

Docente: Orlando Joaqui Barandica
orlando.joaqui@javerianacali.edu.co
www.joaquibarandica.com

Taller 4

- 1) Para los datos de la Liga Nacional de Fútbol.
 - a) Ajustar un modelo de regresión lineal múltiple que relacione la cantidad de juegos ganados con las yardas por aire del equipo (X2), el porcentaje de jugadas por tierra (X7) y las yardas por tierra del contrario (X8).
 - b) Formar la tabla del ANOVA y probar el significado de la regresión
 - c) Calcular el estadístico t para probar las hipótesis $H_0: \beta_2 = 0, H_0: \beta_7 = 0, H_0: \beta_8 = 0$. ¿Qué conclusiones se pueden sacar acerca del papel de las variables X2, X7 y X8 en el modelo?
 - d) Calcular R2 y R2Adj
 - e) Con la prueba F, determinar la contribución de X7 al modelo.
- 2) Con los resultados del punto 1, demostrar calculando en forma numérica que el cuadrado del coeficiente de correlación simple entre los valores observados Y y los valores ajustados \hat{Y} es igual a R2.
- 3) Para el punto 1, Calcular:
 - a) Un intervalo de confianza de 95% para β_7
 - b) Un intervalo de confianza de 95% para la cantidad media de juegos ganados por un equipo cuando $X_2 = 2300, X_7 = 56, X_8 = 2100$
- 4) Para los mismos datos del punto 1, ajustar un modelo a esos datos, usando solo X7 y X8 como regresores.
 - a) Probar la significancia de la regresión
 - b) Calcular R2 y R2Adj, ¿Cómo se comparan esas cantidades con las calculadas para el modelo del punto 1, que tenía un regresor más (X2)?
 - c) Calcular un intervalo de confianza de 95% para β_7 . También un intervalo de confianza de 95% para la cantidad media de juegos ganados por un equipo cuando $X_7 = 56$ y $X_8 = 2100$. Comparar la longitud de esos intervalos de confianza con las longitudes de los correspondientes en el punto 3.
 - d) ¿Qué conclusiones se pueden sacar de este problema, acerca de las consecuencias de omitir un regresor importante de un modelo?

- 5) Utilice los datos sobre precios de viviendas.
- Ajustar un modelo de regresión múltiple que relacione el precio de venta con los nueve regresores.
 - Probar la significancia de la regresión. ¿Qué conclusiones se pueden sacar?
 - Usar pruebas t para evaluar la contribución de cada regresor al modelo.
 - ¿Cuál es la contribución del tamaño de lote y el espacio vital para el modelo, dado que se incluyeron todos los demás regresores?
 - Utilice las herramientas de selección de variables, cuales serían los regresores que deben estar en el modelo.
- 6) ¿Cuál es la mejor estimación en el caso de la Liga Nacional de Fútbol? ¿Cumple los supuestos este modelo para realizar inferencias? Realice un análisis gráfico de las condiciones de correlación, de linealidad, normalidad y varianza constante.
- 7) ¿Cuál es la mejor estimación en el caso de los datos de precios de viviendas? ¿Cumple los supuestos este modelo para realizar inferencias? Realice un análisis gráfico de las condiciones de correlación, de linealidad, normalidad y varianza constante.
- 8) Una empresa fabricante de botellas de vidrio ha registrado datos acerca de la cantidad promedio de defectos por 10000 botellas, debido a piedras (pequeñas piedras embebidas en la pared de la botella) y la cantidad de semanas a partir de la última reparación general del horno. Los datos son los siguientes:

Defectos por 10 000	Semanas	Defectos por 10 000	Semanas
13.0	4	34.2	11
16.1	5	65.6	12
14.5	6	49.2	13
17.8	7	66.2	14
22.0	8	81.2	15
27.4	9	87.4	16
16.8	10	114.5	17

- Ajustar un modelo lineal a los datos y hacer las pruebas usuales de adecuación del modelo.
 - Sugerir una transformación adecuada que elimine los problemas encontrados en la parte a). Ajustar el modelo transformado y comprobar su adecuación.
- 9) El grado de carbonatación de una bebida gaseosa se afecta por la temperatura del producto y por la presión de funcionamiento de la llenadora. Se obtuvieron 12 observaciones y los datos resultantes se presentan a continuación.
- Ajustar un polinomio de segundo orden completo. Es decir, los cuadrados de X_1 y X_2 y su interacción.
 - Probar la significancia de la regresión.
 - ¿Contribuye al modelo el término de interacción en forma significativa?
 - ¿Contribuyen al modelo los términos de segundo orden en forma significativa?

Carbonatación, y	Temperatura, x_1	Presión, x_2
2.60	31.0	21.0
2.40	31.0	21.0
17.32	31.5	24.0
15.60	31.5	24.0
16.12	31.5	24.0
5.36	30.5	22.0
6.19	31.5	22.0
10.17	30.5	23.0
2.62	31.0	21.5
2.98	30.5	21.5
6.92	31.0	22.5
7.06	30.5	22.5

10) En la tabla de datos de vinos.

- Estimar un modelo con todos los regresores. E incorporar la información de variables indicadoras.
- Según el criterio de selección de variables que modelo o modelos podrían ser candidatos para explicar mejor la relación. Investigue la adecuación con los supuestos y gráficos de residuales.
- Evalúe una prueba de hipótesis conjunta, para un par de Betas.

11) Realice un análisis descriptivo a la base de datos de vinos.

- ¿Identifica alguna posible relación que le de indicios de una interacción?
- Estime un modelo con efectos de interacción entre la variable Región y cuantitativas (la que usted identifique). ¿Es significativo? Interprete
- ¿Existen efectos de interacción significativos entre las variables cuantitativas?

Equipo	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Washington	10	2113	1985	38.9	64.7	+4	868	59.7	2205	1917
Minnesota	11	2003	2855	38.8	61.3	+3	615	55.0	2096	1575
New England	11	2957	1737	40.1	60.0	+14	914	65.6	1847	2175
Oakland	13	2285	2905	41.6	45.3	-4	957	61.4	1903	2476
Pittsburgh	10	2971	1666	39.2	53.8	+15	836	66.1	1457	1866
Baltimore	11	2309	2927	39.7	74.1	+8	786	61.0	1848	2339
Los Angeles	10	2528	2341	38.1	65.4	+12	754	66.1	1564	2092
Dallas	11	2147	2737	37.0	78.3	-1	761	58.0	1821	1909
Atlanta	4	1689	1414	42.1	47.6	-3	714	57.0	2577	2001
Buffalo	2	2566	1838	42.3	54.2	-1	797	58.9	2476	2254
Chicago	7	2363	1480	37.3	48.0	+19	984	67.5	1984	2217
Cincinnati	10	2109	2191	39.5	51.9	+6	700	57.2	1917	1758
Cleveland	9	2295	2229	37.4	53.6	-5	1037	58.8	1761	2032
Denver	9	1932	2204	35.1	71.4	+3	986	58.6	1709	2025
Detroit	6	2213	2140	38.8	58.3	+6	819	59.2	1901	1686
Green Bay	5	1722	1730	36.6	52.6	-19	791	54.4	2288	1835
Houston	5	1498	2072	35.3	59.3	-5	776	49.6	2072	1914
Kansas City	5	1873	2929	41.1	55.3	+10	789	54.3	2861	2496
Miami	6	2118	2268	38.2	69.6	+6	582	58.7	2411	2670
Nueva Orleans	4	1775	1983	39.3	78.3	+7	901	51.7	2289	2202
Nueva York Giants	3	1904	1792	39.7	38.1	-9	734	61.9	2203	1988
Nueva York Jets	3	1929	1606	39.7	68.8	-21	627	52.7	2592	2324
Philadelphia	4	2080	1492	35.5	68.8	-8	722	57.8	2053	2550
St. Louis	10	2301	2835	35.3	74.1	+2	683	59.7	1979	2110
San Diego	6	2040	2416	38.7	50.0	0	576	54.9	2048	2628
San Francisco	8	2447	1638	39.9	57.1	-8	848	65.3	1786	1776
Seattle	2	1416	2649	37.4	56.3	-22	684	43.8	2876	2524
Tampa Bay	0	1503	1503	39.3	47.0	-9	875	53.5	2560	2241

y: Juegos ganados (por temporada de 14 juegos)

x_1 : Yardas por tierra (temporada)

x_2 : Yardas por aire (temporada)

x_3 : Promedio de pateo (yardas/patada)

x_4 : Porcentaje de goles de campo (GC hechos/GC intentados, temporada)

x_5 : Diferencia de pérdidas de balón (pérdidas ganadas/pérdidas perdidas)

x_6 : Yardas de castigo (temporada)

x_7 : Porcentaje de carreras (jugadas por tierra/jugadas totales)

x_8 : Yardas por tierra del contrario (temporada)

x_9 : Yardas por aire del contrario (temporada)

y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
25.9	4.9176	1.0	3.4720	0.9980	1.0	7	4	42	0
29.5	5.0208	1.0	3.5310	1.5000	2.0	7	4	62	0
27.9	4.5429	1.0	2.2750	1.1750	1.0	6	3	40	0
25.9	4.5573	1.0	4.0500	1.2320	1.0	6	3	54	0
29.9	5.0597	1.0	4.4550	1.1210	1.0	6	3	42	0
29.9	3.8910	1.0	4.4550	0.9880	1.0	6	3	56	0
30.9	5.8980	1.0	5.8500	1.2400	1.0	7	3	51	1
28.9	5.6039	1.0	9.5200	1.5010	0.0	6	3	32	0
35.9	5.8282	1.0	6.4350	1.2250	2.0	6	3	32	0
31.5	5.3003	1.0	4.9883	1.5520	1.0	6	3	30	0
31.0	6.2712	1.0	5.5200	0.9750	1.0	5	2	30	0
30.9	5.9592	1.0	6.6660	1.1210	2.0	6	3	32	0
30.0	5.0500	1.0	5.0000	1.0200	0.0	5	2	46	1
36.9	8.2464	1.5	5.1500	1.6640	2.0	8	4	50	0
41.9	6.6969	1.5	6.9020	1.4880	1.5	7	3	22	1
40.5	7.7841	1.5	7.1020	1.3760	1.0	6	3	17	0
43.9	9.0384	1.0	7.8000	1.5000	1.5	7	3	23	0
37.5	5.9894	1.0	5.5200	1.2560	2.0	6	3	40	1
37.9	7.5422	1.5	5.0000	1.6900	1.0	6	3	22	0
44.5	8.7951	1.5	9.8900	1.8200	2.0	8	4	50	1
37.9	6.0831	1.5	6.7265	1.6520	1.0	6	3	44	0
38.9	8.3607	1.5	9.1500	1.7770	2.0	8	4	48	1
36.9	8.1400	1.0	8.0000	1.5040	2.0	7	3	3	0
45.8	9.1416	1.5	7.3262	1.8310	1.5	8	4	31	0

y : Precio de venta de la casa/1000

x_1 : Impuestos (locales, escuela, municipal)/1000

x_2 : Cantidad de baños

x_3 : Tamaño del terreno (pies cuadrados \times 1000)

x_4 : Superficie construida (pies cuadrados \times 1000)

x_5 : Cantidad de cajones en cochera

x_6 : Cantidad de habitaciones

x_7 : Cantidad de recámaras

x_8 : Edad de la casa (años)

x_9 : Cantidad de chimeneas

$x_1 =$ claridad	$x_2 =$ aroma	$x_3 =$ cuerpo	$x_4 =$ sabor	$x_5 =$ fuerza	$y =$ calidad	Región
1	3.3	2.8	3.1	4.1	9.8	1
1	4.4	4.9	3.5	3.9	12.6	1
1	3.9	5.3	4.8	4.7	11.9	1
1	3.9	2.6	3.1	3.6	11.1	1
1	5.6	5.1	5.5	5.1	13.3	1
1	4.6	4.7	5	4.1	12.8	1
1	4.8	4.8	4.8	3.3	12.8	1
1	5.3	4.5	4.3	5.2	12	1
1	4.3	4.3	3.9	2.9	13.6	3
1	4.3	3.9	4.7	3.9	13.9	1
1	5.1	4.3	4.5	3.6	14.4	3
0.5	3.3	5.4	4.3	3.6	12.3	2
0.8	5.9	5.7	7	4.1	16.1	3
0.7	7.7	6.6	6.7	3.7	16.1	3
1	7.1	4.4	5.8	4.1	15.5	3
0.9	5.5	5.6	5.6	4.4	15.5	3
1	6.3	5.4	4.8	4.6	13.8	3
1	5	5.5	5.5	4.1	13.8	3
1	4.6	4.1	4.3	3.1	11.3	1
0.9	3.4	5	3.4	3.4	7.9	2
0.9	6.4	5.4	6.6	4.8	15.1	3
1	5.5	5.3	5.3	3.8	13.5	3
0.7	4.7	4.1	5	3.7	10.8	2
0.7	4.1	4	4.1	4	9.5	2
1	6	5.4	5.7	4.7	12.7	3
1	4.3	4.6	4.7	4.9	11.6	2
1	3.9	4	5.1	5.1	11.7	1
1	5.1	4.9	5	5.1	11.9	2
1	3.9	4.4	5	4.4	10.8	2
1	4.5	3.7	2.9	3.9	8.5	2
1	5.2	4.3	5	6	10.7	2
0.8	4.2	3.8	3	4.7	9.1	1
1	3.3	3.5	4.3	4.5	12.1	1
1	6.8	5	6	5.2	14.9	3
0.8	5	5.7	5.5	4.8	13.5	1
0.8	3.5	4.7	4.2	3.3	12.2	1
0.8	4.3	5.5	3.5	5.8	10.3	1
0.8	5.2	4.8	5.7	3.5	13.2	1

La clase de este vino es Pinot Noir. Región se refiere a regiones geográficas distintas.