**Parte A. Problemas teóricos y conceptuales**

**Punto 5.**

**5.1.** Si p=n ¿qué relación hay entre ?

n= número de observaciones

k= número de variables independientes

p=k+1

Si sabemos que p=n, se tendría que

Por lo tanto Y es un vector que surge por la combinación lineal de X y

y ya que entonces

Y sabemos que es el vector en que se puede expresar como una combinación lineal de las columnas de que se encuentran más cercanas al vector .

Ahora, por las propiedades de la matriz proyectada en se tiene que

Entonces tenemos que

Por lo tanto, cuando n=p se tendrá que

**5.2.** Argumentación de por qué debe ser ortogonal a cualquier

Sabemos que para que dos vectores sean ortogonales, su producto punto debe ser igual a cero:

Demostrando esto para todo se tiene que:

Ya que la matriz proyectora es simétrica, entonces , luego

Y ya que , se tiene que

**5.3.** Se hace una transformación lineal para crear nuevas variables Z tal que Z=XA, donde es una matriz de rango completo yse define

En 5.1 se explicó que es el vector en que se puede expresar como una combinación lineal de las columnas de que se encuentran más cercanas al vector . Por lo que la transformación lineal para Z también cumple estas características, luego es el punto más cercano a Y pero ahora sobre.

Conociendo que

Trabajaremos con la matriz de proyección, entonces tenemos que

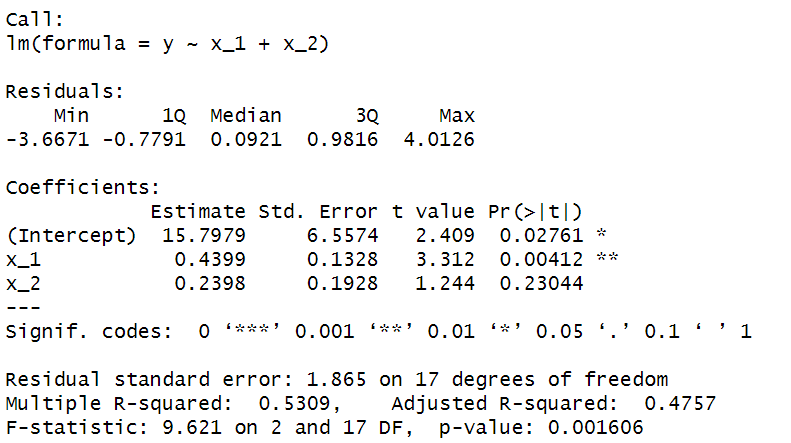
Y ya que y con A tienen el mismo rango, se haría una transformación lineal

Entonces

**5.4.** Tenemos los datos measure.txt que contienen mediciones de los contornos de pecho, cintura y cadera de 20 personas.

**5.4.1** Usando el contorno de pecho como la variable de respuesta (Y) y las otras dos como variables independientes (X1 y X2) estimaremos el modelo de regresión correspondiente.

Donde



Apreciamos que F= 66,18 con p value , por lo tanto el modelo es globalmente significativo.

No obstante, apreciamos por el p-value de la t que, con una significancia del 5%, la variable X\_2 no resulta significativa para explicar la variable de interés, por lo tanto no debería considerarse en el modelo.

**5.4.2** Ahora se quieren cambiar las variables (cintura y cadera) a las variables: tamaño y forma. Donde

**5.4.2.1** Estimando el modelo de regresión

Texto, Carta

Descripción generada automáticamente

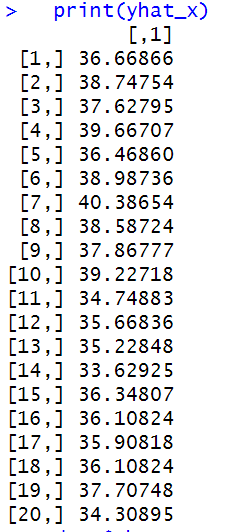
Apreciamos que F= 9,621 con p value , por lo tanto, para una significancia del 5%, el modelo es globalmente significativo.

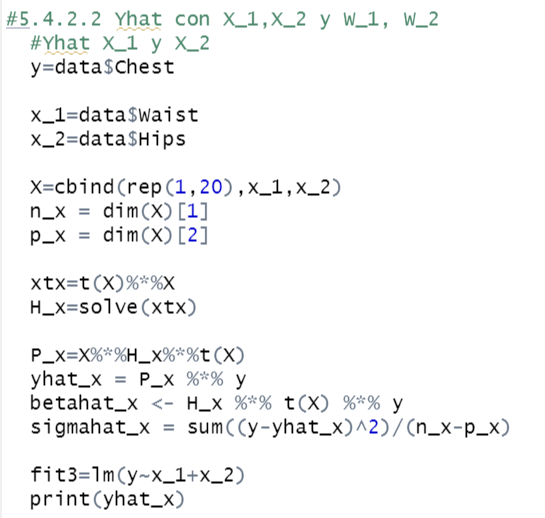
No obstante, apreciamos por el p-value de la t que, con una significancia del 5%, la variable W\_2 no resulta significativa para explicar la variable de interés, por lo tanto no debería considerarse en el modelo.

**5.4.2.2** Encontrando los vectores para cada modelo, con las variables X y con las variables W.

* Modelo con las variables X

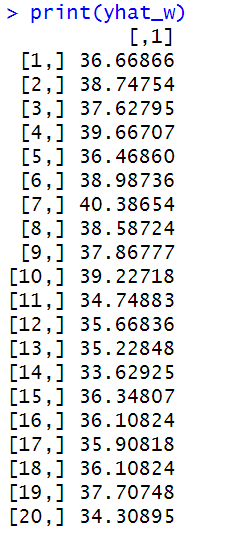
Tabla 1





* Modelo con las variables W

Tabla 2



Texto

Descripción generada automáticamente

Interpretación: se evidencia que los vectores y son iguales, lo cual ejemplifica la demostración realizada en el literal 5.3 con la cual los son iguales aún cuando las columnas de los vectores X y Y no lo son, pero sí son independientes. Lo anterior también ratifica que se construya el mismo subespacio de proyección.

**5.4.2.3** Contrastando los intervalos de confianza del valor esperado del contorno del pecho para ambos casos.

Sabemos que

* Modelo con las variables X

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente



Luego,

En el caso en que la cintura y cadera =30 , con una confianza del 95% y una significancia del 5%, se espera que la media del contorno del pecho esté entre 36,09 y 36,28

* Modelo con las variables W

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Icono

Descripción generada automáticamente

Luego,

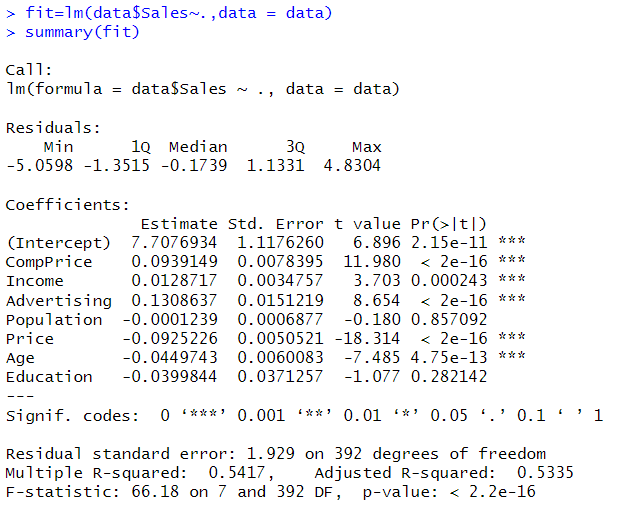
En el caso en que el tamaño=60 y la forma =0 , con una confianza del 95% y una significancia del 5%, se espera que la media del contorno del pecho esté entre 36,09 y 36,28

Interpretación ambos modelos: se evidencia que los intervalos de confianza para y son iguales, lo que indica que la proyección en el subespacio es la misma, con lo cual, el valor esperado de no varía conforme a la transformación que se le haga al vector X.

**Punto 9**

**9.1.**  Estimación modelo de regresión lineal e interpretación de parámetros

Donde



Apreciamos que F= 66,18 con p value , por lo tanto el modelo es globalmente significativo.

**Interpretación parámetros iniciales**  : el valor esperado de ventas en miles de dólares sería de 7,71 en caso de que las demás variables explicativas no sean significativas

: por cada dólar que aumente el precio de la competencia, se estima que el valor esperado de las ventas aumentará en 0.0939 mil unidades.

: por cada mil dólares que aumente el ingreso de la comunidad, se estima que el valor esperado de las ventas aumentará en 0.0129 mil unidades.

: por cada mil dólares que se aumente en la inversión de la publicidad, se estima que el valor esperado de las ventas aumentará en 0.131 mil unidades.

: por cada mil unidades que aumente la población, se estima que el valor esperado de las ventas disminuirá en 0.00012 mil unidades.

: por cada dólar que aumente el precio de las sillas, se estima que el valor esperado de las ventas disminuya en 0.0926 mil unidades.

: por cada año de edad que aumente la población, se estima que el valor esperado de las ventas disminuirá en 0.044 mil unidades.

: por cada nivel que aumente la educación de la población, se estima que el valor esperado de las ventas disminuyan en 0.0399 mil unidades.

No obstante, apreciamos por el p-value de la t que, con una significancia del 5%, las variables de población y educación no son significativas para explicar la variable de interés, por lo tanto no deberían considerarse en el modelo.

**9.2.** ¿Es cierta la hipótesis que afirma que la edad promedio del vecindario no debe influir sobre la venta de sillas?

Sabemos que  y, por el modelo de regresión del anterior punto por prueba t de significancia individual, sabemos que esta variable sí es relevante para explicar la variable de interés, por lo que la hipótesis sería falsa.

No obstante, procederemos a calcularlo manualmente para confirmar:

=0

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Por lo tanto, existe suficiente evidencia estadística para rechazar , lo que significa que la edad del promedio de la población sí es relevante para explicar las ventas.

**9.3.** Un analista cree que la elasticidad del precio es la misma que la del precio de la competencia, pero con signo negativo. ¿Se puede contradecir al analista con los datos?

Se plante la siguiente prueba de hipótesis

Por lo que

Y definimos como la combinación lineal de los parámetros de interés

Y tendremos el siguiente estadístico de prueba

Al calcularlo tenemos que:

Y considerando que

Se concluye que hay suficiente evidencia estadística para no rechazar la hipótesis nula. Por lo cual, no se puede contradecir al analista con los datos.

**9.4.** ¿Es cierta la afirmación de que la edad , la población y la educación del vecindario no son variables que afecten el nivel de ventas de las sillas?

Sabemos que

Se plantea la siguiente prueba de hipótesis

Donde

Al calcularlo tenemos que

Texto, Carta

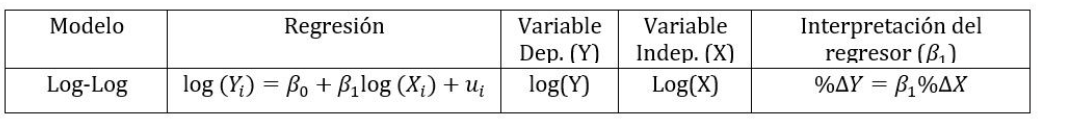
Descripción generada automáticamente

Por lo cual hay suficiente evidencia estadística para rechazar la hipótesis nula. Lo que significa que no es cierta la afirmación.

Lo anterior, puede deberse a que, si revisamos la significancia individual, aunque la población y la educación del vecindario no son variables que afecten el nivel de ventas de las sillas, en el punto 9.2 demostramos que la edad sí es una variable que afecta el nivel de ventas de sillas.

**9.5** Aunque en general que una variable sea significativa no implica causalidad en el cambio de Y (sólamente correlación), es adecuado asumir que las variables precio, precio de la competencia y gasto en publicidad si tienen un efecto directo en el cambio de las ventas de sillas. Suponga que la competencia baja el precio en 10%. La tienda reacciona bajando el precio en 4% e invirtiendo en publicidad un 5% más. ¿Cree que estas acciones son suficientes para no disminuir las ventas de sillas en el largo plazo?

Utilizaremos el modelo de regresión log-log



Ya que con el modelo Log-Log se atribuye a β1 la elasticidad de Y, respecto a X, lo cual se interpreta como un incremento del 1% en X es asociado a un cambio en Y de B1%

Conociendo lo anterior, formularemos el modelo de regresión lineal log-log para hallar los respectivos betas de las variables explicativas.

Sabemos que , y

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Obtenemos los

Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente

Y así, sabemos que

Con lo cual, un incremento del 1% en el precio de la competencia implicará un aumento de las ventas de 1,9%. Un incremento del 1% en el precio de la competencia implicará un aumento de las ventas del 0,106% y un incremento de un 1% en el precio de las sillas implicará un decrecimiento de las ventas de 1,5%

Teniendo en cuenta lo anterior, si la competencia baja el precio en 10% y la tienda reacciona bajando el precio en 4% e invirtiendo en publicidad un 5% más no será suficiente para que no disminuyan las ventas de sillas.