

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
FACULTAD DE CIENCIAS E INGENIERÍA

ESTADÍSTICA PARA INGENIERÍA

1ra Práctica (tipo c)
(Primer Semestre 2021)

Indicaciones Generales:

- Esta práctica debe ser desarrollada por un grupo de máximo 3 alumnos del mismo horario.
- Justifique claramente sus soluciones, presente los códigos y resultados obtenidos en R, interprete estos resultados y escriba respuestas completas teniendo en cuenta el contexto de los ejercicios propuestos; todo esto influirá en la calificación.
- La presentación, la ortografía y la gramática de los trabajos influirán en la calificación.
- La solución debe ser subida en único archivo PDF en la tarea PC1 en la página de Padeia de su correspondiente horario del curso. La primera hoja debe presentar el código PUCP, nombre completo y correo electrónico de todos los integrantes del grupo.

Puntaje debido al cuestionario: 20 puntos

Puntaje total: 20 puntos

Cuestionario:

La base de datos `weatherAUSCanberra.csv` contiene variables de observaciones meteorológicas diarias del 2008 al 2016 de la ciudad de Canberra en Australia. Esta base de datos será utilizada en las preguntas 1 a 5.

Date	Fecha de la observación.
Location	Ciudad de ubicación de la estación meteorológica.
MinTemp	Temperatura mínima en grados centígrados.
MaxTemp	Temperatura máxima en grados centígrados.
Rainfall	Cantidad de lluvia registrada para el día en mm.
Evaporation	Evaporación, mediada en un tanque evaporimétrico clase A (mm) a las 9 a.m. luego de haber pasado 24 hora.
Sunshine	Número de horas de sol brillante en el día.
WindGustDir	Dirección de la ráfaga de viento más fuerte en las 24 horas hasta la medianoche.
WindGustSpeed	Velocidad (km/h) de la ráfaga de viento más fuerte en las 24 horas hasta la medianoche.
WindDir9am	Dirección del viento a las 9 a.m.
WindDir3pm	Dirección del viento a las 3 p.m.
WindSpeed9am	Velocidad del viento (km/h) en promedio durante 10 minutos antes de las 9 a.m.
WindSpeed3pm	Velocidad del viento (km/h) en promedio durante 10 minutos antes de las 3 p.m.
Humidity9am	Humedad (porcentaje) a las 9 a.m.
Humidity3pm	Humedad (porcentaje) a las 3 p.m.
Pressure9am	Presión atmosférica (hectopascascales, hpa) reducida al nivel medio del mar a las 9 a.m.
Pressure3pm	Presión atmosférica (hectopascascales, hpa) reducida al nivel medio del mar a las 3 p.m.
Cloud9am	Fracción del cielo oscurecida por las nubes a las 9 a.m. Esto se mide en "octas", que son una unidad de octavos. Se registra cuantos.
Cloud3pm	Fracción de cielo oscurecida por nubes (en "octas": octavos) a las 3 p.m.
Temp9am	Temperatura (grados Centígrados) a las 9 a.m.
Temp3pm	Temperatura (grados Centígrados) a las 3 p.m.
RainToday	¿Llovió hoy? 1 = Si, la precipitación (mm) en las 24 horas hasta las 9 a.m. excede 1 mm, de lo contrario es 0 = No.
RainTomorrow	¿Llovió mañana? 1 = Si, 0 = No.
Year	Año
Month	Mes (1=Enero, ..., 12=Diciembre)

Season	Estación (summer, spring, winter, autumn)
--------	-------------------------------------------

Pregunta 1 (2.0 puntos)

Se desea comparar gráficamente las distribuciones de frecuencias relativas de la variable ¿Llovió hoy (`RainToday`)? entre las estaciones del año (`Season`). Obtenga también las tablas de distribución de frecuencias de esta variable para cada estación. Interprete los resultados obtenidos.

Pregunta 2 (2.0 puntos)

Para un estudio se van a considerar solamente la información de los días cuya temperatura mínima (`MinTemp`) haya estado dentro del 6% de mayores temperaturas mínimas observadas durante los 9 años. ¿qué temperatura mínima se debe superar como mínimo en un día para que este día sea incluido en el estudio? Justifique su respuesta.

Pregunta 3 (4.0 puntos)

Obtenga la tabla de frecuencias y su histograma considerando 10 intervalos de igual amplitud para la variable Temperatura máxima (`MaxTemp`) y grafique también su polígono de frecuencias.

- (2.0 puntos) ¿Se puede decir que esta es una distribución simétrica? Justifique su respuesta.
- (2.0 puntos) Considere que solamente se tiene la información de la tabla de frecuencias. En este caso, aproxime la proporción de días que tienen una temperatura máxima entre 15 y 25 grados centígrados.

Pregunta 4 (4.0 puntos)

Calcule estadísticas descriptivas de tendencia central, dispersión y forma para temperatura a las 9:00 a. m. (`Temp9am`) para los niveles de la variable Estación (`season`). A partir de sus resultados responda:

- (2.0 puntos) ¿Qué medidas de tendencia central y de dispersión serían las más adecuadas para realizar una comparación de la temperatura diaria a las 9 a. m. en cada estación?
- (2.0 puntos) Considerando solo las estadísticas obtenidas realice una comparación en términos de tendencia central, dispersión y asimetría.

Pregunta 5 (4.0 puntos)

Responda a las siguientes preguntas (presente sus códigos y resultados en R como parte de su justificación)

- (1.0 punto) Para representar la tendencia central de la variable Evaporación (`Evaporation`) es adecuado usar la media. Evalúe la veracidad o falsedad de esta afirmación. Justifique su respuesta.
- (1.0 punto) La variable temperatura máxima en grados centígrados (`MaxTemp`) es más dispersa que la velocidad del viento (km/h) en promedio durante 10 minutos antes de las 3 p.m. (`WindSpeed3pm`). Evalúe la veracidad o falsedad de esta afirmación. Justifique su respuesta.
- (1.0 punto) Se puede afirmar que la temperatura máxima (`MaxTemp`) promedio es la misma para días con lluvia o sin lluvia (`RainToday`). Evalúe la veracidad o falsedad de esta afirmación. Justifique su respuesta.
- (1.0 punto) Se aplicó la función `summary` a la variable temperatura máxima en grados centígrados (`MaxTemp`) y a partir de los resultados obtenidos, se llegó a la conclusión que en el 50% de las temperaturas máxima en grados centígrados es mayor a 20.78 grados centígrados. Evalúe la veracidad o falsedad de esta afirmación. Justifique su respuesta.

Pregunta 6 (4.0 puntos)

Para esta pregunta considere el conjunto de datos WeatherAUSThreeCities.csv, que contiene las mismas variables que el conjunto de datos original, pero medidas en 3 ciudades.

Suponga que se desea comparar las distribuciones de la temperatura mínima alcanzadas en el periodo de estudio entre las ciudades de Adelaide, Canberra y Sidney, para lo cual un investigador propone crear un gráfico tipo boxplot, pero en donde las cajas estén constituidas en su límite inferior por la media menos una desviación estándar, en su límite superior por la media más una desviación estándar, tengan una línea central en la caja a la altura de la media y sus bigotes vayan hasta más y menos 2.5 desviaciones estándar de la media.

a) (2.0 puntos) Realice un boxplot, como fue propuesto en el curso, para hacer esta comparación e indique qué es lo de ella se concluye.

b) (2.0 puntos) Realice, por otro lado, el gráfico propuesto por el investigador y en base a los resultados de a) indique si este le permitiría obtener resultados similares y qué problemas piensa usted tendría el gráfico de este investigador.

Sugerencia: El gráfico pedido en b) lo puede elaborar en forma manual, no necesariamente en R.

Los profesores del curso

San Miguel, 20 de abril del 2021