HW9 – Solution

1. Estimator for $Cov(Y,Z)$

$$\widehat{Cov}(Y,Z) := \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)\cdot(Z_i - \overline{Z}_n)$$

where $\overline{Y}_n := \frac{1}{n}\sum_{i=1}^{n} Y_i$; $\overline{Z}_n := \frac{1}{n}\sum_{i=1}^{n} Z_i$

Same for $Var(Y) = Cov(Y,Y)$; $Var(Z) = Cov(Z,Z)$.

2. $$\hat{\rho} = \frac{\widehat{Cov}(Y,Z)}{\sqrt{\widehat{Var}(Y)\,\widehat{Var}(Z)}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y}_n).(Z_i - \overline{Z}_n)}{\left[\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}(Z_i - \overline{Z}_n)^2\right)\right]^{1/2}}$$

3. By CLT, since

$(Y_i)_{i\geq 1}$; $(Z_i)_{i\geq 1}$; $(Y_iZ_i)_{i\geq 1}$ are sequences of iid random variables

(<u>Note</u> : in the question, we should have assumed existence of second moment for all these quantities), we have

$$\left(\overline{Y}_n - \mathbb{E}[Y]\right)\frac{\sqrt{n}}{\sqrt{Var(Y)}} \xrightarrow[n\to\infty]{\text{distr.}} \text{standard normal r.v.}$$

$$\left(\overline{Z}_n - \mathbb{E}[Z]\right)\frac{\sqrt{n}}{\sqrt{Var(Z)}} \xrightarrow[n\to\infty]{\text{distr.}} \quad //$$

$$\left(\overline{YZ}_n - \mathbb{E}[YZ]\right)\frac{\sqrt{n}}{\sqrt{Var(YZ)}} \xrightarrow[n\to\infty]{\text{distr.}} \quad //$$

4] It means computing

$Var(\overline{Y}_n)$ ~~Cov~~ $Cov(\overline{Y}_n, \overline{Z}_n) = 0$ because $(Y_i)'s$ and $(Z_i)'s$ are $\perp\!\!\!\perp$

$Var(\overline{Z}_n)$ $\quad Cov(\overline{Y}_n, \overline{YZ}_n)$

$Var(\overline{YZ}_n)$ $\quad Cov(\overline{Z}_n, \overline{YZ}_n)$

$$\mathrm{Var}(\overline{Y_n}) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(Y_i) = \frac{\mathrm{Var}(Y)}{n}$$

$$\mathrm{Var}(\overline{Z_n}) = \frac{\mathrm{Var}(Z)}{n}$$

because Y, Z are $\perp\!\!\!\perp$

$$\mathrm{Var}(\overline{YZ_n}) = \frac{\mathrm{Var}(YZ)}{n} = \frac{\mathbb{E}[Y^2]\mathbb{E}[Z^2] - \mathbb{E}[Y]^2\mathbb{E}[Z^2]}{n}$$

$\mathrm{Cov}(\overline{Y_n}, \overline{YZ_n})$ needs a computation

$$= \mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i, \frac{1}{n}\sum_{j=1}^{n} Y_j Z_j\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(Y_i, Y_j Z_j)$$

If $i \neq j$, $\underset{=0}{\mathrm{Cov}(Y_i, Y_j Z_j)}$ because Y_i, Y_j are $\perp\!\!\!\perp$, Y_i, Z_j are $\perp\!\!\!\perp$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Cov}(Y_i, Y_i Z_i)$$

$$= \frac{1}{n}\mathrm{Cov}(Y, YZ) = \frac{1}{n}\left(\mathbb{E}[Y^2 Z] - \mathbb{E}[Y]\mathbb{E}[YZ]\right)$$

$$= \frac{1}{n}\left(\mathbb{E}[Y^2]\mathbb{E}[Z] - \mathbb{E}[Y]^2\mathbb{E}[Z]\right)$$

because Y, Z are $\perp\!\!\!\perp$.

$$= \frac{1}{n}\mathbb{E}[Z]\,\mathrm{Var}(Y).$$

Similarly, $\mathrm{Cov}(\overline{Z_n}, \overline{YZ_n}) = \frac{1}{n}\mathbb{E}[Y]\,\mathrm{Var}[Z]$.

5. Markov (or Bienaymé–Tchebychev) inequality implies

$$\mathbb{P}[|\hat{\rho}| \geq c] = \mathbb{P}[|\hat{\rho}|^2 \geq c^2] \leq \frac{1}{c^2}\mathbb{E}[\hat{\rho}^2]$$

$$\leq \frac{1}{c^2}\cdot\frac{1}{n} \boxed{= \frac{1}{nc^2}}$$

6. Test statistic $\hat{\rho}$

Rejection region $\{|x| \geq c\}$

7. Type I error $\leq \frac{1}{nc^2}$ from Q.5. Type II is impossible to control, because no correlation does not imply $\perp\!\!\!\perp$.