STAT7630 - HW 1

D. Pluta

2024-01-17

Due: February 2nd, 2024 by 5pm

Submission: AU Course Website

Homework 1

STAT 7630, SPRING 2024

1. Complete the sentence: “In the NHST framework, the P-value is the probability of…”
2. *Linear Regression Assumptions.*
    A. Write out the four usual assumptions of linear regression.
    B. Write the complete formal specification of the simple linear regression model.
    C. Suppose that we have two covariates x_1, x_2 and response y. If x_1 and x_2 are known to be correlated, and y is suspected of being correlated with at least one of the covariates, are the assumptions of linear regression still satisfied?
    D. Give an example of real world data set that would likely violate the normality assumption. Explain your choice.
    E. Give an example of a real world data set that would likely violate the independence assumption.
    F. Give an example of a real world data set that would like violate the constant variance assumption.
3. *One Sample Normal Model*
    A. Write down the likelihood for $(y_1, \ldots, y_n) \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2), i = 1, \ldots, n.$
    B. Derive the MLE of μ assuming σ^2 is known.
    C. Derive the MLE of σ^2 when μ is known.
    D. Derive the MLEs for μ and σ^2 when both are unknown.
    E. Compute the bias for both of the MLEs you calculated in the previous part.
4. *One-Way ANOVA*. Consider a sample of observations $y_{ij}, i = 1, \ldots, n; j = 1, 2, 3$, where j is the group indicator for observation i. Assume the usual assumptions of ANOVA are satisfied.
    A. Verify the identity $SST = SSW + SSB$.
    B. Write the null hypothesis for the one-way ANOVA model.
    C. Provide a (brief) intuitive description of what SST, SSW, and SSB are measuring.
5. Suppose a drug x is tested at dosages $10mg, 25mg, 50mg$ against a placebo. Researchers believe the 50mg dose may have a detrimental or neutral effect on the measured outcome y.
    A. What method or model would you use to determine if the 50mg dose produces a lower response than the 25mg dose?
    B. What method or model would you use to determine if the effect of the dose with the highest average response value has a significantly greater effect relative to the other dosages?

C. Construct a linear regression model for the sample, assuming nothing is known regarding the relationship of the average response values at the different dosages.

D. Construct an alternative linear regression model for the sample, now assuming that the average response is quadratic with respect to the dosages.

6. *Simple Linear Regression*

   A. Write the likelihood for SLR.

   B. Derive the MLE for the coefficient β, assuming the error variance is known.

7. Explain why adding irrelevant covariates (i.e., covariates that are not associated with the response) to a linear regression model will strictly increase the R^2 measure.

8. A study wants to measure the average growth rate of a species of bacteria. In the settings and timescale of the experiment, the growth is known to be exponential, but with an unknown rate.

   A. The Principal Investigator (PI) asks you to produce a confidence interval for the average growth rate. The data consists of population counts for one culture, recorded at n evenly spaced intervals, which can be assumed to be scaled to $[0, 1]$. Propose an appropriate method or model.

   B. The PI now has data from 4 additional cultures. They ask you to repeat the analysis you did in the previous part on each of these new samples so that the average growth rates can be averaged together to provide a better estimate of the population growth rate. Explain one or two significant flaws in this method and propose a better one.

9. Consider a two-sample problem with observations $y_{ij} \overset{iid}{\sim} \mathcal{N}(\mu_j, \sigma^2), i = 1, \ldots, n; j = 1, 2.$

   A. Show that the two-sample t-test and ANOVA are identical.

   B. Show that the regression model $y_{ij} = \beta_0 + \beta_1 I(j == 1) + \varepsilon_{ij}$ can also be used to conduct an equivalent test as the two-sample t test and ANOVA.

10. *Modern Regression,* Ch. 2 Exercise 1 (p.38).

11. *Modern Regression*, Ch. 2 Exercise 3 (p. 39).