1 When to use the log formula instead of the regression formula?

log is primarily used when you're interested in the ratio of two parameters. For example,

$$\theta = \frac{\mu_{south}}{\mu_{north}}.$$

Taking the log of both sides, we get

$$\log(\theta) = \log(\mu_{south}) - \log(\mu_{north}).$$

This form is much easier to deal with because we can "trick" any regression function to run this model.

2 Are there differences in regression approaches for μ, π and λ or is it just a difference in the scales?

Yes. We can use the linear model for μ. However, because both π and λ have a restricted domain (π must be between 0 and 1, λ must be greater than 0), care must be taken as to not obtain nonsensical values. In general we use logistic regression for π and poisson regression for λ. These involve transformations of the original parameter so that the domain ranges from $-\infty$ to $+\infty$.

3 Please explain how to assign the baseline with north and south

The baseline is arbitrary. Usually the category of interest is chosen as the "non-reference" category.

4 Not clear why we are doing both a t-test and a regression analysis to test for mean difference? In this case a t test seems to be the appropriate test, are we just applying it to regression to think about it conceptually?

The two-sample t-test is a special case of regression when the only determinant of the parameter is the group. Regression is a much more general approach that allows you to include more determinants (i.e. confounders). The t-test can only handle a single determinant, which is often rarely the case in observational health research.

5 Why is a t-test used to conduct inference for regression, but a z-test for the CI?

t procedures are technically correct, but often we have a large enough sample to assume we have a good estimate of σ, so that we can use the z procedure.

6 Was what we were doing today the same as least squares regression? If not, how was it different?

Today was about writing regression equations with parameters. We cheated by assuming the truth was known. Least-squares is a method to estimate the parameters with data.

7 Is the lm function in R just for difference? How do we adapt it for ratios?

`lm` is just for the difference. We must use the `glm` function and specify the argument `family=gaussian(link=log)` to adapt it for ratios.