IN3401 - Estadística para la Economía y la Gestión

Semestre Primavera 2019

Cátedra -- Martes y Jueves 8.30-10.00 – Sala DII 11

Auxiliares – Martes 16.15-17.45 – Sala B204 (en conjunto con la otra sección)

Tarea 1 – Martes 10 de Septiembre [13.33%]

Control 1 – Jueves 3 de Octubre en horario de clases [30%]

Tarea 2 – Jueves 24 de Octubre [13.33%]

Tarea 3 – Jueves 14 de Noviembre [13.33%]

Examen – Martes 16 de Noviembre [30%]

Fundamentos de Matemática Estadística

Raimundo Undurraga

Ingenieria Industrial, U. de Chile

IN3401-Estadística para la Economía y la Gestión

Contents

Outline

Ejemplo de Motivación

Suponga que una aerolinea tiene que decidir el numero de reservas que debe aceptar para un total de 100 asiento disponibles en el vuelo. Si hay menos de 100 reservas lo conveniente es aceptarlas todas. Pero que pasa si se realizan mas de 100 reservas? Una solución segura es aceptar como máximo 100 reservas. Sin embargo, dado que hay algunas personas que reservan y luego se arrepienten, bajo esta política existe a chance de que el vuelo finalmente no logre ir completo con sus 100 asientos, lo cual es una perdida económica para la empresa. Una estrategia alternativa consiste en permitir que se reserven mas de 100 asientos y esperar que algunas personas se arrepientan de sus reservas. De esa manera el numero final de pasajeros será mas cercano a 100. Esta política implica el riesgo de que la aerolinea tenga que compensar economicamente a aquellos pasajeros que habiendo realizados sus reservas tengan que ser bajados del vuelo por que este está sobrevendido.

Pregunta: Cómo decidir el número óptimo de reservas que la aerolinea debiese poner a disposición? Con información sobre los costos y la frecuencia promedio de reservas que son finalmente utilizadas por los clientes, podemos usar probabilidad básica para obtener una solución.

Variable Aleatoria (v.a.) I

- una v.a. es una variable que toma valores numéricos y tiene un resultado que es determinado por un proceso aleatorio de generación de datos (random DGP).
- La asignación es uno a uno, es decir, cada resultado corresponde a un único valor, y entonces dos resultados no pueden correspinder al mismo valor.
    - Experimento: una moneda es lanzada tres veces. El set de posibles resultados que pueden ser derivados del random DGP está dado por el conjunto $\Omega = \{hhh, hht, htt, hth, ttt, tth, thh, tht\}$. Ahora definamos una v.a. como una función que cuenta el numero de caras, i.e., X puede tomar un valor dentro del conjunto $\{0, 1, 2, 3\}$.
    - Diremos que X es una v.a. porque antes de que los datos sean observados (i.e., antes de que la moneda sea lanzada), es incierto el valor que puede tomar X . De hecho no tenemos como saber si obtendremos 0, 1, 2, o 3 caras.

Variable Aleatoria (v.a.) II

        - Otro ejemplo: el numero de personas que reservaron un ticket para el vuelo y que van a aparecer en el aeropuerto el día del vuelo también es una v.a. pues antes de que llegue ese día no sabemos cuántas personas van a aparecer en el aeropuerto. Hay incertidumbre al respecto.
    - Las probabilidades se asocian a un posible resultado para cuantificar la incertidumbre asociada a ese resultado. En efecto, la probabilidad de que X tome un valor particular x se denota $P(X = x)$.
- Conceptos claves: el experimento o random DGP; Ω, el conjunto de posibles resultados; $X = f(\Omega)$, la v.a.; x, el resultado aleatorio; y $P(X = x)$, la probabilidad de que la v.a. tome un determinado valor.

Variable Aleatoria Discreta I

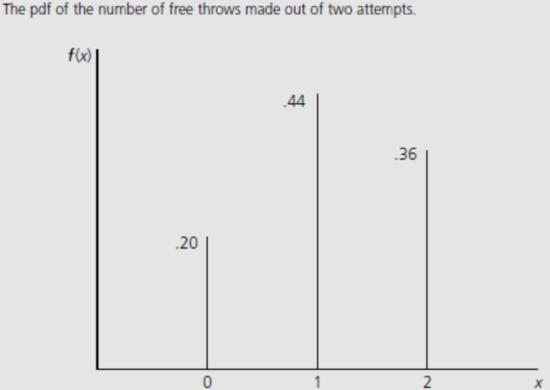
- v.a. discreta: X toma valores contables tal que la probabilidad de que X tome un valor x es siempre diferente de cero.
  - Específicamente, si x puede tomar k valores posibles $\{x_1, x_2, ..., x_k\}$, entonces definimos la **distribución de probabilidad** de X como la probabilidad de que la v.a. X tome un valor x, tal que $f(x_j) = P(X = x_j) = p_j$ para $j = 1, ..., k$. Nótese que: (i) $p_j \epsilon [0, 1]$; y (ii) $\sum p_j = 1$. Además, $f(x) = 0$ para cualquier x distinto de x_j para algún j.
- Ej1: supongamos que X es el numero de aciertos realizados por un basquebolista para un total de dos intentos, tal que $X = \{0, 1, 2\}$. Asumamos que $f(0) = 0.2$; $f(1) = 0.44$; $f(2) = 0.36$. Cuál es la probabilidad de que el jugador meta tenga al menos un acierto? R: $P(X \geq 1) = P(X = 1) + P(X = 2) = 0.8$.
- Ej2: Bernoulli. Experimento: tirar una moneda al aire; $\Omega = \{C, S\}$; $X = 1$ si es cara C, $X = 0$ si es S tal que X toma valores en el conjunto $\{0, 1\}$; con$P(X = 1) = \theta$ y $P(X = 0) = 1 - \theta$.

Variable Aleatoria Discreta II

- Ej3: De nuevo la aerolínea: para un cliente que hace una reserva y es seleccionado aleatoriamente del total de reservistas, defina una v.a. Bernoulli como $X = 1$ si la persona que reservó el asiento aparece el día que sale el vuelo, y $X = 0$ si no, con $\theta = 0.25$ y $1 - \theta = 0.75$.

Función de Distribución de Probabilidad (pdf)

Figure: Función de Distribución de Probabilidad de una v.a. discreta.

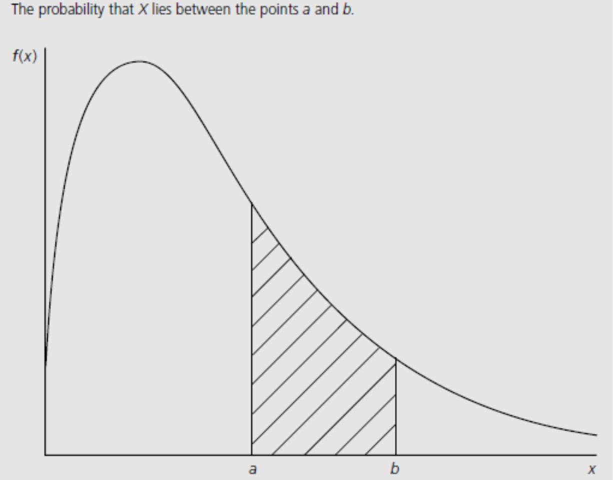


Variable Aleatoria Continua

- v.a. continua: aqui X toma valores reales con probabilidad cero. La v.a. continua puede tomar tantos valores que no podemos contarlos y por tanto matchear a enteros positivos. En efecto, decimos que X toma cada valor con probabilidad cero. (ej, ingreso, tiempo).
- Por lo tanto, para el caso continuo, sólo podemos asignar probabilidades para intervalos de valores en el rango de x. De esa manera, la distribución de probabilidad de una v.a. continua se denomina **función de densidad de probabilidad (pdf)** de X, tal que la pdf es la integral entre dos valores. $P(a \leq x \leq b) = \int_a^b f(x) \cdot dx$, i.e., el area bajo $f(x)$ en el rango de valores a a b.
- Notación: f_x es la pdf de X, f_y es la pdf de Y, etc.
- Nótese que:
  - $\int_{-\infty}^{+\infty} f(x) \cdot dx = 1$;
  - $f(x) = 0$ para valores fuera del rango de x;
  - $P(a \leq x \leq b) = P(a \leq x < b) = P(a < x \leq b) = P(a < x < b)$, ya que la probabilidad asociada con cualquier valor específico siempre será cero.

Función de Densidad de Probabilidad (pdf)

Figure: Función de Densidad de Probabilidad para una v.a. continua.

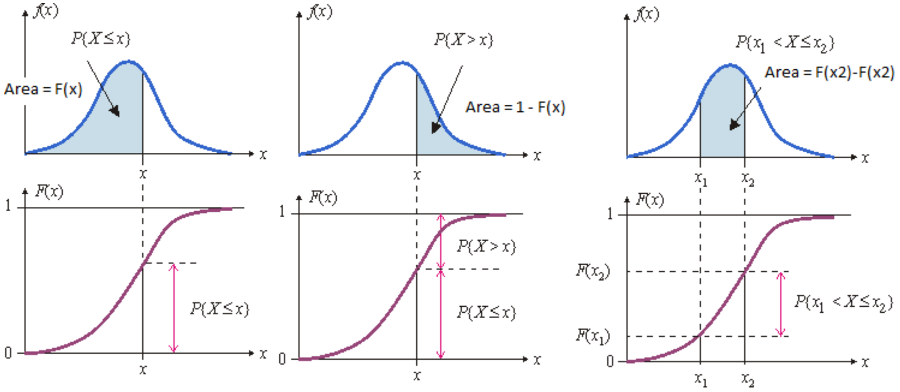


Función de Distribución Acumulada (cdf)

- Para cualquier v.a. X, la probabilidad de que X tome un valor menor o igual a x se denota $F(x)$
- Para una v.a. discreta: $F(x) = \sum_{X \leq x} f(X) = P(X \leq x)$
- Para una v.a. continua: $F(x) = \int_{-\infty}^{x} f(x) \cdot dx$
  - En efecto, $f(x) = \frac{dF(x)}{dx} = lim_{\triangle x \to 0} \frac{(x < X < x + \triangle x)}{\triangle x}$
  - Cuando computamos probabilidades para v.a. continuas es más fácil trabajar con la cdf.
    $P(a \leq x \leq b) = F(b) - F(a)$
    $P(X > x) = 1 - F(x)$
- $F(x)$ satisface las siguientes propiedades:
  1. $0 \leq F(x) \leq 1$
  2. Si $x > y$, entonces $F(x) \geq F(y)$
  3. $F(+\infty) = 1$
  4. $F(-\infty) = 0$

Función de Distribución Acumulada (cdf)

Figure: Relación entre pdf y cdf.



Esperanza de una v.a.

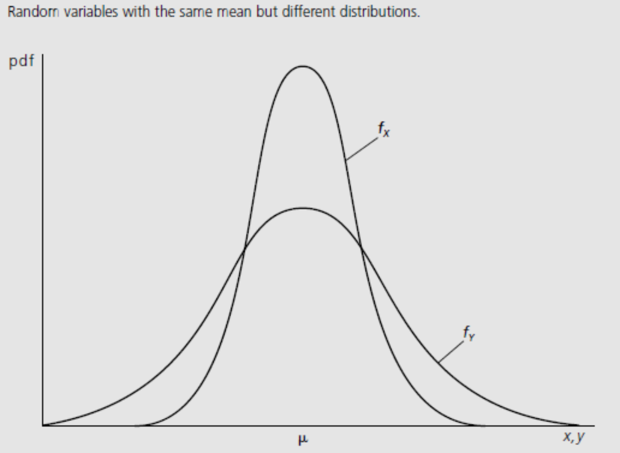
- Para cualquier v.a. X, el valor esperado de X, que se denota como $E[X]$ or μ_x (el promedio poblacional), es un promedio ponderado de los valores que puede tomar x, y en que los ponderadores corresponden a las respectivas probabilidades asociadas a cada valor (i.e., están determinadas por la pdf).
  - Para una v.a. discreta: $\mu = E[X] = \sum_x x \cdot f(x)$
  - Para una v.a. continua: $\mu = E[X] = \int_x x \cdot f(x) \cdot dx$
- Propiedades de $E[X]$
  1. Para cualquier constante c, $E[c] = c$
  2. Sea $g(X)$ una función de X, entonces,
     - ★ para una v.a. discreta, $E[g(X)] = \sum_x g(x) \cdot f(x)$
     - ★ para una v.a. continua, $E[g(X)] = \int_x g(x) \cdot f(x) \cdot dx$
     - ★ Si $g(X) = a + bX$, entonces $E[g(X)] = E[a] + b \cdot E[X]$
  3. Si $\{a_{1,\ldots\ldots} a_n\}$ son constantes y $\{X_{1,\ldots\ldots\ldots} X_n\}$ son v.a., entonces $E[\sum_{i=1}^{n} a_i \cdot X_i] = \sum_{i=1}^{n} a_i \cdot E[X_i]$
     - ★ Si $a = 1$, entonces el valor esperado de la suma es la suma de los valores esperados: $E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$
- Otras medidas de tendencia central: (i) mediana: valor m tal que $P(X \leq m) \geq \frac{1}{2}$ y $P(X \geq m) = \frac{1}{2}$ (no está afectada por valores extremos); (ii) moda: valor de x tal que $\frac{df(x)}{x} = 0 \Rightarrow x = max\{f(x)\}$

Varianza de una v.a.

- No sólo estamos interesados en caracterizar medidas de tendencia central de la distribución, sino también en la forma de la distribución.
- Para cualquier v.a., $Var[X] = \sigma^2 = E[(x-\mu)^2]$, i.e., la distancia esperada (promedio) entre x y μ.
  - Para una v.a. discreta: $\sigma^2 = Var[X] = \sum_x (x-\mu)^2 \cdot f(x)$, en que $\mu = E[X] = \sum_x x \cdot f(x)$
  - Para una v.a. continua: $\sigma^2 = Var[X] = \int_x (x-\mu)^2 \cdot f(x) \cdot dx$, en que $\mu = E[X] = \int_x x \cdot f(x) \cdot dx$
- Propiedades de $Var[X]$
  1. $Var[X]$ es siempre no-negativa:
     $E[(x-\mu)^2] = E[x^2 - 2 \cdot x \cdot \mu + \mu^2] = E[x^2] - \mu^2$
  2. Para cualquier constante c, $Var[c] = 0$
  3. Para cualquier constante a y b, $Var[aX + b] = a^2 Var[X] + b$, i.e., sumar una constante a la v.a. no cambia su varianza, pero multiplicar una v.a. por una constante aumenta su varianza en esa constante al cuadrado.

Varianza de una v.a.

Figure: Variables aleatorias con misma media pero distinta varianza.



Desviación Estándar de una v.a.

- Queremos tener una medida de la forma de la distribución que sea común a X y μ, i.e., que cuando μ aumente en un factor a, la dispersión de la distribución de X aumente en el mismo factor.
- Para cualquier v.a., $SD[X] = \sigma = \sqrt{Var[X]}$.
- Propiedades de $SD[X]$
  1. Para cualquier constante c, $SD[c] = 0$
  2. Para cualquier constante a y b, $SD[aX + b] = |a| \cdot SD[X]$, i.e., sumar una constante a la v.a. no cambia su desviación estándar, y multiplicar una v.a. por una constante aumenta su desviación estándar en el mismo factor.
     - ★ Ej: definamos $Y = 1000 \cdot X$. Supongamos que $E[X] = 20$ y $SD(X) = 6$. Entonces, $E[Y] = 1000 \cdot E[X] = 20000$ y $SD[Y] = 1000 \cdot SD[X] = 6000$. En efecto, tanto la media como la desviación estándar de Y aumentan en el mismo factor 1000. En cambio, $Var[Y] = Var[1000 \cdot X] = (1000)^2 \cdot Var[X]$. Es decir, la varianza de Y aumenta en un factor de un millón.

Estandarizando una v.a.

- Supongamos que dada una v.a. X, definimos una nueva v.a. $Z = \frac{X-\mu}{\sigma}$
- Entonces,
  - $E[Z] = \frac{1}{\sigma}E[X] - \frac{1}{\sigma}E[\mu] = \frac{1}{\sigma}\mu - \frac{1}{\sigma}\mu = 0$; y
  - $Var[Z] = Var[\frac{X}{\sigma}] - Var[\frac{\mu}{\sigma}] = \frac{1}{\sigma^2}\sigma^2 - 0 = 1$
- Luego, la v.a. Z tiene una media 0 y una varianza igual a 1, lo que llamamos una v.a. estandarizada
- Como veremos, esta transformación es frecuentemente utilizada en inferencia estadística.

Otras medidas de Dispersión

Momentos de mayor orden de una distribución:

- Oblicuidad (Skewness)$= E[(x-\mu)^3]$, es una medida del nivel de asimetría de la distribución; para distribuciones simétricas la oblicuidad es cero $\Rightarrow f(x-\mu) = f(x+\mu)$. Si la oblicuidad es positiva, entonces la cola larga de la distribución está en el lado derecho mientras que si esta es negativa entonces se encuentra en el lado izquierdo.
- Kurtosis $= E[(x-\mu)^4]$, es una medida del grosor de las colas de la distribución.

Outline

Ejemplos de Distribuciones de Probabilidad

- Ciertas situaciones experimentales pueden ser estudiadas con reconocidas funciones de probabilidad. En la majoría de los casos, sin embargo, estas funciones de distribución son simplemente modelos para estudiar un determinado fenómeno pero no necesariamente explican fehacientemente su comportamiento.
- Distribuciones que son típícamente utilizadas para modelar la distribución de una v.a. discreta:
  - *Ditribución Bernoulli* (1 intento): $P(X = 1) = \alpha$; $P(X = 0) = 1 - \alpha$, donde $1 \leq \alpha \leq 0$
  - *Distribución Binomial* (n intentos): $P(X = x) = \binom{n}{x}\alpha^x(1 - \alpha)^{n-x}$, donde $x = 0, 1, ..., n$. $E[X] = n\cdot \alpha$ y $Var[X] = n\cdot \alpha\cdot (1 - \alpha)$
    - ⋆ Ej: Una moneda es lanzada 10 veces. Cuál es la probabilidad de obtener 6 caras? Resuelva para $n = 10$, $x = 6$, y $\alpha = P(x = cara) = 0.5$
  - Si el número de intentos crece al mismo tiempo que la probabilidad de éxito decae tenemos que $E[X] = n\cdot \alpha$ es estable, y en efecto la distribución binomial se aproxima a una Distribución *Poisson*: $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$
  - Ej: número de veces que una persona es arrestada en un año; número de patentes suscritas por una firma en un año; etc.

Ejemplos de Distribuciones de Probabilidad

- Distribuciones que son típicamente utilizadas para modelar el comportamiento de v.a. copntinuas:
  - *La Distribución Normal*
  - La Distribución Normal Estándar
  - La Distribución Log-Normal
  - La Distribución Chi-cuadrado
  - La Distribución t
  - La Distribución F
  - ....hay muchas mas: Distribución Exponencial; Distribución Gamma; Distribución Beta; Distribución Logística Distribución Wishart; etc..

La Distribución Normal

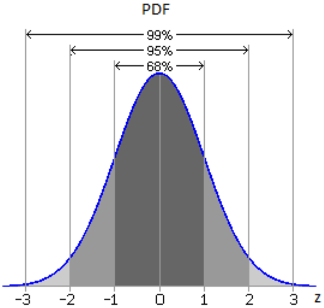
- Defina $x \sim N(\mu, \sigma^2)$ if $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, con $-\infty < x < \infty$
- Propiedades de la Distribución Normal:
  1. Si $x \sim N(\mu, \sigma^2)$, entonces $a + bx \sim N(a + b\cdot \mu, b^2\sigma^2)$⇒cualquier transformación lineal de una v.a. que se distribuye Normal también se distribuye Normal (ej: Distribución Normal Estándar).
     - ★ Ej: si $x \sim N(1, 3)$, entonces $Y = 3 + 2x \sim N(3 + 2E[x], 2^2\cdot 9) \Rightarrow Y = 3 + 2x \sim N(5, 36)$
  2. Es simétrica en torno a μ, es decir (i)$E[X] = median = mode$; (ii) Oblicuidad $= E[(X - \mu)^3] = 0$; y (iii) Kurtosis $= \frac{E[(X-\mu)^3]}{\sigma^3} = 3$
  3. Cualquier combinación lineal de v.a. que se distribuyen Normal tiene una distribución Normal
     - ★ Ej: sea X_1, X_2 y X_3 v.a. independientes y distribuidas $\sim N(\mu, \sigma^2)$. Defina $W = X_1 + 2X_2 - 3X_3$. Luego, $W \sim N(E[X_1] + 2E[X_2] - 3E[X_3], Var(X_1) + 4Var(X_2) + 9Var(X_3)) \Rightarrow$ $W \sim N(0, 14\sigma^2)$
     - ★ Algunas funciones de distribución derivadas de una Distribución Normal: la *Distribución Normal Estándar; Chi-cuadrado; Distribución t; Distribución F*.

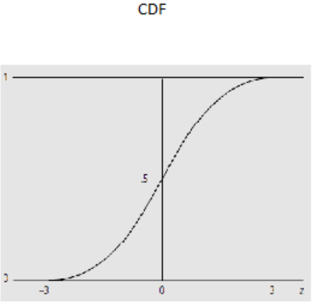
La Distribución Normal Estándar

- Si $x \sim N(\mu, \sigma^2)$, entonces $Z = \frac{x-\mu}{\sigma} \sim N(0,1)$
  - pdf de una Normal Estándar: $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
  - cdf de una Normal Estándar: $\Phi(z) = P(Z \leq z)$
  - Dado que es simétrica en torno a μ, entonces $P(Z < -z) = P(Z > z) \Rightarrow \Phi(-z) = 1 - \Phi(z)$, y entonces podemos usar la cdf de una Normal Estándar para computar la probabilidad de un evento en una v.a. distribuida Normal Estándar.
    - ⋆ Ej1: Computar $P(2 < X \leq 6)$ cuando $X \sim N(4,9)$. Primero estandarizamos X, tal que $Z = \frac{X-4}{3}$, y entonces el problema es encontrar $P(\frac{2-4}{3} < Z \leq \frac{6-4}{3}) = P(Z \leq \frac{2}{3}) - P(Z \leq \frac{-2}{3}) = \Phi(0.67) - \Phi(-0.67) = 0.749 - 0.251 = 0.498$
    - ⋆ Ej2: $P(|X| > 2) = P(X > 2) + P(X < -2) \Leftrightarrow P(\frac{X-4}{3} > \frac{2-4}{3}) + P(\frac{X-4}{3} < \frac{-2-4}{3}) = 1 - \Phi(\frac{-2}{3}) + \Phi(-2) = 1 - 0.251 + 0.023 = 0.772$

La Distribución Normal Estándar

Figure: Distribución Normal Estándar y la CDF de una Normal Estándar (+Tabla).





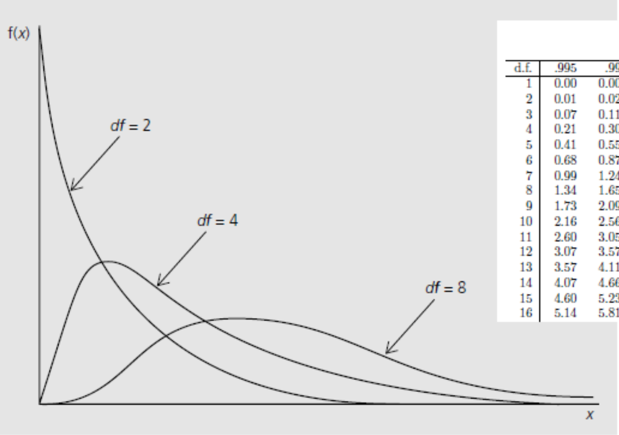
z	0.00	0.01	0.02	0.03
-3.4	0.0003	0.0003	0.0003	0.0003
-3.3	0.0005	0.0005	0.0005	0.0004
-3.2	0.0007	0.0007	0.0006	0.0006
-3.1	0.0010	0.0009	0.0009	0.0009
-3.0	0.0013	0.0013	0.0013	0.0012
-2.9	0.0019	0.0018	0.0018	0.0017
-2.8	0.0026	0.0025	0.0024	0.0023
-2.7	0.0035	0.0034	0.0033	0.0032
-2.6	0.0047	0.0045	0.0044	0.0043
-2.5	0.0062	0.0060	0.0059	0.0057
-2.4	0.0082	0.0080	0.0078	0.0075
-2.3	0.0107	0.0104	0.0102	0.0099
-2.2	0.0139	0.0136	0.0132	0.0129
-2.1	0.0179	0.0174	0.0170	0.0166
-2.0	0.0228	0.0222	0.0217	0.0212
-1.9	0.0287	0.0281	0.0274	0.0268
-1.8	0.0359	0.0351	0.0344	0.0336
-1.7	0.0446	0.0436	0.0427	0.0418
-1.6	0.0548	0.0537	0.0526	0.0516
-1.5	0.0668	0.0655	0.0643	0.0630
-1.4	0.0808	0.0793	0.0778	0.0764
-1.3	0.0968	0.0951	0.0934	0.0918
-1.2	0.1151	0.1131	0.1112	0.1093
-1.1	0.1357	0.1335	0.1314	0.1292
-1.0	0.1587	0.1562	0.1539	0.1515
-0.9	0.1841	0.1814	0.1788	0.1762
-0.8	0.2119	0.2090	0.2061	0.2033
-0.7	0.2420	0.2389	0.2358	0.2327
-0.6	0.2743	0.2709	0.2676	0.2643
-0.5	0.3085	0.3050	0.3015	0.2981
-0.4	0.3446	0.3409	0.3372	0.3336
-0.3	0.3821	0.3783	0.3745	0.3707
-0.2	0.4207	0.4168	0.4129	0.4090
-0.1	0.4602	0.4562	0.4522	0.4483
-0.0	0.5000	0.4960	0.4920	0.4880
0.0	0.5000	0.5040	0.5080	0.5120
0.1	0.5398	0.5438	0.5478	0.5517

La Distribución Chi-cuadrado

- Si $z \sim N(0,1)$, entonces $x = z^2 \sim \chi^2_{[1]}$, i.e., es chi-cuadrado con 1 grado de libertad. En este caso $E[z^2] = 1$ y $Var[z^2] = 2$
- Si $X_1, ..., X_n$ son n v.a. independientes y distribuidas chi-cuadrado [1] , entonces $W = \sum_{i=1}^{n} z_i^2 \sim \chi^2_{[n]}$, con $E[W] = n$ y $Var[W] = 2n$
  - Los grados de libertad corresponden al numero de v.a. normal estándar en la suma
  - Siempre es no-negativa
  - No es simétrica
- Si X_1 y X_2 son v.a. independientes y distribuidas chi-cuadrado con n_1 y n_2 los grados de libertad respectivos, entonces $X_1 + X_2 \sim \chi^2_{[n_1+n_2]}$. Este resultado puede ser generalizado para la suma de un numero arbitrario de v.a. distribuidas chi-cuadrado.
- Para computar los valores críticos de probabilidad de una distribución chi-cuadrado se usa una tabla $[n; prob]$
- La oblicuidad de una chi-cuadrado decrece conforme aumentan los grados de libertad n , y a su vez tiende a una distribución normal.

La Distribución Chi-cuadrado

Figure: Distribución Chi-cuadrado.



Chi-square Distribution Table

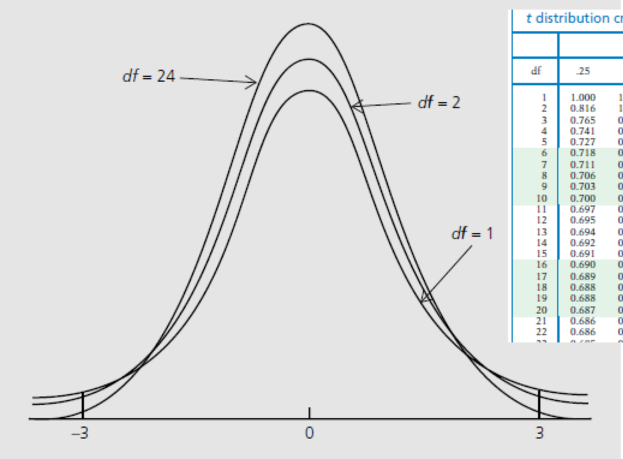
d.f.	.995	.99	.975	.95	.9	.1	.05	.025
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83
6	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85

La Distribución t

- Es una mezcla entre una Normal Estándar y una Chi-cuadrado.
- Si $z \sim N(0,1)$ y $x \sim \chi^2_{[n]}$, entonces $t = \frac{z}{\sqrt{\frac{x}{n}}} \sim t_{[n]}$, con $E[t] = 0$ y $Var[t] = \frac{n}{n-2}$
- La pdf de una distribución t tiene una forma similar a la de una Normal Estándar, pero muestra mayor dispersión en las colas y por tanto tiene colas un poco mas largas. Mientras mayor son los grados de libertad, mas pequeñas serán las colas y por tanto mas parecida será a una Normal Estándar.
- Para computar los valores críticos de probabilidad de una distribución t se usa una tabla [df,prob]
- Esta distribución es típicamente utilizada para hacer inferencia estadística en los modelos de regresión.

La Distribución t

Figure: Distribución t.



t distribution critical values

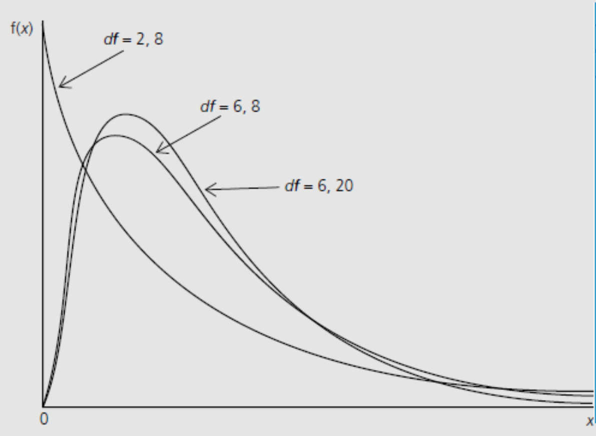
	Upper-tail probability p							
df	.25	.20	.15	.10	.05	.025	.02	.01
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508

La Distribución F

- Combina dos chi-cuadrado independientes
- Sea $X_1 \sim \chi^2_{[n_1]}$ y $X_2 \sim \chi^2_{[n_2]}$, entonces $F = \frac{(X_1/n_1)}{(X_2/n_2)} \sim F(n_1, n_2)$
  - Nótese que si $t \sim t_{[n]}$, entonces $t^2 \sim F(1, n)$. Paso clave: $t = \frac{z}{\sqrt{\frac{x}{n}}}$, luego $t^2 = \frac{z^2}{\frac{x}{n}} = \frac{(z^2/1)}{(x/n)} \sim F(1, n)$
- Para computar los valores críticos de probabilidad de una distribución F se usan los grados de libertad (n_1, n_2) de cada una de las dos chi-cuadrado. De esa manera, la tabla suele presentar valores de probabilidad para valores críticos específicos como el 95% o el 99% de las colas superiores de la distribución.

La Distribución F

Figure: Distribución F.



F critical values

Degrees of freedom in the denominator	p	Degrees of free 1	2	3	4
1	.100	39.86	49.50	53.59	55.83
	.050	161.45	199.50	215.71	224.58
	.025	647.79	799.50	864.16	899.58
	.010	4052.2	4999.5	5403.4	5624.6
	.001	405284	500000	540379	562500
2	.100	8.53	9.00	9.16	9.24
	.050	18.51	19.00	19.16	19.25
	.025	38.51	39.00	39.17	39.25
	.010	98.50	99.00	99.17	99.25
	.001	998.50	999.00	999.17	999.25
3	.100	5.54	5.46	5.39	5.34
	.050	10.13	9.55	9.28	9.12
	.025	17.44	16.04	15.44	15.10
	.010	34.12	30.82	29.46	28.71
	.001	167.03	148.50	141.11	137.10
4	.100	4.54	4.32	4.19	4.11
	.050	7.71	6.94	6.59	6.39
	.025	12.22	10.65	9.98	9.60
	.010	21.20	18.00	16.69	15.98
	.001	74.14	61.25	56.18	53.44
5	.100	4.06	3.78	3.62	3.52
	.050	6.61	5.79	5.41	5.19
	.025	10.01	8.43	7.76	7.39
	.010	16.26	13.27	12.06	11.39
	.001	47.18	37.12	33.20	31.09

Outline

Distribución Conjunta

- En economía, típicamente estamos interesados en estudiar fenómenos que involucran el comportamiento de mas de una v.a.
  - Ej: una aerolinea podría estar interesada en la probabilidad de que una persona que hizo una reserva aparezca en el aeropuerto el día que sale el vuelo y que además sea un pasajero que viaja en business (*probabilidad conjunta*)
- Para dos v.a. X e Y, la pdf conjunta $f(x,y) = P(X = x, Y = y)$
- En particular:
  - Caso discreto: $Prob(a \leq x \leq b,\ c \leq y \leq d) = \sum_{a \leq x \leq b} \sum_{c \leq y \leq d} f(x,y)$
  - Caso continuo: $Prob(a \leq x \leq b,\ c \leq y \leq d) = \int_a^b \int_c^d f(x,y) \cdot dy \cdot dx$
- Ademas,
  - $f(x,y) \geq 0$
  - $\sum_x \sum_y f(x,y) = 1$ y $\int_x \int_y f(x,y) \cdot dy \cdot dx = 1$
- Finalmente, la distribución acumulada de una probabilidad conjunta es:
  - Caso discreto: $F(x,y) = Prob(X \leq x,\ Y \leq y) = \sum_{X \leq x} \sum_{Y \leq y} f(x,y)$
  - Caso cont.: $F(x,y) = Prob(X \leq x,\ Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y) \cdot dy \cdot dx$

Distribución Marginal e Independencia

- Si para dos v.a. X e Y, pdf conjunta es $f(x,y) = f_X(x) \cdot f_Y(y)$ entonces X e Y son *estadísticamente independientes*, donde $f_X(x)$ es la pdf de X y $f_Y(y)$ es la pdf de$Y$.
  - Ej: si la probabilidad de que una persona que tiene una reserva aparezca en el aeropuerto el día que sale el avión es *independiente* de si es un pasajero bussiness o no, entonces $P(aparezca, business) = P(aparezca) \cdot P(business)$.
  - $f_X(x)$ y $f_Y(y)$ son las *pdfs* marginales de $f(x,y)$
  - De igual manera, para el caso de las distribuciones de probabilidad acumuladas, $F(x,y) = F_X(x) \cdot F_Y(y)$
  - Más aún, si $X_1, ...X_n$ son v.a. independientes entonces su distribución de probabilidad conjunta está dada por el producto de las distribuciones de probabilidad marginal para cada $(x_1, ...x_n)$
- Una pdf marginal es definida con respecto a una variable específica y puede ser derivada de la pdf conjunta ya sea sumando o integrando hacia afuera la otra variable.
  - Discreta: $f_X(x) = \sum_y f(x,y)$, i.e., $f_X(x)$ es la suma de las probabilidades de observar x para cada valor y; asi mismo, $f_Y(y) = \sum_x f(x,y)$;
  - Continua: $f_X(x) = \int_y f(x,y) \cdot dy$; $f_Y(y) = \int_x f(x,y) \cdot dx$

Distribución Condicional

- En econometría, típicamente estamos interesados en estudiar como una v.a. Y cambia conforme cambian una o mas v.a. (e.g., X). Esto se puede modelar a partir de una *distribución condicional* de Y dado X, tal que $f(y|x) = \frac{f(x,y)}{f_X(x)}$. De igual manera, $f(x|y) = \frac{f(x,y)}{f_Y(y)}$
- Si X e Y son v.a. independientes, entonces $f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{f_X(x) \cdot f_Y(y)}{f_X(x)} = f_Y(y)$, i.e., nuestro conocimiento sobre los valores que toma X no nos dicen nada con respecto a los valores que puede tomar Y (y viceversa). Esto es, si $f(y|x) = f_Y(y)$, entonces $f(x|y) = f_X(x)$
  - Intuición: Si la probabilidad de observar Y cambia conforme cambian los valores de X, entonces la distribución de probabilidad incondicional de Y (i.e., $f_Y(y)$) cambia cuando es condicionada en X. Por lo tanto, $f(y|x) = \frac{f(x,y)}{f_X(x)} \neq f_Y(y)$. En cambio, si la probabilidad de observar Y no cambia conforme cambian los valores de X, entonces la distribución condicional en X, es igual a su distribución incondicional, esto es, $f_Y(y|x) = f_Y(y)$.

Esperanza de una Distribución Conjunta

- Nótese que para el caso dicreto
  - $E[X] = \sum_x x \cdot f_X(x) = \sum_x x \cdot \sum_y f(x,y) = \sum_x \sum_y x \cdot f(x,y)$
  - $Var[X] = \sum_x (x - E[x])^2 \cdot f_X(x) = \sum_x (x - E[x])^2 \cdot \sum_y f(x,y) = \sum_x \sum_y (x - E[x])^2 f(x,y)$
- Asi mismo, para el caso continuo
  - $E[X] = \int_x x \cdot f_X(x) \cdot dx = \int_x x \cdot \int_y f(x,y) \cdot dy \cdot dx = \int_x \int_y x \cdot f(x,y) \cdot dy \cdot dx$
  - $Var[X] = \int_x (x - E[x])^2 \cdot f_X(x) \cdot dx = \int_x (x - E[x])^2 \cdot \int_y f(x,y) \cdot dy \cdot dx = \int_x \int_y (x - E[x])^2 \cdot f(x,y) \cdot dy \cdot dx$
- Finalmente, para cualquier función $g(x,y) \neq f(x,y)$, tenemos que:
  - Para el caso discreto: $E[g(x,y)] = \sum_x \sum_y g(x,y) \cdot f(x,y)$
  - Para el caso continuo: $E[g(x,y)] = \int_x \int_y g(x,y) \cdot f(x,y) \cdot dy \cdot dx$

Covarianza

- Es útil tener medidas de como, en promedio, dos v.a. varían una con la otra en una distribución conjunta.
  - Tal como es el caso de $E[X]$ y$Var[X]$ para distribuciones individuales, queremos un indicador que resuma la variación de una distribución conjunta.
- Dado dos v.a. X e Y, definamos $Cov(X, Y) = \sigma_{xy} = E[(x - \mu_x)(y - \mu_y)]$ con μ_x y μ_y los promedios poblacionales de X e Y, respectivamente.
  - Si $\sigma_{xy} > 0$ entonces, en promedio, cuando x está sobre su media, y también está sobre su media
  - Si $\sigma_{xy} < 0$ entonces, en promedio, cuando x está sobre su media, y está bajo su media
- Propiedades (demuestre las que mas pueda):
  1. $\sigma_{xy} = E[(x - \mu_x) \cdot (y - \mu_y)] = E[(x - \mu_x) \cdot (y - \mu_y)] = E[(x - \mu_x) \cdot y] = E[x \cdot y] - \mu_x \cdot \mu_y$. Además, $\sigma_{xx} = E[(x - \mu_x)^2] = Var[x] = \sigma_x^2$
  2. Si X e Y son v.a. independientes, entonces $\sigma_{xy} = 0$. Ojo: $\sigma_{xy} = 0$ no implica que X e Y sean independientes
  3. $Cov(a_1 x + b_1,\ a_2 y + b_2) = a_1 a_2 Cov(x, y)$.
  4. Desigualdad de Cauchy-Schzwartz: $|Cov(x, y)| \leq \sigma_x \cdot \sigma_y$

Correlación

- $Cov(a_1 \cdot x + b_1,\ a_2 \cdot y + b_2) = a_1 . a_2 \cdot Cov(x, y) \Rightarrow$la covarianza entre dos v.a. puede ser alterada simplemente por la multiplicación de una o ambas v.a. por una constante, i.e., la covarianza depende de la unidad de medida de X e Y, y eso incomoda el análisis.
  - No queremos que la covarianza entre educación (medida en meses o años) e ingresos (medida en pesos) dependa de como se mide cada una de ellas!
- Solución: Coeficiente de Correlación: $Corr(X, Y) = \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$.

  Algunas propiedades de ρ_{xy}:
  1. Dado que $\sigma_x > 0$ y $\sigma_y > 0$, entonces ρ_{xy} siempre tiene el mismo signo que σ_{xy}.
  2. $-1 \leq \rho_{xy} \leq 1$. Es el indicador de cuan fuerte es la relación entre X e Y.
     - ★ Si $\rho_{xy} = 1$ existe una relación lineal positiva perfecta entre X e Y.
     - ★ Si $\rho_{xy} = -1$ existe una relación lineal negativa perfecta entre X e Y.
     - ★ Si $\rho_{xy} = 0$, entonces $\sigma_{xy} = 0$ y decimos que X e Y *no están correlacionadas*
  3. ρ_{xy} no varía con las unidades de medida de X e Y. Para cualquier valor de las constantes $a_1, a_2, b_1,$ y b_2
     - ★ Si $a_1 \cdot a_2 > 0$, entonces $Corr(a_1 x + b_1,\ a_2 y + b_2) = Corr(x, y)$
     - ★ Si $a_1 \cdot a_2 < 0$, entonces $Corr(a_1 x + b_1,\ a_2 y + b_2) = -Corr(x, y)$

Varianza y Covarianza

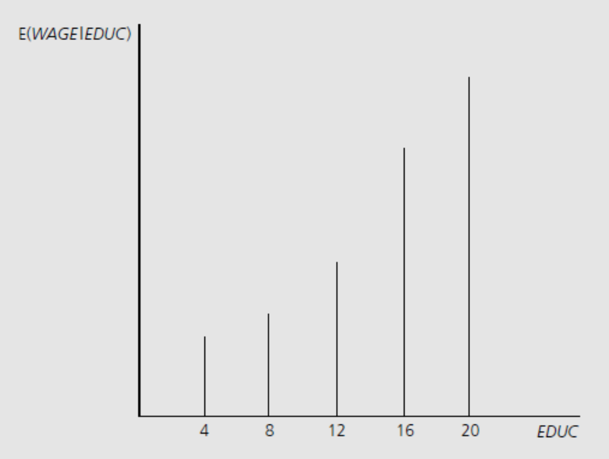
- Sabemos que $Var[ax+b] = a^2 Var[x]$. Cual es la varianza de $ax+by$? Depende de cuan correlacionadas estén X e Y
- $Var[ax+by] = a^2 \cdot Var[x] + b^2 \cdot Var[y] + 2 \cdot a \cdot b \cdot Cov(x,y)$
- En efecto, si X e Y son independientes, $Var[x+y] = Var[x-y] = Var[x] + Var[y]$. En otras palabras, cuando X e Y son independientes, entonces "la varianza de la suma es igual a la suma de las varianzas".
    - Este resultado se extiende a mas de dos v.a., tal que $\{X_1, ..., X_n\}$ son *variables aleatorias no correlacionadas* si cada v.a. en el conjunto no está correlacionada con ninguna de las variables aleatorias del conjunto, i.e., $Cov(x_{i,}x_j) = 0$, para todo $i \neq j$
    - Asi mismo, para constantes $\{a_1, ...a_n\}$, tenemos que $Var[a_1 \cdot x_1 + ... + a_n \cdot x_n] = \sum_{i=1}^n a_i^2 \cdot Var[x_i]$
- Si X e Y están correlacionadas, entonces $Var[a_1 \cdot x_1 + ... + a_n \cdot x_n] = \sum_{i=1}^n a_i^2 \cdot Var[x_i] + \sum_{i=1}^n 2 \cdot a_i \cdot a_j \cdot Cov(x_i, x_j)$
- $Cov(a \cdot x + b \cdot y,\ c \cdot x + d \cdot y)$? Resuelva!

Esperanza Condicional

- *Cov* y *Corr* miden la relación lineal entre dos v.a. y por tanto no sirven para caracterizar relaciones no lineales
- Además, *Cov* y *Corr* tratan a X e Y simétricamente, i.e., es un único indicador que asocia X e Y. Sin embargo, en realidad podríamos estar mas interesados en modelar Y en términos de X, y en efecto explorar como se comporta Y *condicional* en los valores que toma X.
- Esperanza Condicional (también llamada Media Condicional):
  - Caso discreto: $E[Y|X=x] = \sum_y y \cdot f(y|X=x)$ $\rightarrow$Suma de los valores de y_i multiplicado por la probabilidad de observar $Y = y_i$ dado $X = x_i$. Tal como en el caso de $E[Y]$, esto es simplemente un promedio ponderado de Y (donde los ponderadores corresponden a $P(Y = y_i)$), pero condicional en un valor fijo de x. (Caso continuo: $E[Y|X=x] = \int_y y \cdot f(y|X=x) \cdot dy$)
- Nota: estamos condicionando en un valor fijo de X, por lo tanto $E[Y|X=x]$ es un valor fijo. Sin embargo, si computamos $E[Y|X=x]$ para cada valor de X, entonces podemos construir la *función de media condicional* $E[Y|X]$ (la regresión de Y en X).
- Econometría: cómo aproximar $E[Y|X]$ usando datos reales.

Esperanza Condicional

Figure: E[Y=Wage|X=EDUC].



Esperanza Condicional

- Propiedades de la Función de Esperanza Condicional
  1. Para cualquier función $W(X)$, tenemos que $E[W(X)|X] = W(X)$.
     Intuición: si conocemos X, entonces conocemos $W(X)$
  2. Para funciones $W_1(X)$ y $W_2(X)$, tenemos que
     $E[W_1(X) \cdot Y + W_2(X)|X] = W_1(X) \cdot E[Y|X] + W_2(X)$
  3. Si X e Y son v.a. independientes, entonces $E[Y|X] = E[Y]$,i.e., si X e Y son independientes, entonces el valor esperado de Y dado X no depende de X , y entonces el valor esperado condicional de Y en X siempre es igual al valor esperado de Y (es incondicional en X).
  4. Si $E[Y|X] = E[Y]$, entonces $Cov(X, Y) = 0$ y $Corr(X, Y) = 0$, i.e., si el conocimiento de X no cambia el valor esperado de Y, entonces X e Y no están correlacionadas. Lo contrario no es cierto: si X e Y no están correlacionados, $E[Y|X]$ todavía podría depender de X.
     - Independencia implica cero correlación, pero cero correlación no implica independencia.

Ley de Esperanzas Iteradas (LIE)

- SiX es una v.a., entonces $E[Y|X]$ también es una v.a., y en efecto tiene una distribución de probabilidad asociada, y entonces también un valor esperado. La LIE consiste en que $E[E[Y|X]] = E[Y]$. Esto es, si primero obtenemos $E[Y|X]$ como una función de X y tomamos el valor esperado de esta (con respecto a la distribución de X), entonces lo que encontramos es el valor esperado de Y, es decir$E[Y]$.
  - Ej: Sea $Y = WAGE$ and $X = EDUC$. Supongamos que la función de media condicional $E[WAGE|\, EDUC] = 4 + 0.6\cdot EDUC$ y $E[EDUC] = 11.5$. Entonces, por LIE tenemos que $E[WAGE] = E[E[WAGE|\, EDUC]] = E[4 + 0.6\cdot EDUC] = 4 + 0.6\cdot E[EDUC] = 4 + 0.6\cdot 11.5 = 10.90$.
- En particular, $E[Y|X] = E[E[Y|X,Z]|X]$, esto es, podemos encontrar $E[Y|X]$ primero encontrando la función de media condicional de Y dado X y una tercera variable Z (recuerda que esta funcion de media condicional es una v.a.), y luego usarla para computar el promedio de los valores que la funcion de media condicional toma para cada valor de X.

Varianza Condicional

- Varianza Condicional: $Var[Y|X=x] = E[(y-E[Y|X=x])^2|X=x]$
  - Caso discreto: $Var[Y|X=x] = E[(y-E[Y|X=x])^2|X=x] = \sum_y (y-E[Y|X=x])^2 \cdot f(y|X=x)$
  - Caso continuo: $Var[Y|X=x] = E[(y-E[Y|X=x])^2|X=x] = \int_y (y-E[Y|X=x])^2 \cdot f(y|X=x) \cdot dy$
- En particular, $Var[Y|X=x] = E[Var[y^2|X=x]] + (E[Y|X=x])^2$
- La función de varianza condicional $Var[Y|X]$ típicamente se denomina "Función Scedastica", y, tal como es el el caso de la función de media condicional, es una v.a. Cuando la varianza condicional no varía con X decimos que existe *Homoscedasticidad* (misma varianza).
- Descomposición de la Varianza: en una distribución conjunta, $Var[Y] = Var[E[Y|X]] + E[Var[Y|X]]$ $\Rightarrow$la varianza de Y se puede descomponer en la función de media condicional mas el valor esperado de la varianza de la media condicional.

Resúmen Secciones 1 - 3

- Hemos cubierto los conceptos básicos de matemática estadística.
- Primero explicamos la idea de una v.a. y las propiedades de una diustribución de probabilidad y luego dimos algunos ejemplos de distribuciones de probabilidad (Normal, etc.), las cuales sirven para modelar un proceso de generación de datos (GDP) que produce los datos que observamos sobre la población.
- Luego introducimos la idea de distribuciones marginales, conjuntas, y condicionales, i.e., como modelar la relación entre dos o mas variables aleatorias.
- En la próxima parte del curso nos concentraremos en la meta de la inferencia estadística: usar los principios de la matemática estadística para combinar funciones de probabilidad conocidas y la data observada con el objetivo de modelar el comportamiento económico de individuos, firmas, instituciones, etc.