Assignment reminder

• Literature review due October 27 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: $55/6 = 9.17$
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).

![Graph showing regression line and distance from line to data points.](image)
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: Jorge Arango and Ernesto Valenzuela
Reading questions: Pitfalls of Big Data

“The very best way to minimize both types of error is simply to gather more evidence. This is the principal reason why collecting data is important. To the extent that minimizing knowledge-related error can contribute to health, justice, environmental well-being, and other good things, the drive toward Big Data is not merely some obsession with things numerical, or a kleptophilic compulsion to collect, but a proper moral imperative.”

While Huron does note some of the potential pitfalls of Big Data, there are many he leaves out: apart from the commercial uses and snooping BD encourages, the shift to data comes at the expense of examining specific cases, the fine-grained rather than some bigger picture, looking for likelihoods instead of numerical certainties; BD also has the problem of being presented as indisputable fact – who can argue with numbers? – but the algorithms used to sift through large amounts of information and the way questions are posed can often be bent to favor certain results (even without post-hoc "theorizing"). There are also major issues for preservation (just ask Kent over in Bobst!), the economics of music, or the Neil Postman argument that the medium changes the way we see and do things to a far greater extent than the content (the internet isn’t just a “tool” that we control, it guides and limits us in countless ways, shapes and determines our thinking far more than we generally acknowledge). So I suppose my question is: is Huron balanced enough, or is he overstepping in calling Big Data – even with the condition of use for “good things” – “a proper moral imperative”? (Michael)
Reading questions: Internet vs. other sources for music research

• While most popular music can be found on the internet, there is a movement within the independent music scene to not only press in vinyl, but to avoid the mainstream marketing strategies involving social media marketing, digital libraries, and electronic music store-houses. Don’t such trends toward countercultural movement stand to undermine the integrity and reliability of the internet as a sole source of music in any type of music research? (John)
Reading questions: The internet effect

• David Huron writes, “from the beginning there has been a strong bias toward English-language materials. As much as the web feels like a global community, it remains something of a privileged club.” What does this mean for musicians. Is the future of music been shaped somehow by the type of musical information pushed to the top of the web via "big internet"? (Victoria)

• "Within the next decade one can reasonably expect nearly all of the history of commercially recorded music to become available online.” Will future performers and composers use this collection as THE listening springboard for ideas. Will ideas become less original and more based on this sample size? Will the music of cultures that are not commercially recorded die out as a result? (Victoria)
Reading questions: Post hoc theories

• When performing an exploratory study, how does that not then lead to post hoc theories? (Jason R.)
• The author underlines the fact that researchers are excited to find any statistically significant relationship among data. So is it impossible to form a theory (post hoc) from one datum and not from the relationships among data? (Federico)
• On the other hand… can the shift “from prediction to hermeneutics” be in some ways useful? I wouldn't advocate this as the main direction for academic work, but it does seem very possible that data collection could reveal some patterns that scholars might not have even thought to ask about… In other words, are the post-hoc “theories” Huron’s condemning all bad, or could there be some redeeming value for scholars? (Michael)
Reading questions: Ethics

• One of the latest Hot Topics in computing has been the Internet of Things, with the goal of linking not only one's computer, but everyday objects, to the internet. What ramifications could this have for the field of musicology / music in general? How might researchers employ such a scheme towards publishing new findings about human processing of music? Would this even be ethical? This ties into Huron's discussion of privacy concerns and the pitfalls of commercial “spying.” (Kyle)

• The description of the "drive toward Big Data" as being a "proper moral imperative" is an interesting claim, especially given his comments about privacy. Where do we draw the line between that which benefits researchers—making music and associated data available through databases, rather than locked into publishers' copyrights—and that which can come at the loss of individual privacy? Where does data about music in general end, and data about individuals begin, especially as more and more individuals make their music—and, through Facebook and other platforms, their habits associated with that music—available for free using the internet? (Kyle)
On page 8, Huron mentions the difficulty a researcher would have in getting any IRB/UCAIHS/etc approval for a program that monitors subjects' music listening patterns without their explicit consent. The implication is that he is referring to services like Spotify that make recommendations based on history, and presumably those services have privacy policies that would forbid the release of an individual's data to a researcher -- but is there any reason to think that over a long period of time, a user would change their listening habits because of the knowledge that they were being willfully monitored? Theoretically a given individual may stray away from his or her 'guilty pleasures' of music in a short-term study, but over the course of something like a year-long or even month-long period of time, I would expect listening habits to normalize.

So, as long as the researcher can get their music-tracking program the exposure to achieve a statistically significant sample size, wouldn't such a study be approvable by a human subjects board? (Sean)
Reading questions: Proprietary information

• What would a musical trade secret be, exactly? Ideal chord progressions, melodic techniques? Which lyrics or instrumentations sell the best? It seems that they would be concentrated around popular music, so could research take a focus on older, non-popular music in order to bypass the secrecy problems and disincentives outline in the paper? (Ned)
Reading questions: Internet and data bias

- Regarding Internet data, Huron states, "The Internet is not a level playing field; not all cultures and periods are represented. From the beginning, there has been a strong bias toward English-language materials." (5) Obviously, the Internet is fallible when it comes to gathering reliable sources, but is this statement accurate? Or does it depend on the topic on which you are focusing? (Sarah)
Reading questions: Critical editions

• Many of the author's claims regarding future availability of information (i.e. recorded music and critical editions) are loosely substantiated, but particularly there's something of a disagreement between the notion that critical editions will continue to be unavailable to the public whilst nearly all recorded music will become universally accessible. What evidence would substantiate this claim? And in its absence, is there a possibility that today's composers will make their original scores publicly available? (Jason S.)
In the paragraph of quality control, the author suggests that the effect of errors might get increased when the processing types move on. The example he gave was that a 1% pitch error rate may be a 2% melody error rate. However, when it comes to analyzing music with data science, it is common to break the features down to different levels. For instance, timbre is one of the lower features and the level can go up to other forward analysis such as mood recognition or genre classification. Other than trying to reduce error rates with advanced technology, could breaking the information down even more help handling errors? (Taihua)
Reading questions: Listener preferences

- It seems like Huron is trying to say that with the internet, the era of big data has begun, and therefore we now have access to infinitely larger data streams than ever before, creating new research options and therefore with them, new research pitfalls. But it seems like internet-based music cognition research would venture more into music sociology than psychology, dealing with larger groups of people or even entire music subcultures. Music listening is often an intensely personal activity, which would support the idea that the data might be a great multitude, but it is a great multitude of individuals. But even given that, can such giant sample sizes yield any hard, objective data about how an individual experiences music? (Sean)
Reading questions: Technology

• "With regard to notated scores, the irksome task of turning pages is likely to have a facilitating effect in bringing materials online...Automatic (audio-driven) pageturning may become the norm, and so data will need to include symbolic representations of notes, not simply image files."

When Ebook came out, many people still chose to love paper book (though many scientists predicted that Ebook would take the place of paper book). Will it be same thing that performers prefer paper score rather than electronic score? (maybe because of less electronic-breakdown problem or others) (Vincent)
The paper discusses big behavioral databases utilized by music services like iTunes or Amazon. These algorithmic, statistical models contrast with Pandora's Music Genome project, which involves humans categorizing individual tracks. Sometimes it will take over a half hour for Pandora's trained specialists to classify a track, leading to a much smaller database of a few million songs, compared to Spotify's 30 million. Is Pandora's music catalogue superior or inferior to the behavioral/statistically arranged catalogues within Spotify/iTunes? Does the manual human review enhance or potentially flaw Pandora's service? (Ned)
Reading questions: Future possibilities

- This paper covers many grounds in musicological research, but how does Big Data benefit music cognition research in particular? Perhaps we could, possibly with supercomputers and through the database of information collected through past research, create a simulation of the brain and the auditory system, and continuously compare and improve it through numerous world-wide subjective tests? (Chi)