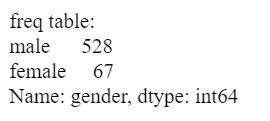
* **Frequency table in pandas**

table1\_pd = PSID1982['gender']**.value\_counts()**

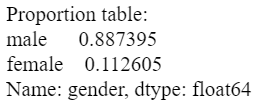
#Print frequency table

print(f'freq table: \n{table1\_pd}\n')



#Print proportion table

print(f'Proportion table: \n{**table1\_pd/table1\_pd.sum()**}\n')

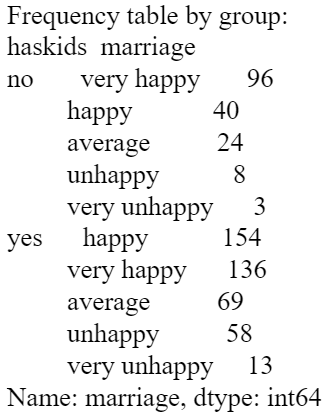
****

# frequency table with groupby

ft\_pd2 = affairs['marriage'].**groupby**(affairs['haskids']).**value\_counts()**

print(f'Frequency table by group: \n{ft\_pd2}\n')

Output:

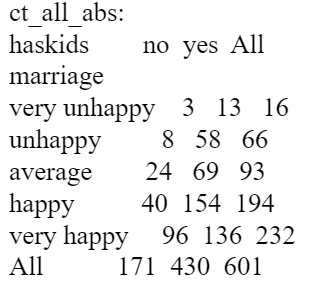
****

* **Contingency Table**

# **contingency table** in pandas:

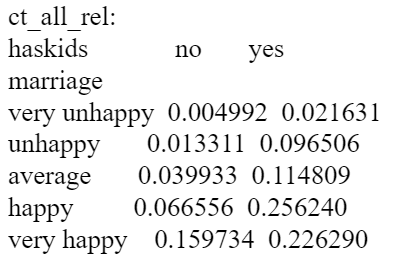
ct\_all\_abs = pd.**crosstab**(affairs['marriage'], affairs['haskids'], margins = 3)

print(f'ct\_all\_abs: \n{ct\_all\_abs}\n')



ct\_all\_rel = pd.crosstab(affairs['marriage'], affairs['haskids'], **normalize = "all"**)

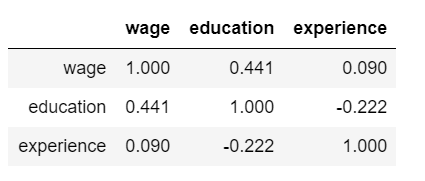
print(f'ct\_all\_rel: \n{ct\_all\_rel}\n')



* **Correlation Table** between wages, education, and experience

cormat = PSID1982[["wage", "education", "experience"]]

cormat.corr().round(decimals=3)



* Correlation Graphs

pd.plotting.scatter\_matrix(cormat, alpha=0.5, figsize=(10, 8));

* Pivot Table

#Calculate the averages of each variable, grouped by gender

bygender = PSID1982.**pivot\_table**(**values=**["wage", "experience", "education"], **index**="gender", **aggfunc**=[np.mean])

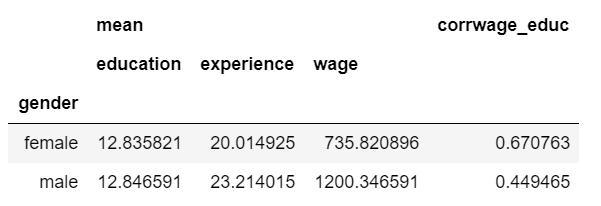
#Create correlation coefficient between wage and education, grouped by gender

bla = PSID1982.groupby(['gender'])[['wage','education']].corr().iloc[0::2,-1].values[0::1]

#Add a column

bygender['corrwage\_educ'] = bla

Bygender



# **summary information**

# remember using [[]] double bracket returns the results as a dataframe format

des = ceosall[['salary']].**describe()**

**# sample average:**

meansal = round(ceosall[['salary']]**.mean()**, 2)

# **correlation** with ROE (2 ways)

1. roe\_corr = np.corrcoef(ceosall['salary'], ceosall['roe'])
2. corr = ceosall['salary'].corr(ceo['reo'])

* **Different way to present stats**

table1 = ceosall[['salary']]**.agg**(['mean', 'median', 'std'])

# This can also calculate correlation coefficient

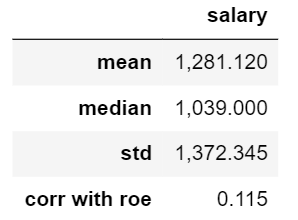
corr = ceosall['salary'].corr(ceosall['roe'])

newrow = pd.Series(data = {'salary': corr}, name = 'corr with roe')

table1 = table1**.append**(newrow, ignore\_index = False)

# present a neat table

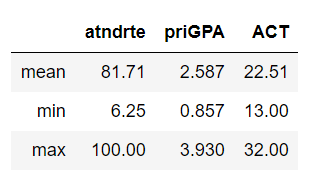
table1.style.format('{:,.3f}' .format)



# Obtain the minimum, maximum, and average values for the variables atndrte, priGPA, and ACT using two ways:

(1)attend = woo.dataWoo('attend')

tablestats = attend[["atndrte", "priGPA", "ACT"]].***agg(["mean",min,max])***



(2)attend[["atndrte", "priGPA", "ACT"]]***.describe()***.round(3)

