente a $H_1 : \mu_1 \neq \mu_2$, se contrasta

$$H_0 : \mu = 0 \quad \text{frente a} \quad H_1 : \mu \neq 0,$$

donde μ es el valor esperado de las diferencias $d_i = x_i - y_i$.

Dosis	1	2	3	4	5
x_i	2.0	2.0	2.3	2.1	2.4
y_i	2.2	1.9	2.5	2.3	2.4
d_i	-0.2	0.1	-0.2	-0.2	0

Con estos datos: $\bar{d} = -0.1$ y $S_d = 0.1414$.

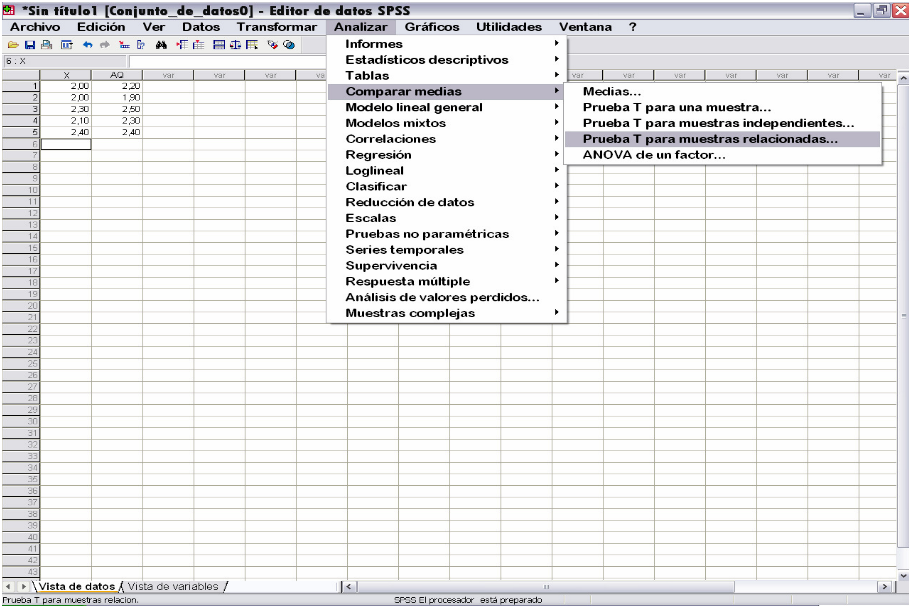
Región crítica (formulario):

$$R = \left\{ \frac{|\bar{d}|}{S_d/\sqrt{n}} > t_{n-1;\alpha/2} \right\}.$$

Mirando en las tablas $t_{4;0.025} = 2.776$.
Por otra parte,

$$\frac{|\bar{d}|}{S_d/\sqrt{n}} = \frac{0.1}{0.1414/\sqrt{5}} = 1.5811.$$

Como $1.5811 < 2.776$, los datos disponibles no permiten afirmar a nivel 0.05 que los dos métodos proporcionan cantidades medias de hierro diferentes.


*Sin título1 [Conjunto_de_datos0] - Editor de datos SPSS
Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?
6 : X
X AQ var var var
1 2,00 2,20
2 2,00 1,90
3 2,30 2,50
4 2,10 2,30
5 2,40 2,40
Informes
Estadísticos descriptivos
Tablas
Comparar medias
Modelo lineal general
Modelos mixtos
Correlaciones
Regresión
Loglineal
Clasificar
Reducción de datos
Escalas
Pruebas no paramétricas
Series temporales
Supervivencia
Respuesta múltiple
Análisis de valores perdidos...
Muestras complejas
Medias...
Prueba T para una muestra...
Prueba T para muestras independientes...
Prueba T para muestras relacionadas...
ANOVA de un factor...
Vista de datos
Vista de variables
Prueba T para muestras relacion.
SPSS El procesador está preparado

Estadísticos de muestras relacionadas

		Media	N	Desviación típ.	Error típ. de la media
Par 1	X	2,1600	5	,18166	,08124
	AQ	2,2600	5	,23022	,10296

Correlaciones de muestras relacionadas

		N	Correlación	Sig.
Par 1	X y AQ	5	,789	,113

Prueba de muestras relacionadas

	Diferencias relacionadas					t
				95% Intervalo de confianza para la diferencia		
	Media	Desviación típ.	Error típ. de la media	Inferior	Superior	
	-,10000	,14142	,06325	-,27560	,07560	-1,581

Prueba de muestras relacionadas

	gl	Sig. (bilateral)
	4	,189

Contrastes para una proporción

En un estudio, 1000 personas siguieron una dieta de adelgazamiento durante 3 meses. De las 1000 personas, 791 perdieron más de 3 kg de peso. ¿Permiten los datos afirmar, con el nivel de significación $\alpha = 0.01$, que más del 70% de la población perdería más de 3 kg de peso de seguir la misma dieta durante el mismo tiempo?

Hipótesis:

$$H_0 : \ p \leq 0.7 \ \text{ frente a } \ H_1 : \ p > 0.7,$$

donde p es la proporción poblacional que pierde peso.

Región crítica (formulario):

$$R = \left\{ \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > z_\alpha \right\}$$

En este caso, $n = 1000$, $p_0 = 0.7$, $\hat{p} = 0.791$ y $z_{0.01} = 2.33$.

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.791 - 0.7}{\sqrt{\frac{0.7 \times 0.3}{1000}}} = 6.28$$

Por lo tanto, podemos rechazar H_0 y afirmar que más del 70% de la población perdería más de 3 kg de peso de seguir la misma dieta durante el mismo tiempo.

Cuestiones

- ¿Es el p-valor menor o mayor que 0.02?
- Con los mismos datos, ¿podemos afirmar a nivel $\alpha = 0.01$ que menos del 90% de la población perdería más de 3 kg?
- La misma pregunta si la muestra es de 10 personas y hay 8 de ellas que pierden más de 3 kg.

Comparación de dos proporciones

Se ha llevado a cabo un estudio para determinar si un medicamento dirigido a reducir el nivel de colesterol reduce también la probabilidad de sufrir un infarto. Para ello, a hombres de entre 45 y 55 años se les asignó aleatoriamente uno de los dos tratamientos siguientes:

- 2051 hombres tomaron un medicamento para reducir el nivel de colesterol
- 2030 hombres tomaron un placebo

Durante los cinco años que duró el estudio, 56 de los hombres que tomaron el medicamento, y 84 de los que tomaron el placebo, sufrieron infartos.

¿Podemos afirmar a nivel 0.05 que el medicamento es efectivo?

Parámetros:

- p_1: Probabilidad de sufrir un infarto si se toma el medicamento.
- p_2: Probabilidad de sufrir un infarto si se toma el placebo.

Hipótesis: $H_0 : p_2 \leq p_1$ frente a $H_1 : p_2 > p_1$.

Estimadores de los parámetros:

$$\hat{p}_1 = \frac{56}{2051} = 0.0273 \text{ y } \hat{p}_2 = \frac{84}{2030} = 0.0414$$

Estimación de la probabilidad de infarto si fuese $p_1 = p_2$ (es decir, cuando H_0 es cierta pero la probabilidad de error es más alta):

$$\bar{p} = \frac{\textsc{Numero total de infartos}}{\textsc{Numero total de personas}} = \frac{56+84}{2051+2030} = 0.0343$$

Región crítica (formulario)

$$R = \left\{ \frac{\hat{p}_2 - \hat{p}_1}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} > z_\alpha \right\}$$

Con los datos del ejemplo:

$$\frac{\hat{p}_2 - \hat{p}_1}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.0141}{\sqrt{0.0343 \times 0.9657 \times \left(\frac{1}{2051} + \frac{1}{2030}\right)}} = 2.47$$

$$z_{0.05} = 1.64$$

Como $2.47 > 1.64$, podemos rechazar H_0 y afirmar que el medicamento es efectivo a nivel $\alpha = 0.05$.

Contrastes para la varianza de una población normal

Contrastes unilaterales:

- Hipótesis: $H_0 : \sigma \leq \sigma_0$ frente a $H_1 : \sigma > \sigma_0$
- Región crítica:

$$R = \left\{ \frac{(n-1)s^2}{\sigma_0^2} \geq \chi^2_{n-1;\alpha} \right\}.$$

- Hipótesis: $H_0 : \sigma \geq \sigma_0$ frente a $H_1 : \sigma < \sigma_0$
- Región crítica:

$$R = \left\{ \frac{(n-1)s^2}{\sigma_0^2} \leq \chi^2_{n-1;1-\alpha} \right\}.$$

Contraste bilateral:

- Hipótesis: $H_0 : \sigma = \sigma_0$ frente a $H_1 : \sigma \neq \sigma_0$
- Región crítica:

$$R = \left\{ \frac{(n-1)s^2}{\sigma_0^2} \notin (\chi^2_{n-1;1-\alpha/2}\,, \chi^2_{n-1;\alpha/2}) \right\}.$$

Ejemplo

Ejemplo: El calcio se presenta normalmente en la sangre de los mamíferos en concentraciones de alrededor de 6 mg. por cada 100 ml. de sangre. La desviación típica habitual de esta variable es 1 mg. de calcio por cada 100 ml. Una variabilidad mayor que ésta puede ocasionar graves trastornos en la coagulación de la sangre. Una serie de nueve pruebas sobre un paciente revelaron una media muestral de 6.2 mg. de calcio por 100 ml. de sangre, y una cuasidesviación típica muestral de 2 mg. de calcio por cada 100 ml. de sangre. ¿Hay alguna evidencia, para un nivel $\alpha = 0.05$, de que el nivel medio de calcio para este paciente sea más alto de lo normal? ¿Hay alguna evidencia, para un nivel $\alpha = 0.05$, de que la desviación típica del nivel de calcio sea más alta de lo normal?

Contrastes de bondad de ajuste

Durante el curso hemos manejado suposiciones del tipo:

- X es una variable aleatoria normal.
- X es de Poisson (o exponencial, etc.)

Sin embargo en la práctica no se conoce la distribución de la población.

Los contrastes de bondad de ajuste proporcionan criterios objetivos para decidir si estas hipótesis son razonables a partir de los datos muestrales.

Las medidas en las que se basan los contrastes tienen, aproximadamente, una distribución χ^2.

Caso 1: Distribución totalmente especificada bajo H_0

Objetivo: Decidir, en base a una muestra observada $(x_1, \ldots, x_n)$, si es razonable admitir que X sigue una determinada distribución P_0.

$$\begin{cases} H_0 : \text{la distribución de } X \text{ es } P_0, \\ H_1 : \text{la distribución de } X \textit{ no} \text{ es } P_0. \end{cases}$$

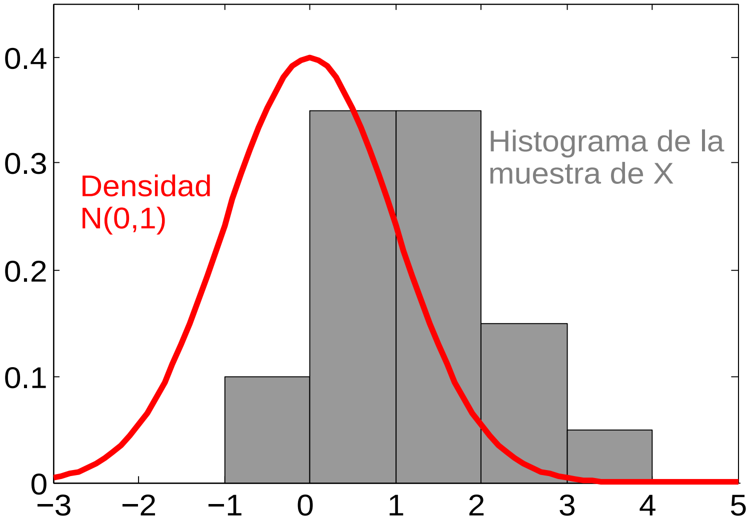
Ejemplo: Consideremos la siguiente muestra de tamaño 20 de X:

0.57	-0.67	1.13	1.29	-0.15	2.20	2.19	0.96	1.33	1.17
0.81	1.73	0.41	3.18	0.86	1.11	2.07	1.06	0.90	0.17

Queremos contrastar:

$$\begin{cases} H_0 : X \sim \mathrm{N}(0; 1), \\ H_1 : \text{la distribución de } X \textit{ no} \text{ es } \mathrm{N}(0;1). \end{cases}$$

Ejemplo



0.4
0.3
0.2
0.1
0
−3
−2
−1
0
1
2
3
4
5
Densidad
N(0,1)
Histograma de la
muestra de X

Procedimiento

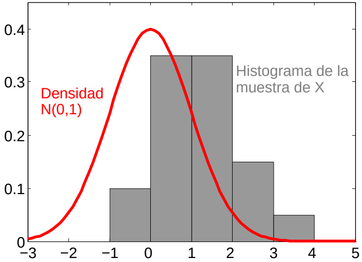
Paso 1: Se divide el rango de valores de X en k clases $A_1, \ldots, A_k$.

Paso 2: Para cada clase A_i, se consideran las frecuencias absolutas observadas y esperadas:

$$
\begin{aligned}
o_i &\equiv \textbf{ Frecuencia observada} \text{ en } A_i \\
&= \text{ número de elementos de la muestra } X_1, X_2, \ldots, X_n \text{ en } A_i.
\end{aligned}
$$

$$
\begin{aligned}
e_i &\equiv \textbf{ Frecuencia esperada} \text{ en } A_i \text{ si } H_0 \text{ es cierto} \\
&= n\, P_0(A_i).
\end{aligned}
$$

Ejemplo (cont.):



	$A_1 = (-\infty, -1]$	$A_2 = (-1, 0]$	$A_3 = (0, 1]$	$A_4 = (1, 2]$	$A_5 = (2, \infty)$
o_i	0	2	7	7	4
e_i					

Procedimiento

Paso 3: Bajo H_0, medimos la discrepancia entre los datos y la distribución de H_0 mediante el *estadístico de Pearson*:

$$T = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} = \sum_{i=1}^{k} \frac{o_i^2}{e_i} - n,$$

sigue aproximadamente (bajo H_0 y para n grande) una distribución χ^2_{k-1}.

Contraste: Cuanto mayor es T, peor es el ajuste, ya que hay más diferencia entre lo observado y lo esperado bajo H_0. La región de rechazo del contraste, a nivel de significación α, es:

$$R = \left\{ T > \chi^2_{k-1;\alpha} \right\}.$$

Ejemplo (cont.):

	$A_1 = (-\infty, -1]$	$A_2 = (-1, 0]$	$A_3 = (0, 1]$	$A_4 = (1, 2]$	$A_5 = (2, \infty)$
o_i	0	2	7	7	4
e_i					

$T =$

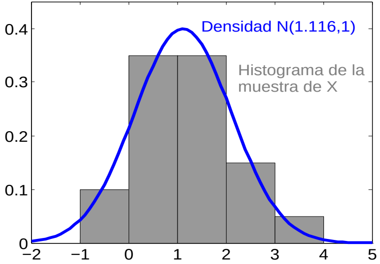
$\chi^2_{k-1;\alpha} =$

Conclusión:

Caso 2: pertenencia a una familia paramétrica

Objetivo: Podemos plantearnos el problema más general de contrastar si X sigue una distribución dada por un modelo en el que los valores de r parámetros no se especifican.

Ejemplo (cont.): $\begin{cases} H_0 : X \sim \text{N}(\mu, 1) \text{ para algún } \mu \in \mathbb{R}, \\ H_1 : \text{La distribución de } X \text{ no es } \text{N}(\mu, 1). \end{cases}$



$$\hat{\mu} = \bar{x} = 1.116$$

Procedimiento

Paso 1: Se divide el rango de valores de X en k clases $A_1, \ldots, A_k$.

Paso 2: Para cada clase A_i, se consideran las frecuencias absolutas observadas y esperadas:

$o_i \equiv$ **Frecuencia observada** en A_i

$e_i \equiv$ **Frecuencia esperada** en A_i si H_0 es cierto

$= n\,P_{\boldsymbol{\theta}_0}(A_i) \simeq n\,P_{\hat{\boldsymbol{\theta}}}(A_i),$ donde $\hat{\boldsymbol{\theta}}$ es el E.M.V. de $\boldsymbol{\theta}$.

Paso 3: Bajo H_0,

$$T = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} = \sum_{i=1}^{k} \frac{o_i^2}{e_i} - n \approx \chi^2_{k-1-r}.$$

Contraste: La región de rechazo del contraste, a nivel de significación α, es: $R = \left\{ T > \chi^2_{k-1-r;\alpha} \right\}$.

Ejemplo (cont.):

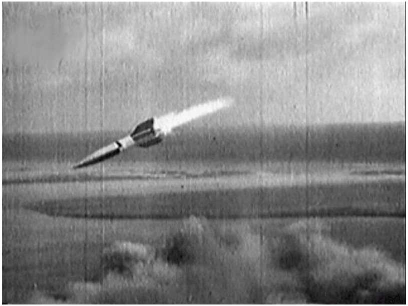
	$(-\infty,-1]$	$(-1,0]$	$(0,1]$	$(1,2]$	$(2,3]$	$(3,\infty)$
o_i	0	2	7	7	3	1
e_i						

$T =$

$\chi^2_{k-1-r;\alpha} =$

Conclusión:

Ejemplo: los bombardeos de Londres

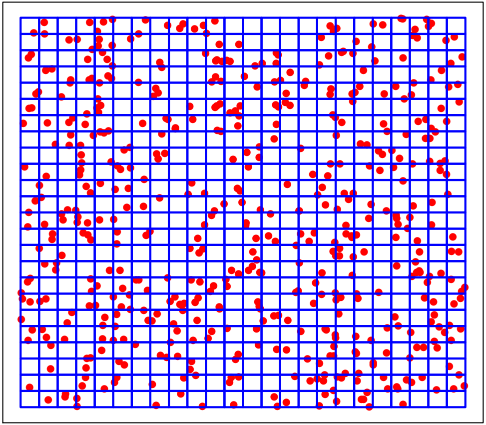


- Desde el 13 de junio de 1944 y durante nueve meses fueron lanzados contra Inglaterra unos 2.300 misiles V-1.
- ¿Eran los lugares de impacto en Londres aleatorios?

Ejemplo: los bombardeos de Londres

- Clarke, en 1946, dividió un área de 144 km^2 del sur de Londres en 576 cuadrados de 0.25 km^2 cada uno.
- La zona había registrado 537 impactos, por lo que la media era de $537/576 \approx 0.9323$ misiles por cuadrado.
- ¿Se ajusta la fórmula de Poisson a la distribución de impactos?

Ejemplo: los bombardeos de Londres



Ejemplo: los bombardeos de Londres

The actual results were as follows:

No. of flying bombs per square	Expected no. of squares (Poisson)	Actual no. of squares
0	226·74	229
1	211·39	211
2	98·54	93
3	30·62	35
4	7·14	7
5 and over	1·57	1
	576·00	576

Test basado en la distribución χ^2

Hipótesis nula: La distribución es Poisson de parámetro λ.

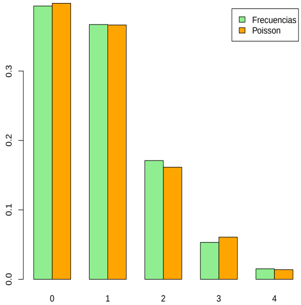
Bajo H_0, el estimador de máxima verosimilitud de λ es:

$$\hat{\lambda} = \frac{0 \times 229 + 1 \times 211 + \cdots + 1 \times 7}{576} \approx 0.9323$$

Frecuencias esperadas estimadas bajo H_0:

$$\hat{E}_k = n\hat{p}_k = 576 \times e^{-\hat{\lambda}} \frac{\hat{\lambda}^k}{k!}$$

k	O_k	$\hat{p}_k$	$\hat{E}_k$
0	229	0.39	226.74
1	211	0.37	211.34
2	93	0.17	98.54
3	35	0.05	30.62
≥ 4	8	0.02	8.71



Estadístico χ^2 de Pearson:

$$T = \sum_k \frac{(O_k - \hat{E}_k)^2}{\hat{E}_k} = 1.0176$$

Bajo H_0, $T \equiv \chi^2_3$ (5 clases - 1 parámetro estimado -1).

Si $Y \cong \chi^2_3$, $\mathrm{P}\{Y > 1.0176\} \approx 0.797$.

El p-valor del contraste es aproximadamente 0.797.

Al nivel habitual $\alpha = 0.05$ no se puede rechazar que los datos procedan de una distribución de Poisson.