

# Homework assignment

## L3: Validity

---

**Assignment date:** 16.10.2018  
**Deadline:** 22.10.2018 23:59  
**Slides:** <http://www.cs.cas.cz/martinkova/NMST570>  
**Note:** Send answers and R script to drabinova@cs.cas.cz  
**Name:**

---

## 1 Reading and ShinyItemAnalysis

**Ex. 1.1** Read following article and answer questions:

<https://doi.org/10.1187/cbe.16-10-0305>

1. What is a conceptual test? [0.25]
2. Development of which conceptual test is described in the paper? [0.25]
3. In how many iterations was the HCI test developed? Who was involved in development and on whome was the test pre-tested? (**Methods**) [0.75]
4. How many students from how many institutions participated in the large-scale testing of the HCI test? (**Methods/Homeostasis Concept Inventory**) [0.5]
5. What methods were used for validation of the HCI test? (**Table 4**) [1.25]

**Ex. 1.2** Run `ShinyItemAnalysis` (online or locally), change data to HCI and answer following questions:

1. Which items do correlate with item 9 the most? (**Validity/Correlation structure**) [0.5]
2. Read wording of these items in the supplement of the paper and try to explain why is the correlation between them increased? [0.75]
3. HCI dataset includes criterion variable - indicator whether student plans to major in the life sciences. What is the correlation between criterion variable and total score? Briefly interpret. (**Validity/Criterion validity**) [0.5]

## 2 Correction for restriction of range

**Ex. 2.1** 2,254 examinees were given theoretical test with 65 dichotomous items. Those examinees who passed the theoretical test (had score at least 52 points) were given practical test which consists of five parts. To pass the practical test, examinee needed to succeed in all five parts (i.e., gain score of 5 points).

Download data available at

[http://www.cs.cas.cz/drabinova/documents/test\\_theory\\_practice.RData](http://www.cs.cas.cz/drabinova/documents/test_theory_practice.RData)

and create R script to answer following questions.

1. How many examinees passed theoretical test? What is the pass rate of the theoretical test? [0.5]
2. How many examinees passed practical test? What is the pass rate of the practical test in restricted sample (i.e., examinees who passed theoretical test)? [0.5]
3. Calculate observed correlation between scores of theoretical and practical tests in restricted sample (i.e., examinees who passed theoretical test) [0.5]
4. Use formula to correct for restriction of range to estimate correlation between scores of theoretical and practical tests in unrestricted sample:

$$\text{cor}(X, Y) = \frac{\sigma_X \text{cor}(x, y)}{\sqrt{\sigma_X^2 \text{cor}(x, y)^2 + \sigma_x^2 - \sigma_x^2 \text{cor}(x, y)^2}},$$

where  $X$  and  $Y$  are scores of unrestricted sample,  $x$  and  $y$  are scores of restricted sample,  $\sigma_X^2$  and  $\sigma_x^2$  are variances of  $X$  and  $x$  (score in the first test of unrestricted and restricted sample). [0.75]

HINT: You can use `attach(data)` to simply use variables by calling their name (e.g. `score_theory` instead of `data$score_theory` or `data[, "score_theory"]`). When missing values are present, you can add `na.rm = T` in some functions to account for them (e.g. `mean(x, na.rm = T)`). When calculating covariance or correlation, you need to use argument `use = "complete.obs"` (e.g. `cov(x, y, use = "complete.obs")`).

**Ex. 2.2** Assume that score of the first test  $X$  is normally distributed with expected value  $EX = 5$  and standard deviation  $\text{sd}(X) = 1$ , i.e.  $X \sim \mathcal{N}(5, 1)$ . Score of the second test  $Y$  is linearly dependent on the score of the first test by formula  $Y = 2X - 1 + e$ , where  $e$  is error term with expected value  $Ee = 0$  and standard deviation  $\text{sd}(e) = 2$ , i.e.  $e \sim \mathcal{N}(0, 2)$ , where  $Y$  and  $e$  are independent. Theoretical correlation between scores  $X$  and  $Y$  is 0.707.

1. Using R create script to generate  $X$  and  $Y$  of sample size 1000. Use `set.seed(123)` for reproducibility. [0.25]
2. Calculate estimate of correlation between generated scores  $X$  and  $Y$ . Compare to theoretical correlation. [0.25]
3. Consider only those respondents who are
  - at least in 85th percentile
  - at most in 15th percentile
  - at least in 15th percentile but at most in 85th percentile
  - at most in 15th percentile or at least in 85th percentile

Recalculate the estimate of correlation, apply formula for correction for restriction of the range, compare results and briefly comment. [2.5]

Scenario	Correlation in restricted sample	Correlation after correction
$\geq 85\text{th percentile}$		
$\leq 15\text{th percentile}$		
$\geq 15\text{th percentile} \ \& \ \leq 85\text{th percentile}$		
$\leq 15\text{th percentile} \ \text{or} \ \geq 85\text{th percentile}$		

### 3 Provide feedback

Here you can provide feedback on lecture, lab session and/or materials (slides, HW assignment, ShinyItemAnalysis manual) [1pt bonus] :)