NRES 746, Week 14, Lecture 1: Neighbors (take-home)

Name: _____

Due Monday, Nov. 30. Work on your own. Parts 1 and 2 are on paper; you'll need R for Parts 3 and 4.

A water district measures snow water equivalent (SWE, the depth of water in the snowpack, in cm) on April 1 at 60 snow courses spread over a 60 by 40 km mountain range. It wants to know how fast SWE rises with elevation, and how much snow to expect at a site where it is thinking of adding a station. Storms don't deliver snow evenly, so stations near each other tend to be alike in ways elevation doesn't explain.

Part 1: Two stations (on paper)

Two stations have SWE values y_1 and y_2. Each has variance σ^2, and their correlation is ρ.

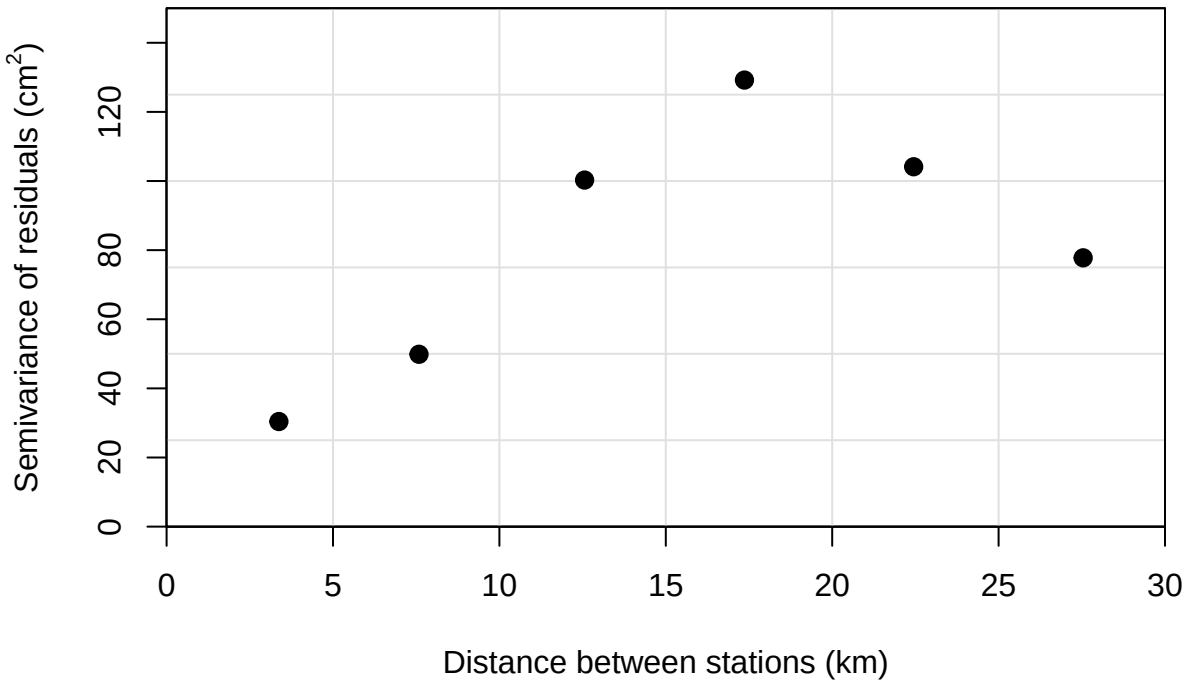
1a. Find the variance of their mean, $\bar{y} = (y_1 + y_2)/2$, in terms of σ^2 and ρ.

1b. Fill in the table. n independent stations have $\mathrm{Var}(\bar{y}) = \sigma^2/n$; the last column asks what n would give the same variance as these two correlated stations.

ρ	$\mathrm{Var}(\bar{y})/\sigma^2$	**equivalent number of independent stations**
0		
0.5		
0.9		

Part 2: Reading a variogram (on paper)

A regression of SWE on elevation leaves residuals. The figure shows their semivariance against the distance between stations, in 5-km bins.



2a. Sketch a smooth variogram curve through the points, and read off rough values:

nugget _____ sill _____ range _____ km

2b. If each residual is a spatially smooth part plus independent noise, then $\gamma(d) = \sigma_e^2 + \sigma_s^2\,(1 - \rho(d))$. From your curve, about what is the correlation between the residuals of two stations 5 km apart? Show how you got it.

Part 3: The spatial model, by hand

Each coding question has a label:

- *Recommended: no AI*: this week's core skill. R's help pages, web search, the Week 14 page, and classmates are fine.
- *AI optional/encouraged*: the results matter, but the code isn't the point.

After each coding question, circle how you used AI.

What to turn in: your filled-in answers and **Figure 3a** as an image on WebCampus. You don't need to submit code beyond the two lines asked for in 3b and 4b.

Read the data with

```
snow <- read.csv(
  "https://kevintshoemaker.github.io/NRES-746/data/SnowCourses.csv")
snow$elev_km <- (snow$elev_m - 2400) / 1000      # km above 2400 m
```

The columns are the station, its easting and northing in km (`x_km`, `y_km`), elevation (`elev_m`), and April 1 SWE (`swe_cm`).

3a. *(AI optional/encouraged.)* Fit `lm(swe_cm ~ elev_km, data = snow)`. Make a map of the stations with the symbol size or color showing each residual. Name it **Figure 3a**.

SWE rises _____ cm per km of elevation (SE _____)

AI use (circle one): no AI / AI helped (e.g., debugging) / AI wrote the code

The spatial model is

$$\mathbf{y} \sim \text{MVN}(X\boldsymbol{\beta},\ V), \qquad V_{ij} = \sigma_s^2 \exp(-d_{ij}/\ell) + \sigma_e^2\,[i = j]$$

where d_{ij} is the distance in km between stations i and j, and $[i = j]$ is 1 on the diagonal and 0 elsewhere.

3b. *(Recommended: no AI.)* Write a function `nll_spatial(par)` that returns the negative log-likelihood, where `par` holds the intercept, the elevation slope, $\log \sigma_s$, $\log \ell$, and $\log \sigma_e$, in that order. Check it: `nll_spatial(c(40, 40, log(8), log(10), log(3)))` should return 209.42.

Your value: _____ Copy the line of your function that builds V:

_____

_____

AI use (circle one): no AI / AI helped (e.g., debugging) / AI wrote the code

3c. *(Recommended: no AI.)* Minimize it with `optim()` (`method = "BFGS"`, `hessian = TRUE`), starting from the values in the check above. Confirm `$convergence` is 0, and fill in the table (spatial SD is $\hat{\sigma}_s$, length-scale is $\hat{\ell}$ in km, noise SD is $\hat{\sigma}_e$).

	slope	SE of slope	spatial SD	length−scale	noise SD	logLik
lm()			(none)	(none)		
spatial						

AI use (circle one): no AI / AI helped (e.g., debugging) / AI wrote the code

3d. The slope changed more than its standard error did. Use your Figure 3a to explain why the ordinary regression got a different slope. How do $\hat{\sigma}_e^2$, $\hat{\sigma}_s^2 + \hat{\sigma}_e^2$, and $2\hat{\ell}$ compare with the nugget, sill, and range you read in 2a?

_____

_____

_____

_____

_____

_____

_____

Part 4: Predicting at a new site

The district is considering a new station at easting 30 km, northing 20 km, elevation 2450 m.

4a. From the spatial model's intercept and slope alone, the predicted SWE at the new site is _____ cm.

The model can do better, because nearby stations say something about the unmeasured part of the snowpack there. Let $\mathbf{k}_0$ hold the covariance between the field at the new site and at each station, $k_{0i} = \hat{\sigma}_s^2 \exp(-d_{0i}/\hat{\ell})$, with d_{0i} the distance from the new site to station i. The prediction is

$$\hat{y}_0 = \mathbf{x}_0^T \hat{\boldsymbol{\beta}} + \mathbf{k}_0^T V^{-1} (\mathbf{y} - X\hat{\boldsymbol{\beta}})$$

4b. *(Recommended: no AI.)* Compute $\hat{y}_0$ in R (`solve(V, b)` gives $V^{-1}\mathbf{b}$).

$\hat{y}_0 =$ _____ cm Copy the line that computes it:

_____

_____

AI use (circle one): no AI / AI helped (e.g., debugging) / AI wrote the code

4c. The entries of $\mathbf{k}_0^T V^{-1}$ are the weights each station's residual gets. Which three stations have the largest weights, and how far is each from the new site? Why did the prediction move the way it did from 4a?

_____

_____

_____

_____

_____

_____

4d. *(AI optional/encouraged.)* Fit the same model with `glmmTMB` (the syntax is in section 5 of the Week 14 page) and compare log-likelihoods.

Yours: _____ `glmmTMB`: _____

AI use (circle one): no AI / AI helped (e.g., debugging) / AI wrote the code

4e. Where did you use AI, if anywhere, and what did you have to check or fix? If you didn't use it, what was hardest to get working?

_____

_____

_____

_____

_____

Bonus (optional). The prediction's variance is $\hat{\sigma}_s^2 + \hat{\sigma}_e^2 - \mathbf{k}_0^T V^{-1} \mathbf{k}_0$. Compute its square root for the new site, and again for a site at easting 100 km, northing 20 km, far outside the network. What does the prediction itself become at the far site, and why?