Exercises 2

Björn Andersson

University of Oslo

May 24, 2022

L4 Task 1

Consider a random variable X which takes values 1, 2, 3 and 4 with corresponding probabilities 0.1, 0.2, 0.4 and 0.3.

a What is the median?

b What is the mode?

L4 Task 2

Which of the following are symmetric distributions?

a A $t(49)$-distribution.

b A $\chi^2(1)$-distribution.

c For the R.V. X, we have $P(X = 1) = 0.3$, $P(X = 2) = 0.15$, $P(X = 3) = 0.1$, $P(X = 4) = 0.15$ and $P(X = 5) = 0.3$.

L5 Task 1

Two alternate-form tests were given to the same population, yielding estimated standard deviations of $s_1 = 7.5$ and $s_2 = 8$ and an estimated covariance of $s_{12} = 12$. Calculate the estimated standard error of measurement for the test.

L5 Task 2

- We have autumn (A) and spring (S) test-takers of the SweSAT, with A having a $N(\mu_A = 60, \sigma_A^2 = 12)$ true score distribution and S having a $N(\mu_S = 70, \sigma_S^2 = 16)$ true score distribution.
- The test takers are given the same test which is defined by the classical true score model with an error score distribution of $N(0, \sigma_E^2 = 4)$.

Calculate the reliability of the test for populations A and S.

L6 Task 1

The following estimated factor loadings and error variances were obtained from a single factor model for five mathematics items.

	Item 1	Item 2	Item 3	Item 4	Item 5
$\hat{\lambda}_j$	0.109	0.202	0.322	0.068	0.148
$\hat{\Psi}_j^2$	0.196	0.173	0.140	0.173	0.147

Calculate the estimated coefficient omega.

L6 Task 2

A restricted model with the factor loadings set equal was also estimated. The estimated factor loading was $\lambda = 0.164$ and the error variances are given in the table below.

	Item 1	Item 2	Item 3	Item 4	Item 5
$\hat{\Psi}_j^2$	0.181	0.188	0.217	0.151	0.142

Calculate the estimated coefficient alpha/omega.

Extra task 1

Consider a random variable X which takes values 1, 2, 3, 4, 5 with corresponding probabilities 0.1, 0.1, 0.2, 0.2 and 0.4.

a What is $E(X)$?

b What is $E(X^2)$?

c What is $\mathrm{Var}(X)$?

Extra task 2

For two binary random variables X_1 and X_2 we have $P(X_1 = 1) = 0.5$, $P(X_2 = 1) = 0.4$ and $P(X_1 = 1, X_2 = 1) = 0.25$. Are X_1 and X_2 independent?

Extra task 3

Two random variables X and Y have variances $\sigma_X^2 = 25$ and $\sigma_Y^2 = 100$ and covariance $\sigma_{X,Y} = 30$.

a What is $\mathrm{Cor}(X, Y)$?

b What is $\mathrm{Cov}(2X, 5Y)$?

c What is $\mathrm{Cor}(2X, 5Y)$?

Extra task 4

Let $x_1, \ldots, x_n$ be a random sample from a population, where $n > 2$. We want to estimate the expected value $\mu = E(X)$ and use an estimator based on the first two observations in the sample:

$$\hat{\mu} = \frac{\sum_{i=1}^{2} x_i}{2}.$$

a Derive the bias of $\hat{\mu}$ with respect to μ.

b Derive the variance of $\hat{\mu}$ and compare it to the variance of the sample mean $\bar{x} = \sum_{i=1}^{n} x_i / n$. Which estimator do you prefer?

L4 Task 1: solution

a) Since $P(X \leq 3) > 0.5$ and $P(X \geq 3) > 0.5$, we have that the median is 3.
b) Since $P(X = 3) > P(X = 1)$, $P(X = 3) > P(X = 2)$ and $P(X = 3) > P(X = 4)$, we have that the mode is 3.

L4 Task 2: solution

a) Yes.
b) No.
c) Yes.

L5 Task 1: solution

We first estimate the reliability:

$$\hat{\rho}_{Y,Y'} = \frac{s_{12}}{s_1 \times s_2} = \frac{12}{7.5 \times 8} = 0.2$$

We then obtain the standard error of measurement from:

$$\widehat{\text{SEM}}(Y) = \sqrt{s_Y^2 \times (1 - \hat{\rho}_{Y,Y'})} = \sqrt{7.5 \times 8 \times (1 - 0.2)} \approx 6.93$$

L5 Task 2: solution

Remember that the reliability coefficient is $\rho_{Y,Y'} = \sigma^2_T/\sigma^2_Y$. We have, for population A,

$$\rho^A_{Y,Y'} = \frac{\sigma^2_A}{\sigma^2_{Y(A)}} = \frac{\sigma^2_A}{\sigma^2_A + \sigma^2_E} = \frac{12}{12+4} = \frac{12}{16} = 0.75$$

and, for population S,

$$\rho^S_{Y,Y'} = \frac{\sigma^2_S}{\sigma^2_{Y(S)}} = \frac{\sigma^2_S}{\sigma^2_S + \sigma^2_E} = \frac{16}{16+4} = \frac{16}{20} = 0.8.$$

L6 Task 1: solution

Coefficient omega is equal to the reliability of the sum score when a single factor model holds. We obtain the following estimate of the reliability with the estimated factor model parameters:

$$\hat{\omega} = \frac{(\sum_{j=1}^{5} \hat{\lambda}_j)^2}{(\sum_{j=1}^{5} \hat{\lambda}_j)^2 + \sum_{j=1}^{5} \hat{\Psi}_j^2} \approx \frac{0.721}{0.721 + 0.829} \approx 0.465.$$

If the factor model is the true model, the estimator of coefficient omega is an unbiased estimator of the reliability of the sum score.

L6 Task 2

$$\hat{\omega} = \frac{5^2 \times \hat{\lambda}^2}{5^2 \times \hat{\lambda}^2 + \sum_{j=1}^{5} \hat{\Psi}_j^2} = \frac{5^2 \times 0.164^2}{5^2 \times 0.164^2 + 0.879} \approx 0.433.$$

Extra task 1: solution

Consider a random variable X which takes values 1, 2, 3, 4, 5 with corresponding probabilities 0.1, 0.1, 0.2, 0.2 and 0.4.

a What is $E(X)$?

b What is $E(X^2)$?

c What is $\mathrm{Var}(X)$?

Answer:

a $E(X) = 0.1 + 0.2 + 0.6 + 0.8 + 2 = 3.7$

b $E(X^2) = 0.1 \times 1^2 + 0.1 \times 2^2 + 0.2 \times 3^2 + 0.2 \times 4^2 + 0.4 \times 5^2 = 15.5$

c $\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = 15.5 - 3.7^2 = 1.81$

Extra task 2: solution

For two binary random variables X_1 and X_2 we have
$P(X_1 = 1) = 0.5$, $P(X_2 = 1) = 0.4$ and $P(X_1 = 1, X_2 = 1) = 0.25$.
Are X_1 and X_2 independent?
Answer: Independence means that
$P(X_1 = 1, X_2 = 1) = P(X_1 = 1) \times P(X_2 = 1)$. We have that

$$\begin{aligned}
P(X_1 = 1) \times P(X_2 = 1) =& 0.5 \times 0.4 \\
=& 0.2 \\
\neq& P(X_1 = 1, X_2 = 1) \\
=& 0.25
\end{aligned}$$

Thus, we have that X_1 and X_2 are not independent.

Extra task 3: solution

Two random variables X and Y have variances $\sigma_X^2 = 25$ and $\sigma_Y^2 = 100$ and covariance $\sigma_{X,Y} = 30$.

a What is $\mathrm{Cor}(X, Y)$?

b What is $\mathrm{Cov}(2X, 5Y)$?

c What is $\mathrm{Cor}(2X, 5Y)$?

$$\mathrm{Cor}(X, Y) = \frac{\mathsf{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{30}{5 \times 10} = 30/50 = 0.6$$

$$\mathrm{Cov}(2X, 5Y) = 2 \times 5 \times \mathsf{Cov}(X, Y) = 10 \times 30 = 300$$

$$\mathrm{Cor}(2X, 5Y) = \frac{2 \times 5 \times \mathsf{Cov}(X, Y)}{2 \times \sigma_X \times 5 \times \sigma_Y} = \mathrm{Cor}(X, Y) = 0.6$$

Extra task 4: solution

Consider a sample of size $n > 2$. We have two estimators of the expected value: the regular sample mean $\bar{x} = \sum_{i=1}^{n} x_i/n$ and the sample mean for the first two observations $\bar{x}_2 = \sum_{i=1}^{2} x_i/2$. Which estimator do you prefer?

Answer: We can compare the mean squared error (MSE) of the estimators. Recall that MSE $=$ Variance $+$ Bias^2. From a previous exercise we obtain

$$\mathrm{MSE}(\bar{x}) = \mathrm{Var}(\bar{x}) + (\mathrm{Bias}(\bar{x}))^2 = \frac{\mathrm{Var}(X)}{n} + 0^2 = \frac{\mathrm{Var}(X)}{n}.$$

We can note that $\mathrm{Var}(\bar{x}_2) = \frac{\mathrm{Var}(X)}{2}$ and $\mathrm{Bias}(\bar{x}_2) = 0$. We thus have

$$\mathrm{MSE}(\bar{x}_2) = \frac{\mathrm{Var}(X)}{2} + 0^2 = \frac{\mathrm{Var}(X)}{2}.$$

Since $\frac{\mathrm{Var}(X)}{n} < \frac{\mathrm{Var}(X)}{2}$ if $n > 2$, we prefer $\bar{x}$ over $\bar{x}_2$ based on the MSE.

Some further notes

$$
\begin{aligned}
\operatorname{Cov}\left(\sum_{j=1}^{J} X_j, \sum_{j=1}^{J} X_j\right) =& \operatorname{Cov}(X_1 + X_2 + \cdots + X_J, X_1 + X_2 + \cdots + X_J) \\
=& \operatorname{Cov}(X_1, X_1 + X_2 + \cdots + X_J) + \cdots + \\
& \operatorname{Cov}(X_J, X_1 + X_2 + \cdots + X_J) \\
=& \operatorname{Cov}(X_1, X_1) + \cdots + \operatorname{Cov}(X_1, X_J) + \\
& \operatorname{Cov}(X_2, X_1) + \cdots + \operatorname{Cov}(X_2, X_J) \\
& + \cdots + \\
& \operatorname{Cov}(X_J, X_1) + \cdots + \operatorname{Cov}(X_J, X_J) \\
=& \sum_{j=1}^{J} \sum_{k=1}^{J} \operatorname{Cov}(X_j, X_k).
\end{aligned}
$$