

Clase Auxiliar 11: Coeficiente y Razón de Correlación y Test de Anova

Profesora Nancy Lacourly
Auxiliares: Andrés Iturriaga J. - Héctor Olivero Q.

20 de junio de 2009

El coeficiente de Correlación Lineal

Clase Auxiliar
9

MA34B-01

Coficiente y
Razón de
Correlación

Test de Anova

Definición:

Sea $\{(x_i, y_i)\}_{i=1}^n$ una muestra aleatoria bivariada del par de variables (X, Y) . se define el coeficiente de correlación lineal como sigue:

$$r_{X,Y} = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y}_n)^2}} \quad (1)$$

Este coeficiente mide el grado de relación de tipo lineal que existe entre X e Y .

Interpretación del coeficiente de Correlación

Clase Auxiliar
9

MA34B-01

Coefficiente y
Razón de
Correlación

Test de Anova

En este punto hay que ser cuidadoso, ya que el coeficiente de correlación no entrega información suficiente, como se vé en los graficos que se encuentran en el archivo Excel correspondiente a esta clase. De todas formas se tiene lo siguiente.

$r_{X,Y} = -1$	Relación estrictamente lineal negativa.
$-1 < r_{X,Y} < 0$	Tendencia lineal con pendiente negativa.
$r_{X,Y} = 0$	Ausencia de tendencia lineal.
$0 < r_{X,Y} < 1$	Tendencia lineal con pendiente positiva.
$r_{X,Y} = 1$	Relación estrictamente lineal positiva.

Cuadro: Interpretación Coeficiente de Correlación

Razón de Correlación

Clase Auxiliar
9

MA34B-01

Coefficiente y
Razón de
Correlación

Test de Anova

Seguimos en el problema de detectar una relación funcional entre dos variables. Con el coeficiente de correlación encontramos una respuesta para el caso de dos variables cuantitativas. Abordaremos ahora el caso en que una de las variables es cualitativa.

Consideremos dos variables aleatorias X e Y donde la variable X es cualitativa y puede tomar q modalidades o categorías (e.g. Color de Pelo, Raza de un perro, etc.). Supongamos ahora que tenemos n_j observaciones de la variable Y en la categoría j (e.g. Tomo diez perros de la misma raza y mido su peso). Sean $\{y_{ij}\}_{i=1}^{n_j}$ los valores de estas observaciones.

Razón de Correlación

Clase Auxiliar
9

MA34B-01

Coefficiente y
Razón de
Correlación

Test de Anova

Seguimos en el problema de detectar una relación funcional entre dos variables. Con el coeficiente de correlación encontramos una respuesta para el caso de dos variables cuantitativas. Abordaremos ahora el caso en que una de las variables es cualitativa.

Consideremos dos variables aleatorias X e Y donde la variable X es cualitativa y puede tomar q modalidades o categorías (e.g. Color de Pelo, Raza de un perro, etc.). Supongamos ahora que tenemos n_j observaciones de la variable Y en la categoría j (e.g. Tomo diez perros de la misma raza y mido su peso). Sean $\{y_{ij}\}_{i=1}^{n_j}$ los valores de estas observaciones.

Razón de Correlación

Clase Auxiliar
9

MA34B-01

Coefficiente y
Razón de
Correlación

Test de Anova

La idea es separar la variabilidad total de la variable Y sobre la muestra en dos partes: una que represente la variabilidad entre categorías y otra que represente la variabilidad dentro de las categorías.

De esta manera si tuvieramos que la variabilidad entre categorías es nula, podríamos concluir que el conocimiento de la categoría para X en que se encuentra una observación de Y no nos sirve para determinar su valor. Es decir, concluiríamos que no hay indicios de una relación funcional entre las variables X e Y .

Si por el contrario encontramos que la variabilidad dentro de cada categoría es nula, podríamos concluir que saber en que categoría de la variable X esta la observación de la variable Y determina el valor de esta.

Razón de Correlación

Clase Auxiliar
9

MA34B-01

Coefficiente y
Razón de
Correlación

Test de Anova

La idea es separar la variabilidad total de la variable Y sobre la muestra en dos partes: una que represente la variabilidad entre categorías y otra que represente la variabilidad dentro de las categorías.

De esta manera si tuvieramos que la variabilidad entre categorías es nula, podríamos concluir que el conocimiento de la categoría para X en que se encuentra una observación de Y no nos sirve para determinar su valor. Es decir, concluiríamos que no hay indicios de una relación funcional entre las variables X e Y .

Si por el contrario encontramos que la variabilidad dentro de cada categoría es nula, podríamos concluir que saber en que categoría de la variable X esta la observación de la variable Y determina el valor de esta.

Razón de Correlación

Clase Auxiliar
9

MA34B-01

Coefficiente y
Razón de
Correlación

Test de Anova

La idea es separar la variabilidad total de la variable Y sobre la muestra en dos partes: una que represente la variabilidad entre categorías y otra que represente la variabilidad dentro de las categorías.

De esta manera si tuvieramos que la variabilidad entre categorías es nula, podríamos concluir que el conocimiento de la categoría para X en que se encuentra una observación de Y no nos sirve para determinar su valor. Es decir, concluiríamos que no hay indicios de una relación funcional entre las variables X e Y .

Si por el contrario encontramos que la variabilidad dentro de cada categoría es nula, podríamos concluir que saber en que categoría de la variable X esta la observación de la variable Y determina el valor de esta.

Razón de Correlación

Clase Auxiliar

9

MA34B-01

Coficiente y
Razón de
Correlación

Test de Anova

Notemos que si definimos:

$$\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij} \quad (2)$$

$$w_j^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2 \quad (3)$$

$$w^2 = \sum_{j=1}^q \frac{n_j}{n} w_j^2 \quad (4)$$

$$b^2 = \sum_{j=1}^q \frac{n_j}{n} (\bar{y}_j - \bar{y}_n)^2 \quad (5)$$

Razón de Correlación

Clase Auxiliar
9

MA34B-01

Coefficiente y
Razón de
Correlación

Test de Anova

Entonces:

$$S_n^2(Y) = \frac{1}{n} \sum_{j=1}^q \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_n)^2 \quad (6)$$

$$S_n^2(Y) = w^2 + b^2 \quad (7)$$

Donde b^2 captura la variabilidad entre los grupos y w^2 captura la variabilidad dentro de los grupos o categorías. Notemos que si $b^2 = 0$ entonces las medias de todas las categorías son iguales a la media de la muestra completa y por lo tanto no hay relación funcional entre X e Y . Si por el contrario $w^2 = 0$ tenemos que $w_j^2 = 0$ para todo $j = 1 \dots q$ y por lo tanto dentro de cada categoría todas las observaciones son iguales. De lo que concluimos que existe relación funcional entre X e Y .

Razón de Correlación

Clase Auxiliar
9

MA34B-01

Coefficiente y
Razón de
Correlación

Test de Anova

Entonces:

$$S_n^2(Y) = \frac{1}{n} \sum_{j=1}^q \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_n)^2 \quad (6)$$

$$S_n^2(Y) = w^2 + b^2 \quad (7)$$

Donde b^2 captura la variabilidad entre los grupos y w^2 captura la variabilidad dentro de los grupos o categorías. Notemos que si $b^2 = 0$ entonces las medias de todas las categorías son iguales a la media de la muestra completa y por lo tanto no hay relación funcional entre X e Y . Si por el contrario $w^2 = 0$ tenemos que $w_j^2 = 0$ para todo $j = 1 \dots q$ y por lo tanto dentro de cada categoría todas las observaciones son iguales. De lo que concluimos que existe relación funcional entre X e Y .

Definición:

Se define la razón de correlación entre X e Y como:

$$\eta_{Y|X}^2 = \frac{b^2}{S_n^2(Y)} \quad (8)$$

Se tiene lo siguiente:

- Este coeficiente toma valores entre 0 y 1.
- Si es igual a cero no hay relación funcional entre X e Y .
- Si es igual a uno hay relación funcional estricta.

Definición:

Se define la razón de correlación entre X e Y como:

$$\eta_{Y|X}^2 = \frac{b^2}{S_n^2(Y)} \quad (8)$$

Se tiene lo siguiente:

- Este coeficiente toma valores entre 0 y 1.
- Si es igual a cero no hay relación funcional entre X e Y .
- Si es igual a uno hay relación funcional estricta.

Definición:

Se define la razón de correlación entre X e Y como:

$$\eta_{Y|X}^2 = \frac{b^2}{S_n^2(Y)} \quad (8)$$

Se tiene lo siguiente:

- Este coeficiente toma valores entre 0 y 1.
- Si es igual a cero no hay relación funcional entre X e Y .
- Si es igual a uno hay relación funcional estricta.

Definición:

Se define la razón de correlación entre X e Y como:

$$\eta_{Y|X}^2 = \frac{b^2}{S_n^2(Y)} \quad (8)$$

Se tiene lo siguiente:

- Este coeficiente toma valores entre 0 y 1.
- Si es igual a cero no hay relación funcional entre X e Y .
- Si es igual a uno hay relación funcional estricta.

Definición:

Se define la razón de correlación entre X e Y como:

$$\eta_{Y|X}^2 = \frac{b^2}{S_n^2(Y)} \quad (8)$$

Se tiene lo siguiente:

- Este coeficiente toma valores entre 0 y 1.
- Si es igual a cero no hay relación funcional entre X e Y .
- Si es igual a uno hay relación funcional estricta.

Anova

Clase Auxiliar
9

MA34B-01

Coficiente y
Razón de
Correlación

Test de Anova

Aunque estemos en el caso de que exista una relación funcional estricta entre dos variables que estamos estudiando, en una muestra dada es posible que la razón de correlación no nos dé exactamente 1. El test de anova nos da un criterio para decidir cuando la variabilidad que observamos se debe a verdaderas diferencias entre categorías y cuando se debe a la aleatoriedad propia del proceso.

Anova

Clase Auxiliar
9

MA34B-01

Coefficiente y
Razón de
Correlación

Test de Anova

Debemos tener en cuenta algunos puntos con respecto a este test:

- Se asume que la variable cuantitativa estudiada sigue una distribución normal
- Se asume que dentro de todas las categorías se tiene la misma varianza.
- La hipótesis nula es $\mu_i = \mu_j \forall i, j$.
- El test es robusto frente a pequeñas variaciones en el tamaño de las categorías, aunque para las deducciones se suponen todas del mismo tamaño.

Anova

Clase Auxiliar
9

MA34B-01

Coefficiente y
Razón de
Correlación

Test de Anova

Debemos tener en cuenta algunos puntos con respecto a este test:

- Se asume que la variable cuantitativa estudiada sigue una distribución normal
- Se asume que dentro de todas las categorías se tiene la misma varianza.
- La hipótesis nula es $\mu_i = \mu_j \forall i, j$.
- El test es robusto frente a pequeñas variaciones en el tamaño de las categorías, aunque para las deducciones se suponen todas del mismo tamaño.

Anova

Clase Auxiliar
9

MA34B-01

Coefficiente y
Razón de
Correlación

Test de Anova

Debemos tener en cuenta algunos puntos con respecto a este test:

- Se asume que la variable cuantitativa estudiada sigue una distribución normal
- Se asume que dentro de todas las categorías se tiene la misma varianza.
- La hipótesis nula es $\mu_i = \mu_j \forall i, j$.
- El test es robusto frente a pequeñas variaciones en el tamaño de las categorías, aunque para las deducciones se suponen todas del mismo tamaño.

Anova

Clase Auxiliar
9

MA34B-01

Coefficiente y
Razón de
Correlación

Test de Anova

Debemos tener en cuenta algunos puntos con respecto a este test:

- Se asume que la variable cuantitativa estudiada sigue una distribución normal
- Se asume que dentro de todas las categorías se tiene la misma varianza.
- La hipótesis nula es $\mu_i = \mu_j \forall i, j$.
- El test es robusto frente a pequeñas variaciones en el tamaño de las categorías, aunque para las deducciones se suponen todas del mismo tamaño.

Anova

Clase Auxiliar
9

MA34B-01

Coefficiente y
Razón de
Correlación

Test de Anova

Debemos tener en cuenta algunos puntos con respecto a este test:

- Se asume que la variable cuantitativa estudiada sigue una distribución normal
- Se asume que dentro de todas las categorías se tiene la misma varianza.
- La hipótesis nula es $\mu_i = \mu_j \forall i, j$.
- El test es robusto frente a pequeñas variaciones en el tamaño de las categorías, aunque para las deducciones se suponen todas del mismo tamaño.

Pivotes Test Anova

Clase Auxiliar
9

MA34B-01

Coficiente y
Razón de
Correlación

Test de Anova

Se muestra que bajo H_0 :

■ Si $W = nw$ entonces $\frac{W}{\sigma^2} \sim \chi_{n-q}^2$

■ Si $B = nb$ entonces $\frac{B}{\sigma^2} \sim \chi_{q-1}^2$

■ Finalmente

$$\frac{\frac{B}{q-1}}{\frac{W}{n-q}} \sim F_{(q-1), (n-q)}$$

Pivotes Test Anova

Clase Auxiliar
9

MA34B-01

Coficiente y
Razón de
Correlación

Test de Anova

Se muestra que bajo H_0 :

■ Si $W = nw$ entonces $\frac{W}{\sigma^2} \sim \chi_{n-q}^2$

■ Si $B = nb$ entonces $\frac{B}{\sigma^2} \sim \chi_{q-1}^2$

■ Finalmente

$$\frac{\frac{B}{q-1}}{\frac{W}{n-q}} \sim F_{(q-1), (n-q)}$$

Pivotes Test Anova

Clase Auxiliar
9

MA34B-01

Coficiente y
Razón de
Correlación

Test de Anova

Se muestra que bajo H_0 :

- Si $W = nw$ entonces $\frac{W}{\sigma^2} \sim \chi_{n-q}^2$
- Si $B = nb$ entonces $\frac{B}{\sigma^2} \sim \chi_{q-1}^2$
- Finalmente

$$\frac{\frac{B}{q-1}}{\frac{W}{n-q}} \sim F_{(q-1), (n-q)}$$

Pivotes Test Anova

Clase Auxiliar
9

MA34B-01

Coficiente y
Razón de
Correlación

Test de Anova

Se muestra que bajo H_0 :

■ Si $W = nw$ entonces $\frac{W}{\sigma^2} \sim \chi_{n-q}^2$

■ Si $B = nb$ entonces $\frac{B}{\sigma^2} \sim \chi_{q-1}^2$

■ Finalmente

$$\frac{\frac{B}{q-1}}{\frac{W}{n-q}} \sim F_{(q-1), (n-q)}$$