DAE 7 (z-Test and t-Test)

Dr. Hendrik Purwins, Associate Professor
Dept. Architecture, Design and Media Technology, Aalborg University Copenhagen
A.C. Meyers Vænge 15, DK-2450 Copenhagen SV, Denmark

1. *(Singing performance of Chinese and American English native speakers)* Investigate, whether Chinese native speakers are better singers than American English native speakers. A group of Chinese (25) and American English native speakers (25) are tested with respect to their ability to sing tunes correctly. For each tune they get a performance score. The performance score is chosen so that an equal differences between performance scores corresponds to an equal differences in the actual singing ability. The average performance scores per person (`x_us, x_ch`) are saved in `ChineseEnglish.mat`.

   (a) Perform descriptive statistics on the data.
      i. Calculate sample mean and estimate the standard error each for Chinese and American English native speakers individually.
      ii. Plot the histograms for the performance scores of Chinese and American English native speakers in the same figure.

   (b) Check if the data is parametric.
      i. What is the level of measurement of the data?
      ii. Are the samples drawn from a normal distribution (use significance level $\alpha = 5\%$)? Plot the qq-plot and do an appropriate test.
      iii. Are the samples drawn from a distribution with the same variance (use significance level $\alpha = 5\%$)?

   (c) If the criteria are fulfilled to apply a parametric test, perform the right test to find out whether Chinese native speakers are better singing performers than American English native speakers.
      i. Formulate H_0 and H_1.
      ii. Do you use a one-sample, paired-sample, two-sample, or more-than-two sample test?
      iii. Do you use a dependent or a an independent test?
      iv. Do you use a two-tailed, right-tailed or left-tailed test?
      v. Perform the adequate test. Can you reject H_0 with significance level $\alpha = 0.01$?

   (d) Calculate the effect size. Is it a small, medium, or large effect?

   (e) Report the results as you would have to report them in your project report.

   (f) Plot the means with standard errors as error bars. (*Hint:* In Matlab, you can use `errorbar(y,e)` where `y` is plotted with error bars at y-e and y+e.)

2. A diet product is given to 25 subjects. The subjects' weights before and after the diet are saved as variable names `w_before, w_after` in `weights.mat`.

   (a) Perform descriptive statistics on the data.

      i. Calculate sample mean and the sample standard deviation for the weights before and after the diet.

      ii. Plot the histograms for the weight differences after - before the diet in the same figure.

   (b) Check if the data is parametric.

      i. What is the level of measurement of the data?

      ii. Are the samples drawn from a normal distribution (use significance level $\alpha = 5\%$)? Do an appropriate test.

      iii. Are the samples drawn from a distribution with the same variance (use significance level $\alpha = 5\%$)?

   (c) If the criteria are fulfilled to apply a parametric test, perform the right test to find out whether the diet product decreased the weight of the subjects.

      i. Formulate H_0 and H_1.

      ii. Do you use a one-sample, paired-sample, two-sample, or more-than-two sample test?

      iii. Do you use a dependent or a an independent test?

      iv. Do you use a two-tailed, right-tailed or left-tailed test?

      v. Perform the adequate test. Can you reject H_0 with significance level $\alpha = 0.0005$?

   (d) Report the results.

   (e) Plot the means with standard errors as error bars.

3. Is there an effect of nationality (Dutch or Danish) on the heights of males of age 20-25? Samples of 25 height measurements of Dutch and Danish males are stored as variables `x_du, x_da` in `Dutchdanish.mat`.

   (a) Perform descriptive statistics on the data.

      i. Calculate sample mean and estimate the standard error for Dutch and Danish heights individually.

      ii. Plot the histograms of the Dutch and the Danish heights in the same figure.

   (b) Check if data is parametric.

      i. What is the level of measurement of the data?

      ii. Are the samples drawn from a normal distribution (use significance level $\alpha = 5\%$)? Plot the qq-plot and do an appropriate test.

      iii. Are the samples drawn from a distribution with the same variance (use significance level $\alpha = 5\%$)?

   (c) If the criteria are fulfilled to apply a parametric test, perform the right test to find out whether Dutch or Danish males of age 20-25 differ in height on average.

      i. Formulate H_0 and H_1.

ii. Do you use a one-sample, paired-sample, two-sample, or more-than-two sample test?

iii. Do you use a dependent or a an independent test?

iv. Do you use a two-tailed, right-tailed or left-tailed test?

v. Perform the adequate test. Can you reject H_0 with significance level $\alpha = 0.05$?

(d) Calculate the effect size. Is it a small, medium, or large effect?

(e) Report the results.

(f) Plot the means with standard errors as error bars.

4. On the same 20 Med4 students, the following experiment is performed. You measure the galvanic skin response of subjects before and 3 s after watching a video. In Condition n, a video about the natural surroundings in Denmark is shown. In Condition t, a video about the terrorist attacks on Feb 14-15 2015 in Copenhagen is shown. For each subject, the conditions n and t are shuffled in their order of presentation. Now compare the measurement differences after and before watching the videos for n and for video t. `attack_data.csv` includes the measurements before ('nb') and after ('na') watching a video about Denmark's nature . 'tb','ta' are the measurements before and after watching a video about the terrorist attacks. Galvanic skin response measurements are given in kOhm. Find a test to answer the question: " Does the Galvanic skin response increase after watching the video on the terrorist attacks compared to a watching a video about Denmark's nature?" (*Hint:* How to read .csv data, a popular data format for tables, e.g. used by Excel? In Matlab: `M = csvread('FILENAME',R,C)`, where R is the first row and C the first column to start reading, where counting rows and columns starts with 0. First inspect `attack_data.csv` to see where the actual numbers start in the file. The columns in 'attack_data.csv' are to be interpreted as follows: 1 (nb), 2(na), 3(tb), 4(ta).)

(a) Perform descriptive statistics on the data.

i. Calculate sample mean and estimate the standard error for the differences in Galvanic skin response (after - before) of each video individually.

ii. Plot the histograms for the differences (after - before) of each video in the same figure.

(b) Check if data is parametric.

i. What is the level of measurement of the data?

ii. Are the samples drawn from a normal distribution (use significance level $\alpha = 5\%$)? Do an appropriate test.

iii. Are the samples drawn from a distribution with the same variance (use significance level $\alpha = 5\%$)?

(c) If the criteria are fulfilled to apply a parametric test, perform the right test to find out whether Galvanic skin response increases more after watching the video on the terrorist attacks compared to watching a video about Denmark's nature.

i. Formulate H_0 and H_1.

ii. Do you use a one-sample, paired-sample, two-sample, or more-than-two sample test?

iii. Do you use a dependent or a an independent test?

iv. Do you use a two-tailed, right-tailed or left-tailed test?

v. Perform the adequate test. Can you reject H_0 with significance level $\alpha = 0.05$?

(d) Calculate the effect size. Is it a small, medium, or large effect?

(e) Report the results.

(f) Plot the means with standard errors as error bars.