
```

data poptot popnorko popsouko;

infile '/home/u63324141/sasuser.v94/onecolumndata.txt';
do year=1998,2004,2010,2016,2022;
* Set the variable "year" to represent the 5 survey years;

do state='Baden-Württemberg','Bayern','Berlin','Brandenburg','Bremen','Hamburg','Hessen',
'Mecklenburg-Vorpommern','Niedersachsen','Nordrhein-Westfalen','Rheinland-Pfalz','Saarland',
'Sachsen','Sachsen-Anhalt','Schleswig-Holstein','Thüringen';
* Set the variable "state" to represent the 16 states in Germany;

do nationality='north-k','south-k';
* Set the variable "nationality" to represent the nationality of origin (North or South);

do gender='male','female';
* Set the variable "gender" to represent the gender;

if state in ('Brandenburg','Mecklenburg-Vorpommern','Sachsen','Sachsen-Anhalt','Thüringen')
then sphere='Eastgermany';
else if state='Berlin' then sphere='Berlin';
else sphere='Westgermany';
* Include the five states, including Brandenburg, under the variable "sphere" as 'East Germany.'
Save Berlin as 'Berlin' under "sphere."
Save the remaining 10 states under "sphere" as 'West Germany.';

input lessthan1 from1to4 from4to10 from10to25 over25 @@;
count=lessthan1+from1to4+from4to10+from10to25+over25;
shortstay=lessthan1+from1to4;
* The population was categorized into five groups based on the duration of residence:
less than 1 year, 1 year or more but less than 4 years, 4 years or more but less than 10 years, 10 years or more but less
The total population for each category was stored in variables named "count."
Additionally, the population with a residence duration of less than 4 years was stored in a variable named "shortstay.";

if nationality='north-k' then output poptot popnorko;
else output poptot popsouko;
* The population from North Korea is stored in the variable "popnorko,"
the population from South Korea is stored in "popsouko,"
and the total combined population (North + South) is stored in a dataset named "poptotal.";
end;
end;
end;
end;
run;

```

```

proc means data=poptot mean std clm alpha=0.05;
class nationality year;
var count;
run;
* The "means procedure" is used to calculate the mean, standard deviation, and a 95% confidence interval for the variable
This estimation is performed with respect to different nationalities
and survey years and outputs the estimated values for each combination of these two variables.;

```

```

proc freq data=popsouko order=data;
weight count;
tables gender/binomial (p=0.422) alpha=0.05;
run;
* The test using the freq procedure under the South Korean population dataset is used to test
whether the proportion of males in the overall population is 42.2%.
Additionally, a 95% confidence interval for the population proportion is calculated.;

```

```

proc ttest data=popnorko h0=5;
var over25;
run;
* A test is conducted using the North Korean population dataset to check whether
the mean population of individuals who have resided for 25 years or more is at least 5.;

```

```

proc ttest data=poptot;
class nationality;
var count;
run;
* A t-test is conducted to compare the population mean of counts for two independent samples:
the South Korean population and the North Korean population.;

```

```

proc freq data=popsouko order=data;
weight count;
tables state*gender/nocol nopercnt expected chisq measures;
run;
* A chi-squared test of independence is conducted to assess the relationship between
"state" (region) and "gender" in the South Korean population dataset.
The null hypothesis (H0) is that state and gender are independent.
Additionally, the "measures" command provides a measure of association or correlation between these variables,

```

which can help interpret the relationship between them.;

```
proc freq data=popnorko;
weight shortstay;
tables year/nocum testp=(0.24 0.50 0.10 0.14 0.02);
run;
```

* A chi-squared goodness-of-fit test is conducted using the "year" and "shortstay" variables in the North Korean population dataset. The null hypothesis (H0) states that the proportions of short-stay individuals in each year category are 24%, 50%, 10%, 14%, and 2%, respectively. This test assesses whether the observed proportions differ significantly from the expected proportions under the null hypothesis.;

```
proc glm data=popnorko;
class sphere;
model count=sphere;
means sphere/lines;
means sphere/hovtest=bartlett;
contrast 'east vs west' sphere 0 1 -1;
run;
```

* A one-way analysis of variance (ANOVA) is conducted using the North Korean population dataset. Specifically, the ANOVA examines whether there are significant differences in population counts based on the "sphere" variable which categorizes individuals into East Germany, West Germany, or Berlin based on their place of residence. The test assesses whether there are statistically significant differences in population counts among these categories.;

```
proc glm data=popsouko;
class gender sphere;
model count=gender sphere gender*sphere;
means gender sphere gender*sphere;
run;
```

* A two-way analysis of variance (ANOVA) is performed using the South Korean population dataset. This analysis evaluates the impact of two independent variables, "sphere" (place of residence) and "gender" (male or female). It assesses whether there are significant interactions between these variables and their effects on population counts.;

```
proc reg data=popnorko;
model shortstay=year/p clm cli dw;
plot shortstay*year;
run;
```

* A simple linear regression analysis is conducted on the North Korean population dataset. In this analysis:
Dependent variable (Y): "shortstay" (the number of short-term residents).
Independent variable (X): "years" (number of years of residence).;

```
proc reg data=popsouko;
model count=lessthan1 from1to4 from4to10 from10to25 over25/stb
selection=stepwise slstay=0.10 slentry=0.10;
run;
```