Assignment 3

Brian Sohn

The distribution of 'number of successful impressions' follows a binomial distribution *Bin(n,p)*, with the following parameters:

n = number of ads served

p (probability of successful impression) = $\frac{1}{1 + exp(-p_0)}$

$p_0 \sim N(X\beta, 1)$, $\beta \sim N(0,1)$: priors

X is a matrix that contains the bias term and dummy variables created by gender and age_grp. Now we will estimate β and p.

This model allows us to use demographic information to estimate probability of success, but also allows for variation among individuals with same demographics according to observed behaviors.

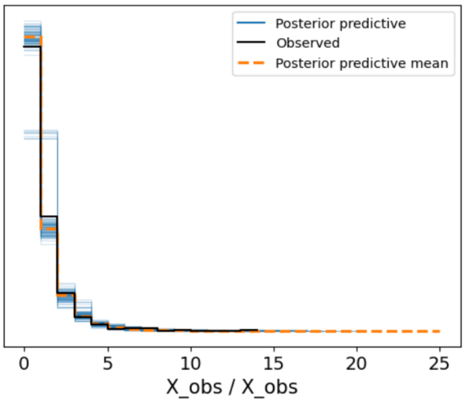
We can code up this model using Pymc as follows:

```
model = pm.Model()

with model:
    beta = pm.Normal('beta', mu=0, sigma=1, size=7)
    p_0 = pm.Normal('p_0', mu=pm.math.dot(df.iloc[:,:7], beta), sigma=1)
    p = pm.Deterministic('p', 1 / (1 + pm.math.exp(-1 * p_0)))

    X_obs = pm.Binomial('X_obs', n=df['number_of_ads_served'], p=p, observed=df['number_of_sucessful_impressions'])

    idata=pm.sample()
    pm.sample_posterior_predictive(idata,extend_inferencedata=True)

```

And we get the following coefficients (β) as a result. The posterior predictive looks close to the actual data's distribution, so our coefficients should be trustworthy.

	Coefficient
constant	-3.33
gender	1.06
age_grp_1	0.23
age_grp_2	0.55
age_grp_3	0.66
age_grp_4	0.92
age_grp_5	0.60

	Probability of Success
0	0.080
1	0.107
2	0.064
3	0.050
4	0.065
...	...
995	0.060
996	0.061
997	0.211
998	0.171
999	0.152



For an individual in the data, we can directly access his/her probability of successful impression by idata.posterior.p since we used pm.Deterministic. For a new individual, we can calculate the expected probability of successful impression using the coefficients of β.