TEMA 3:
DISTRIBUCIONES EN EL MUESTREO PARA POBLACIONES NORMALES

Estimación I

Grado en Estadística Aplicada

Curso 2019-2020

Como ya hemos visto, la distribución teórica **normal es muy frecuente** en estadística. Además, aunque la distribución teórica de los datos no sea normal, para **muestras grandes**, la distribución de la media muestral también se **aproxima** a una normal (por el Teorema Central del Límite).

Por dicho motivo, en este tema vamos a estudiar algunas **distribuciones relacionadas con la normal**, así como la **distribución en el muestreo** de varios estadísticos cuando se asume que la distribución poblacional es normal.

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma)$.

- $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ es el estimador máximo verosimil, así como el EIMV, para μ.
- Además, $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$.
- Si $\mu = \mu_0$ es conocido, $\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_0)^2$ es el estimador máximo verosimil, así como el EIMV, para σ^2.
- Si μ es desconocido,
  - $S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$ es el estimador máximo verosimil para σ^2.
  - $S_c^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ es el EIMV para σ^2.

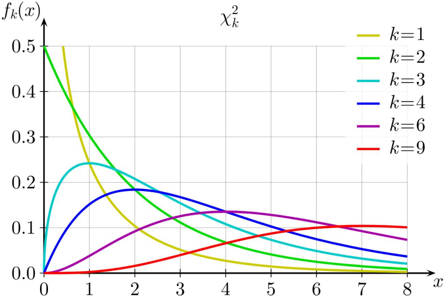
Definición: distribución χ^2

Sea $X \sim N(0,1)$ y sea $Y = X^2$. Por definición, diremos que Y sigue una distribución **chi-cuadrado** con un grado de libertad χ^2_1.

> Si $Y \sim \chi^2_1$, entonces $Y \sim \Gamma(\frac{1}{2}, \frac{1}{2})$.

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(0,1)$ y sea $Y = \sum_{i=1}^{n} X_i^2$. Por definición, diremos que Y sigue una distribución **chi-cuadrado** con n grados de libertad χ^2_n.

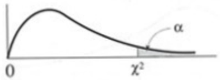
> Si $Y \sim \chi^2_n$, entonces $Y \sim \Gamma(\frac{n}{2}, \frac{1}{2})$.



PROPIEDADES DE LA DISTRIBUCIÓN χ^2

- La distribución χ^2 es **reproductiva**.
- $E\left[\chi^2_n\right] = n$.
- $Var\left[\chi^2_n\right] = 2n$.
- No hay una expresión sencilla para su función de distribución, por lo que (como ocurre con la normal) se hace necesario el uso de **tablas**.
- Para valores de n **grandes**, $\chi^2_n \approx N(n, \sqrt{2n})$.

DISTRIBUCIÓN χ^2



dl	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.8
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.82	9.35	11.35	12.8
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.8
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.7
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.5
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.2
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.54	20.09	21.9
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.66	23.5
10	2.15	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.1
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.7
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.21	28.3
13	3.56	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.8
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.69	26.12	29.14	31.3
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.8
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.2

$$P\left\{\chi_6^2 > 7.84\right\} = 0.25$$

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(0, \sqrt{\theta})$. Calcula la cota de Cramer-Rao para θ y demuestra que $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ es insesgado y su varianza alcanza la cota.

Ejercicio 1

La cantidad de vitamina C (en gr.) contenida en una naranja se distribuye según χ^2_1. Si la cantidad de vitamina C recomendada son 3 gr., ¿con que probabilidad no se alcanza esa cantidad al consumir una naranja? Suponiendo que la cantidad de vitamina C que contienen las naranjas es independiente, ¿cuántas naranjas ha de consumir para asegurar que se han ingerido 3 gr. con una probabilidad del 95 %?

DEFINICIÓN: DISTRIBUCIÓN t DE STUDENT

Sea $X \sim N(0,1)$ e $Y \sim \chi^2_n$ **independientes**. Por definición, diremos que

$$T = \frac{X}{\sqrt{Y/n}}$$

sigue una distribución **t de Student** con n grados de libertad t_n.

Así, si dividimos una normal tipificada entre la raiz de una χ^2 corregida por sus grados de libertad, siendo ambas independientes, la distribución resultante es t.

PROPIEDADES DE LA DISTRIBUCIÓN t DE STUDENT

- $E\left[t_n\right] = 0$, si $n > 1$.
- $Var\left[t_n\right] = \frac{n}{n-2}$, si $n > 2$.

Propiedades de la distribución t de Student

- Se trata de una distribución **simétrica**, similar a la de la normal, pero con colas **más pesadas**.
- No hay una expresión sencilla para su función de distribución, por lo que se hace necesario el uso de **tablas**.
- Para valores de n **grandes**, $t_n \approx N(0,1)$.

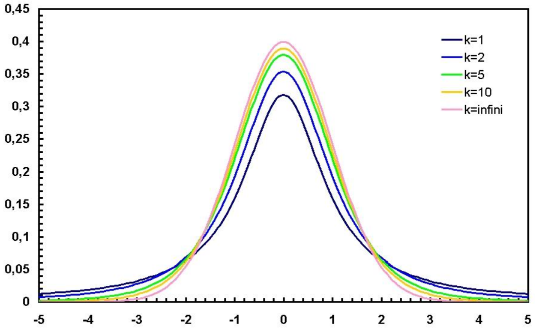
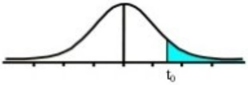


Tabla t-Student



Grados de libertad	0.25	0.1	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.8165	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.7407	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467

$$P\{t_3 > 2.3534\} = P\{t_3 < -2.3534\} = 0.05$$

EJERCICIO 2

Sea $X_1, \ldots, X_5$ una m.a.s. de una $N(0,1)$. Determina el valor de c para que la siguiente v.a. tenga distribución t de Student:

$$c \frac{X_1 + X_2}{\sqrt{X_3^2 + X_4^2 + X_5^2}}$$

Definición: distribución F de Snedecor

Sea $X \sim \chi^2_n$ e $Y \sim \chi^2_m$ **independientes**. Por definición, diremos que

$$F = \frac{X/n}{Y/m}$$

sigue una distribución **F de Snedecor** con n y m grados de libertad $F_{n,m}$.

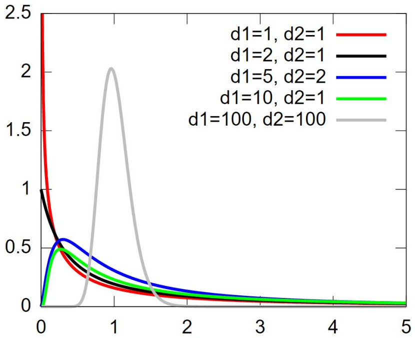
Así, si dividimos dos chi-cuadrados independientes corregidas por sus grados de libertad, la distribución resultante es F.

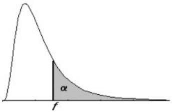
Propiedades de la distribución F de Snedecor

- $E\left[F_{n,m}\right] = \frac{m}{m-2}$, si $m > 2$.
- Si $T \sim t_n$, entonces $T^2 \sim F_{1,n}$.

Propiedades de la distribución F de Snedecor

- Si $F \sim F_{n,m}$, entonces $\frac{1}{F} \sim F_{m,n}$.
- No hay una expresión sencilla para su función de distribución, por lo que se hace necesario el uso de **tablas**. En este caso, además, existe una tabla para cada valor de probabilidad (pues las filas y las columnas se emplean para los grados de libertad).





$\alpha = 0{,}05$

$m \setminus n$	1	2	3	4	5	6	7	8	9	10	11	12	15
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	242,98	243,90	245,95
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40	19,41	19,43
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,70
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,86
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,62
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,94
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,51
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,22
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,01
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,85

Las tablas sólo existen para valores de α pequeños, para α grandes, usamos:

$$\alpha = P\{F_{n,m} > F_{n,m,\alpha}\} = P\left\{\frac{1}{F_{n,m}} < \frac{1}{F_{n,m,\alpha}}\right\}$$

$$= P\left\{F_{m,n} < \frac{1}{F_{n,m,\alpha}}\right\} = 1 - P\left\{F_{m,n} > \frac{1}{F_{n,m,\alpha}}\right\}.$$

$$P\{F_{6,3} > 8.94\} = 0.05 \quad \rightarrow \quad P\left\{F_{3,6} > \frac{1}{8.94}\right\} = 0.95$$

EJERCICIO 3

El gasto mensual en alimentación y los ingresos totales de una familia (en cientos de euros) tienen distribución χ^2 con medias 5 y 12, respectivamente. ¿Qué proporción de las familias destina más del 13 % de sus ingresos en alimentación?

Una vez estudiadas las distribuciones más importantes asociadas con la distribución normal, vamos a estudiar la **distribución en el muestreo** de algunos estadísditos cuando la distribución poblacional es normal.

Teorema de Fisher

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma)$.

- I $\bar{X} \sim N(\mu, \sigma/\sqrt{n}) \Leftrightarrow \frac{\bar{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1)$
- II $\frac{nS^2}{\sigma^2} = \frac{(n-1)S_c^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$
- III $\bar{X}$ y S^2 son v.a. **independientes**.

Nótese que I y II especifican la distribución de la media y la varianza muestral, lo que nos será de utilidad para obtener **intervalos de confianza** en el siguiente tema.

EJERCICIO 4

Consideremos $X_1, \dots, X_{20}$ m.a.s. de $X \sim N(\mu, \sqrt{3})$. Calcula $P\left\{S_c^2 > 1'5\right\}$.

EJERCICIO 5

Consideremos $X_1, \ldots, X_{100}$ m.a.s. de $X \sim N(2'5, 6)$. Calcula $P\ \{\ 1'5 < \overline{X} < 3'5\ ,\ 30 < S^2 < 44\ \}$.

Distribución de la media muestral con σ desconocida

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma)$, con σ **desconocido**. Entonces,

$$T = \frac{\bar{X} - \mu}{S}\sqrt{n-1} = \frac{\bar{X} - \mu}{S_c}\sqrt{n} \sim t_{n-1}$$

Diferencia de medias en dos poblaciones independientes normales

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu_X, \sigma_X)$ y $Y_1, \ldots, Y_m$ una m.a.s. de $Y \sim N(\mu_Y, \sigma_Y)$, **independientes** entre sí. Entonces,

- si σ_X y σ_Y son **conocidas**,

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}} \sim N(0,1)$$

- si σ_X y σ_Y son **desconocidas**, pero iguales, $\sigma_X = \sigma_Y = \sigma$,

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{nS_X^2 + mS_Y^2}{n+m-2}}\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

Ejercicio 6

Demuestra los puntos anteriores.

COCIENTE DE VARIANZAS EN DOS POBLACIONES INDEPENDIENTES NORMALES

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu_X, \sigma_X)$ y $Y_1, \ldots, Y_m$ una m.a.s. de $Y \sim N(\mu_Y, \sigma_Y)$, **independientes** entre sí, con μ_X y μ_Y desconocidos. Entonces,

$$\frac{S^2_{c,X}/\sigma^2_X}{S^2_{c,Y}/\sigma^2_Y} \sim F_{n-1,m-1}$$

EJERCICIO 7

Consideremos dos m.a.s. de tamaños $n = 5$ y $m = 4$ de dos poblaciones normales independientes, con $\sigma_X = \sigma_Y$. Calcula:

a) $P\left\{\frac{S^2_{c,X}}{S^2_{c,Y}} > 6'25\right\}$ b) $P\left\{\frac{S^2_{c,X}}{S^2_{c,Y}} < 0'192\right\}$.

DIFERENCIA DE MEDIAS EN UNA POBLACIÓN BIDIMENSIONAL NORMAL

Sea $(X_1, Y_1), \ldots, (X_n, Y_n)$ una m.a.s. de $(X, Y) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, con

$$\boldsymbol{\mu} = (\mu_X, \mu_Y) \qquad \text{y} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}$$

Entonces,

- si $\boldsymbol{\Sigma}$ es **conocida**,
$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}}\sqrt{n} \sim N(0,1)$$
- si algún elemento de $\boldsymbol{\Sigma}$ es **desconocido**, definimos $Z = X - Y$, con lo que $(Z_1, \cdots, Z_n) = (X_1 - Y_1, \cdots, X_n - Y_n)$ es una m.a.s. de $Z \sim N(\mu_Z, \sigma_Z) \equiv N(\mu_X - \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}})$ y, por tanto,
$$\frac{\bar{Z} - \mu_Z}{S_Z}\sqrt{n-1} = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i - (\bar{X} - \bar{Y}))^2}}\sqrt{n-1} \sim t_{n-1}$$

Ejercicio 8

Sean $(X_1, Y_1), \ldots, (X_6, Y_6)$ m.a.s. de $X \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, con

$$\boldsymbol{\mu} = (7, 8) \quad \text{y} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 4 & -2 \\ -2 & 3 \end{bmatrix}.$$

Calcula: $P\left\{\overline{X} - \overline{Y} > 0'41\right\}$.