exams

February 17, 2020

[1]:
```
import pandas as pd
import numpy as np
import statsmodels.api as sm
```

```
/nfs/kshedden/python3/lib/python3.7/site-
packages/statsmodels/compat/pandas.py:23: FutureWarning: The Panel class is
removed from pandas. Accessing it from the top-level namespace will also be
removed in the next version
  data_klasses = (pandas.Series, pandas.DataFrame, pandas.Panel)
```

Exam scores data from this page: http://www.bristol.ac.uk/cmm/learning/support/datasets/

[2]:
```
colspecs = [(0, 5), (6, 10), (11, 12), (13, 16), (17, 20)]
```

[3]:
```
df = pd.read_fwf("../data/exam_scores/SCI.DAT", colspecs=colspecs, header=None)
df.columns = ["schoolid", "subjectid", "gender", "score1", "score2"]
df["female"] = 1*(df.gender == 1)
df = df.dropna()
```

[4]:
```
# A school-clustered model for exam score 1 with no correlation.
model1 = sm.GEE.from_formula("score1 ~ female", groups="schoolid", data=df)
rslt1 = model1.fit()
```

[5]:
```
# A school-clustered model for exam score 1 with exchangeable correlations.
model2 = sm.GEE.from_formula("score1 ~ female", groups="schoolid",
                             cov_struct=sm.cov_struct.Exchangeable(), data=df)
rslt2 = model2.fit()
```

[6]:
```
# A subject-clustered model for exam score 1 with exchangeable correlations.
model3 = sm.GEE.from_formula("score1 ~ female", groups="subjectid",
                             cov_struct=sm.cov_struct.Exchangeable(), data=df)
rslt3 = model3.fit()
```

[7]:
```
# Prepare to do a joint analysis of the two scores.
dx = pd.melt(df, id_vars=["subjectid", "schoolid", "female"],
             value_vars=["score1", "score2"], var_name="test",
             value_name="score")
```

[8]:
```
# A nested model for subjects within schools, having two scores per subject.
model3 = sm.GEE.from_formula("score ~ female + test", groups="schoolid",␣
 ↪dep_data="0 + subjectid",
```

```
                                cov_struct=sm.cov_struct.Nested(), data=dx)
rslt3 = model3.fit()
```