Leisure Activities - Meths & Stats

Melissa Schneiderová, Josef Mana, and Hana Georgi

OVV, PVSPS

Author Note

Correspondence concerning this article should be addressed to Melissa SchneiderováHana Georgi

Leisure Activities - Meths & Stats

Methods

Statistical analyses

VLS-ALQ

The dimensionality of the VLS-ALQ questionnaire scale in our sample was first checked via the confirmatory factor analysis (CFA) as implemented in the *lavaan* R package (R Core Team, 2024; Rosseel, 2012). A three-factor model with separate 'Private,' 'Public,' and 'Religious' independent dimensions, a second order factor model with separate 'Private,' 'Public,' and 'Religious' and a superordinate 'Social' factor, and separate unidimensional models for the 'Private' and the 'Public' subscales. Models were evaluated by Tucker Lewis Index (TLI), Comparative Fit Index (CFI) and root-mean-square-error-aproximation (RMSEA) with values of TLI > .9, CFI > .9, and RMSEA < .08 considered indicating adequate fit. Next, internal consistency of the full questionnaire and the 'Public' and the 'Private' subscales were estimated via the Cronbach's α and Guttman's Lambda 6 (G6). Internal consistency indexes were estimated in the *psych* R package using the default settings (R Core Team, 2024; Revelle, 2024). Following the classical recommendations for interpretation of coefficient α (Nunnally & Bernstein, 1994; Streiner, 2003), we consider values above 0.7 as sufficient for early research, values above 0.8 as sufficient for basic research, and values above 0.9 as necessary for clinical use.

Sum scores of the full VLS-ALQ questionnaire as well as its 'Private' and 'Public' subscales were described by their in-sample means ± standard deviations for SA and non-SA groups separately. The null hypothesis that SA and non-SA groups' means are equal was tested via the independent sample *t*-test with Welch modification to the degrees of freedom and effect size of mean differences characterised by Cohen's d. The null hypothesis that regarding test scores in SA and non-SA groups, neither distribution is stochastically greater than the other, was tested via the Mann-Whitney U test with effect size characterised by Vargha and Delaney A as implemented in the *effsize* R package (Torchiano, 2020; Vargha & Delaney, 2000). Decision threshold for claiming statistically significant deviance from a null hypothesis in these comparisons was set at p < .05.

COBRA-A

To be added, by someone else hopefully :-)

Leisure activities

The self-reported leisure activities' distributions were described in three ways: (i) as counts of activity type per participant, and (ii) as a number of participants reporting at least one activity of each activity type, and (iii) as a time-series of expected means of self-reported intensity ratings.

The counts of activity type were then described separately for SA and non-SA groups stratified by modality (physical, mental, both) and seasoness (seasonal, non-seasonal) by their in-sample means ± standard deviations. Null hypotheses of equal means and stochastical equality of SA versus non-SA count distributions were tested via *t*-test with Welch modification to the degrees of freedom and the Mann-Whitney U test respectively. Effect sizes of differences between SA and non-SA means or distributions were described by Cohen's d and Vargha and Delaney A as implemented in the *effsize* R package (Torchiano, 2020; Vargha & Delaney, 2000). Decision threshold for claiming statistically significant deviance from a null hypothesis in these comparisons was set at p < .05.[1]

The number of participants reporting at least one activity belonging to a given activity type was described by contingency tables separately for each modality (physical, mental, both), seasoness (seasonal, non-seasonal), and SA status (SA, non-SA). Within each modality and seasoness combination, the Pearson's χ^2 test of the null hypothesis that SA status and activity

---

[1] *Note from Pepa, not to be included in the final text. I used the 'vanilla approach' of testing nulls about mean differences and stochastical equality, however, the very viable alternative would be a set of GLMs (regressions) with Poisson or Negative Binomial likelihood which are good candidates for modelling count variables. The choice between methods depends on the type of answer we are looking for (i.e., the estimand). Do we want just compare distributions of these counts with not many assumptions? Then the vanilla approach might work well. Or do we want a way to compare distributions and derive sensible predictions? Then the 'regression approach' would be better. Do we want causal instead of descriptive estimates? Then we need to put more work into the pre-statistic scientific model building. Pick your poison I guess :-D*

type are independent of each other was performed. The strength of association between SA status and activity type was described by Cramérs' V. Decision threshold for claiming statistically significant deviance from a null hypothesis in these comparisons was set at p < .05.

The expected means of self-reported intensities of single leisure activities across lifetime was modelled via a set of Bayesian Generalised Linear Mixed Models (GLMMs) with monotonic effect of time (Bürkner, 2017a; Bürkner & Charpentier, 2020). The models specified below and in the Appendix were fitted to data and used to generate posterior draws of the expected value of the intensity conditional on different combinations of life decade, SA status, seasoness and modality as predictors of interest. These posterior draws were then described by their medians and 95% equal tailed posterior intervals (ETIs).

Since there was no mental activity that would also be seasonal, two GLMMs were fitted. In the first model, the self-reported intensity of non-seasonal activities was regressed on the life decade (monotonic ordered predictor), SA status, modality, and their two- and three-way interactions on group-level as well as correlated participant-level and activity-level intercept and slopes across decades. This model was used to estimate expectations of intensity ratings conditional on decade of life, SA status and activity modality. In the second model, the self-reported intensity of physical activities was regressed on the life decade (monotonic ordered predictor), SA status, seasoness, and their two- and three-way interactions on group-level as well as correlated participant-level and activity-level intercept and slopes across decades. This model was used to estimate expectations of intensity ratings conditional on decade of life, SA status and seasoness of physical activities. See the Appendix for full model specifications.

All GLMMs were fitted using Stan's (version 2.35.0) build-in Hamiltonian Monte Carlo sampler accessed via R version 4.3.3 using package "brms" (Bürkner, 2017b; R Core Team, 2024; Stan Development Team, 2024). Four parallel chains were run each for 2,000 iterations for each GLMM. The first 1,000 iterations served as a warm-up and were discarded. Convergence was checked numerically by inspection of the $\hat{R}s$ and visually by inspection of trace plots. Posterior predictive checks for each combination of predictors was used to ensure the models reliably

reproduce in-sample means. This study's design and its analysis were not pre-registered. The data are not publicly available due to privacy or ethical restrictions. The computer code used in the data analysis can be accessed at https://github.com/josefmana/cosactiw-retro-activities.git.

Results

Overall, there was a total of 106 women categorised as SA(82.50 ± 2.56 years old, 14.17 ± 3.00 years of education), and 119 women categorised as non-SA (83.11 ± 3.21 years old, 13.80 ± 3.36 years of education). The null hypothesis of equal means was not rejected for neither age ($t(220.34) = -1.580$, $p = .116$) nor education ($t(220.88) = 0.880$, $p = .380$).

VLS-ALQ

The total VLS-ALQ scale ($\alpha = 0.588$, G6 = 0.619), as well as the private subscale ($\alpha = 0.582$, G6 = 0.557) and public subscale ($\alpha = 0.576$, G6 = 0.556) showed low levels of internal consistency in the current sample (see also Table A1). Consequently, the following results ought to be interpreted with the low internal consistency in mind as this could severely lower statistical power. The statistical comparisons of means and distributions of VLS-ALQ scores in non-SA versus SA participants is presented in Table 1. No null hypothesis was rejected based on our data.

Leisure activities

Counts of activity types

Number of participants reporting an activity

Intensity ratings

Table 1

Description and statistical comparison of VLS-ALQ scale scores

| | Descriptive statistics | | Inferential statistics | | | | | | |
| | | | t-test | | | | Mann Whitney U test | | |
Scale	non-SA	SA	Cohen's d	t	df	p	VD.A	W	p
Total	47.51 ± 8.89	48.07 ± 8.21	-0.06 [-0.33, 0.20]	-0.48	220.92	0.630	0.48	5930	0.574
Private	33.54 ± 5.89	34.06 ± 5.17	-0.09 [-0.36, 0.17]	-0.70	221.89	0.487	0.47	5833	0.384
Public	11.72 ± 5.63	11.47 ± 5.34	0.05 [-0.22, 0.31]	0.34	222.11	0.732	0.51	6444	0.779
Religious	2.51 ± 2.09	2.54 ± 2.14	-0.01 [-0.28, 0.25]	-0.10	218.28	0.918	0.49	6159	0.828

Appendix

Table A1

Confirmatory factor analyses and internal consistency estimates of the VLS-ALQ scale
N: number of participants; K: number of parameters of the confirmatory factor analysis model; RMSEA: root-mean-square error approximation; CI: confidence interval; CFI: comparative fit index, TLI: Tucker-Lewis Index; raw: Cronbach's alpha; based upon the covariances; stand.: Cronbach's alpha; based upon the correlations, G6: Guttman's lambda; 6 reliability; r: the average inter-item correlation; for RMSEA, values less than 0.08 indicating an adequate model fit (especially values of upper 90% CI bound lower than 0.08), for CFI and TLI, values exceeding 0.90 are considered to indicate a good model fit.

Bürkner, P.-C. (2017a). *Advanced bayesian multilevel modeling with the r package brms*.
https://arxiv.org/abs/1705.11123

Bürkner, P.-C. (2017b). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C., & Charpentier, E. (2020). Modelling monotonic effects of ordinal predictors in bayesian regression models. *British Journal of Mathematical and Statistical Psychology*, *73*(3), 420–451. https://doi.org/10.1111/bmsp.12195

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw-Hill Companies,Incorporated. https://books.google.cz/books?id=r0fuAAAAMAAJ

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Revelle, W. (2024). *Psych: Procedures for psychological, psychometric, and personality research*. Northwestern University. https://CRAN.R-project.org/package=psych

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Stan Development Team. (2024). *Stan modeling language users guide and reference manual, version 2.21.0.* http://mc-stan.org/

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103. https://doi.org/10.1207/S15327752JPA8001/_18

Torchiano, M. (2020). *Effsize: Efficient effect size computation.* https://doi.org/10.5281/zenodo.1480624

Vargha, A., & Delaney, H. D. (2000). A critique and improvement of the CL common language effect size statistics of McGraw and wong. *Journal of Educational and Behavioral Statistics*, *25*(2), 101–132. https://doi.org/10.3102/10769986025002101

| Scale | N | K | Confirmatory factor analysis | | | | Internal consistency | | | |
| | | | RMSEA | | | | Cronbach's alpha | | | |
			estimate	90% CI	CFI	TLI	raw	stand.	G6	r
Total	225	36	0.088	[0.070, 0.105]	0.57	0.47	0.57	0.59	0.62	0.11
Multidimensional	225	38	0.056	[0.035, 0.076]	0.83	0.78	-	-	-	-
Second order	225	38	0.056	[0.035, 0.076]	0.83	0.78	-	-	-	-
Private	225	18	0.071	[0.013, 0.121]	0.89	0.82	0.58	0.58	0.56	0.19
Public	225	15	0.113	[0.062, 0.169]	0.86	0.72	0.56	0.58	0.56	0.21