|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Question** | **Hypothesis** | **Sampling plan (e.g. power analysis)** | **Analysis Plan** | **Interpretation given to different outcomes** |
| 1. Do objective and subjective measures of performance reflect an increase in task load with increasing n-back level? | 1a) The signal detection measure d’ declines with increasing n-back level. | F tests - ANOVA: Repeated measures, within factors  Analysis: A priori: Compute required sample size  Input:  Effect size f = 0.8685540  α err prob = 0.05  Power (1-β err prob) = 0.95  Number of groups = 1  Number of measurements = 4  Corr among rep measures = 0.5  Nonsphericity correction ε = 1  Output:  Noncentrality parameter λ = 30.1754420  Critical F = 3.4902948  Numerator df = 3.0000000  Denominator df = 12.0000000  Total sample size = 5  Actual power = 0.9824202 | Repeated measures ANOVA with three linear contrasts, comparing the d’ value of two n-back levels (2, 3, 4) at a time.  The ANOVA is calculated using aov\_ez() of the afex-package, estimated marginal means are calculated using emmeans() from the emmeans-package, and pairwise contrasts are calculated using pairs(). | ANOVA yields p < .05 is interpreted as d’ changing significantly with n-back levels. Values of d’ are interpreted as equal between n-back levels if p > .05.  Each contrast yielding p < .05 is interpreted as d’ being different between those levels, magnitude and direction are inferred from the respective estimate. Values of d’ are interpreted as equal between n-back levels if p > .05. |
| 1b) Reaction time increases with increasing n-back level. | F tests - ANOVA: Repeated measures, within factors  Analysis: A priori: Compute required sample size  Input:  Effect size f = 0.2041241  α err prob = 0.05  Power (1-β err prob) = 0.95  Number of groups = 1  Number of measurements = 4  Corr among rep measures = 0.5  Nonsphericity correction ε = 1  Output:  Noncentrality parameter λ = 17.6666588  Critical F = 2.6625685  Numerator df = 3.0000000  Denominator df = 156  Total sample size = 53  Actual power = 0.9506921 | Repeated measures ANOVA with three linear contrasts, comparing the median reaction time of two n-back levels (2, 3, 4) at a time.  The ANOVA is calculated using aov\_ez() of the afex-package, estimated marginal means are calculated using emmeans() from the emmeans-package, and pairwise contrasts are calculated using pairs(). | ANOVA yields p < .05 is interpreted as the median reaction time changing significantly with n-back levels. Median reaction times are interpreted as equal between n-back levels if p > .05.  Each contrast yielding p < .05 is interpreted as the median reaction time being different between those levels, magnitude and direction are inferred from the respective estimate. Median reaction times are interpreted as equal between n-back levels if p > .05. |
| 1c) Ratings on all NTLX subscales increase with increasing n-back level. | From Kramer et al.:  F tests - ANOVA: Repeated measures, within factors  Analysis: A priori: Compute required sample size  Input:  Effect size f = 0.7071068  α err prob = 0.05  Power (1-β err prob) = 0.95  Number of groups = 1  Number of measurements = 4  Corr among rep measures = 0.5  Nonsphericity correction ε = 1  Output:  Noncentrality parameter λ = 24.0000013  Critical F = 3.2873821  Numerator df = 3.0000000  Denominator df = 15.0000000  Total sample size = 6  Actual power = 0.9620526 | A repeated measures ANOVA for each NASA-TLX subscale, with six linear contrasts comparing the subscale score of two n-back levels (1, 2, 3, 4) at a time.  The ANOVA is calculated using aov\_ez() of the afex-package, estimated marginal means are calculated using emmeans() from the emmeans-package, and pairwise contrasts are calculated using pairs(). | ANOVA yields p < .05 is interpreted as the subscale score changing significantly with n-back levels. The subscale scores are interpreted as equal between n-back levels if p > .05.  Each contrast yielding p < .05 is interpreted as the subscale score being different between those levels, magnitude and direction are inferred from the respective estimate. The subscale scores are interpreted as equal between n-back levels if p > .05. |
| 2. Is the effort required for higher n-back levels less attractive, regardless of how well a person performs? | 2a) Subjective values decline with increasing n-back level. | F tests - ANOVA: Repeated measures, within factors  Analysis: A priori: Compute required sample size  Input:  Effect size f = 0.9229582  α err prob = 0.05  Power (1-β err prob) = 0.95  Number of groups = 1  Number of measurements = 4  Corr among rep measures = 0.5  Nonsphericity correction ε = 1  Output:  Noncentrality parameter λ = 27.2592588  Critical F = 3.8625484  Numerator df = 3.0000000  Denominator df = 9.0000000  Total sample size = 4  Actual power = 0.9506771 | Repeated measures ANOVA with six linear contrasts, comparing the subjective values of two n-back levels (1, 2, 3, 4) at a time.  The ANOVA is calculated using aov\_ez() of the afex-package, estimated marginal means are calculated using emmeans() from the emmeans-package, and pairwise contrasts are calculated using pairs(). | ANOVA yields p < .05 is interpreted as subjective values changing significantly with n-back levels. Subjective values are interpreted as equal between n-back levels if p > .05.  Each contrast yielding p < .05 is interpreted as subjective values being different between those levels, magnitude and direction are inferred from the respective estimate. Subjective values are interpreted as equal between n-back levels if p > .05. |
| 2b) Subjective values decline with increasing n-back level, even after controlling for declining task performance measured by signal detection d’ and reaction time. | t tests - Linear multiple regression: Fixed model, single regression coefficient  Analysis: A priori: Compute required sample size  Input:  Tail(s) = One  Effect size f² = 0.34  α err prob = 0.05  Power (1-β err prob) = 0.95  Number of predictors = 3  Output:  Noncentrality parameter δ = 3.4000000  Critical t = 1.6972609  Df = 30  Total sample size = 34  Actual power = 0.9532263 |  |  |
| 2c) SVs decline stronger with increasing task load for individuals with low compared to high NFC scores. |  |  |  |
| 3. Is there a discrepancy between perceived task load and subjective value of effort depending on a person’s Need for Cognition? | 3a) Subjective values positively predict individual NFC scores. | t tests - Linear multiple regression: Fixed model, single regression coefficient  Analysis: A priori: Compute required sample size  Input:  Tail(s) = One  Effect size f² = 0.33  α err prob = 0.05  Power (1-β err prob) = 0.95  Number of predictors = 1  Output:  Noncentrality parameter δ = 3.3985291  Critical t = 1.6923603  Df = 33  Total sample size = 35  Actual power = 0.9537894 | Subjective values are regressed on NFC scores using the lm() function from the stats-package. | Subjective values are interpreted as predicting NFC scores if the slope yields p < .05. Direction and magnitude are inferred from the slope estimate. |
| 3b) NASA-TLX scores negatively predict individual NFC scores. | Westbrook et al. have only reported the p-value here, so we used the regression results of our pilot study, which included NASA-TLX scores and subjective values as predictors of NFC scores.  t tests - Linear multiple regression: Fixed model, single regression coefficient  Analysis: A priori: Compute required sample size  Input:  Tail(s) = One  Effect size f² = 1.10  α err prob = 0.05  Power (1-β err prob) = 0.95  Number of predictors = 2  Output:  Noncentrality parameter δ = 3.6331804  Critical t = 1.8331129  Df = 9  Total sample size = 12  Actual power = 0.9552071 | Subjective values and the area under the curve of each subject’s NASA-TLX scores are regressed on NFC scores using the lm() function from the stats-package. | Subjective values and NASA-TLX scores are interpreted as predicting NFC scores if their slope yields p < .05. Direction and magnitude are inferred from the slope estimate. |