

Análisis Estadístico de Datos Climáticos
Práctico 5 (2015)
Regresión lineal, composites y ANOVA

Se pide la entrega de uno de los primeros dos ejercicios resueltos. Si se entregan dos o los tres, se otorgará mayor calificación.

1) El Índice de Oscilación Sur (IOS o SOI por sus siglas en inglés) se calcula como la diferencia de presiones a nivel del mar (estandarizadas) entre la isla de Tahití en el Océano Pacífico y la ciudad de Darwin en el norte de Australia. El archivo ios_5107 contiene datos mensuales de este índice entre 1951 y 2007. En el archivo nino34.dat se dan, para el mismo período, anomalías mensuales del índice Niño3.4 (N3.4) medidas en °C (la primera columna es el año, la segunda el mes, y la tercera el valor de la anomalía N3.4).

Se pide:

- 1) Construir la serie de anomalías de N3.4 para los meses de diciembre entre 1951 y 2007.
- 2) Estimar los coeficientes de correlación (Pearson) entre cada serie mensual del IOS y la construida en 1). (Es decir entre el IOS de cada mes del año con la serie de anomalías de N3.4 en diciembre siguiente; en total se piden 12 valores.)
- 3) Utilizando un test de significancia estadística para correlaciones dado en clase, determinar para qué meses las correlaciones son significativas a) al 5%, b) al 1 % (test de 1 cola o extremo).
- 4) Indicar cuál es el mes del IOS para el cual la relación con las anomalías de N3.4 en diciembre es más aproximadamente lineal. Para ese caso, construir un scatterplot ubicando en abscisas (\underline{x}), los valores del IOS del mes (“predictor”), y en ordenadas (\underline{y}) las anomalías de N3.4 en diciembre (“predictando”).
- 5) Ajustar una línea recta entre ambas variables por el método de mínimos cuadrados, dando su ecuación y graficándola superpuesta al scatterplot de 4).
- 6) Graficar los residuos del ajuste en función del predictor. Evaluar visualmente si los residuos cumplen razonablemente la hipótesis de varianza constante. Adicionalmente, realizar un test de normalidad de los residuos con un nivel de significancia del 5%.
- 7) Construir y graficar intervalos de confianza al 95% para las anomalías de N3.4 en diciembre (superpuestos en la gráfica anterior). En particular, sabiendo que el valor del IOS en setiembre 2008 fue de 1.5, utilizando esta aproximación lineal, dar una “predicción retrospectiva” para la anomalía de N3.4 en diciembre 2008, con un intervalo de confianza del 95%.

2) El fenómeno de El Niño que ocurre en el Pacífico ecuatorial tiene manifestación no solamente en la TSM, sino también por debajo de la superficie. De hecho, la dinámica oceánica por debajo de la superficie es fundamental para el desarrollo de este fenómeno. En este ejercicio veremos la evolución de las anomalías de temperatura ecuatorial entre superficie y 500m de profundidad.

El archivo temp_ecuatorial_Pacifico_E1982_D2007.mat contiene datos de temperatura ecuatorial (latitud 0) desde superficie hasta 500m de profundidad desde enero de 1982 hasta diciembre de 2007 en °C.

Se pide:

1. Graficar la media de temperatura en todo el período. La zona de gran gradiente vertical de temperatura se define como la termoclina y juega un papel fundamental en la dinámica de El Niño. (Sug.: para graficar se pueden usar los comandos `contourf` o `contour`, y `clabel`)
2. Superponer en la misma figura anterior la desviación estandar de las anomalías de temperatura para todo el período. ¿Cual es la región de máximas anomalías?
3. Los eventos Niño tienden a alcanzar su máximo en el mes de Diciembre. Para observar su evolución, graficar los composites (Años Niño – Años Neutros) de anomalías de temperatura para los bimestres MJ JA SO y ND.
Considerar: Años El Niño [1982 1986 1987 1991 1997 2002]
Años La Niña [1984 1988 1995 1998 1999 2000]
4. Calcular la significancia estadística de los composites al 5% usando el test de significancia explicado en clase.
5. Describir la evolución de las anomalías de temperatura en superficie y en subsuperficie.
6. El archivo `clorofila_ND_1997_2001.mat` contiene datos de clorofila-a en (mg/m^3) para el bimestre ND de los años 1997-2001 en el Pacífico ecuatorial medidos por Sea-viewing Wide Field-of-view Sensor (SeaWiFS). Comparar la concentración de clorofila entre los años Niño y Niña de este período (sin calcular significancia). ¿Cómo se interpreta la diferencia de concentraciones?

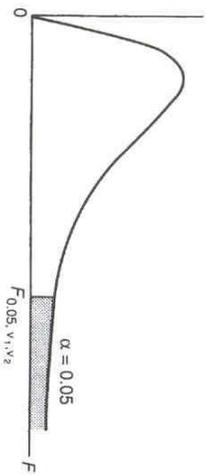
3) Se quiere comparar los registros de temperatura realizados con 3 tipos de CTD (instrumentos conductivity-temperature-depth) colocados en el mismo lugar en idénticas condiciones muestrales. Cuatro países toman parte en esta intercomparación y cada uno trae los mismos 3 tipos de CTD. Los resultados del test son los siguientes:

Medida	CTD-1	CTD-2	CTD-3
1	15.001	15.004	15.002
2	14.999	15.002	15.003
3	15.000	15.001	15.002
4	14.998	15.004	15.002

Determinar, usando ANOVA, si diferentes CTDs (diferentes tratamientos) miden la misma temperatura dentro del error de los instrumentos. Testear hipótesis al 5% y al 1%. Se adjuntan tablas de la distribución F para dichos niveles.

Table D.4. Critical values of the F -distribution for (a) $\alpha = 0.05$ and (b) 0.01 . The distributions represent the area exceeding the value of $F_{0.05, v_1, v_2}$ and $F_{0.01, v_1, v_2}$ as shown by the shaded area in the figure for different degrees of freedom, v . For example, if $v_1 = 15$ and $v_2 = 20$, then the critical value for $\alpha = 0.05$ is 2.20. [Adapted from Introductory Statistical Analysis by D. L. Harnett and J. L. Murphy, Addison-Wesley, 1976; abridged from tables of percentage points of the inverted beta (F) distribution by M. Merrington and C. M. Thompson Biometrika, Vol. 33 (1943).]

$P(F > 2.20) = 0.05$,
 $P(F < 2.20) = 0.95$.



Values of $F_{0.05, v_1, v_2}$
 v_1 = degrees of freedom for numerator

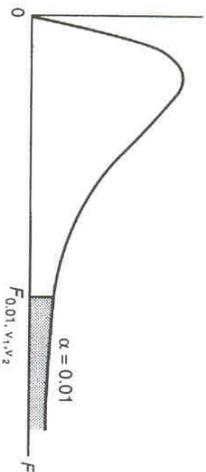
v_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54

v_2 = degrees of freedom for denominator

Table D.4 (b) Critical values of the F-distribution

$$P(F > 3.09) = 0.01,$$

$$P(F < 3.09) = 0.99.$$



Values of $F_{0.01, v_1, v_2}$
 v_1 = Degrees of freedom for numerator

v_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	5000	5403	5625	5764	5859	5928	5982	6023	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.4	26.3	26.2	26.1
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.7	13.6	13.5
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91