

DSC 530 - Week 9 Assignment

Name: Barbara Gaskins

Exercises 11-3 and 11-4

Exercise 11-3: Poisson Regression

Predicting the **number of children a woman has** based on demographic factors.

```
In [ ]: import nsfg
import pandas as pd
import statsmodels.formula.api as smf

# Load and prepare data
live = live[live.prglngth > 30]
resp = nsfg.ReadFemResp()
resp.index = resp.caseid
join = live.join(resp, on='caseid', rsuffix='_r')

join['age2'] = join.age_r**2

# Define Poisson regression model
formula = 'numbabes ~ age_r + age2 + C(race) + totincr + educat'
model = smf.poisson(formula, data=join)
results = model.fit()
print(results.summary())

# Prediction for a woman who is:
# - 35 years old
# - Black
# - College graduate
# - Household income > $75,000
columns = ['age_r', 'age2', 'race', 'totincr', 'educat']
prediction_df = pd.DataFrame([[35, 35**2, 1, 14, 16]], columns=columns)

# Predict number of children
predicted_children = results.predict(prediction_df)
print("Predicted number of children:", predicted_children)
```

Exercise 11-4: Multinomial Logistic Regression

Predicting a **woman's marital status** based on demographic factors.

```
In [ ]: import statsmodels.formula.api as smf

# Define multinomial logistic regression model
model = smf.mnlogit('rmarital ~ age_r + C(race) + totincr + educat', data=
results = model.fit()
print(results.summary())

# Prediction for a woman who is:
# - 25 years old
# - White
# - High school graduate
# - Household income of ~$45,000
columns = ['age_r', 'race', 'totincr', 'educat']
new = pd.DataFrame([[25, 2, 11, 12]], columns=columns)

# Predict probabilities for marital status categories
predicted_marital_status = results.predict(new)
print("Predicted probabilities for marital status:
", predicted_marital_status)
```