Subjective Data Analysis

Mikołaj Leszczuk, Jakub Nawała, Margaret Pinson, Łukasz Zieliński

# Goal

The goal of this exercise is to give you a tentative understanding of subjective data analysis. You are going to analyse the personal opinions of people regarding the quality of images. Through those analyses, you will acquire a preliminary intuition of what information can be extracted from such data.

# Data

To perform any kind of analysis, you need to first understand the data you are dealing with. In this case three components are important: (i) what methodology was used to gather the personal opinions (e.g.., how many images were shown? Did everyone see the same ordering of images?), (ii) what type of images was presented (e.g., do they contain any post-processing? Do they contain any artificial distortions like blur or noise?) and (iii) in what format do you have the data accessible (i.e., is it a CSV file? If yes, how many columns it contains? What is their purpose?).

## Methodology

Below is an excerpt from the publication describing the two data sets. If you have any questions please do not hesitate to ask them. This text may contain domain-specific keywords that may not be immediately obvious to you.

“*The experiment itself used the 5-levels Absolute Category Rating (ACR) scale. (...) The presentation order of images was randomised for each participant.*

*Due to many subjects taking the experiment simultaneously, we were not able to control all of them at once. This resulted in non-identical subject pools in the training session and the two target sessions (VIME1 and CCRIQ2). Eighteen (18) subjects took part in the training session. Twenty-one (21) subjects took part in the VIME1 session. Nineteen (19) subjects took part in the CCRIQ2 session. This means some testers participated in VIME1 or CCRIQ2 session but had not gone through the training session.*

*The test started with a training session. It consisted of 11 images (four from CCRIQ2 and seven from VIME1). After this, we presented the testers with 101 images (the VIME1 session). Again, after those, we showed them 88 images (the CCRIQ2 session). Importantly, 11 images from the training session did appear in VIME1 and CCRIQ2 sessions.*”

## Data Sets

The two subsections below provide an overview of two data sets we are going to use: VIME1 and CCRIQ2.

### [VIME1 Data Set](https://www.dropbox.com/sh/t2i1x73e1nuyxxf/AABR5N8MXyb1uzAwbzZCPcB8a?dl=0)

VIME stands for “Video and Image Models for consumer content Evaluation” and taken its name from [one of VQEG’s (Video Quality Experts Group) working groups](https://www.its.bldrdoc.gov/vqeg/projects/vime.aspx). As its name suggests, it aims to develop models for comparing, evaluating and testing consumer devices used in conditions typical for an average user (meaning everyone not being a professional image/video producer). This goal is reflected in the database’s content. For discussion about file naming convention used, one is referenced to [the *Readme\_vime1.txt* file](https://docs.google.com/document/d/1-w3Zzgevgq0IIpxMntzrAaZReGZ633MAELmR8PRYeHU/edit?usp=sharing) residing in the home folder of this exercise.

### [CCRIQ2 Data Set](https://www.dropbox.com/sh/8ozcjzm3i0htl0x/AABTgCqQMv2P23LPVR3gzfD5a?dl=0)

CCRIQ stands for “Consumer Content Resolution and Image Quality” and similarly as VIME, is related to one of many activities of the VQEG group. In particular, it tries to address the problem of the limited resolution of displays, as compared to enormous resolutions offered by the various image capturing devices. Those ranging from smartphones to DSLRs (Digital Single-Lens Reflex Cameras). Content of this database tries to build a foundation for future work related to the evaluation of quality for images captured with consumer devices and displayed in normal viewing conditions (meaning 1920x1080 px or similar monitor, TV, etc.). For discussion about file naming convention used, one is referenced to [the *Readme\_ccriq2.txt* file](https://docs.google.com/document/d/184dxhTm0DVtbWb17dr7WD8QyOBk-HF7bBO-_OYbkF9o/edit?usp=sharing) residing in the home folder of this exercise.

## Data Format

You are provided with a CSV file. It is called *ccriq2\_vime1\_tidy.csv* and resides in the home folder of this exercise. It contains 3,745 opinion scores. Each row corresponds to a single opinion score of a single study participant. The CSV file has 10 columns:

1. *lab* - the location the subjective study has been performed (i.e., AGH UST),
2. *Tester\_id* - a unique identification number of each study participant,
3. *Exp* - a name of a test session (either “vime1” or “ccriq2”),
4. *PVS* - a name of an image presented to a study participant,
5. *Scene* - a scene, which a photograph of constitutes a test material (a.k.a. PVS),
6. *Camera* - an identification string of a camera used to capture the image (a.k.a. PVS or test material),
7. *OS* - an opinion score of the study participant expressing his/her judgment of the overall image quality,
8. *Timestamp* - timestamp of the opinion score submission,
9. *Camera\_w\_ver* - a string combining the camera identifier with a version of the test material (versions other than “ver1” mean the image is post-processed),
10. *Orientation* - the test image orientation (horizontal or vertical; applicable only to the VIME 1 session).

# Data Analysis

You are encouraged to perform some data processing using MATLAB or LibreOffice Calc (or any other environment you prefer[[1]](#footnote-0)).

Tasks:

1. Each person has different rating patterns. Compare the average score for each subject (i.e., mean of all images rated by that subject). Plot these subject averages with a histogram (they should look normally distributed). Also, look at each subject’s histogram of votes (1...5 in a 5-level histogram). Does everyone use the entire scale? Do you see any differences, like one person being likely to use part of the scale?
2. Look at *ccriq2*’s subjective data. Which scene (i.e., *BouquetPastel*, *PipesNight* and so on) does the best job at showing the differences between the 23 cameras?
3. Look at *vime1*’s subjective data. Which scene (i.e., *alley*, *black white* and so on) does the best job at showing the differences between the cameras?
4. Just consider the subject matter that was photographed for each experimental design. The *vime1* photographs show buildings and statues in one city. The *ccriq2* dataset has four very different subject matters (flowers, garden, the building at night, and a woman). Which is better for cameras comparison? Why?
5. Let us assume that you do not care about the cameras in either experiment, but simply want to understand people’s opinions about various image impairments. Create a boxplot for each scene in *ccriq2* and each scene in *vime1*. What can you learn from these box plots? Dataset *vime1* has seven (7) scenes, while *ccriq2* has only four (4) scenes. Can you learn something new from each of *vime1*’s scenes? Are any of these scenes redundant (that is, the scene tells you no new information)?.

Compute the MOS of each file, across all subjects (done by averaging quality assessment scores for each file). With that data:

1. Look at the distribution of MOS for each scene for both databases (4 scenes in *ccriq2* and 7 scenes in *vime1*). What do you see? What is the average quality, overall scenes? What fraction of images are good or better? What fraction are poor or bad? Is there a difference between qualities for the same device used to shoot different scenes? Why is that?

1. In fact, most data analysts use [the *R* programming language](https://www.r-project.org/) or Python (with [the *pandas’* package](https://pandas.pydata.org/)) [↑](#footnote-ref-0)