WEEK 8: MAKING SENSE OF YOUR DATA

Dr. T. Kody Frey

Assistant Professor | School of Information Science

OVERVIEW

- Q&A: Where are the Rakes?
- Making Sense of Your Data
- Continuing to Make Sense of Your Data
- Workshop: Survey Construction in Qualtrics

WHAT'S NEXT?

Week 9 **3/6**	**In-Class Midterm Examination** The midterm will resemble a traditional comprehensive exam environment. Students wil have three hours to complete the midterm together in the classroom. Students will be allowed to create a one-page cheat sheet with as much information they can strategically place on the page.
Week 10 **3/13**	**No CLASS – Spring Break**

THE MIDTERM

What are the *key* areas that need more explanation?

MAKING SENSE OF YOUR DATA

The content this week is designed to help you understand what to do with data once it is collected. Before you can start to analyze, you have to get a better idea of what you are working with.
We have a whole WORKDAY planned for your actual data in Week 14.

KEY TERMS

Let’s make sure we are all on the same page regarding some key terms.

TYPES OF STATISTICS

Descriptive Statistics **Inferential Statistics**

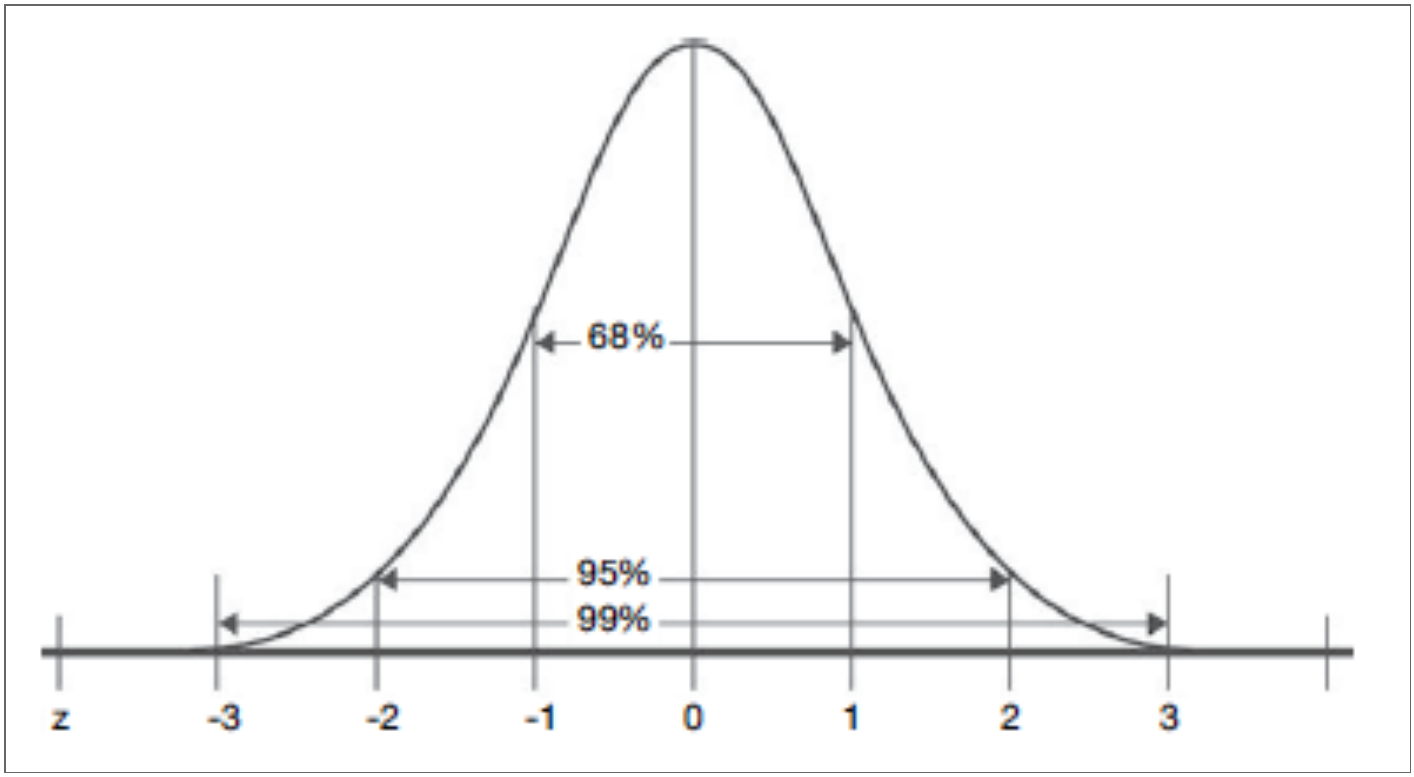
Summarize data from the current sample of participants *without* making inferences about the larger population of interest.
Averages, percentages, frequencies
Make inferences about a larger group, the population, from the group studied,
t-tests, *F*-tests, correlation, regression

TYPES OF TESTING

Nonparametric Tests **Parametric Tests**

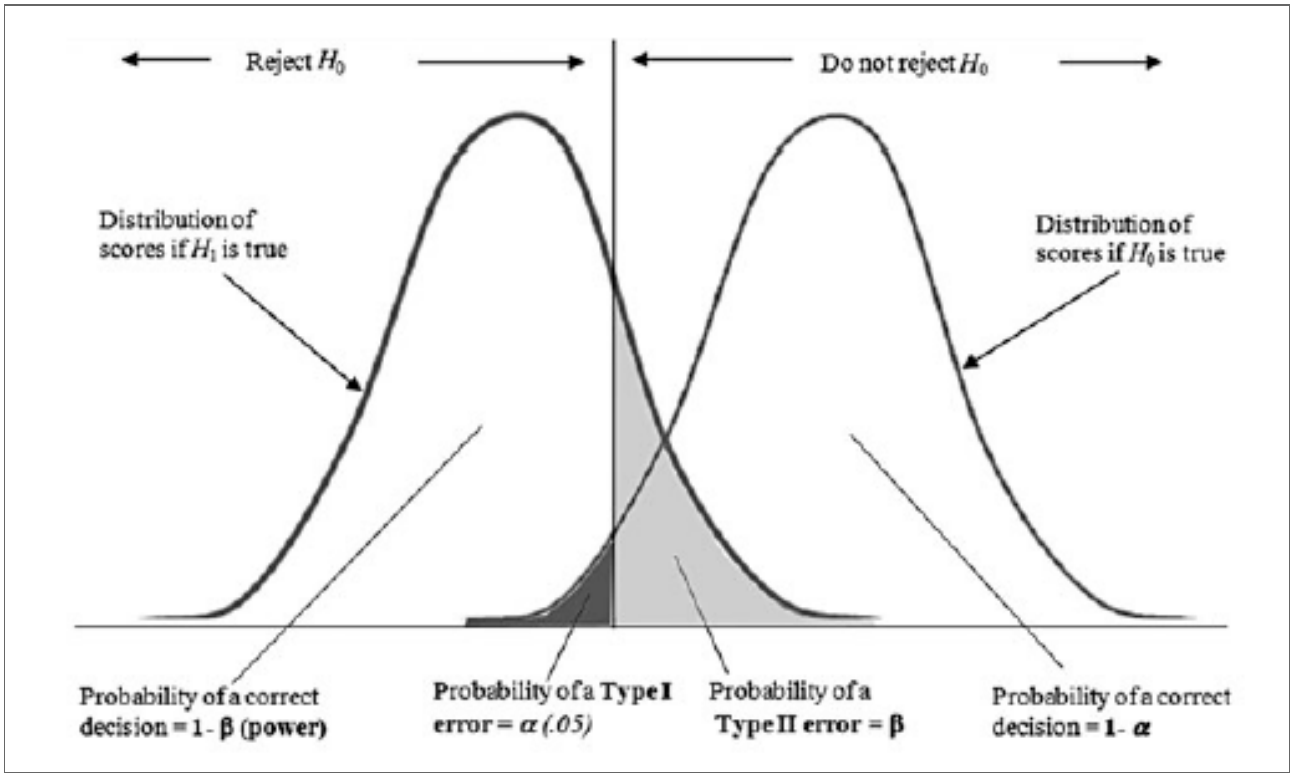
Inferential statistics that are used when the data do not meet the assumption normality.
Chi-square test of independence, Mann-Whitney U Test, Fisher Exact Test
Inferential statistics that assume data are normally distributed.
t-tests, *F*-tests

NORMALITY?



The normal distribution

NORMALITY!



The normal distribution applied to hypothesis testing

ASSUMPTIONS

To use parametric statistics, you have to meet three assumptions:

Normality **Homogeneity of Variance** **Independence**

The DV (or IV if continuous) need to come from a population that is *normally distributed*
Most tests tolerate a good degree of violation, so approximately normal is fine.
Samples in study have equal variation among members.
Some violation is okay, but too much can lead to Type I error.
Observations must be independent!
Scores cannot be influenced or contingent on one another.
Ex: Students in the same class answering questions about their teacher.

MANIPULATION CHECKS

Used to ensure that your control of the *active* independent variable was successful.
For example, if test messages, we need to ensure that participants noticed the difference they were supposed to notice.
Can often be accomplished with a few questions (Ex: What was the name of the influencer who made the post?)

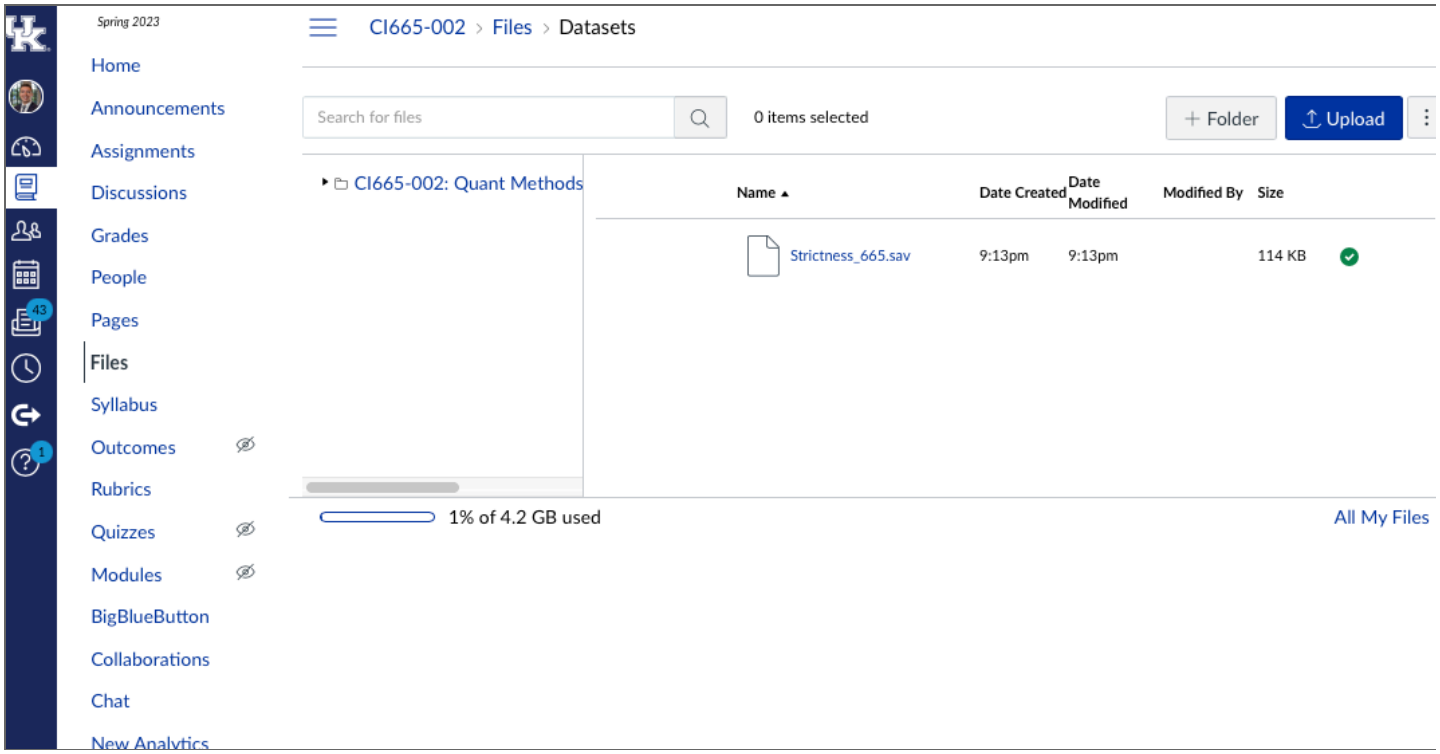
PILOT TESTING

To add more assurance that our manipulations and measures will work, we often want to pilot test them.
A form of qualitative checking!
How do similar samples interpret the different conditions or understand the measures?

LOOKING AT SOME DATA

Context: Working study on the effects perceived *instructor strictness* on motivation, interest, engagement, cognitive learning.

UNDERSTANDING THE SPACE



IMPORTANT PIECES

- Data View vs. Variable View
- Cases and Columns
- Variable Names, Labels, and Values
- Analyze: Descriptive Statistics
- Pairwise vs. Listwise

GETTING AN ACCURATE FILE

When you pull your data, there will be a lot there. Some considerations:

- Eliminate data you do not plan to use like geographical conditions
- Screen for participants who did not complete survey
- Use time as a criterion
  - My rule of thumb is under 4 minutes or over 18 hours
- Recode early
- Dichotomous data is meaningful!

Transform > Recode Into Same Variables > Select Cases > Old and New Values > Enter Values > Add > Repeat

CHECKING FOR ACCURACY

Look at the frequency distributions for your items
Analyze > Descriptive Statistics > Descriptives > Enter Variables > Options (Mean, Std. deviation, minimum, maximum) > OK

MISSING DATA

The key is *consistency*; establish rules and apply them
MAR, MNAR, MCAR

EXPLORING MISSING DATA

Let the software do the work for you.
Descriptive Statistics -> Explore -> Add Variables to Dependent List -> Statistics (Check Outliers) ->Plots (Histogram & Normality plot) -> Options (Exclude cases pairwise) -> OK

DEALING WITH MISSING DATA

- Mean Imputation
- Series Mean
- Leaving 'em blank
- Replacing with 999

CALCULATING COMPOSITES

We create summed scores (or composites) of variables to reduce the number of items we have to analyze (among other reasons).

Transform > Compute Variable > Target Variable > Numeric Expression > Create formula for average of total items used in instrument > OK
(SUM(EvalStrict1 to EvalStrict_12)/12)

EXPLORING COMPOSITES

We want to examine *mean*, *variance*, *skewness*, *kurtosis*
Using the composite variables, we run this code again:
Descriptive Statistics -> Explore -> Add *Composite* Variables to Dependent List -> Statistics (Check Outliers) ->Plots (Histogram & Normality plot) -> Options (Exclude cases pairwise) -> OK

INTERPRETATION

What patterns do the histograms reveal? Does the data look normally distributed?
Are skewness and kurtosis > +2 or < -2? These are *extreme* scores.
Want plotted observed value to closely resemble expected value.
Significant Shapiro-Wilk and Kolmogorov-Smirnov test of normality mean data are NOT distributed normally.

CONVERTING TO Z-SCORES

This will help us determine univariate outliers.
Analyze -> Descriptive statistics -> Descriptives -> Select Variables -> Select "Save standardized values as variables" -> Options (select Kurtosis and Skewness) -> OK
Z scores automatically added to dataset

ASSESSING STANDARDIZED MEANS

Analyze -> Descriptive Statistics -> Frequencies -> Move newly created Zvariables to Variables -> Statistics (mean, standard deviation, skewness, kurtosis) -> Charts (Histogram & Show normal curve on histogram) -> OK
Does the distribution follow the normal curve?

IDENTIFYING THE OUTLIERS

Look at the minimum and maximum scores for each standardized variable. Values greater than 3.29 indicates +3 standard deviations and indicate outliers.
If present, options include:

- Deleting the whole case
- Identifying the response as missing to preserve other cases

IDENTIYING OUTLIERS AS MISSING

First, sort the z scores so higher or lower scores appear first.
Variable View > Find Respective Variable > Missing > Select ellipses within box > Range plus one optional discrete missing value > Enter observed score for first corresponding z score above +/- 3.29 > OK
When this is done, assess the data again with the explore function.

CAREFUL!

My rule of thumb is that you can do this once. If still contains skewness, platykurtic/ leptokurtic distribution, or is not normal, have to move on. Cannot massage the data to fit your needs.

WHAT ELSE FOR OUTLIERS AND ASSUMPTIONS?

At this point, you can:

- Check for *multivariate normality* if questions are concerned with combinations of IVs or DVs
- Check for *linearity* is questions are concerned with associations between variables. ## Reliability

For composite measures, alpha and omega are the standard statistics
If items need to be changed for reliability purposes, run descriptives on your composites again

MEANS AND STANDARD DEVIATIONS

Very important to include as a description of your data.
I like to include in the composite variable LABEL.
Question: What do they mean?

SUMMARY

- Key terms
- Cleaning and understanding data
- SPSS!

FREY FACTS

There is a YouTube video for *everything*
If you learn to do this in R, you have a line of code for every step. All you have to do is hit “run”.

WHAT'S NEXT

The Midterm!

CREATING SURVEYS IN QUALTRICS

Qualtrics is the default platform for UK, but many other online survey platforms exist.

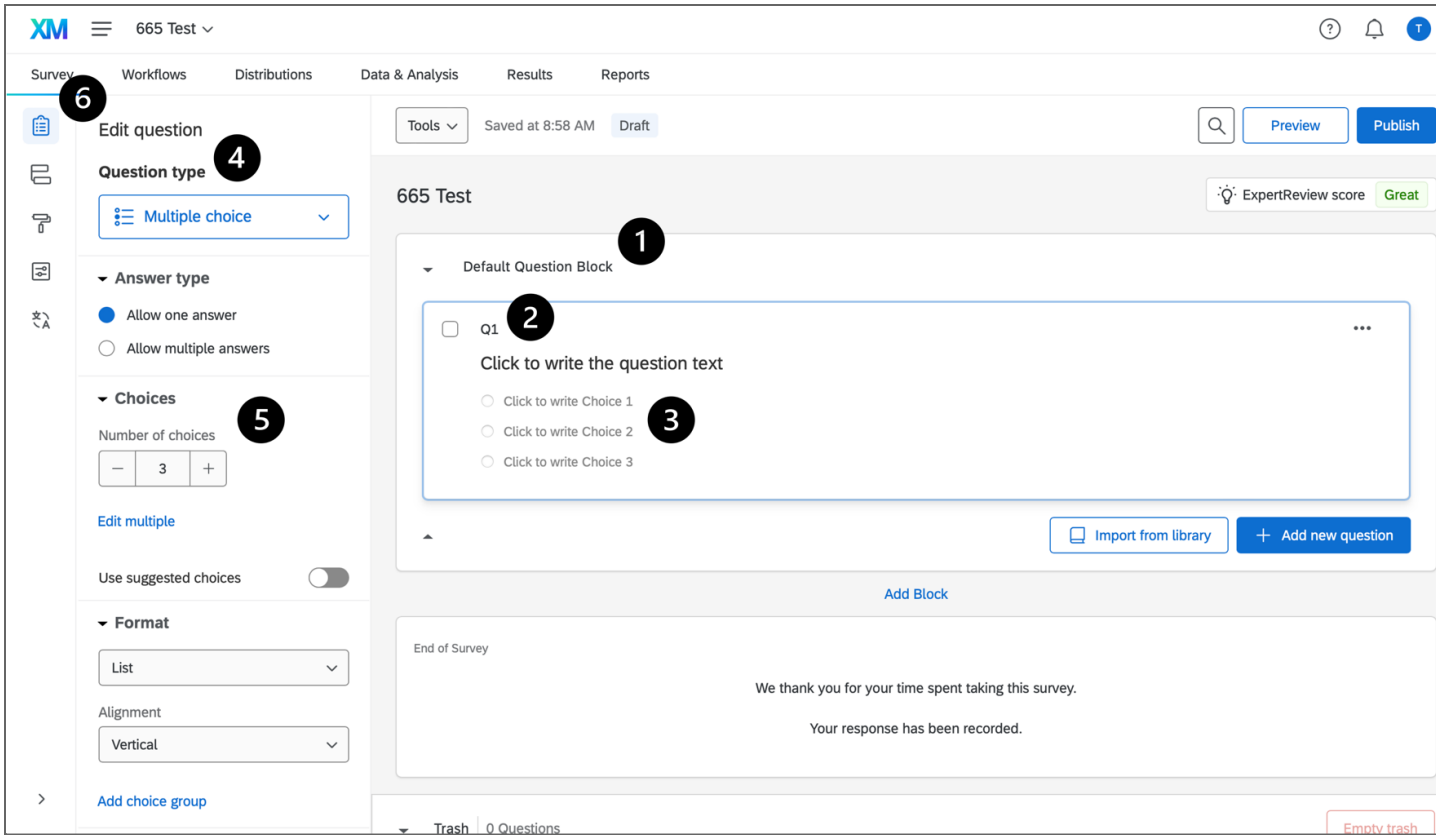
GETTING STARTED

- Log in with linkblue
- Create Project > From Scratch > Survey > Get Started
- 4 options:
  - Starting with blank survey
  - Insert QSF (contains data from existing Qualtrics survey)
  - Copy survey from existing project
  - Use survey from your library

THE INTERFACE

- Blocks
- Question Types
- Survey Flow

KEY POINTS

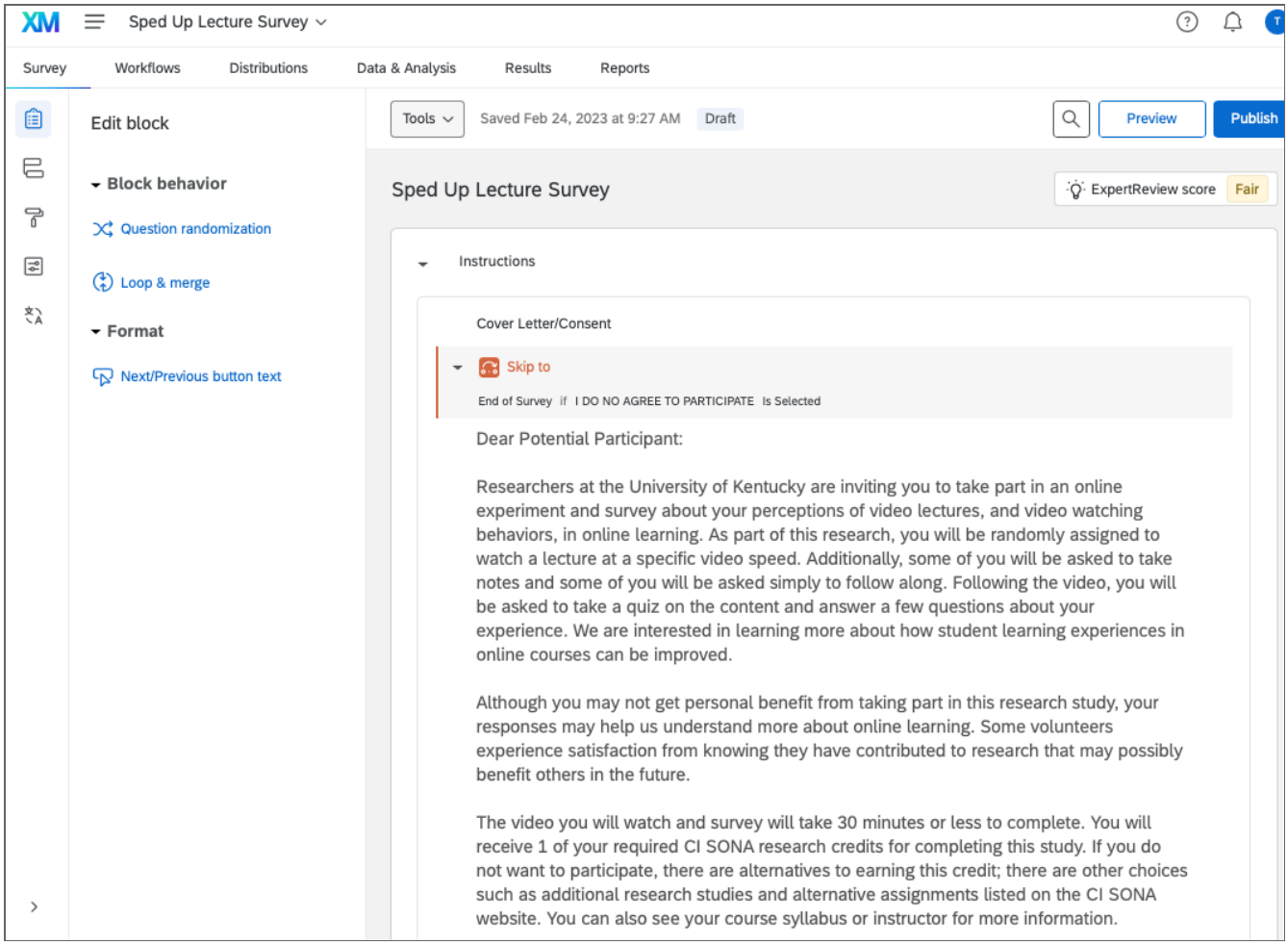

XM
665 Test
Survey
Workflows
Distributions
Data & Analysis
Results
Reports
6
Edit question
Question type
4
Multiple choice
Answer type
Allow one answer
Allow multiple answers
Choices
5
Number of choices
3
Edit multiple
Use suggested choices
Format
List
Alignment
Vertical
Add choice group
Tools
Saved at 8:58 AM
Draft
Preview
Publish
665 Test
ExpertReview score
Great
1
Default Question Block
Q1
2
Click to write the question text
Click to write Choice 1
Click to write Choice 2
Click to write Choice 3
3
Import from library
Add new question
Add Block
End of Survey
We thank you for your time spent taking this survey.
Your response has been recorded.
Trash
0 Questions
Empty trash

OUR GUIDE

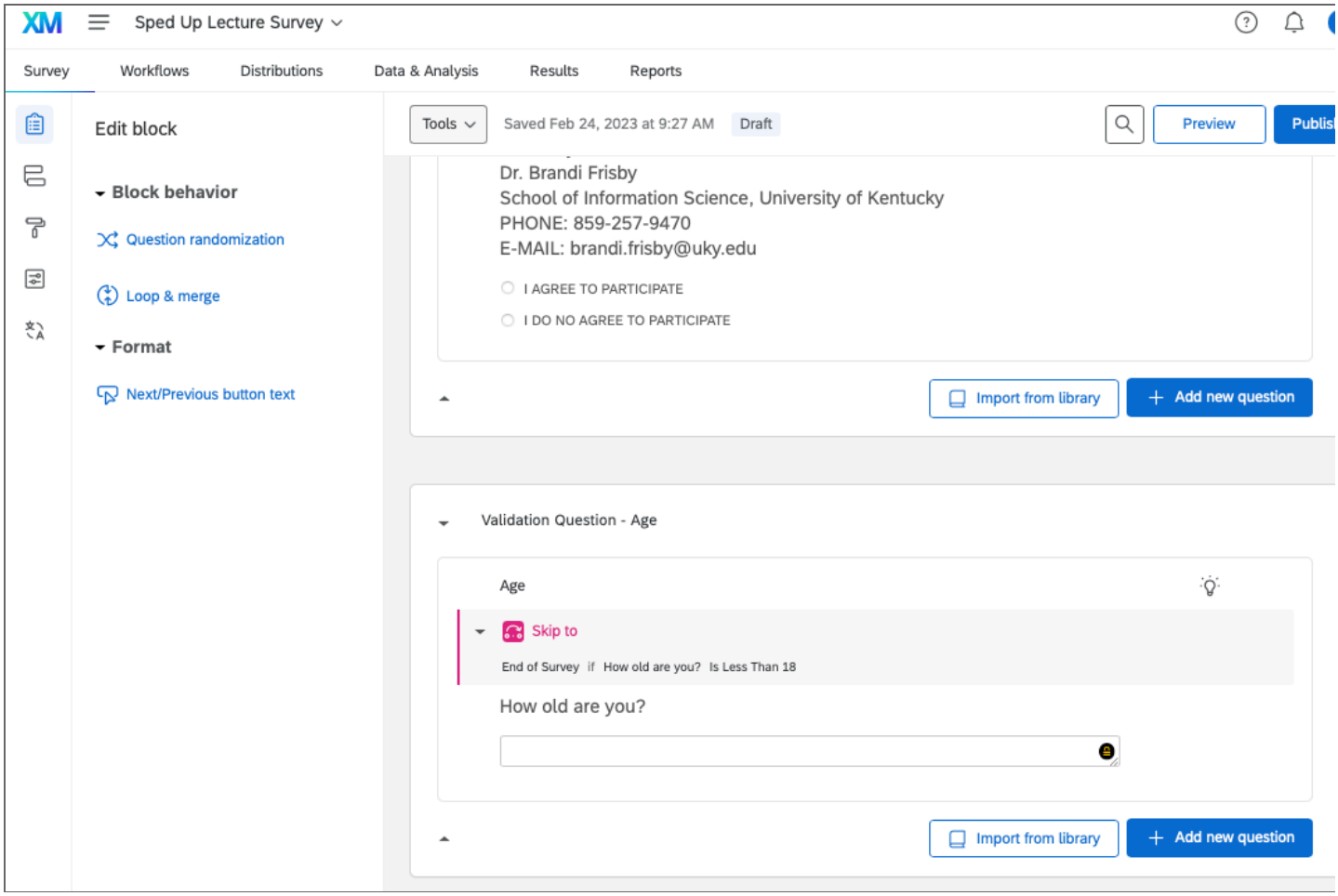
Typical Workflow:

- Save a space for consent (add once approved by IRB)
- After consent, include questions with participation criteria (e.g., age)
  - *If necessary, include directions for manipulation
  - *If necessary, include manipulation / realism checks
- Provide questions for continuous IV / DVs
- Collect demographic information
- Code variable values
- Adjust survey flow
- Edit look and feel
- Preview survey
- Publish and Distribute

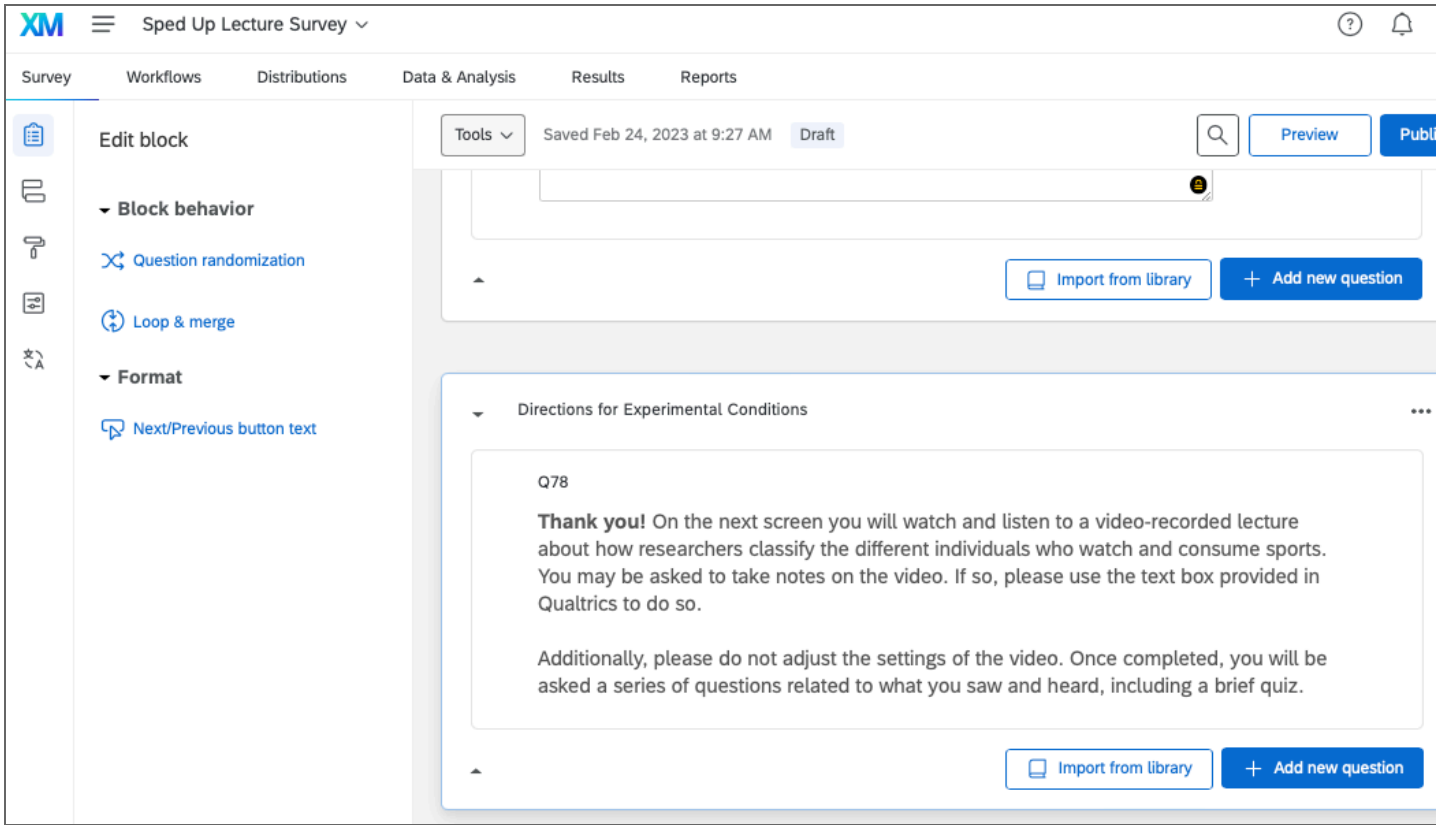
CONSENT


Sped Up Lecture Survey
Survey
Workflows
Distributions
Data & Analysis
Results
Reports
Edit block
Block behavior
Question randomization
Loop & merge
Format
Next/Previous button text
Tools
Saved Feb 24, 2023 at 9:27 AM
Draft
Preview
Publish
Sped Up Lecture Survey
ExpertReview score
Fair
Instructions
Cover Letter/Consent
Skip to
End of Survey if I DO NO AGREE TO PARTICIPATE Is Selected
Dear Potential Participant:
Researchers at the University of Kentucky are inviting you to take part in an online experiment and survey about your perceptions of video lectures, and video watching behaviors, in online learning. As part of this research, you will be randomly assigned to watch a lecture at a specific video speed. Additionally, some of you will be asked to take notes and some of you will be asked simply to follow along. Following the video, you will be asked to take a quiz on the content and answer a few questions about your experience. We are interested in learning more about how student learning experiences in online courses can be improved.
Although you may not get personal benefit from taking part in this research study, your responses may help us understand more about online learning. Some volunteers experience satisfaction from knowing they have contributed to research that may possibly benefit others in the future.
The video you will watch and survey will take 30 minutes or less to complete. You will receive 1 of your required CI SONA research credits for completing this study. If you do not want to participate, there are alternatives to earning this credit; there are other choices such as additional research studies and alternative assignments listed on the CI SONA website. You can also see your course syllabus or instructor for more information.

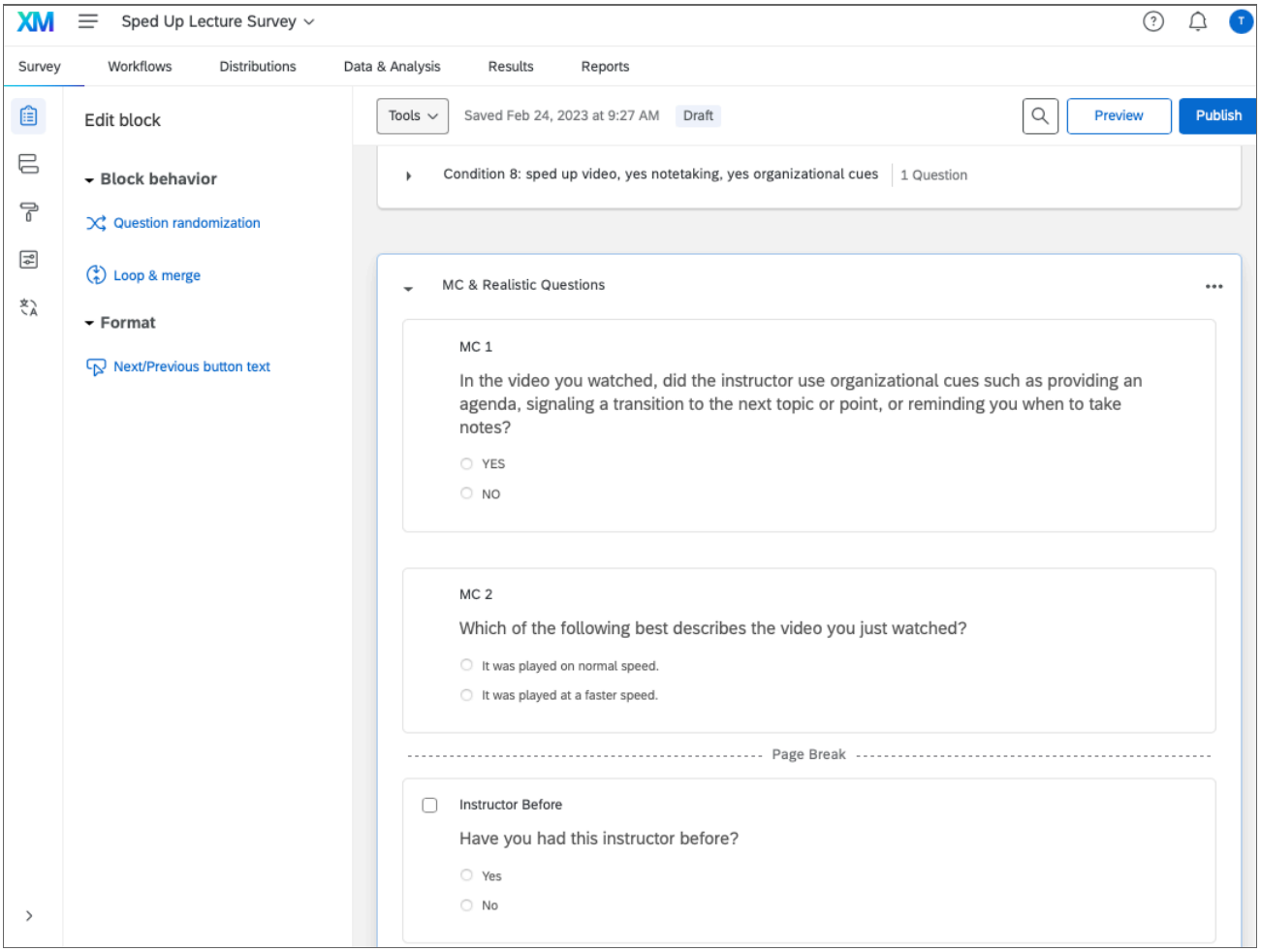
ELIGIBILITY



TEXT


Sped Up Lecture Survey
Survey
Workflows
Distributions
Data & Analysis
Results
Reports
Edit block
Block behavior
Question randomization
Loop & merge
Format
Next/Previous button text
Tools
Saved Feb 24, 2023 at 9:27 AM
Draft
Preview
Import from library
Add new question
Directions for Experimental Conditions
Q78
Thank you! On the next screen you will watch and listen to a video-recorded lecture about how researchers classify the different individuals who watch and consume sports. You may be asked to take notes on the video. If so, please use the text box provided in Qualtrics to do so.
Additionally, please do not adjust the settings of the video. Once completed, you will be asked a series of questions related to what you saw and heard, including a brief quiz.
Import from library
Add new question

MANIPULATION CHECKS



SOME QUALTRICS BEST PRACTICES

- Use colors and italics to highlight or provide key information (e.g., time to complete)
- Check ExpertReview for tips
- Avoid forcing responses unless necessary
- Edit multiple statements at once!
- Be careful with preset values
- Preview often
- Test if on yourself
- Be mindful or time
- Find the recode values button

TEXTBOOK RULES FOR CODING VARIABLES {.SMALLER)}

- Each level of a variable must be mutually exclusive
- Each variable should be coded to obtain maximum information
- For each participant, there must be a code or value for each variable
- Apply coding rules consistently
- High numbers should be used to Agree, Good, or Positive
- All Data Should be Numeric
- Each variable for each case or participant must occupy the same column in the spreadsheet or data editor
- Check for problems!

tkfr222@uky.edu | **@TKFrey** | **tkodyfrey.github.io**

University of Kentucky