

6.12 Exercise: Inference with iNZight (*Lite version*)

This exercise will enable you to use iNZight to put intervals around the estimates of a population mean and proportions to allow for sampling error and compare groups.

The skills addressed are:

- Use iNZight Lite to get confidence interval for estimates of the population mean and make comparisons across sub-groups.
- Use iNZight Lite to build confidence intervals around estimates of population proportions and make comparisons between groups.

INSTRUCTIONS

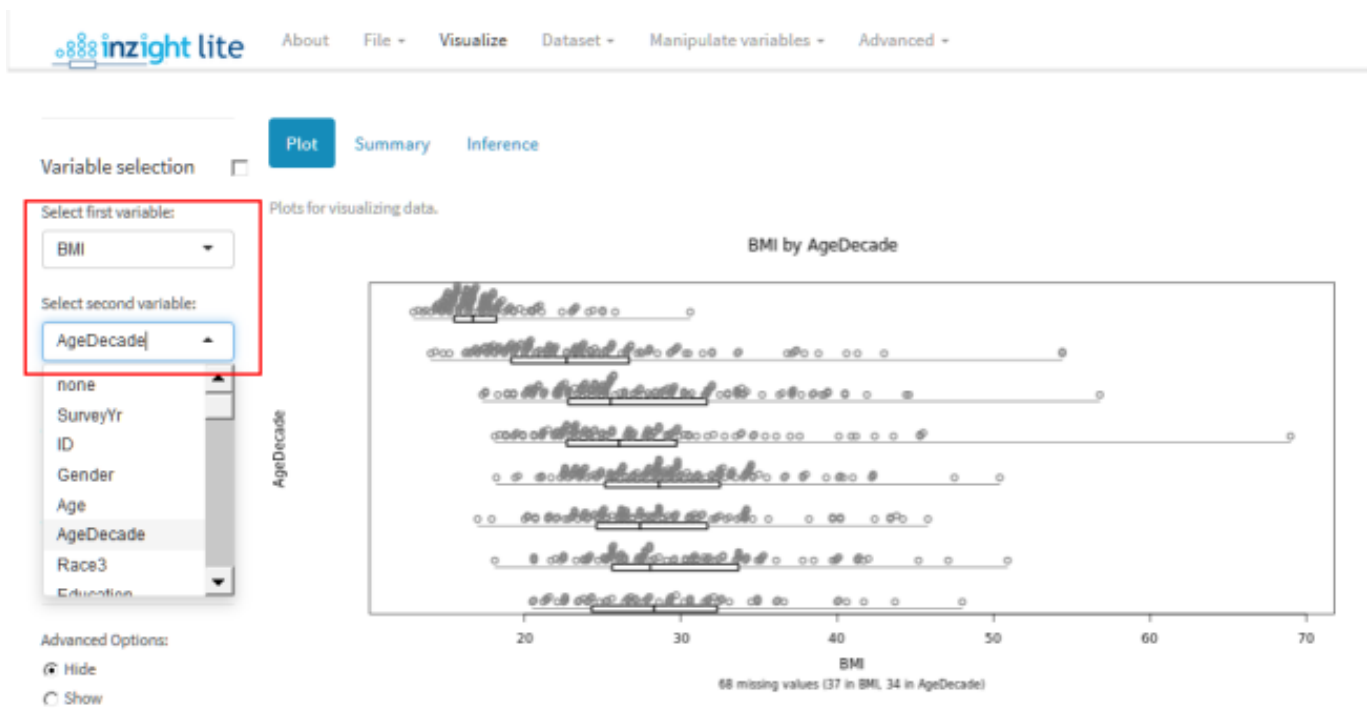
Start iNZight Lite and import the **NHANES-1000** dataset into iNZight using **File > Dataset Examples**. Select the **NHANES-1000** dataset from Data set Category **Future-Learn**.

If you have any problems during this exercise, see the **Common Questions** on the last page.

Construct confidence intervals for sub-groups of a numeric outcome

We are revisiting **BMI** (Body Mass Index) for different age groups and using the **NHANES-1000** data, a sample of the American population. We will use our sample to estimate the mean BMI for different age groups. To do this we need to construct intervals around our estimates in order to allow for sampling error.

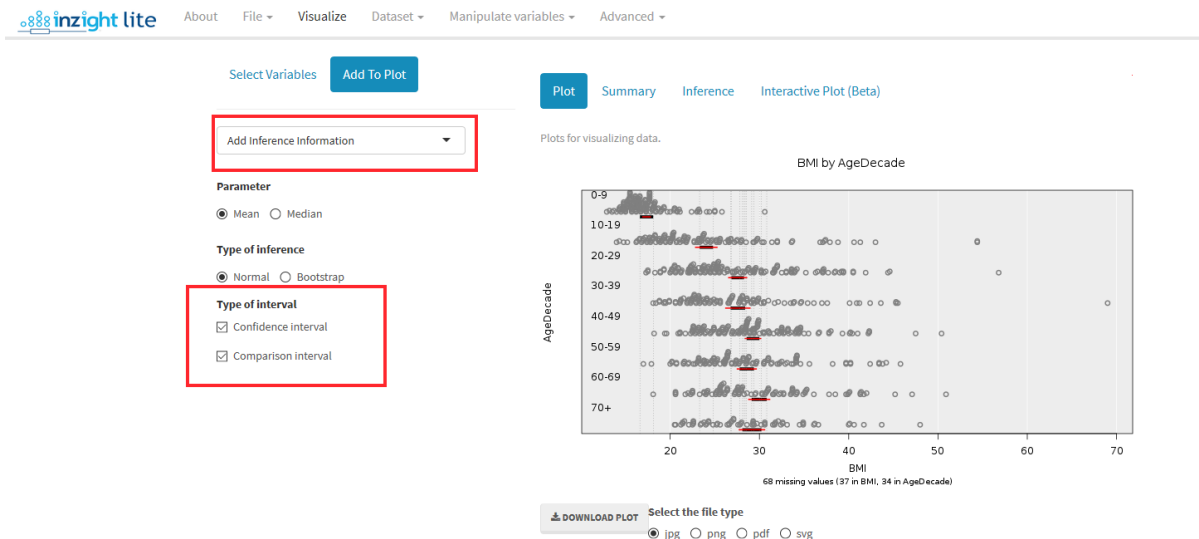
- Select **BMI** in the **first variable** slot
- Select **AgeDecade** in the **second variable** slot.



You should have a series of dot plots of BMI for different age groups in your plot window.

To get confidence and comparison intervals:

- On the **Add to Plot** tab select **Add Inference Information**
- Under **Type of Interval** select both **Confidence** and **Comparison Intervals**



What do you see here? Remember that the blue lines are called 'comparison intervals' and are lines that we look at when observing any overlap.

Post a comment if you have any interesting observations about the estimation of the population means for each age group and the differences between them.

For the actual (numerical) confidence intervals for the group means/medians and also confidence intervals for the differences between them:

- click the **Inference** tab towards the top right of the page.

Clicking the **Summary** tab will show you sets of summary statistics for each group.

Variable selection

Select first variable:
BMI

Select second variable:
AgeDecade

Subset by:
none

Subset by:
none

Advanced Options:
 Hide
 Show

RESET ALL

Plot Summary **Inference**

Select type of inference
normal

Statistical Inference for the data.

```
=====
                          inZight Inference using Normal Theory
-----
Primary variable of interest: BMI (numeric)
Secondary variable: AgeDecade (factor)

Total number of observations: 1000
Number omitted due to missingness: 68 (37 in BMI, 34 in AgeDecade)
Total number of observations used: 932
=====

Inference of BMI by AgeDecade:
-----

Group Means with 95% Confidence Intervals

      Lower   Mean   Upper
0-9    16.86   17.36   17.85
10-19  22.78   24.05   25.31
20-29  26.49   27.57   28.64
30-39  26.16   27.58   29.01
```

and more ...

EXPLORE (~5 min)

Find another numeric variable in the NHANES-1000 dataset that you might like to explore across age groups, e.g. **Height**, and move it to the **first variable** slot. Use iNZight Lite to explore the estimates for the American population and build intervals around your estimates.

Post a comment if you see anything interesting.

Confidence and comparison intervals for proportions

Now we'll use the NHANES-1000 dataset to form intervals around estimates of how people in the American population rate their general health at the time of the survey (to convey uncertainty). Are there any differences between age groups?

First, filter out people below the age of 20:

- Click **Dataset > Filter Dataset**
- Select levels of a categorical variable
- Filter data by: Select **AgeDecade**
- Select all of the age groups below the age of 20 to remove the unwanted levels.
- Click **Perform Operation**

The screenshot shows the Inzight Lite interface. On the left, a 'Filter Dataset' dialog box is open, showing the following settings:

- Select Filter to apply: levels of categorical variable
- Select a categorical variable to filter the data on: AgeDecade
- Select levels to remove from the data: 0-9, 10-19

On the right, a summary of the data is displayed:

Number of rows in data: 1000
Number of columns in data: 41

SurveyYr:
2011_12
1000

ID
Min. 1st Qu. Median Mean 3rd Qu. Max.
62170 64680 67100 67060 69540 71910

Gender
female male
498 508

Age
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.00 17.00 36.00 36.55 56.00 80.00

AgeDecade
0-9 10-19 20-29 30-39 40-49 50-59 60-69 70+ NA's
157 127 146 118 144 111 105 66 54

Races
Asian Black Hispanic Mexican Other white
63 120 70 30 17 632

Education
8thGrade 9_11thGrade CollegeGrad HighSchool SomeCollege NA's
45 79 229 133 229 285

MaritalStatus
Divorced LivePartner Married NeverMarried Separated widowed
73 43 381 161 14 44
NA's
284

HHIncome
0-4999 5000-9999 10000-14999 15000-19999 20000-24999 25000-34999
36 21 56 53 75 182
35000-44999 45000-54999 55000-64999 65000-74999 75000-99999 more 99999
37 68 53 60 116 204
NA's
79

HHIncomeId
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

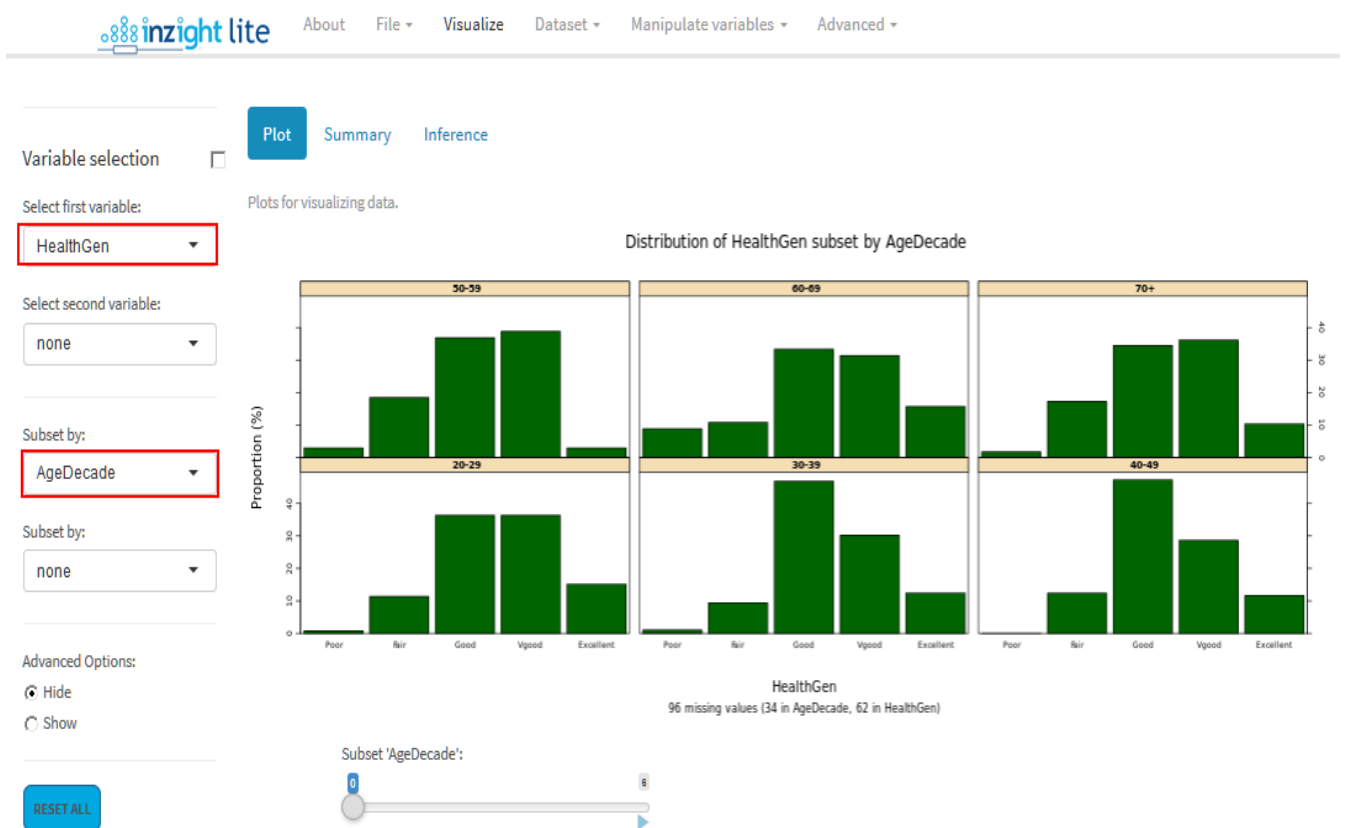
And

- Re-order the values for Health Gen setting from 1=Poor to 5=Excellent using **Manipulate Variables > Categorical Variables > Reorder Levels**.

Now construct the series of plots for **HealthGen** and **AgeDecade**:

- Select **HealthGen** in the first variable slot
- Select **AgeDecade** in the first subset by slot.

A series of graphs will appear in the plot window. It will help you see more of the detail if you enlarge your plot window.



To add confidence and comparison intervals to the graph:

- Click **Add to Plot** and select **Add Inference Information**
- Under **Type of Interval** select both **Confidence** and **Comparison Intervals** (they may still be selected from your work above).

Add Confidence Intervals and Comparison Intervals to the plot

Select Variables

Add To Plot

Plot

Summary

Inference

Interactive Plot (Beta)

Add Inference Information

Parameter

Proportions

Type of inference

Normal Bootstrap

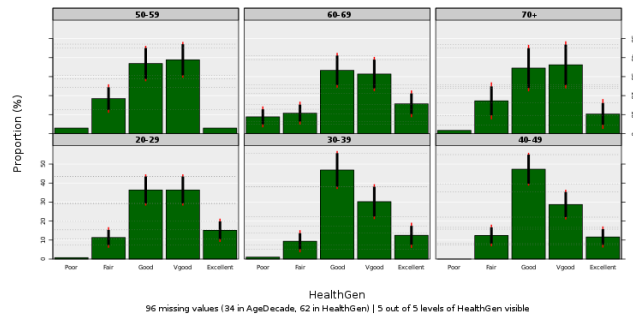
Type of interval

Confidence interval

Comparison interval

Plots for visualizing data.

Distribution of HealthGen subset by AgeDecade



DOWNLOAD PLOT

Select the file type

jpg png pdf svg

What does this graph tell you? What can you infer about true differences between the percentages in each health category?

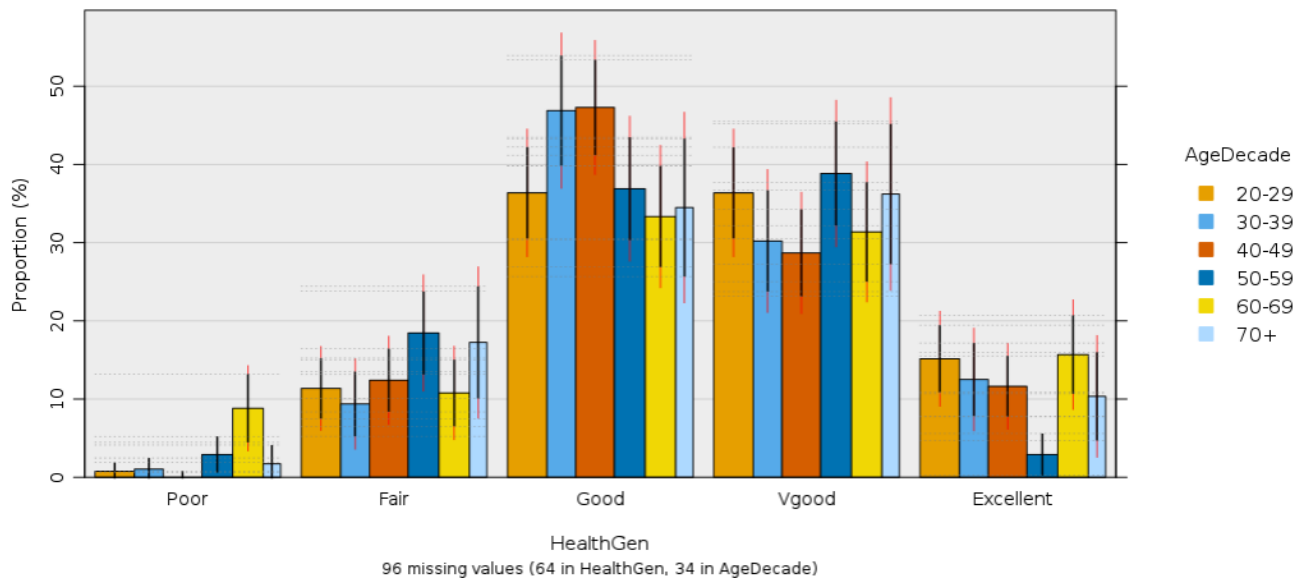
Select the **Inference** tab to see the actual confidence intervals around each proportion and for the differences between them.

Now construct the series of plots for **HealthGen** by **AgeDecade**:

- Select **HealthGen** in the **first variable** slot
- Select **AgeDecade** in the **second variable** slot.

A graph will appear in the plot window. (see over page).

Distribution of HealthGen by AgeDecade



If the intervals do not appear intervals use **Add Inference Information** as before.

What does this graph tell you about age-differences in the percentages for each health category (e.g. differences between age groups in the percentages saying they are in very good health)?

Use the **Inference** tab to see the exact numerical confidence intervals for differences between the true proportions.