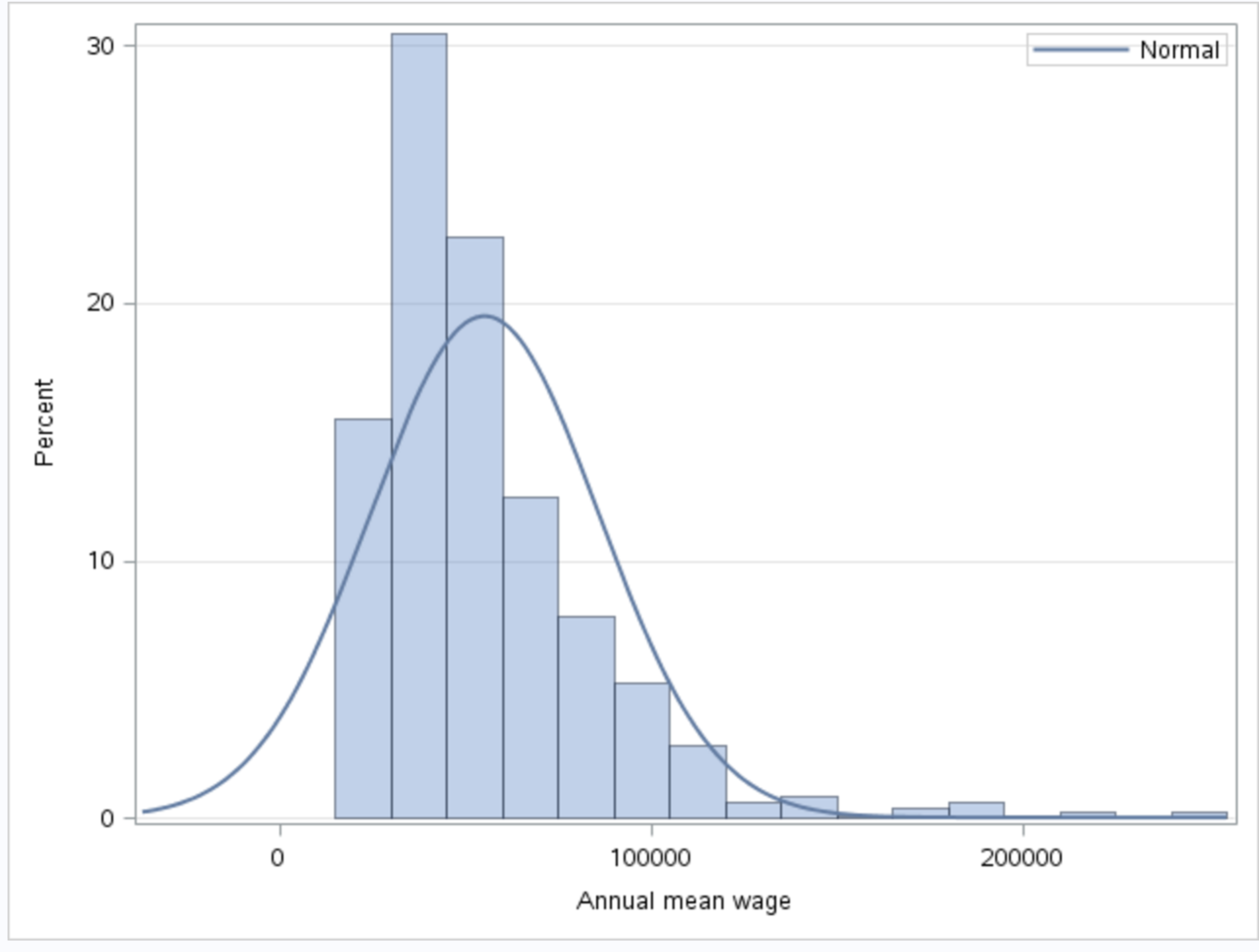
Analysis of Occupation and Pay dataset.

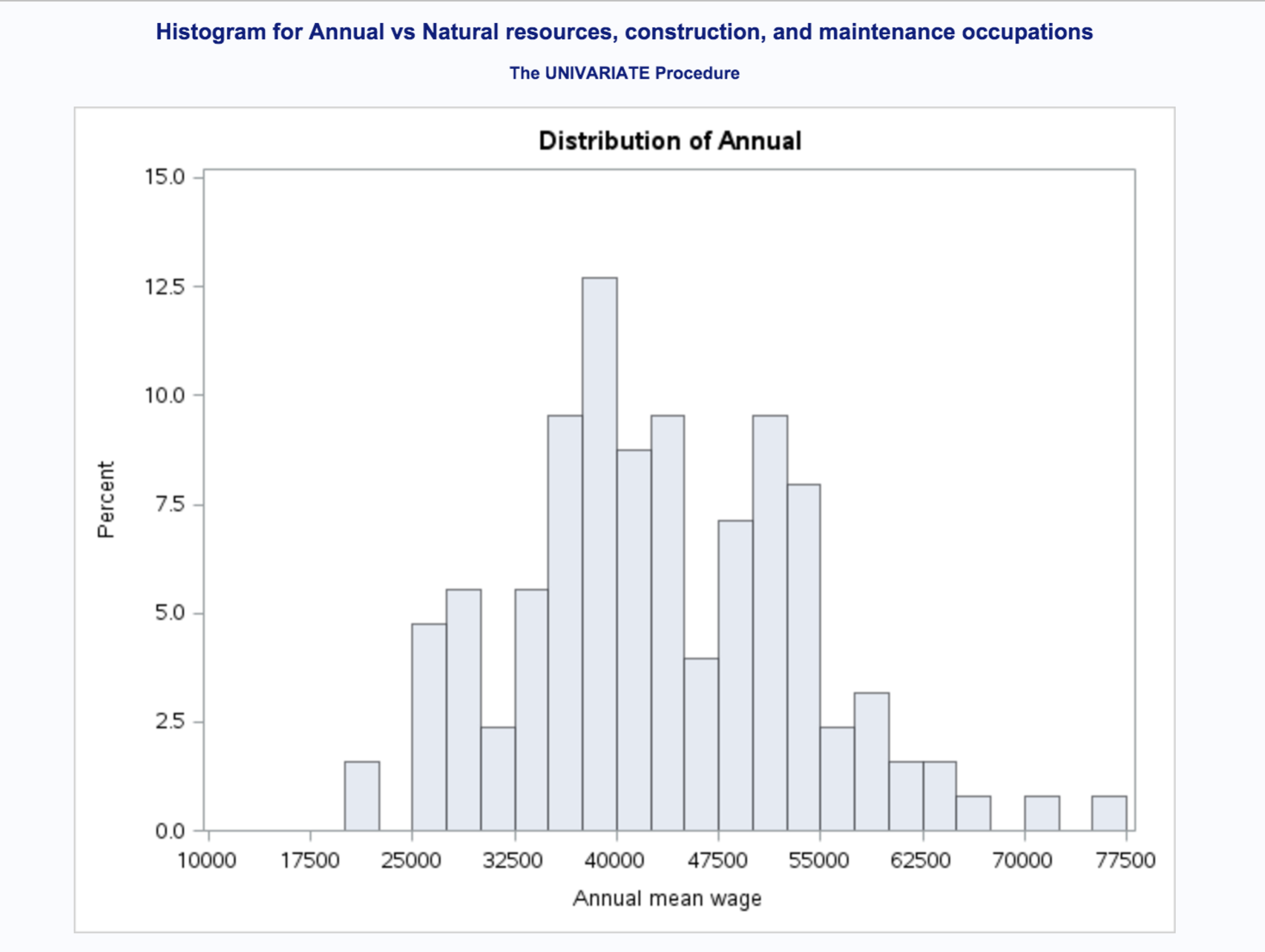
Occupation and Pay dataset does contain 840 observations and 4 variables, no missing data present in the dataset.

Following Histogram does provide information about Annual mean wage and we detect that is not distributed uniformly the data. Looking at percentage distribution of Annual age is mostly concentrated to left meaning is right skewed, basically almost all data is incorporated till 100000. We see few wages between 100000 to 200000.



Histogram for Annual vs Natural resources, construction, and maintenance occupations

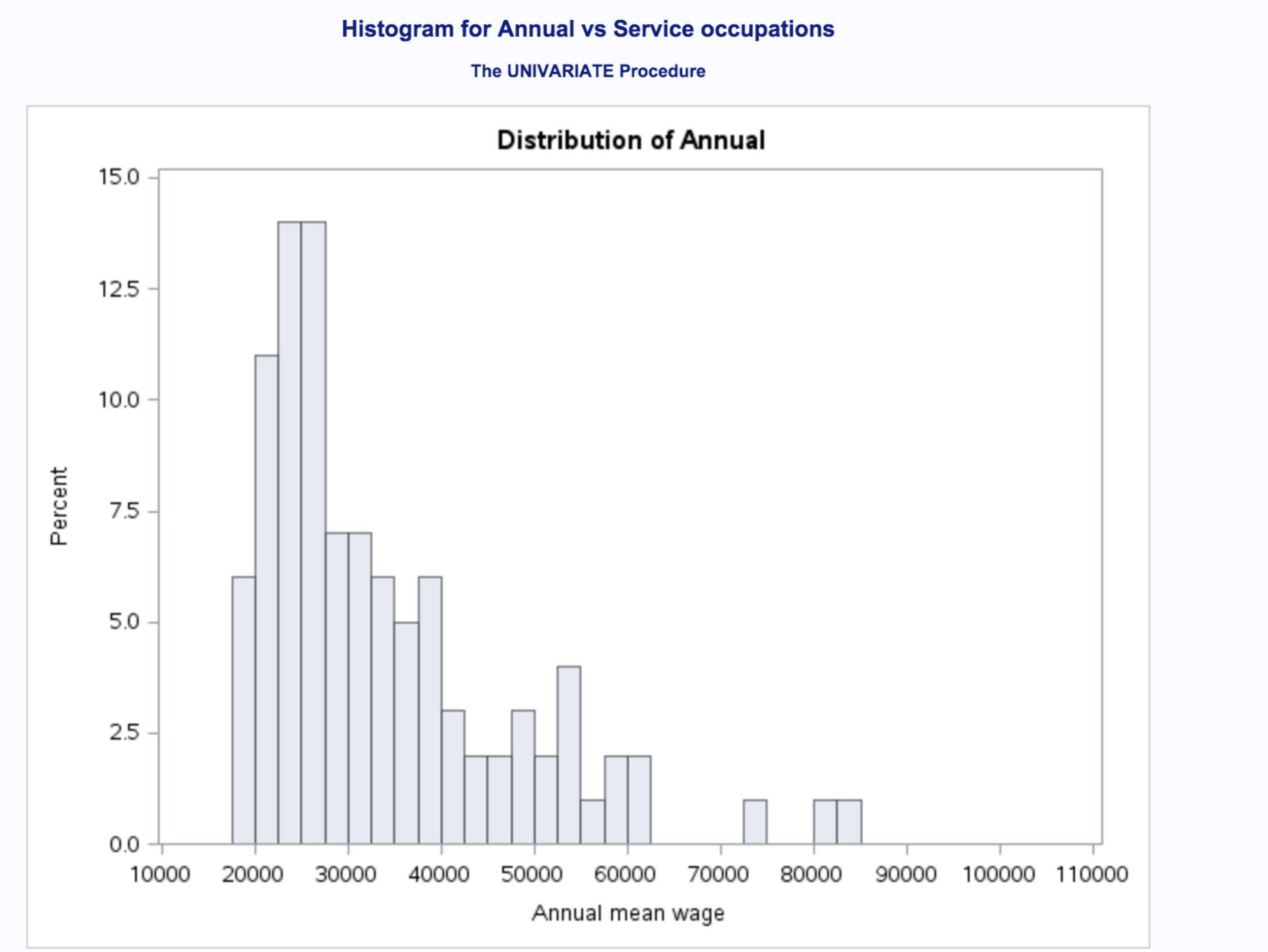
We do see uniform distribution of annual mean wage for Natural resources, construction, and maintenance occupations.



Histogram for Annual vs Natural resources, construction, and maintenance occupations

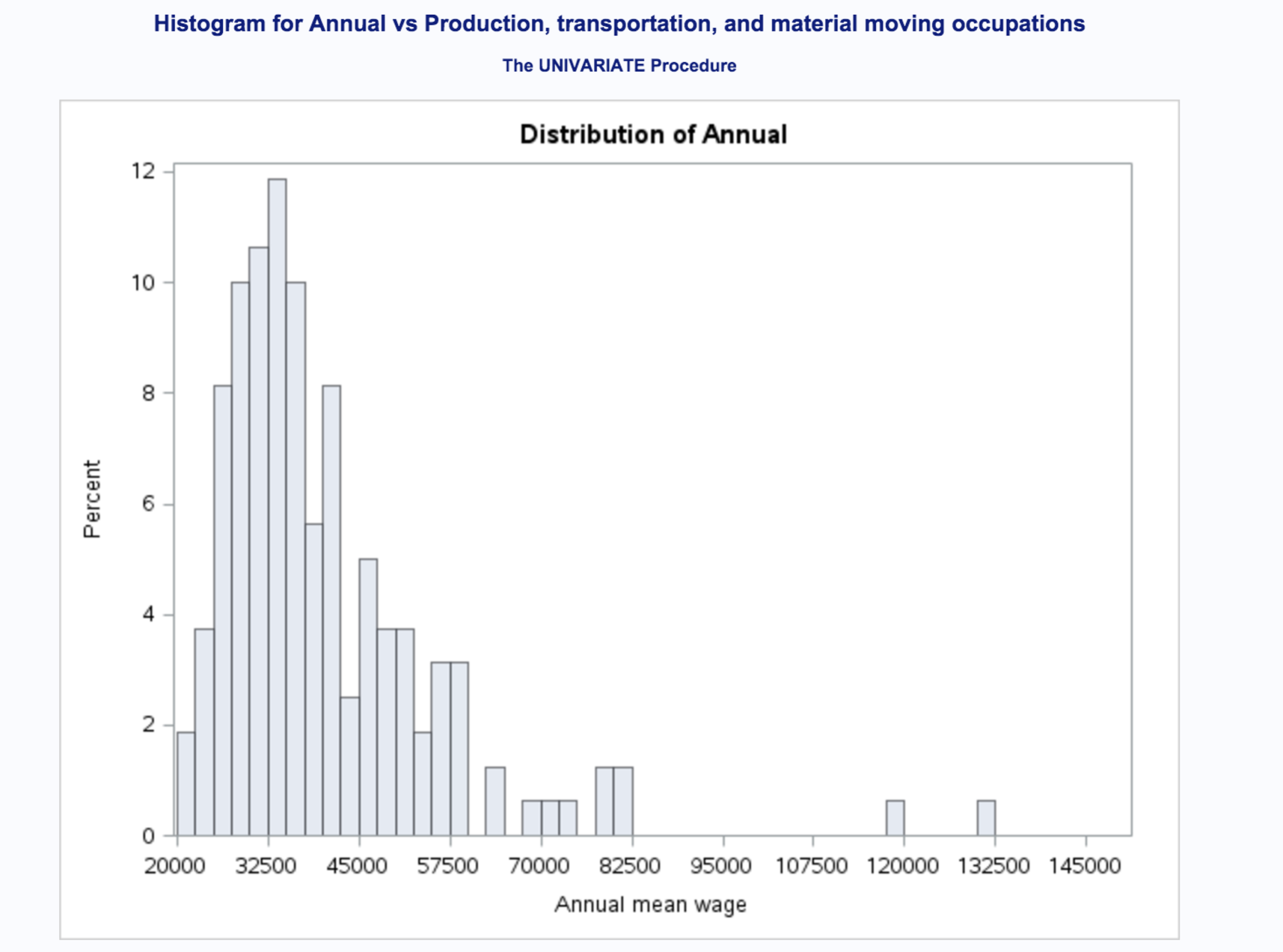
Is the only variable basically what has a normal distribution of Annual mean wage for Natural resources , construction and maintenance occupation. Most of the data is concentrated between 25000 and 62500 but more of the wage are 35000 and 55000. No missing data detected and not need for imputation or data transformation.

Histogram for Annual vs Service occupations does show distribution of the Annual mean wage not normal.



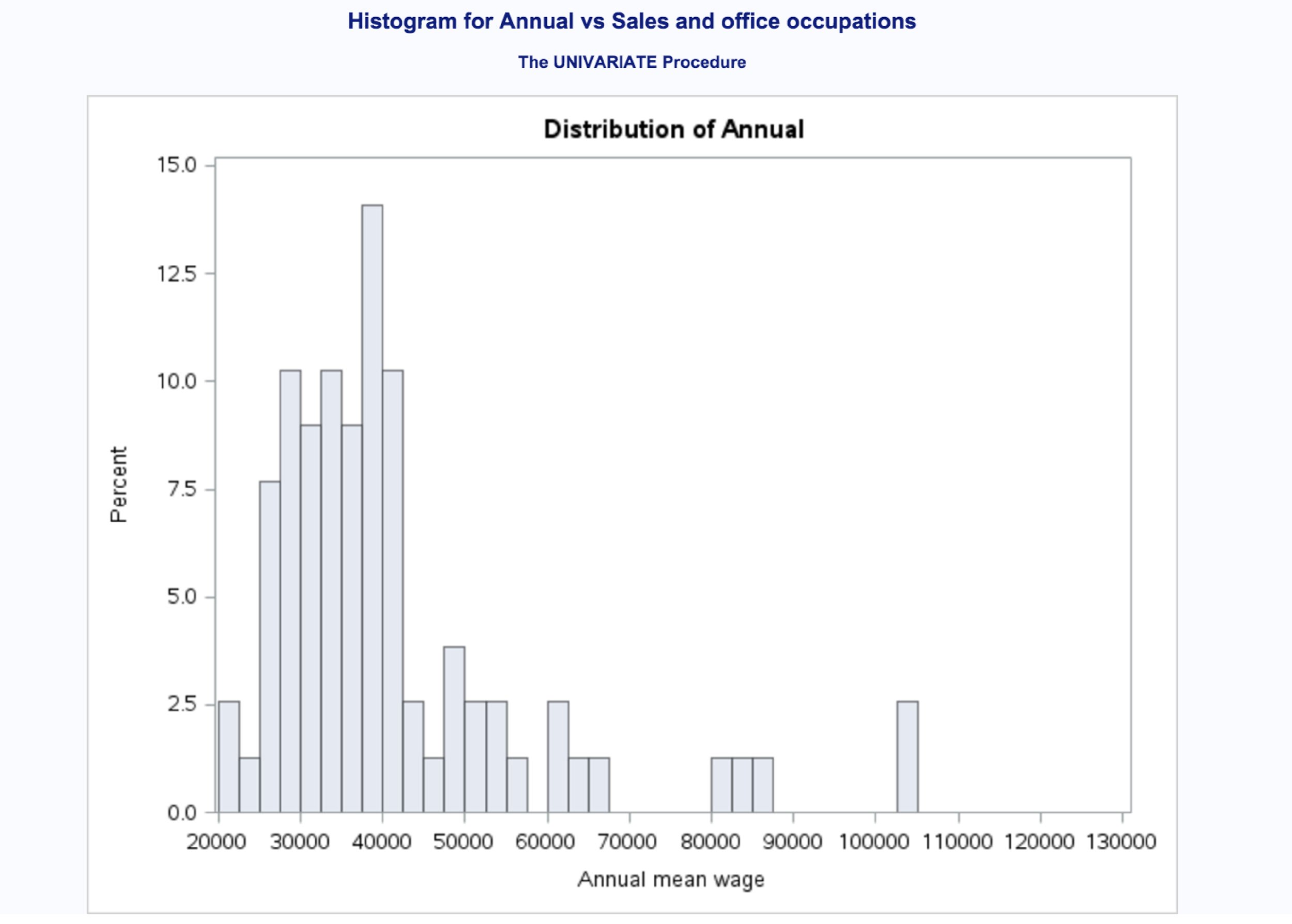
Histogram for Annual wages vs Service occupations are mostly distributed on left, meaning is skewed to the right. Looking at distribution we detect that most data is explained between 20000 and 60000, we have some outliers what can be transformed and create a normal distribution in case of using variable in model and to focus on employers what Annual wage maximum is 60000.

Production, transport, and material moving occupations does point us that distribution of Annual mean wage is not normal distributed.



Histogram for Annual vs Production, transportation and material moving occupation is pointing us to conclusion that data is mostly distributed on the left and skewed on right. We do see some outliers and can transform data located between 107500 to 145000 by imputing with median and create a normal distribution. Most of the data is explained from 20000 to 82500, we can consider that is explained at about 98 % of the data.

Histogram for Sales and office occupations does not have normal distribution of the Annual mean wage.



Histogram for Sales and office occupations also is mostly concentrated on the left y being skewed on right. Most data about Annual wages is explained between 20000 to 65000. Outliers between 65000 to 130000 can be transformed by imputing with the median to can have a normal distribution and can fit a model in predicting salary grow over time.