AST 0212 – Introducción al análisis de datos

Semestre 2016-1

Profesor:
Alejandro Clocchiatti
Ayudantes:
Francisco Aros – Nicolás Castro

Tarea 2:
Compensación de errores estadísticos y sistemáticos.
Propagación de errores. Correlación.
Comparación cuantitativa de observaciones y modelos de la realidad.

Entrega: Lunes 09 de Mayo hasta 17:20hs
Texto y gráficos: En casilla del profesor – Instituto de Astrofísica.
Programas desarrollados y tablas de datos usadas: En buzón de tareas en web de AST0212.

Problema 1: Construcción de la "Tabla grande" y transformación a pesa GAMA .

En la página web del curso, sección "Recursos", hay un archivo tipo "tar" donde Francisco Aros transcribió todas y cada una de las mediciones de peso que ustedes hicieron el día 18 de Marzo (las mediciones de peso de ustedes mismos, repetidas en todas las balanzas). Están organizadas en archivos de texto simple, con un número identificatorio para cada individuo, de 1 a 32, y dos o tres mediciones en cada una de las balanzas. Nos interesa obtener dos elementos de esos datos, (1) la incerteza de la medición en cada balanza (el llamado "error de lectura") y (2) el peso de cada persona en la escala de la balanza de mejor calidad (la GAMA, usada por el grupo 7). Esos dos elementos nos permitirán conocer el peso de cada persona que midieron en la primera semana del curso, y su incerteza, expresados en una única escala de peso, la de la balanza GAMA.

1: Tabla Grande .

Tenemos que construir la "Tabla Grande" que contenga el peso promedio de cada uno de los 32 estudiantes que participó de la pesatón del 18 de Marzo y la incerteza de cada medida. Esto se hace directamente con los datos que copió Francisco. Para el caso de balanzas que hicieron dos mediciones estime el "error de lectura" como la diferencia cuadrática media entre estos pares. Para el caso de balanzas en que se hicieron tres mediciones estime el "error de lectura" como la desviación cuadrática media (en sentido usual, con relación al promedio de las tres medidas). Organice la tabla con 17 columnas (ID del sujeto, y peso e incerteza en cada balanza), y 32 líneas (una por sujeto), más lo que le parezca conveniente como encabezado.

NOTA: Como es usual en estos casos, va a encontrar números que no hacen sentido dentro de la lógica de los datos (afortunadamemte son sólo dos, una medición del peso de un individuo, por un lado, y el peso completo de otro individuo en otro). Tendrá que decidir qué hacer con esos datos. Sea lo que sea que haga, debe informarlo y justificarlo.

2: Correlación con los datos de la balanza GAMA .

Usando las columnas de datos de la "Tabla Grande", obtenga la correlación lineal de cuadrados mínimos que le permita transformar los pesos medidos en la balanza 7 a la escala de las balanzas de los grupos 1, 2, 3, 4, 5, 6, y 8. Calcule las constantes de transformación con sus respectivas incertezas, para lo cual necesitará como dato las incertezas de lectura calculadas en el problema anterior. Si las incertezas en la "Tabla Grande" son inferiores al error de lectura esperado en principio (por ejemplo en el caso de las balanzas digitales), utilice el error de lectura esperado. Prepare una tabla con los parámetros de cada ajuste (constantes de la recta, sus incertezas, RMS global del ajuste, y coeficiente de correlación). Grafique cada uno de los ajustes en un gráfico separado.

3: Transformación lineal a la escala de la balanza GAMA .

Invierta las relaciones lineales anteriores para poder pasar de la escala de estas balanzas a las del Grupo 7. Tenga especial cuidado en la transformación de las incertezas. Explique por qué es que seguimos este camino (es decir, ¿por qué no buscamos directamente la correlación que tranforma el peso de la balanza GAMA a cada una de las otras?) (¿pregunta de evaluación?).

Problema 2: .

Habiendo transformado los pesos a la escala de la balanza GAMA, nos preocupa ahora entender cuan exacta es ésta. De la "pesatón" hecha en la Clase 2 resulta claro que es la balanza más precisa. Pero ¿da el peso exacto? ¿lo da para todos los pesos? Podría ser que las medidas de la balanza GAMA tuvieran una desviación sistemática con relación a la escala absoluta de peso. Para contestar esto, el profesor hizo la rutina que describió en clase, pesando volúmenes cuidadosamente medidos de agua.

En la sección "Recursos" de la página web del curso usted encontrará un archivo llamado calibracion_GAMA_0.dat donde el profesor registró los datos necesarios para la calibración de las lecturas de la balanza GAMA en una escala absoluta. Estudie las distintas posibilidades de ajustar estos datos con polinomios de cuadrados mínimos (probando con distintos grados y grupos de puntos, en clase discutimos algunas estrategias) y explique por qué usó el ajuste que usó. Grafique los datos y el polinomio, o los polinomios, de ajuste.

Problema 3: Transformación de pesos a escala absoluta .

Con las relaciones lineales que calculó en el Problema 1 (para pasar las lecturas de todas las balanzas a la balanza GAMA), y la ecuación de calibración que obtuvo en el Problema 2 (para pasar las lecturas de la balanza GAMA a la escala absoluta), usted puede pasar todos los pesos tomados en las cafeterías, cantinas y casinos del campus a esta escala absoluta. Escriba las ecuaciones combinadas de transformación para las balanzas 1, 2, 3, 4, 5, 6, y 8, (tenga en cuenta las incertezas) y utilícelas para pasar todos los pesos a la escala absoluta. Note que para la balanza GAMA (Grupo 7), sólo necesita la ecuación de calibración del Problema 2. Reconstruya las tablas de datos originales de cada grupo, las que tenían las columnas tipo "E 22 F AC 50 12 59" pero cambie la columna del peso por dos, una con el peso en escala absoluta y otra con su incerteza (le quedarán como "E 22 F AC 51.6 0.3 12 59"). Preste atención al número de dígitos que reporta (use sólo los que sean significativos y justifíquelos). Habiendo completado este último paso, usted tiene los pesos calibrados de tal manera que podría compararlos con cualquier otro investigador de la Tierra y utilizarlos en análisis combinados.

Problema 4: Comparación final: Realidad, modelos & imaginación .

Trabajaremos a partir de ahora con histogramas para poblaciones diferentes estadísticamente significativas observadas en los distintos escenarios lo que implica concentrarse solamente en los estudiantes. Construya un histograma combinado para todos los estudiantes varones y otro para las mujeres, asuma como modelo de la distribución subyacente una gaussiana $N(\mu, \sigma)$, donde μ y σ son el valor medio y la desviación estándar medidos para cada histograma. Calcule el número de casos esperados en cada bin y produzca otros dos gráficos como los anteriores, incluyendo ahora con línea de otro tipo, o color, los histogramas de valores esperados (teóricos). Comparando el histograma observado con el teórico calcule el valor de χ^2 para cada caso y su probabilidad. ¿Son consistentes las observaciones con la hipótesis sobre la población subyacente?

Poblaciones aisladas: Datos reales .

Asumiendo que los valores medios y dispersiones medidos en el item previo son los representativos de la población subyacente, obtenga los χ^2 y sus probabilidades para las distribuciones de peso de estudiantes varones y mujeres medidos por los 8 grupos. ¿Cómo se comparan con los anteriores?

Datos imaginarios .

- Repita el ejercicio para los histogramas de datos inventados, comparándolos con la distribución subyacente medida antes (Calcule nuevos χ^2 y sus probabilidades para los histogramas inventados, individuales y colectivos).
- Vuelva ahora a la tabla final de la Tarea 1, que contiene los χ^2 de los histogramas de datos inventados, y calcule la probabilidad de que esos valores de χ^2 puedan haber sido obtenidos por azar. Asegúrese de entender la diferencia conceptual (profunda) entre estos χ^2 y los obtenidos en el item previo (¿pregunta de evaluación?).

Ahora tiene en sus manos las herramientas y datos necesarios para poder hacer una interpretación cuantitativa de todo este proceso. Hay una población subyacente de "variables" (peso de personas en este caso), que sigue una cierta FDP. Hicimos observaciones y decidimos caracterizarla como una $N(\mu, \sigma)$. Tenemos los resultados que nos permiten evaluar cuán buena es esta hipótesis para los datos obtenidos en total, para cada una de las poblaciones aisladas, y para las poblaciones inventadas. A partir de todo esto, evalue el experimento y el valor de las hipótesis (siga las líneas de análisis descritas por el profesor en la clase) (¿pregunta de evaluación?).

Problema 5: ¿Se notará la comida? .

Divida el total de poblaciones de estudiantes hombres y mujeres en aquellos que fueron pesados antes de la comida y después, calcule el valor medio del peso y el error del valor medio. ¿Hay alguna diferencia? ¿Cuál es el peso de la comida promedio que comió cada población? ¿Es significativa la medida? Si ese fuera el peso real de la comida ¿Cuán grande debería ser la población estudiada para que la medición del peso de la comida resultara significativa? (¿pregunta de evaluación?)