Sheet 8

1. Learning more about the relationship between dose and response and find relevant dosages
2. Estimating a dose-response-profile under the restriction of montonicity. The algorithm pools together mean responses, that violate the assumption
3. Firstly the CSD is a graph of points, the x coordinates are the cumulative sum of the weights and the y coordinates the cumulative sum of the product of mean responses and weights (starting in (0,0)). The GCM is then the greatest convex function that is smaller than the points of the CSD. Both help with interpreting the results of the PAVA.
4. The slope of the lines in the CSD are the mean values of each dose-response. The points-that lie above the GCM are pooled and the number of straight lines in the GCM are the final number of pooling samples.
5. The predictor-variable is always the dose. It is the variable, that is controlled and exogenous in the lab-experiments
6. The definition is:

$$f(x, b, c, d, e) = c + \frac{d - c}{1 + \exp(b(\log(x) - \log(e)))}$$

   - b is the slope parameter. Positive means a decreasing sigmoid curve and negative the opposite.
   - c, d are asymptotes which direction of asymptotes depends on the slope parameter b.
   - e is the $ED50$ dose
7. The parameter θ_4 can be the highest possible overall response, while the E_{max} is always refering to the maximal effect → $E_{max} = |\theta_1 - \theta_4|$
8. The strength of the hormesis effect
9. ML estimation simplifies to OLS:

$$\min_{\theta \in \Theta} \sum_{i,j} (f(x_i|\theta) - y_{ij})^2$$

10. Scaled inverse of the information-matrix M. Which is the squared jacobian matrix on evaluated for each dosage. Scaled by the varaince of the residuals. The diagonal elements are the variance of the parameter estimation itself and can be used for evaluating significant differnces from 0
11. $\hat{\beta} \pm K\sqrt{Var(\hat{\beta})}$ while K is the $(1-\alpha)$ quantile of the t-distribution with $n-k$ degrees of freedom $n =$ sample size and k = parameter-count.
12. Absolute is a fixed effect while relative are percentages of the maximal effect
13. In the context of a continuous responsevariabel, it means that the dosage has no effect on the variance of the residuals. The residual plot can help visualize the relationship between dose and variance. If we see a noticeable variety of variance, the assumption of variance homogeneity is not valid
14. Sandwich estimation of the variance and the box-cox-transformation
    - Sandwich estimation → scales by magnitude of residuals relative to standarderror
    - box-cox → transformes data such that variance-homogeneity is preserved
15. T test which compares the residual sum of squares (RSS). $H_0 : f = g$ and $H_1 : f \neq g$
16. The weights parameter determines the n parameter of the estimation. Without this parameter, the estimator does not know at which proportion the event ocured at
17. We can do a standard T-test with `compParm(….)` function.
18. The upperbound n of the total occurences of the event is not specified in counting data (for example leaves on a tree)
19. Variance of data is not represented in the variance of the model. We can choose a model with more variaty or we can $\log$ scale the data. The latter could be a negative binomial distribution in the case of counting data
20. Usually the weights are given by time. The responsevariable is then normalized to a common time frame.
21. Usually a background risk p_0 is given and then the BMR can be defined by:

$$BMR_{\text{additional}} = f(BMD|\theta) - p_0$$
$$BMR_{\text{excess}} = \frac{f(BMD|\theta) - p_0}{1 - p_0}$$

22. This is usually given by an expert and not by the statistician. However in a relative BMR the background risk is the defined by the left asymptote of the model.

23. We assume the predictions of the model are normally distributed with variance homogeneity. We then define:

$$Z_n = \frac{x - f(x_i|\theta)}{\sigma^2}$$

    Then $P(Z_n \leq z) = \Phi(z)$. We can then express the BMR as a probability.

24. The $ALEC$ is the first time the model exceeds a fixed value λ. So:

$$f(ALEC_\lambda|\theta) = \lambda \quad f(ALEC_\lambda|\theta) = 100\lambda$$

    In a way it then holds that $ALEC_\lambda = f^{-1}(\lambda|\theta)$ and then:

$$ALEC_\lambda = h(\lambda) = e\sqrt[b]{\frac{d-\lambda}{\lambda - c}} = f^{-1}(x, b, c, d, e)$$

25. We use the delta method for this. In short $X_n \sim \mathcal{N}(\hat{\mu}_i, \hat{\sigma}_i^2) \implies f(X_n) \sim \mathcal{N}(f(\hat{\mu}_i), f'^2(\hat{\mu}_i)\hat{\sigma}_i^2)$ (using taylor this can be obtaiend)

26. The first dosage where the effect is exceeded and significantly exceeded (testing problem)

27. A design is D-Optimal if the information matrix M is maximized by determinant (so the covariance of the parameters is minimized). If we denote the efficiency of a given design D as E_D then we need $1/E_D$ times the sample-size to get the same performance as the optimal design

28. Look question 27

29. The Goal is first to sort out flat profiles and then fit only relevant models to the data.

    The procedure can be divided into 4 steps:

    1. Guestimate coefficients of chosen models

2. Compute best contrast coefficients (maximizing the power of a t-contrast-test)
3. Test for flat profiles (proof of concept)
4. Average (AIC based) Models that got proof of concept

30. Given a model $f(x|\theta)$ we can disect the model into: $f(x|\theta) = \theta_0 + \theta_1 f_0(x|\theta)$. Then $f_0(x|\theta)$ is the standardized model. We guestimate the coefficients for the standardized model and optimize the contrast-coefficients for the resulting functions.
31. Optimal contrast coefficients maximize the power of the t-contrast test. We say $\sum_{i=0}^{k} c_i = 0$ because we are trying to test for flat profiles.
32. Contrast t-test:

$$T_m = \frac{\sum_{i=1}^{k} c_{mi}\bar{Y}_i}{S\sqrt{\sum_{i=1}^{k} c_{mi}/n_i}}$$

33. Using the modelaverages (based on AIC) or the model with the lowest AIC

    One thing is important:

$$P(D|f_m) \propto \exp\left(-\frac{1}{2}\Delta AIC_m\right)$$

    After that we can apply Bayes rule and the law of total probabilty:

$$P(f_m|D) = \frac{P(D|f_m)P(f_m)}{\sum_{i=1}^{M} P(D|f_i)P(f_i)} = \frac{\exp(-\frac{1}{2}\Delta AIC_m)P(f_m)}{\sum_{i=1}^{M} \exp(-\frac{1}{2}\Delta AIC_i)P(f_i)}$$

34. …..
35. What profile curves (sigmoid, linear or polinomial) are most prevelant in the dataset?
36. We denote α_i the columnwise effect (mean - μ) and β_j the rowwise effect and μ the overall mean. Then:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

37. The measure H_Y is defined as: $\sum_{ij} \epsilon_{ij}^2$. Which is the sum of squared residuals of a model that assumes the same slope for each profile. So we assume the slope is the same for the model and see how well the model fits.
38. The two methods which were shown in the lecture are ORIOGEN and ORICC. The core of which are assuming different inequalities between mean responses and estimating with these inequalities as restrictions.
39. ORIOGEN clusters dose-response profiles for gen-expression data. It works by estimating restricted means based on pre-specified profiles and assessing goodness of fit by testing for difference.
40. The maximum difference of means in the largest subgraph of the profile. And then relative to the variance of the estimate.
41. Assessing how well the model fits the data while at the same time taking into account how “unrestricted” the profile was and how big the sample is.
42. The first stages are the as in ORIOGEN → pre-specify profiles, apply restricted estimation of means. Then we compute the ORIC (Order-restricted information criterion) and then choose the one profile with the highest information criterion.