MPATC-GE 2042:
Psychology of Music

Statistics Part I and Big Data
Assignment reminder

• Literature review due **October 28** at **11:55pm**
  
  – The literature review is an in-depth discussion of prior work relevant to your research topic. It cites the references you have already compiled in your annotated bibliography, summarizing and reflecting on their main findings, problems, and open questions.
  
  – It should be a cogent, well structured outline of the major prior work in your topic area. Your review should point out divergent perspectives and try to tie together prior work in a way that concisely summarizes the main issues for an unfamiliar reader.
  
  – You should try to group papers by theme; for example, if you find there are two competing schools of thought about your research topic, you should highlight this in your review and summarize the relevant articles in light of this larger context.
  
  – The literature review will serve as a major portion of the introduction section to your final research report.
Literature review requirements

• The general structure of the literature review must be discussed and agreed upon by the group.
• Include a short (a few sentences at most) topic summary at the beginning of your document.
• Include the bibliography (without annotations), in alphabetical order by author last name, at the end of the document.
• Use the introduction/background sections in the journal articles you’ve read as a model for how you should write your review.
• Make sure you use correct APA citation style (both within the main text and in parentheses).
Literature review requirements

• Group participation expectations
  – You should discuss the structure of the literature review with your group and develop an outline together
  – All students in the group should read and be familiar with all the work cited
  – Indicate who did the actual writing

• *IMPORTANT NOTE*: Some groups have bibliographies that were too general and/or scattered; if this is the case, you must make sure to update your references instead of trying to write a lit review from your original, problematic list of references
Statistics

• We use statistics for many reasons:
  – To mathematically describe/depict our findings
  – To draw conclusions from our results
  – To test hypotheses
  – To test for relationships among variables

• Things to consider when using statistics:
  – Be certain that the data is valid and reliable
  – Make sure it’s the right type of data
  – Make sure the statistical tests are applied appropriately
  – Make sure the results are interpreted correctly
Two types of statistics

- Descriptive statistics – summarizes data
  - Can be applied to any measurements (quantitative or qualitative)
  - Offers a summary/overview/description of data
  - *Does not explain or interpret anything*

- Inferential statistics - used to make inferences and draw conclusions on the basis of sample data about a population
  - Statistics used to interpret the meaning of descriptive statistics
  - Allows researchers to infer or generalize observations made with samples of the larger population from which they were selected
Types of experimental variables

• Dependent variable - the measured variable in an experiment
  – Can often be measured in many ways, and therefore requires an operational definition
  – e.g., the operational definition of an exam performance is a score between 0 and 100

• Independent variable – variable that the researcher has control over and manipulates
Types of measurement

• Nominal (categorical) – variables have no numerical or quantitative properties
  – Examples: gender, eye color, college major
• Ordinal – Exhibit minimal quantitative distinctions
  – Variables can be in ranked order; levels of the variable can be studied from lowest to highest
  – Examples: finishing order in a competition, education level, and rankings
• Interval – More quantitative than ranked measurements; distance between variables is meaningful and the intervals between the levels are equal in size.
  – Common example are rating scales
  – e.g., mood rating on a 7-point scale ranging from a “very negative” to a “very positive” mood. There is no absolute zero point that indicates an "absence" of mood
• Ratio – have both equal intervals and an absolute zero point that indicates the absence of the variable being measured.
  – Examples: time, weight, length
PART I: Descriptive statistics
Central tendency

- **Mode**: most frequently occurring value in a distribution (any scale, most unstable)
- **Median**: midpoint in the distribution below which half of the cases reside (ordinal and above)
- **Mean**: arithmetic average - the sum of all values in a distribution divided by the number of cases (interval or ratio)
Median

• Insensitive to extremes

3, 3, 7, 10, 12, 15, 200

Median in this case is 10
Mean: Arithmetic Average

- Mean the sum of a set of values divided by the number of values:
- Scores: 5, 6, 7, 10, 12, 15
- Sum: 55
- Number of scores: 6
- Computation of Mean: $\frac{55}{6} = 9.17$
Mean

• Influenced by extremes
• Only appropriate with interval or ratio data
• Is this five-point scale ordinal or interval?

1 = Strongly Agree
2 = Agree
3 = Neutral
4 = Disagree
5 = Strongly Disagree
Mode: Frequency

- Mode is the most frequently occurring value in a set
- Best used for nominal data
Normal distribution (Gaussian)

Normal curve

Negative skew

Positive skew
Variability

• Variability is a measure of the differences among scores; shows how data vary
• Measures of variability:
  – Variance and standard deviation: spread of scores in a distribution; the greater the scatter, the larger the variance
  – Standard deviation (abbreviated as $SD$ in scientific reports) – square root of the variance
  – Range - simply the difference between the highest score and the lowest score
• Used for interval or ratio data
Standard deviation

- Standard deviation: measures how much subjects differ from the mean of their group
- The more spread out the subjects are around the mean, the larger the standard deviation
- Sensitive to extremes or “outliers”
Example from Schulkind & Davis, 2012

Method

Participants. Forty-six undergraduate volunteers received partial credit towards a course requirement or $5 for their participation. Data on music training was available for forty-five of the forty-six participants. On average, the participants had studied 1.7 instruments ($SD = 1.33$; range 0–5) for an average of 6.3 years ($SD = 4.57$; range 0–16). They averaged 1.0 years ($SD = 2.47$; range 0–12) of formal voice training and 2.6 years ($SD = 3.87$; range 0–14) of performance in organized singing groups. Three subjects were excluded from the study because they claimed to be unfamiliar with at least one third of the target songs.
Fig. 5. Soloists’ swing ratio as a function of tempo. The error bars show the standard deviation.
# Summary table

<table>
<thead>
<tr>
<th>OK to compute....</th>
<th>Nominal</th>
<th>Ordinal</th>
<th>Interval</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency distribution</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Median and percentiles</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Add or subtract</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mean, standard deviation, standard error</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Ratio, or coefficient of variation*</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*ratio of standard deviation to mean (signal-to-noise ratio)
Correlation

- The *correlation coefficient* is a statistic that describes how strongly variables are related to one another.
- The most commonly used is Pearson’s $r$ (Pearson product-moment correlation coefficient):
  - Used for interval and ratio variables.
  - Provides information about the strength of the relationship and the direction of the relationship.
  - Value ranges from 0 (no correlation) to +1 or -1 (perfect correlation).
  - The size of the value (in absolute magnitude), not the sign, represents the strength of the relationship.
  - The plus or minus signs indicate whether there is a positive or negative linear relationship between the two variables.
Correlation visualized

(Perfect correlation)
**Effect size**

- *Effect size* refers to the strength of association between variables
- The Pearson $r$ correlation coefficient is an indicator of effect size
  - A general guide is that correlations near .15 (about .10 to .20) are considered small, those near .30 are medium, and correlations above .40 are large.
- Sometimes $r^2$ instead of $r$ is reported. Thus, if the obtained $r = .50$, the reported $r^2 = .25$.
  - This is useful because the transformation changes the obtained $r$ to a percentage
  - This value represents the percent of variance in one variable that is accounted for by the second variable
  - $R^2$ values are always reported in regression analysis
but there were indications that musical training affected response accuracy. Accuracy scores were calculated by subtracting ratings of structure-preserving alterations from ratings of structure-violating alterations. Accuracy scores and years of musical training were correlated for simple-meter stimuli, $r(48) = .43$, $p < .01$, but not for complex-meter stimuli, $r(48) = .18$, $p = .21$. In other words, more extensive musical training was associated with more differentiated responding to structure-violating and structure-preserving stimuli only in the context of familiar metrical structure.
Regression

- Regression equations are calculations used to predict one variable when the value of another variable is already known.
- They are essentially “prediction equations” that are based on known information about the relationship between the two variables.
- X is the independent variable; Y is the dependent variable (what we want to predict).

![Diagram showing regression line and coordinates]
Multiple regression

• Instead of one predictor variable, we have multiple.

• $a$ is a constant, $b$ are the weights of the predictor variables

Predicted grade point average = $a + b_1$ (college grades)  
$ + b_2$ (score on GRE Aptitude Test)  
$ + b_3$ (score on GRE Psychology Test)  
$ + b_4$ (favorability of recommendation letters)
An aside: Z-scores

- Z-scores, also known as standard scores, are normalized data.
- A set of values are converted to z-scores by subtracting by the mean of all the values (creating a zero mean) and then dividing by the standard deviation (unit standard deviation/variance).
- Can be useful for comparing ratings provided by different subjects or data sets.
PART II: Inferential statistics
NEXT WEEK
Article discussion: Huron (2013)

• Discussion leader: Sripathi Sridhar
Reading question: Ethical issues

• In his article Huron mention that it is not ethical for a researcher to request “human subjects approval for distributing a piece of software that spies on people’s listening habits without their permission. However, this practices are the norm in the commercial world.” Why do you believe people accept one and not the other? Would you consider the statistics obtained by those softwares trust worthy for research purposes, even if they might have been influenced by marketing and other types of enforcement? (Julian T.)
Reading question: Recorded music

- Huron, D. (2014) says that "Within the next decade, one can reasonably expect nearly all of the history of commercially recorded music to become available online." What questions could we answer if we had access to all recorded music for testing? How could we test said questions in an empirical way? What questions would still be unanswerable and unknowable even with access to all of recorded music? (Johnny)
Reading question: Current trends

- The article mentions digitization to be the future of music. From potentially eliminating publishing companies to creating more efficient ways to perform music real-time (through tablets/audio page-turning), have any of Huron’s predicted trends started to manifest more prominently in the last five years since the article was published? What are some other ways, both beneficial and detrimental, that digitization could affect the future of music? (Shannon)
Reading question: Sampling biases

- Huron mentions a few methodological caveats of Big Data initiatives, including sampling biases due to unequal representations on the internet. Since not all cultures and time periods are represented equally on the internet, the population sample used in a statistical test will not likely be representative of the population of interest to the researcher. Does this imply that Big Data initiatives using the internet should only test claims about specific subpopulations (those that are known to have a presence on the internet)? In addition, which topics in music research (score analysis, MIR, MIDI data analysis, survey response, etc) would be most affected by sampling biases using a Big Data approach through internet data? Which music cultures would be most affected? (Henry)
Reading question: Annotating data

• Even in the time of big data, the amount of symbolic data is still not enough to support researches that require large amount of data. Different from labeling image data, annotating musical data is labor-intensive and also need strong expertise in music. The copyright problem which is also mentioned in the paper is another hindrance. Are there any good ways to acquire more musical data except hiring people to label them? Ex, sometime people will design games to collect data from players. (Wei-Tsung)
I’m curious why the article didn’t talk about how readily having so much music at our finger tips could potentially limit us as a society. Because I think that if you constantly have access to information without really having to work/study that the critical thinking skills you learn young could potentially go out of the window. By having this technology at your fingertips one could simple look up reviews for an article/composition and base their opinion off of that instead of thinking about it on their own. (Mark)
Reading question: Bad data?

• In the era of Big Data, information is manipulated, people behavior is influenced by sporadic trends that last less than a week, opinion is tricked by a fake article posted on Facebook, marketing campaigns or a video that went viral. What mechanisms do researchers have in order to reduce the risks of error when conducting exploratory research through this kind of media? Do you think we live in the era of information or rather disinformation? How do you usually corroborate the veracity of the information you read on social media? (Julian T.)
Reading question: Enough music?

As Huron notes, “many questions regarding music will prove to be intractable because of limitations in the amount of pertinent music, even assuming all of the pertinent sources become available online.” Are there certain questions about music that will eventually become answerable as the amount of pertinent music increases and music evolves and certain genres or styles develop and increase in popularity? If so, when can one conclude that enough pertinent music exists to draw conclusions? (Tyler)
Reading question: Quality control

• How do you know the error rate of a database you are researching on? (Rebekah)

• "The very best way to minimize both types is simply to gather more evidence." What amount of data would be required to minimize the errors to the minimum possible or even make them disappear? Is that amount of data even possible to be found? (Karolos)
Reading question: Potential research

• In what ways can Big Data help understand listener's technical understanding of music? For e.g. analysing musical complexity and people's perception of that music?

• In which ways can Big Data be used to understand people's listening habits, their perception of a particular music, and their mental health? (Mihir)

• Since we’re moving towards a more statistical era in research, will that affect research on music as it’s very emotional and personal? Can we still test how music affects people with Big Data? (Jess)
Reading question: Proprietary research

• Huron states that there are musical breakthroughs that already exist as trade secrets. What policies can be enacted to ensure that future discoveries/breakthroughs are made accessible as public domain or intellectual property? (Eugenio)

• Huron paints a bleak picture of the future when it comes to musicologists' abilities to compete with the private sector in terms of resources and discoveries. What does this mean for the future of our society's musicological growth? Are we nearing a peak with regards to what we can possibly know without industry sharing the findings it holds so secretively? (Max)
Reading question: Assumptions

- Huron, D. (2013) says that "Methodologists point to a general trade-off between two kinds of scholarly errors: claiming something to be true, useful, or knowable that is in fact false, useless, or unknowable, and claiming something to be false, useless, or unknowable that is in fact true, useful, or knowable." What current ideas about music do we currently share that fall into these two problems? What assumptions do we hold about music that are thought to be true, useful, and knowable but might be false, useless, and unknowable? What assumptions do we hold about music that are thought to be false, useless, and unknowable but might be true, useful, and knowable? How might big data help in solving these problems? (Johnny)
On the subject of copyright protection, the author seems pessimistic about the future, but is the problem of creating a critical edition of the works of a composer actually a fixable one? As we approach a future of Dataism, wherein we worship simply the mass collection of knowledge and data, can musicologists use this philosophy to champion copyright changes to better suit the general public? Or is waiting the only solution? (Julian C.)
Reading questions: Complete databases

• Looking towards the future, it is mentioned in the article that once the population parameters are known, almost all of statistics becomes irrelevant. Depending on the subject, how will empirical science be replaced by hermeneutics when the population is accessible? If that was the case, does it mean that a lot of research currently being done is an act of counter-productiveness, will amount to nothing? (Tommy)