Proyecto completo: análisis bayesiano de escuelas

El proyecto aplica todo el flujo del curso a un problema nuevo. Tienes los puntajes de 1028 alumnos de 30 escuelas, con entre 5 y 60 alumnos por escuela. Para cada alumno conoces las horas de estudio semanales (0 a 10) y su nivel socioeconómico estandarizado (`nse`). La pregunta de negocio: ¿cuánto sube el puntaje por hora de estudio, cambia ese efecto entre escuelas y qué puntaje esperar en una escuela que todavía no está en los datos?

Datos

Genera los datos con esta función y la semilla 7. Así tus números serán comparables con los de referencia:

```
import numpy as np, pandas as pd
def make_escuelas(seed=7):
    rng = np.random.default_rng(seed)
    tam = rng.integers(5, 61, 30)
    u0, u1 = rng.normal(0, 6, 30), rng.normal(0, 0.8, 30)
    filas = []
    for e in range(30):
        n = tam[e]
        horas = rng.uniform(0, 10, n).round(1)
        nse = rng.normal(0, 1, n).round(2)
        y = (60 + u0[e] + (2.5 + u1[e]) * horas
             + 4 * nse + rng.normal(0, 8, n))
        y = np.clip(y, 0, 100).round(1)
        for h, z, v in zip(horas, nse, y):
            filas.append({"escuela": f"E{e+1:02d}",
                          "horas_estudio": float(h),
                          "nse": float(z), "puntaje": float(v)})
    df = pd.DataFrame(filas)
    df["escuela"] = pd.Categorical(df["escuela"])
    return df
```

Trata los parámetros del generador como desconocidos: solo sirven para que, al final, compruebes si tu análisis los recupera.

Preguntas guía

- ¿Qué proporción de la variación del puntaje se debe a diferencias entre escuelas?
- ¿El efecto de `horas_estudio` es el mismo en todas las escuelas o varía?
- ¿Qué pasa con la estimación de una escuela con solo 5 alumnos frente a una con 60?
- ¿Qué tan incierta es la predicción para una escuela nueva comparada con una conocida?

Pasos

1. **EDA por escuela.** Tamaño, media y dispersión del puntaje por escuela; gráfica de puntaje contra horas para 6 escuelas. Revisa el rango de `puntaje`: hay 8 valores recortados en 100.
2. **Tres modelos**, ajustados con `draws=1000, tune=1000, chains=2, random_seed=42, inference_method="nutpie"`:
   - pooled: `puntaje ~ horas_estudio + nse`
   - intercepto aleatorio: `puntaje ~ horas_estudio + nse + (1|escuela)`
   - pendiente aleatoria: `puntaje ~ horas_estudio + nse + (horas_estudio|escuela)`
3. **Diagnósticos.** R-hat máximo, ESS bulk mínimo y divergencias de cada modelo, sobre todas las variables. Si algo falla, reajusta con más draws o cadenas y documenta qué cambió.
4. **PPC** con `predict(kind="response")` y `az.plot_ppc`. Comenta qué fracción de los puntajes simulados supera 100.
5. **LOO.** Llama a `compute_log_likelihood` en cada modelo y compara con `az.compare`. Reporta `elpd_diff`, `dse` y el máximo de Pareto $\hat{k}$.
6. **Interpretación.** HDI de los efectos comunes, $P(\beta_{horas} > 2 \mid y)$, coeficiente de correlación intraclase del modelo de intercepto aleatorio, $\text{ICC} = \sigma^2_{escuela} / (\sigma^2_{escuela} + \sigma^2)$, y shrinkage de la escuela más pequeña.
7. **Predicción** para una escuela nueva (`E99`), un alumno con 5 horas y `nse = 0`, con `sample_new_groups=True`. Distingue el HDI de μ del HDI predictivo.

Valores de referencia

Úsalos para comprobar tu código, no como conclusiones:

- El modelo de pendiente aleatoria estima `horas_estudio` = 2.27 (HDI 94% de 1.95 a 2.61), `nse` = 3.59, `sigma` = 7.91 y `horas_estudio|escuela_sigma` = 0.78. Sin divergencias.
- `az.compare`: pendiente aleatoria con elpd_loo = −3609.4; intercepto aleatorio a 34.9 puntos (dse 8.2); pooled a 227.1 puntos (dse 19.6). Ningún Pareto $\hat{k}$ supera 0.7.

- Con 2 cadenas y 1000 draws, el modelo de intercepto aleatorio da R-hat = 1.02 y ESS bulk = 146 en `Intercept`. Este aviso es intencional: decide cómo resolverlo antes de interpretar ese modelo.

Entregables

- Un notebook que corra de principio a fin tras reiniciar el kernel.
- Una tabla de diagnósticos (R-hat, ESS, divergencias) y la tabla de `az.compare`.
- Gráficas: PPC del modelo elegido, `az.plot_forest` de los efectos por escuela y predicciones con `bmb.interpret.plot_predictions` para tres escuelas.
- Un informe de 5 a 8 frases que responda las preguntas guía, cada una con un número y su intervalo, más una frase sobre limitaciones.

Criterios de evaluación

- **Especificación:** fórmulas correctas y priors revisados con `print(modelo)` o con un prior predictive check.
- **Diagnóstico:** ningún resultado se interpreta sin R-hat < 1.01, ESS suficiente y cero divergencias.
- **Comparación:** la elección del modelo se justifica con `elpd_diff` en relación con `dse`, no solo con el rank.
- **Interpretación:** intervalos de credibilidad leídos como probabilidades sobre parámetros y efectos de grupo explicados con shrinkage.
- **Predicción:** el sujeto nuevo usa `sample_new_groups=True` y se reporta la incertidumbre predictiva, no solo la media.

Errores comunes

- Olvidar `compute_log_likelihood` con nutpie y obtener un error en `az.compare`.
- Comparar la media cruda de la escuela más pequeña con su intercepto del modelo: el intercepto corresponde a `horas_estudio = 0` y `nse = 0`.
- Ignorar el recorte en 100: un modelo gaussiano puede simular puntajes imposibles.

Para ir más allá

Agrega una pendiente aleatoria para `nse` y compara con LOO si mejora la predicción.