HRI Questionnaire Analysis - Technical Documentation

Overview

Analysis of Dum-E Robot Interaction Study with 5 hypotheses tested using appropriate statistical methods.

---

1. Hypotheses and Sample Sizes

Hypothesis	Description	Sample	Rationale
H1	Cognitive load difference	n=6 (both-mode testers)	Paired design requires same participants in both conditions
H2	SUS by preference	n=7 (all participants)	SUS measured once for all
H3	Learnability vs tech background	n=7 (all participants)	Tech background measured for everyone
H4	Control-seeking → On-demand pref	n=6 (on-demand testers)	Control only measured for on-demand
H5	Safety → Sequence preference	n=7 (all participants)	Safety measured for everyone

---

2. Technical Background Score

Components

- **Self-assessment (70%)**: Mean of 7 familiarity items (1-5 scale)
- **Objective test (30%)**: DoF question (correct=1, incorrect=0)

Formula

```
tech_composite = 0.7 × self_assess_mean + 0.3 × dof_correct × 5
```

Results

Participant	Self-Assess	DoF	Composite
P0	2.86	✓	3.50
P1	1.43	✓	2.50

Participant	Self-Assess	DoF	Composite
P2	5.00	✓	5.00
P3	2.86	✓	3.50
P4	1.57	✗	1.10
P5	5.00	✓	5.00
P6	4.93	✗	3.45

3. SUS Score Computation

Formula (9 items, Q10 missing)

Positive items (Q1,Q3,Q5,Q7,Q9): score = raw - 1
Negative items (Q2,Q4,Q6,Q8): score = 5 - raw
SUS_total = Σ(scores) × 2.7778

Results

Participant	SUS Score	Grade
P0	80.56	B
P1	69.44	C
P2	94.44	A
P3	44.44	F
P4	86.11	A
P5	80.56	B
P6	77.78	B
Mean	**76.19**	**B**

4. Hypothesis Results

H1: Cognitive Load (Paired Wilcoxon)

- **Sequence attention** (n=7): M=4.14, SD=0.69
- **On-demand attention** (n=6): M=4.17, SD=0.98
- **Paired differences** (n=6): [+1, 0, +1, -1, 0, 0]
- **Wilcoxon**: W=2.0, p=1.0
- **Result**: NOT SIGNIFICANT

H2: SUS by Preference (Mann-Whitney)

- **On-demand pref** (n=4): M=78.47
- **Sequence pref** (n=1): M=94.44
- **Cannot test**: Need n≥2 per group

H3: Learnability vs Technical Background (Spearman)

- **Sample**: n=7 (ALL participants)
- **Spearman ρ** = 0.272, p=0.555
- **Result**: NOT SIGNIFICANT (weak positive trend)

H4: Control → On-demand Preference (Point-Biserial)

- **Sample**: n=6 (on-demand testers)
- **Point-biserial r** = -0.500, p=0.312
- **Result**: NOT SIGNIFICANT
- **Note**: Negative correlation suggests opposite direction

H5: Safety → Sequence Preference (Point-Biserial)

- **Sample**: n=7 (all participants)
- **Point-biserial r** = -0.091, p=0.846
- **Result**: NOT SIGNIFICANT

---

5. Reliability Analysis (Cronbach's α)

Scale	Items	α	Interpretation
SUS	9	0.829	Good

Scale	Items	α	Interpretation
Technical Background	7	0.973	Excellent
Godspeed Safety	4	0.670	Questionable
Safety Perception	4	0.466	Poor

Notes:

- H1 & H4 use single-item measures (α not applicable)
- Safety Perception scale (H5) has poor reliability → interpret with caution

---

6. Multi-Value Cell Handling

Cells with comma-separated values (e.g., "4, 5") are averaged:

```
"4, 5" → (4 + 5) / 2 = 4.5
```

Found: P6, Col 10: "4, 5" → 4.50

---

7. Key Findings

1. **All hypotheses NOT SUPPORTED** at α=0.05
2. **Small sample** (n=7) severely limits statistical power
3. **SUS reliability good** (α=0.829) despite missing Q10
4. **Technical background scale excellent** (α=0.973)
5. **Safety perception scale poor** (α=0.466) → items may not form coherent construct
6. **Overall usability rated Good** (SUS=76.2)

---

8. Script Usage

```
bash

python hri_analysis.py <excel_file.xlsx>
```

Output Files

- `analysis_report.txt` - Complete analysis log

- `processed_data.csv` - All computed variables
- `results.json` - Machine-readable results
- `hypothesis_plots.png` - 6-panel hypothesis visualization
- `distribution_plots.png` - 6-panel score distributions

---

9. Statistical Tests Used

Hypothesis	Test	Rationale
H1	Wilcoxon signed-rank	Paired data, non-parametric for small n
H2	Mann-Whitney U	Independent groups, non-parametric
H3	Spearman correlation	Continuous vs continuous, ordinal-appropriate
H4	Point-biserial correlation	Continuous vs binary
H5	Point-biserial correlation	Continuous vs binary