Setting up your experiment

• Readings:
  • Chapter 5.9 of textbook
  • Chapters 9 and 12 of Doing Psychology Experiments (DW Martin)
  • Chapter 3 of Improving Survey Questions by Floyd Fowler, Jr.
Procedure

• The *procedure* encompasses everything that occurs with participants

• The procedure includes the experiment task (obviously), but everything else as well…
  – Arriving, welcoming
  – Signing a consent form
  – Instructions given to participants about the experiment task (next slide)
  – Demonstration trials, practice trials
  – Rest breaks
  – Administering of a questionnaire or an interview
Instructions

- Very important (best to prepare in advance; write out)
- For example, the goal in the experiment task is “to proceed as quickly and accurately as possible but at a pace that is comfortable”
- Other instructions are fine, as per the goal of the experiment or the nature of the tasks, but…
- Give the same instructions to all participants
- If a participant asks for clarification, do not change the instructions in a way that may cause the participant to behave differently from the other participants
Participants

• Researchers want experimental results to apply to people not actually tested – a population

• Population examples:
  – Computer-literate adults, teenagers, children, people with certain disabilities, left-handed people, engineers, musicians, etc.

• For results to apply generally to a population, the participants used in the experiment must be…
  – Members of the desired population
  – Selected at random from the population

• True random sampling is rarely done (consider the number and location of people in the population examples above)

• Some form of *convenience sampling* is typical
How Many Participants?

How Many Participants?

- Too few $\rightarrow$ experimental effects fail to achieve statistical significance

---

How Many Participants?

- Too few ➔ experimental effects fail to achieve statistical significance
- Too many ➔ statistical significance for effects of no practical value

How Many Participants?

• Too few $\rightarrow$ experimental effects fail to achieve statistical significance
• Too many $\rightarrow$ statistical significance for effects of no practical value
• Statistically sound method: Power analysis
  • If you know the size of the effect you would consider practically significant, and the acceptable probabilities of Type I and Type II errors, then you can determine the sample size that will allow you to detect this effect
• R packages: [https://www.statmethods.net/stats/power.html](https://www.statmethods.net/stats/power.html)

How Many Participants?

- Too few → experimental effects fail to achieve statistical significance
- Too many → statistical significance for effects of no practical value
- Statistically sound method: Power analysis
  - If you know the size of the effect you would consider practically significant, and the acceptable probabilities of Type I and Type II errors, then you can determine the sample size that will allow you to detect this effect
- R packages: [https://www.statmethods.net/stats/power.html](https://www.statmethods.net/stats/power.html)
- The correct number… (drum roll please)
  - Use the same number of participants as used in similar research

---

Questionnaires
Questionnaires

• Two purposes
  – Collect information about the participants
    • Demographics (gender, age, first language, handedness, visual acuity, etc.)
    • Prior experience with interfaces or interaction techniques related to the research
  – Solicit feedback, comments, impressions, suggestions, etc., about participants’ use of the experimental apparatus
Questionnaires

• Two purposes
  – Collect information about the participants
    • Demographics (gender, age, first language, handedness, visual acuity, etc.)
    • Prior experience with interfaces or interaction techniques related to the research
  – Solicit feedback, comments, impressions, suggestions, etc., about participants’ use of the experimental apparatus

• Questionnaires, as an adjunct to experimental research, are usually brief
Questionnaires

• Two purposes
  – Collect information about the participants
    • Demographics (gender, age, first language, handedness, visual acuity, etc.)
    • Prior experience with interfaces or interaction techniques related to the research
  – Solicit feedback, comments, impressions, suggestions, etc., about participants’ use of the experimental apparatus

• Questionnaires, as an adjunct to experimental research, are usually brief

• We will discuss some brief guidelines. For an in-depth discussion, see additional readings.
Questionnaires

• Two purposes
  – Collect information about the participants
    • Demographics (gender, age, first language, handedness, visual acuity, etc.)
    • Prior experience with interfaces or interaction techniques related to the research
  – Solicit feedback, comments, impressions, suggestions, etc., about participants’ use of the experimental apparatus

• Questionnaires, as an adjunct to experimental research, are usually brief

• We will discuss some brief guidelines. For an in-depth discussion, see additional readings.

Available through course reserves: Doing Psychology Experiments Chapter 9 and 12, by David W. Martin,
Improving Survey Questions Design and Evaluation from Questions to measure subjective states by Floyd J. Fowler, Jr.
Information Questions

• Questions constructed according to how the information will be used

Please indicate your age: ________

Ratio-scale data

Please indicate your age?

☐ < 20  ☐ 20-29  ☐ 30-39
☐ 40-49  ☐ 50-59  ☐ 60+

Ordinal-scale data

Which browser do you use? ________________

Open-ended

Which browser do you use?

☐ Mozilla Firefox  ☐ Google Chrome
☐ Microsoft IE  ☐ Other (______________________ )

Closed
Information Questions

• Questions constructed according to how the information will be used

Please indicate your age: ________

Adv: Means and standard deviations can be computed in ratio-scale data.

Please indicate your age?

☐ < 20  ☐ 20-29  ☐ 30-39
☐ 40-49  ☐ 50-59  ☐ 60+

Which browser do you use? ________________

Which browser do you use?

☐ Mozilla Firefox  ☐ Google Chrome
☐ Microsoft IE  ☐ Other ( ______________________)
Information Questions

- Questions constructed according to how the information will be used

Please indicate your age: ________

Adv: Means and standard deviations can be computed in ratio-scale data.

Please indicate your age?
- [ ] < 20
- [ ] 20-29
- [ ] 30-39
- [ ] 40-49
- [ ] 50-59
- [ ] 60+

Adv: Grouping is useful if there are a large number of respondents.

Which browser do you use? ________________

Which browser do you use?
- [ ] Mozilla Firefox
- [ ] Google Chrome
- [ ] Microsoft IE
- [ ] Other (_______________________)
Information Questions

- Questions constructed according to how the information will be used

Please indicate your age: ________

Adv: Means and standard deviations can be computed in ratio-scale data.

Adv: Grouping is useful if there are a large number of respondents.

Which browser do you use? ________________

Adv: Closed-ended questions simplify follow-up analyses

Which browser do you use?
- Mozilla Firefox
- Google Chrome
- Microsoft IE
- Other (__________________ )
Participant Feedback

• Opinions and feelings:

**Frustration:** I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>Neutral</td>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• Subjective self-reports:

**Eye fatigue:**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very high</td>
<td>Very low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Likert Items**

**Frustration**: I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>Neutral</td>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Likert Items

- Respondent is asked to evaluate a statement (usually agree or disagree)

**Frustration**: I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>Neutral</td>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Likert Items

• Respondent is asked to evaluate a statement (usually agree or disagree)

• Responses are given a numeric code along with a visualization, usually a horizontal line on which they circle or check mark

<table>
<thead>
<tr>
<th>Frustration: I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>Strongly disagree</td>
</tr>
</tbody>
</table>


Likert Items

- Respondent is asked to evaluate a statement (usually agree or disagree)
- Responses are given a numeric code along with a visualization, usually a horizontal line on which they circle or check mark

**Frustration:** I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>Neutral</td>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Likert Items

- Respondent is asked to evaluate a statement (usually agree or disagree)
- Responses are given a numeric code along with a visualization, usually a horizontal line on which they circle or check mark

**Frustration**: I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>Neutral</td>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

32
Likert Items

- Respondent is asked to evaluate a statement (usually agree or disagree)
- Responses are given a numeric code along with a visualization, usually a horizontal line on which they circle or check mark

**Frustration:** I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>Neutral</td>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Image of Likert scale]

[Image of Likert scale]

[Image of Likert scale]
Likert Items

• Respondent is asked to evaluate a statement (usually agree or disagree)

• Responses are given a numeric code along with a visualization, usually a horizontal line on which they circle or check mark

Frustration: I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>Neutral</td>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

32
Likert Items

• Respondent is asked to evaluate a statement (usually agree or disagree)

• Responses are given a numeric code along with a visualization, usually a horizontal line on which they circle or check mark

<table>
<thead>
<tr>
<th>Frustration: I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1   2   3   4   5   6   7</td>
</tr>
<tr>
<td>Strongly disagree   Neutral   Strongly agree</td>
</tr>
</tbody>
</table>

• Central value is neutral with the two ends being extremes (bipolar)
Likert Items

• Respondent is asked to evaluate a statement (usually agree or disagree)

• Responses are given a numeric code along with a visualization, usually a horizontal line on which they circle or check mark.

| Frustration: I felt a high level of insecurity, discouragement, irritation, stress, or annoyance. |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Strongly disagree | Neutral | Strongly agree |

• Central value is neutral with the two ends being extremes (bipolar)

• Gradations are considered equally spaced
**Likert Items**

- Respondent is asked to evaluate a statement (usually agree or disagree)
- Responses are given a numeric code along with a visualization, usually a horizontal line on which they circle or check mark

<table>
<thead>
<tr>
<th>Number</th>
<th>Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>2</td>
<td>Neutral</td>
</tr>
<tr>
<td>3</td>
<td>Strongly agree</td>
</tr>
</tbody>
</table>

**Frustration:** I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.

- Central value is neutral with the two ends being extremes (bipolar)
- Gradations are considered equally spaced
- The verbal tags attached to the numbers can influence how people think about the items
Likert Items

- Respondent is asked to evaluate a statement (usually agree or disagree)
- Responses are given a numeric code along with a visualization, usually a horizontal line on which they circle or check mark

```
Frustration: I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>Neutral</td>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

- Central value is neutral with the two ends being extremes (bipolar)
- Gradations are considered equally spaced
- The verbal tags attached to the numbers can influence how people think about the items

# Likert Scale

**Frustration:** I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>Neutral</td>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Likert Scale

• Potential problems:
  • Central tendency (people tend to avoid extreme opinions)
  • Social desirability (portray themselves more favorably, or in accordance with majority opinion)
  • Agreement bias (people tend to agree with a statement rather than disagree)

<table>
<thead>
<tr>
<th>Frustration: I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>Strongly disagree</td>
</tr>
</tbody>
</table>
Likert Scale

• Potential problems:
  • Central tendency (people tend to avoid extreme opinions)
  • Social desirability (portray themselves more favorably, or in accordance with majority opinion)
  • Agreement bias (people tend to agree with a statement rather than disagree)

• How to alleviate these problems:
  • Choose the wording of the question carefully
  • Ask about the same underlying concept in different ways
  • Combine multiple items into a scale

**Frustration:** I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>Neutral</td>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Likert Scale

• Potential problems:
  • Central tendency (people tend to avoid extreme opinions)
  • Social desirability (portray themselves more favorably, or in accordance with majority opinion)
  • Agreement bias (people tend to agree with a statement rather than disagree)

• How to alleviate these problems:
  • Choose the wording of the question carefully
  • Ask about the same underlying concept in different ways
  • Combine multiple items into a scale

<table>
<thead>
<tr>
<th>Frustration: I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>Strongly disagree</td>
</tr>
</tbody>
</table>

Likert Scale
Likert Scale

• Interpretation:
  • Data is ordinal (1 is more disagreement than 2)
  • If participants perceive equal distance between categories, then it can be considered interval data (selecting 3 or 5 means respondent is as much in disagreement as in agreement with a statement)
Interpretation:
- Data is ordinal (1 is more disagreement than 2)
- If participants perceive equal distance between categories, then it can be considered interval data (selecting 3 or 5 means respondent is as much in disagreement as in agreement with a statement)

Analysis:
- Individual items are ordinal data: Report median, percentiles, Visualize as bar charts
- Scales measure an underlying concept by combining items, and thus, they can be thought of as interval data: Report mean, do parametric tests
- Researchers disagree!

Likert Scale
Likert Scale

• Interpretation:
  • Data is ordinal (1 is more disagreement than 2)
  • If participants perceive equal distance between categories, then it can be considered interval data (selecting 3 or 5 means respondent is as much in disagreement as in agreement with a statement)

• Analysis:
  • Individual items are ordinal data: Report median, percentiles, Visualize as bar charts
  • Scales measure an underlying concept by combining items, and thus, they can be thought of as interval data: Report mean, do parametric tests
  • Researchers disagree!

•
Likert Scale

- Interpretation:
  - Data is ordinal (1 is more disagreement than 2)
  - If participants perceive equal distance between categories, then it can be considered interval data (selecting 3 or 5 means respondent is as much in disagreement as in agreement with a statement)

- Analysis:
  - Individual items are ordinal data: Report median, percentiles, Visualize as bar charts
  - Scales measure an underlying concept by combining items, and thus, they can be thought of as interval data: Report mean, do parametric tests
  - Researchers disagree!

“Ten Common Misunderstandings…” by Carifio and Perla

Alternatives

• With online testing, more continuous measurements are possible

Classroom Exercise

- What is a well-written consent form?
- Project groups get together to write your consent forms.
- Gentle start to peer grading: Get consent form approved by a classmate who is not in your group. You are OK to leave once you have an approved consent form for use in your experiments.
- Bonus activity: Create an IRB application using myIRB.