STA302/STA1001, Weeks 8-9

Mark Ebden, 31 October & 2 November 2017

With grateful acknowledgment to Alison Gibbs

Plan for Tuesday 31 October

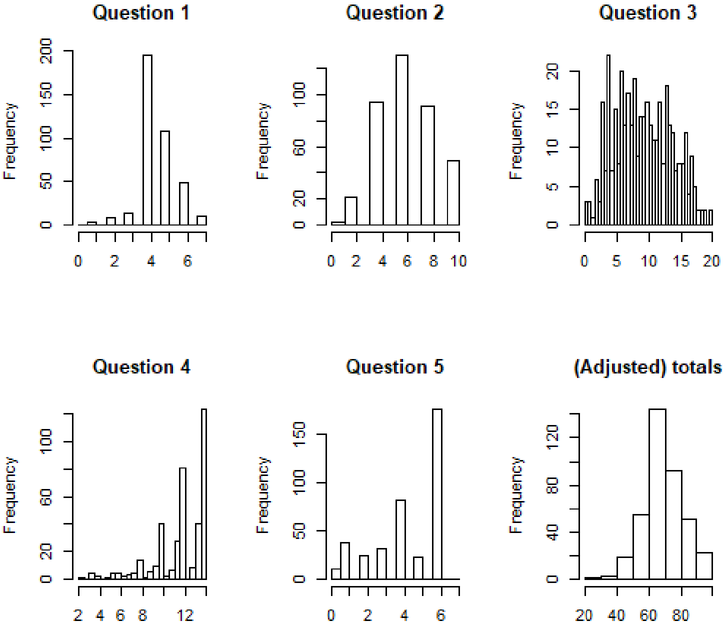
- Section 1 can pick up midterms and digest them for a few minutes
- Midterm discussion
- Chapter 5
- One-to-one discussion about any midterm issues

Midterms: Section 1

The number of test takers was 406 in Section 1: STA302/1001 LEC0101 & LEC2001.

Marking has been completed for Section 1 only.

Midterm marks: Based on 96% of test papers in Section 1



Midterm marks: Section 1

Raw scores:

- Top students got 55/57
- Average of 64%

Adjusted scores:

- Two marks of 100%
- Average of 68.3%

Questions:

- Questions 1, 2, and 5 had averages of 62 to 64%
- Hardest question was #3: average was 47%
- Easiest question was #4: average was 84%
- We'll review some of these in a moment

To request a re-grade: Section 1

By 9 November, please email `sta302sec1@gmail.com` with a description of the problem, including whether or not you spoke with me during class on 31 October.

You should receive a reply within a week of sending your request, and your mark may go up or down.

For efficiency, you may wish to include a picture of the problem. This is encouraged but optional. If the picture you take isn't found to match what we have on record, when verified at a later date, then you may be subject to an academic offence and any previous mark adjustment would be moot.

Post-midterm work so far

- §3.3 (Transformations) except for Box-Cox transformations and inverse-response plots
- §5.2 (Estimation and Inference in MLR) is what we've been heading towards, via the RMA

A closer look at $\mathbf{X}'\mathbf{X}$ and $(\mathbf{X}'\mathbf{X})^{-1}$

Previously we simplified $\mathbf{X}'\mathbf{X}$, for $\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$

We began to consider $(\mathbf{X}'\mathbf{X})^{-1}$ (Week 8, slide 12), because of its appearance in $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

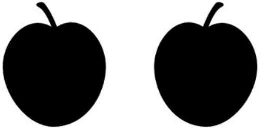
This $\hat{\boldsymbol{\beta}}$ expression is a concise way to write the estimators for linear regression, compared to $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$ and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$.

Properties of least squares estimates

Recall that $\mathrm{E}(\hat{\beta}_0) = \beta_0$ and $\mathrm{E}(\hat{\beta}_1) = \beta_1$, and that

$$\mathrm{var}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{S_{xx}}, \qquad \mathrm{var}(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right] \quad \text{and} \quad \mathrm{cov}(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\sigma^2 \bar{x}}{S_{xx}}$$

Let's now confirm that our new equations give this as well.



Next steps

- HW2 (not for credit) is on Portal
- We'll continue in Chapter 5
  - More parallels between the old- and new expressions
  - MLR (multiple linear regression)