Stats 401 Lab 7

401 GSI team

2/15/2018 and 2/16/2018

Outline

- Exam information
- Review of newest material (by examples)
- Review of old material (by examples)
- Lab Ticket

Exam information

- Material covered in class up to 2/14, homeworks 1-6, and labs 1-6.
- Closed book and closed notes.
- You will be provided the following formulas:
  - $\mathbf{b} = (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \mathbf{y}$
  - $Var(\hat{\beta}) = \sigma^2 (\mathbb{X}^T \mathbb{X})^{-1}$
  - $Var(\mathbb{A}\mathbf{Y}) = \mathbb{A} Var(\mathbf{Y}) \mathbb{Y}^T$
- R coding will be tested by multiple choice questions with the possible exception being 'pnorm()' and 'qnorm()'
- Key: "writing" a model means fully defining all the necessary quantities and writing out the necessary equations

- No lab next week!

Review - Properties of Expected Value and Variance

Last lab we had the following lab ticket:

- Calculate $\mathbb{E}(\hat{\mathbf{Y}})$ and $Var(\hat{\mathbf{Y}})$, where $\hat{\mathbf{Y}} = \mathbb{X}\hat{\beta}$.
- What about $\mathbb{E}(\hat{\mathbf{y}})$ and $Var(\hat{\mathbf{y}})$, where $\hat{\mathbf{y}} = \mathbb{X}\mathbf{b}$. (You can actully answer this question without any computation)

Review - Properties of Expected Value and Variance (lab ticket solutions)

- $\mathbb{E}(\hat{\mathbf{Y}}) = \mathbb{E}(\mathbb{X}\hat{\beta})$
- $\mathbb{E}(\mathbb{X}\hat{\beta}) = \mathbb{X}\mathbb{E}(\hat{\beta})$ (since $\mathbb{X}$ is a constant)
- $\mathbb{X}\mathbb{E}(\hat{\beta}) = \mathbb{X}\beta$
- Therefore, $\mathbb{E}(\hat{\mathbf{Y}}) = \mathbb{X}\beta$

Review - Properties of Expected Value and Variance (lab ticket solutions cont.)

- $Var(\hat{\mathbf{Y}}) = Var(\mathbb{X}\hat{\beta})$
- $Var(\mathbb{X}\hat{\beta}) = \mathbb{X} Var(\hat{\beta})\mathbb{X}^T$ (from properties of variance)
- $\mathbb{X} Var(\hat{\beta})\mathbb{X}^T = \mathbb{X}\sigma^2(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T$ (from class)
- $\mathbb{X}\sigma^2(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T = \sigma^2\mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T$ (since σ^2 is a constant)

- Note 1: we cannot make $(\mathbb{X}^T\mathbb{X})^{-1} = (\mathbb{X}^{-1}(\mathbb{X}^T)^{-1})$ because $\mathbb{X}$ is not a square matrix (recall: this is a requirement for invertibility).

Review - Properties of Expected Value and Variance (lab ticket solutions cont.)

- $Var(\hat{\mathbf{Y}}) = Var(\mathbb{X}\hat{\beta})$
- $Var(\mathbb{X}\hat{\beta}) = \mathbb{X}Var(\hat{\beta})\mathbb{X}^T$ (from properties of variance)
- $\mathbb{X}Var(\hat{\beta})\mathbb{X}^T = \mathbb{X}\sigma^2(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T$ (from class)
- $\mathbb{X}\sigma^2(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T = \sigma^2\mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T$ (since σ^2 is a constant)

- Note 1: we cannot make $(\mathbb{X}^T\mathbb{X})^{-1} = (\mathbb{X}^{-1}(\mathbb{X}^T)^{-1})$ because $\mathbb{X}$ is not a square matrix (recall: this is a requirement for invertibility).

- Note 2: $\mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T$ is often referred to as the Hat matrix and written as $\mathbb{H}$.

Review - Properties of Expected Value and Variance (lab ticket solutions cont.)

- What about $\mathbb{E}(\hat{\mathbf{y}})$ and $Var(\hat{\mathbf{y}})$, where $\hat{\mathbf{y}} = \mathbb{X}\mathbf{b}$. (You can actully answer this question without any computation)
- Note that $\hat{\mathbf{y}}$ is the fitted values of the equation which have $\mathbb{E}(\hat{\mathbf{y}}) = \mathbf{y}$ and $Var(\hat{\mathbf{y}}) = Var(\mathbf{y}) = \sigma^2$

Review - Properties of Expected Value and Variance

Let X and Y be independent bivariate random variables with mean $\mu = (0, 0)^T$ and variance $\mathbb{I}$.

- Calculate $E(XY)$.
- Calculate $Var(2X + 3Y)$.

Review - Properties of Covariance

- Let X, Y, and Z be random variables (not related to the ones in the previous question)
- Calculate Cov(aX + bY, X - Y + Z)

Review - Probability Calculations

Suppose X is a binomial random variable with $n = 200$ and $p = .4$

- Use the normal approximation to the binomial find the value of α such that $P(Z \leq \alpha) = .85$.
- Find the probability that X is greater than or equal to 150.

Review - Fitting a linear model by least squares

Recall from lab 6, we had briefly examined the birthweight dataset. Below is the first five observations of the dataset.

```
library(MASS)
data(birthwt)
head(birthwt, n = 5)
```

```
##    low age lwt race smoke ptl ht ui ftv  bwt
## 85   0  19 182    2     0   0  0  1   0 2523
## 86   0  33 155    3     0   0  0  0   3 2551
## 87   0  20 105    1     1   0  0  0   1 2557
## 88   0  21 108    1     1   0  0  1   2 2594
## 89   0  18 107    1     1   0  0  1   0 2600
```

Suppose in R, we calculate the following model {r, eval = F}lm(log_bwt ~ age + log_lwt + smoke + racew + raceb, data = birthwt) where log_bwt is the log birth weight of the child, log_lwt is the log weight of the mother, smoke is the smoking status of the mother and racew and raceb are indicator variables for whether the mother was white or black respectively. Given that the dataset contains 189 observations, (i) write the corresponding sample version of the probability model, in subscript form (ii) write the design matrix of this model.

Write the code that you would use to input the design matrix that you came up with in the previous slide into R. (Note: this is harder than what would be expected on the exam so it's good practice!)

Describe a suitable probability model, in matrix form, to give a population version of a linear model from the birthweight data.

Review - Summation Exercises

Evaluate $\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ where $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$

▶ Note: This is the mean squared error. You will see more of this later.