(possibly incomplete list of)

Topics Covered in Sections 10.1-10.2 and Chapter 11

Sections 10.1-10.2 - One-way Analysis of Variance

The null and alternate hypothesis for one-way analysis of variance

The assumptions for the one-way ANOVA are the same as for a two-sample t-test: samples are random and independent (need to know how it was sampled), the populations are normally distributed (make a separate q-q plot for each group), that the variances of the populations are the same (use side by side box-plots or calculate the standard deviations)

How the ANOVA table is constructed for the one-way analysis of variance

- MS = SS/df
- MSB = MST is based on the variance between the treatment averages
- MSW = MSE is based on the variance within each group separately
- TMS is the estimated variance when H_0 is true
- If the assumptions are met we can use the MS to construct an F test
- How the F is constructed from the MS, and what null and alternate hypotheses it tests
- That we reject the null hypothesis when the F statistic is large
- Using the formulas for the SS to construct the ANOVA table
- Finding the mean squares given the sum of squares and degrees of freedom

How the degrees of freedom are found

- for total and error/within it is the sample size minus the number of parameters estimated
- for the treatment/between it is the number of treatments minus 1

For two-samples, the one-way ANOVA and the two-sample t-test (with equal variances) give the same p-value and that $F=t^2$

That the one-way ANOVA can be written with a model equation $y_{ij} = \mu_i + \varepsilon_{ij}$

How we write the assumptions for a one-way ANOVA using the model equation (same as for regression, see page 523)

Chapter 11 - Linear Regression

The regression model equation (as given in the middle of page 508)
The four assumptions for linear regression (as given on page 523)
Independent vs. dependent variable
Correlation vs. Causality
Extrapolation
The estimates of β_0 and β_1 are gotten by minimizing the sum of the squared residuals
What is meant by regression to the mean
The analysis of variance table

  TSS is the total amount of error/variation in Y
  SSR = TSS - SSE is the amount of error/variation we explained by using the regression line
  MSE is the estimated variance (σ^2) of the points around the regression line
  TMS=Var(Y) is the estimated variance (σ^2) when $\beta_1 = 0$
  Because we assumed the residuals are normal we can use the MS to do an F test

Why must SSE and SSR each be less than TSS?
Finding the mean squares given the sum of squares and degrees of freedom
How the degrees of freedom are found

  for total and error it is the sample size minus the number of parameters estimated
  for the regression it is df for the total minus the df for the error

F = MSR/MSE
When we accept or reject $\beta_1 = 0$ based on the ANOVA table
Using the formulas for the SS to construct an ANOVA table
How the MS relate to the variances of the normal distributions in figures like 11.7 (page 524)
The t-test and confidence interval for β_1 (section 11.5)
The predicted value of y given x
The prediction interval for a single y at a fixed x value (section 11.8)
The confidence interval for the mean of the y at a fixed x value (section 11.8)
The range of values the correlation coefficient can take.
The correlation as a single number summary of regression: with sample size it determines the F-statistic, r^2 = coefficient of determination = percent of variation/error explained by the model, and r is like the slope of the regression line after adjusting for the scale of y and x.
What to look for in residual plots: is there a pattern in the residual vs. predicted plot? (violation of independence), is there a "funnel shape" in the residual vs. predicted plot? (error variances are not constant), does the q-q plot look like a straight line? (are the errors normal)
That taking the log of the dependent variable can solve the problem of the variance being larger for large values of the independent variable