Homework 12 Solutions

Stat061-F23

Prof Amanda Luby

---

1. For the simple linear regression case ($y = \beta_0 + \beta_1 x + \epsilon$), show that $\hat{\beta}_1 = r\frac{\hat{\sigma}_y}{\hat{\sigma}_x}$.

Solution

Note that $\hat{\sigma}_x = \sqrt{\frac{1}{n}\sum(X_i - \bar{X})^2}$ (and likewise for $\hat{\sigma}_y$) and so:

$$r\frac{\hat{\sigma}_y}{\hat{\sigma}_x} = \frac{\frac{1}{n}\sum x_i y_i - \frac{1}{n}\sum x_i \frac{1}{n}\sum y_i}{\hat{\sigma}_x \hat{\sigma}_y}\frac{\hat{\sigma}_y}{\hat{\sigma}_x} = \frac{\frac{1}{n}\sum x_i y_i - \frac{1}{n}\sum x_i \frac{1}{n}\sum y_i}{\hat{\sigma}_x^2}$$

We'll then multiply numerator and denominator by n^2:

$$\frac{n\sum x_i y_i - \sum x_i \sum y_i}{n^2 \hat{\sigma}_x^2} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum(X_i - \bar{X})^2} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n(\sum x_i^2 - n\bar{X}^2)} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

which is the form of $\hat{\beta}_1$ from the notes.

2. Assuming the standard multiple linear model ($Y = X\beta + \epsilon$, where X is an $n \times p$ design matrix):
   (a) Show that $\sigma^2 I = \Sigma_{\hat{y}} + \Sigma_{\hat{\epsilon}}$
   (b) Using (a), conclude that $n\sigma^2 = \sum Var(\hat{Y}_i) + \sum Var(\hat{\epsilon}_i)$

Solution

Note that:

$$\hat{y} = Hy = HX\beta + H\epsilon$$

and so since $HX\beta$ is a constant,

$$\Sigma_{\hat{y}} = H\Sigma_\epsilon H^T = H\sigma^2 I H^T = \sigma^2 HH^T = \sigma^2 H.$$

In class, we showed that $\Sigma_{\hat{\epsilon}} = \sigma^2(I - H) = \sigma^2 I - \sigma^2 H$. So:

$$\Sigma_{\hat{y}} + \Sigma_{\hat{\epsilon}} = \sigma^2 H + \sigma^2 I - \sigma^2 H = \sigma^2 I$$

(b) Note that the trace (sum of the diagonal) of $\sigma^2 I$ is $n\sigma^2$ (add up the n diagonal entries and multiply by σ^2). Since we showed

$$\Sigma_{\hat{y}} + \Sigma_{\hat{\epsilon}} = \sigma^2 I$$

, this means the trace of $\Sigma_{\hat{y}} + \Sigma_{\hat{\epsilon}}$ must also equal $n\sigma^2$. Since each of these are covariance matrices, the diagonal entries are all variances. So $Tr(\Sigma_{\hat{y}}) = \sum Var(\hat{Y}_i)$ and $Tr(\Sigma_{\hat{\epsilon}}) = \sum Var(\epsilon_i)$. So $\sum Var(\hat{Y}_i) + \sum Var(\epsilon_i) = n\sigma^2$.

3. Consider a multiple linear regression problem with design matrix $\mathbf{X}$ and observations $\mathbf{Y}$. Let $\mathbf{X}_1$ be the matrix remaining when at least one column is *removed* from $\mathbf{X}$. (So $\mathbf{X}_1$ is the design matrix for a linear regression on $\mathbf{Y}$ but with fewer predictors). Show that R^2 (non-adjusted) for the regression model calculated using design matrix $\mathbf{X}$ is *at least as large* as the R^2 for the regression model using design matrix $\mathbf{X}_1$.

Solution

Since R^2 is a decreasing function of the residual sum of squares, we shall show that the residual sum of squares is at least as large when using $\boldsymbol{Z}'$ as when using $\boldsymbol{Z}$. Let $\boldsymbol{Z}$ have p columns and let $\boldsymbol{Z}'$ have $q < p$ columns. Let $\hat{\boldsymbol{\beta}}_*$ be the least-squares coefficients that we get when using design matrix $\boldsymbol{Z}'$. For each column that was deleted from $\boldsymbol{Z}$ to get $\boldsymbol{Z}'$, insert an additional coordinate equal to 0 into the q-dimensional vector $\hat{\boldsymbol{\beta}}_*$ to produce the p-dimensional vector $\tilde{\boldsymbol{\beta}}$. This vector $\tilde{\boldsymbol{\beta}}$ is one of the possible vectors in the solution of the minimization problem to find the least-squares estimates with the design matrix $\boldsymbol{Z}$. Furthermore, since $\tilde{\boldsymbol{\beta}}$ has 0's for all of the extra columns that are in $\boldsymbol{Z}$ but not in $\boldsymbol{Z}'$, it follows that the residual sum of squares when using $\tilde{\boldsymbol{\beta}}$ with design matrix $\boldsymbol{Z}$ is identical to the residual sum of squares when using $\hat{\boldsymbol{\beta}}_*$ with design matrix $\boldsymbol{Z}'$. Hence the minimum residual sum of squares available with design matrix $\boldsymbol{Z}$ must be no larger than the residual sum of squares using $\hat{\boldsymbol{\beta}}$ with design matrix $\boldsymbol{Z}'$.

4. ~~Problem from Monday~~

5. Suppose we observe data $(x_1, Y_1), (x_2, Y_2), ..., (x_n, Y_n)$, where each Y_i represents a count and has mean μ_i. (e.g. the answer to "How many devices do you own that can access the internet?" or "How many children do you expect to have in your lifetime?). Since counts are always positive, we often use a *log-linear model* to model the mean of the Y_i's: $\log \mu_i = \beta_0 + \beta_1 x_i$. The Poisson loglinear model additionally assumes that the counts are independent poisson random variables: $Y_i \sim Pois(\mu_i)$.

   (a) Show that the log-likelihood (in terms of the Poisson parameters μ_i) is $l(\mu) = \sum[y_i \log(\mu_i) - \mu_i - log(y_i!)]$
   (b) Substitute the linear model component to show that part (a) is equivalent to $l(\beta) = \beta_0 \sum y_i + \beta_1 \sum y_i x_i - \sum \exp(\beta_0 + \beta_1 x_i) - \sum \log(y_i)!$.
   (c) Explain why the sufficient statistics for the model parameters (β_0 and β_1) are $\sum y_i$, $\sum x_i y_i$.
   (d) Show that the likelihood equation solutions have the form $\sum y_i = \sum \mu_i$ and $\sum y_i x_i = \sum \mu_i x_i$
   (e) We could also think about simply transforming the Y variables and fitting a linear regression for $\log(Y)$. The transformed-data approach uses a linear predictor for $E(\log(Y))$ whereas the GLM approach uses a linear predictor for $\log E(Y)$. Explain why these are not the same, and state an advantage of using the GLM approach if we are truly interested in modeling $E(Y)$.

Solution

(a) Recall that the Poisson likelihood is $\prod \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} = \frac{e^{-\sum \mu_i}\prod \mu_i^{y_i}}{\prod y_i!}$

Then,

$$l(\mu) = \ln e^{-\sum \mu_i} + \ln \prod \mu_i^{y_i} - \ln \prod y_i! = \sum[y_i \log(\mu_i) - \mu_i - log(y_i!)]$$

(b) Since $\ln \mu_i = \beta_0 + \beta_1 x_i$, $\mu_i = e^{\beta_0 + \beta_1 x_i}$, which we will substitute in:

$$l(\beta) = \sum[y_i \log(e^{\beta_0 + \beta_1 x_i}) - e^{\beta_0 + \beta_1 x_i} - log(y_i!)]$$

The first term simplifies to $y_i(\beta_0 + \beta_1 x_i)$. Rearrange to obtain $l(\beta) = \beta_0 \sum y_i + \beta_1 \sum y_i x_i - \sum \exp(\beta_0 + \beta_1 x_i) - \sum \log(y_i)!$.

(c) If we undo the log in (b):

$$L = \exp(\sum e^{\beta_0+\beta_1 x_i}) \exp(\beta_0 \sum y_i) \exp(\beta_1 \sum x_i y_i) \exp(-\sum \ln_i) = r_1(\beta, \sum y_i) r_2(\beta \sum x_i y_i) b(y_i)$$

Since x_i's are constants in the regression setting. So by the factorization theorem, $\sum x_i y_i$ and $\sum y_i$ are jointly sufficient for β_0 and β_1.

(d) First, note that

$$\frac{\partial l}{\partial \beta_0} = -\sum e^{\beta_0+\beta_1 x_i} + \sum y_i$$

To find the solution to this equation, we set equal to zero and rearrange. Then we replace $e^{\beta_0+\beta_1 x_i}$ with μ_i to obtain $\sum \mu_i = \sum y_i$.
Following the same process, we find:

$$\frac{\partial l}{\partial \beta_1} = -\sum e^{\beta_0+\beta_1 x_i} x_i + \sum x_i y_i$$

and then $\sum \mu_i x_i = \sum x_i y_i$.

(e) For the Poisson case, $E(Y_i) = \mu_i$, where $Y_i \sim Pois(\mu_i)$. When we transform and then do linear regression, we get $E(\log Y_i) = \beta_0 + \beta_1 x_i$. There is no way to "work backwards" from the linear equation to $E(Y_i)$. On the other hand, using a GLM, we get $\log E(Y_i) = \beta_0 + \beta_1 x_i$ and so we can obtain $E(Y_i) = \mu_i = e^{\beta_0+\beta_1 x_i}$.