STA 2101/442 Assignment 1 (Mostly Review)[1]

Questions 6 and 7 are to be done with R; please bring your printouts to the quiz and be prepared to hand them in if requested. Remember, the computer assignments in this course are *not group projects*.

The other questions are practice for the quiz on Friday September 30th, and are not to be handed in. For the linear algebra part starting with Question 8, there is an excellent review in Chapter Two of Renscher and Schaalje's *Linear models in statistics*. The chapter has more material than you need for this course. Note they use $\mathbf{A}'$ for the transpose, while in this course we'll use $\mathbf{A}^\top$.

1. Let $X_1, \ldots, X_n$ be a random sample (meaning independent and identically distributed) from a distribution with density $f(x) = \frac{\theta}{x^{\theta+1}}$ for $x > 1$, where $\theta > 0$.

    (a) Find the maximum likelihood estimator of θ. Show your work. The answer is a formula involving $X_1, \ldots, X_n$.

    (b) Suppose you observe these data: `1.37, 2.89, 1.52, 1.77, 1.04, 2.71, 1.19, 1.13, 15.66, 1.43`. Calculate the maximum likelihood estimate. My answer is 1.469102.

2. Let $Y_1, \ldots, Y_n$ be a random sample from a distribution with mean μ and standard deviation σ.

    (a) Show that the sample variance $S^2 = \frac{\sum_{i=1}^n (Y_i - \overline{Y})^2}{n-1}$ is an unbiased estimator of σ^2.

    (b) Denote the sample standard deviation by $S = \sqrt{S^2}$. Assume that the data come from a continuous distribution, so it's easy to see that $Var(S) \neq 0$. Using this fact, show that S is a *biased* estimator of σ.

3. Let $Y_1, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2, so that $T = \frac{\sqrt{n}(\overline{Y} - \mu)}{S} \sim t(n-1)$. This is something you don't need to prove, for now.

    (a) Derive a $(1-\alpha)100\%$ confidence interval for μ. "Derive" means show all the high school algebra. Use the symbol $t_{\alpha/2}$ for the number satisfying $Pr(T > t_{\alpha/2}) = \alpha/2$.

    (b) A random sample with $n = 23$ yields $\overline{Y} = 2.57$ and a sample variance of $S^2 = 5.85$. Using the critical value $t_{0.025} = 2.07$, give a 95% confidence interval for μ. The answer is a pair of numbers.

[1] This assignment was prepared by Jerry Brunner, Department of Statistics, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. Use any part of it as you like and share the result freely. The LaTeX source code is available from the course website: `http://www.utstat.toronto.edu/~brunner/oldclass/appliedf16`

(c) Test $H_0 : \mu = 3$ at $\alpha = 0.05$.

   i. Give the value of the T statistic. The answer is a number.

   ii. State whether you reject H_0, Yes or No.

   iii. Can you conclude that μ is different from 3? Answer Yes or No.

   iv. If the answer is Yes, state whether $\mu > 3$ or $\mu < 3$. Pick one.

(d) Show that using a t-test, $H_0 : \mu = \mu_0$ is rejected at significance level α if and only the $(1 - \alpha)100\%$ confidence interval for μ does not include μ_0. The problem is easier if you start by writing the set of T values for which H_0 is *not* rejected.

(e) In Question 3b, does this mean $Pr\{1.53 < \mu < 3.61\} = 0.95$? Answer Yes or No and briefly explain.

4. Label each statement below True or False. Write "T" or "F" beside each statement. Assume the $\alpha = 0.05$ significance level. If your answer is False, be able to explain why.

   (a) _____ The p-value is the probability that the null hypothesis is true.

   (b) _____ The p-value is the probability that the null hypothesis is false.

   (c) _____ In a study comparing a new drug to the current standard treatment, the null hypothesis is rejected. This means the new drug is ineffective.

   (d) _____ We observe a Pearson correlation coefficient of $r = -0.70$, $p = .009$. We conclude that high values of X tend to go with low values of Y and low values of X tend to go with high values of Y.

   (e) _____ The greater the p-value, the stronger the evidence against the null hypothesis.

   (f) _____ If $p > .05$ we reject the null hypothesis at the .05 level.

   (g) _____ If $p < .05$ we reject the null hypothesis at the .05 level.

   (h) _____ In a study comparing a new drug to the current standard treatment, $p > .05$. We conclude that the new drug and the existing treatment are not equally effective.

   (i) _____ The 95% confidence interval for β_3 is from -0.26 to 3.12. This means $P\{-0.26 < \beta_3 < 3.12\} = 0.95$.

5. For $i = 1, \ldots, n$, let $Y_i = \beta_0 + \beta_1 x_i + \epsilon_1$, where β_0 and β_1 are unknown constants, $x_1, \ldots, x_n$ are known observable constants, and $\epsilon_1, \ldots, \epsilon_n$ are random variables with expected value zero.

   (a) What is $E(Y_i)$?

   (b) Clearly $E(Y_i)$ is a function of the unknown parameters β_0 and β_1. The *least squares estimates* of β_0 and β_1 are the numbers that make the Y_i values closest to their expected values in the sense of minimizing the quantity $Q = \sum_{i=1}^{n} (Y_i - E(Y_i))^2$. Find the least squares estimates of β_0 and β_1. Show your work.

(c) Now let $\epsilon_1, \ldots, \epsilon_n$ be independent normal random variables with expected value zero and common variance σ^2.

i. What is the distribution of Y_i? Just write down the answer.

ii. Find the maximum likelihood estimates of β_0 and β_1. You may stop as soon as you realize that you have already done the problem.

(d) Calculate your estimates for the following data:

```
x y
---
5 6
4 7
3 4
4 5
2 4
```

If you wish you can check your answer with R, but you will not turn in any R output for this question.

6. In the United States, admission to university is based partly on high school marks and recommendations, and partly on applicants' performance on a standardized multiple choice test called the Scholastic Aptitude Test (SAT). The SAT has two sub-tests, Verbal and Math. A university administrator selected a random sample of 200 applicants, and recorded the Verbal SAT, the Math SAT and first-year university Grade Point Average (GPA) for each student. The data are given SAT data file. The university administrator knows that the Verbal and Math SAT tests have the same number of questions, and the maximum score on both is 800. But are they equally difficult on average for this population of students?

Using R, do a reasonable analysis to answer the question. You can read the data into a data frame with

```
sat = read.table(http://www.utstat.toronto.edu/~brunner/data/legal/openSAT.data.txt
```

Bring your printout to the quiz; you may be asked to hand it in. Be ready to

- State your model.
- Justify your choice of model. Would you expect Verbal and Math scores from the same student to be independent?
- State your null and alternative hypotheses, in symbols.
- Express your conclusion (if any) in plain, non-statistical language that could be understood by someone who never had a Statistics course. Your answer is something about which test is more difficult for these students. Marks will be deducted for use of technical terms like null hypothesis, significance level, critical value, p-value, and so on even if what you say is correct.

There is more than one correct answer to this question. I did the analysis several different ways, and I consider all of them correct. I can think of about five more acceptable ways that I did not try. The number of bad ways to analyze the data is virtually unlimited.

7. Please use R for the numerical parts of the following question. Bring your printout to the quiz. It may be handed in.

   As in Question 3, let $Y_1, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2. This implies that $W = \frac{\sum_{i=1}^n (Y_i - \overline{Y})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2}$ has a chi-squared distribution with $n-1$ degrees of freedom, a fact you may use without proof for now.

   (a) Suppose we want to test $H_0 : \sigma^2 \leq \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$. Give the formula for a reasonable test statistic.

   (b) We are interested in testing $H_0 : \sigma^2 \leq 4$ against $H_1 : \sigma^2 > 4$.

      i. What is the critical value at $\alpha = 0.05$ for $n = 25$? The answer is a number. Get it with R.

      ii. A sample of size 25 yields a sample mean of 10.08 and a sample variance of $S^2 = 6.82$. Using R, calculate the test statistic and the p-value. Do you reject H_0 at $\alpha = 0.05$?

      iii. Suppose that the true variance is $\sigma^2 = 5$. What is the probability of rejecting the null hypothesis? Do a bit of hand calculation and then use R to obtain a numerical answer. This is the power of the test for $\sigma^2 = 5$ when $n = 25$.

      iv. That was pathetic. The sample size is far too small. What is the smallest n that yields a power of at least 0.8 when $\sigma^2 = 5$? The answer is a number.

8. Which statement is true? (Quantities in boldface are matrices of constants.)

   (a) $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$

   (b) $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{BA} + \mathbf{CA}$

   (c) Both a and b

   (d) Neither a nor b

9. Which statement is true?

   (a) $a(\mathbf{B} + \mathbf{C}) = a\mathbf{B} + a\mathbf{C}$

   (b) $a(\mathbf{B} + \mathbf{C}) = \mathbf{B}a + \mathbf{C}a$

   (c) Both a and b

   (d) Neither a nor b

10. Which statement is true?
    (a) $(\mathbf{B}+\mathbf{C})\mathbf{A} = \mathbf{AB}+\mathbf{AC}$
    (b) $(\mathbf{B}+\mathbf{C})\mathbf{A} = \mathbf{BA}+\mathbf{CA}$
    (c) Both a and b
    (d) Neither a nor b
11. Which statement is true?
    (a) $(\mathbf{AB})^\top = \mathbf{A}^\top\mathbf{B}^\top$
    (b) $(\mathbf{AB})^\top = \mathbf{B}^\top\mathbf{A}^\top$
    (c) Both a and b
    (d) Neither a nor b
12. Which statement is true?
    (a) $\mathbf{A}^{\top\top} = \mathbf{A}$
    (b) $\mathbf{A}^{\top\top\top} = \mathbf{A}^\top$
    (c) Both a and b
    (d) Neither a nor b
13. Suppose that the square matrices $\mathbf{A}$ and $\mathbf{B}$ both have inverses and are the same size. Which statement is true?
    (a) $(\mathbf{AB})^{-1} = \mathbf{A}^{-1}\mathbf{B}^{-1}$
    (b) $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
    (c) Both a and b
    (d) Neither a nor b
14. Which statement is true?
    (a) $(\mathbf{A}+\mathbf{B})^\top = \mathbf{A}^\top+\mathbf{B}^\top$
    (b) $(\mathbf{A}+\mathbf{B})^\top = \mathbf{B}^\top+\mathbf{A}^\top$
    (c) $(\mathbf{A}+\mathbf{B})^\top = (\mathbf{B}+\mathbf{A})^\top$
    (d) All of the above
    (e) None of the above

15. Which statement is true?

    (a) $(a+b)\mathbf{C} = a\mathbf{C} + b\mathbf{C}$

    (b) $(a+b)\mathbf{C} = \mathbf{C}a + \mathbf{C}b$

    (c) $(a+b)\mathbf{C} = \mathbf{C}(a+b)$

    (d) All of the above

    (e) None of the above

16. Let $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ $\mathbf{B} = \begin{pmatrix} 0 & 2 \\ 2 & 1 \end{pmatrix}$ $\mathbf{C} = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}$

    (a) Calculate $\mathbf{AB}$ and $\mathbf{AC}$

    (b) Do we have $\mathbf{AB} = \mathbf{AC}$? Answer Yes or No.

    (c) Prove $\mathbf{B} = \mathbf{C}$. Show your work.

17. Let $\mathbf{A}$ be a square matrix with the determinant of $\mathbf{A}$ (denoted $|\mathbf{A}|$) equal to zero. What does this tell you about $\mathbf{A}^{-1}$? No proof is required here.

18. Recall that an inverse of the square matrix $\mathbf{A}$ (denoted $\mathbf{A}^{-1}$) is defined by two properties: $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ and $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. Prove that inverses are unique, as follows. Let $\mathbf{B}$ and $\mathbf{C}$ both be inverses of $\mathbf{A}$. Show that $\mathbf{B} = \mathbf{C}$.

19. Suppose that the square matrices $\mathbf{A}$ and $\mathbf{B}$ both have inverses. Prove that $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$. You have two things to show.

20. Let $\mathbf{X}$ be an n by p matrix with $n \neq p$. Why is it incorrect to say that $(\mathbf{X}^\top\mathbf{X})^{-1} = \mathbf{X}^{-1}\mathbf{X}^{\top-1}$?

21. Let $\mathbf{A}$ be a non-singular square matrix. Prove $(\mathbf{A}^{-1})^\top = (\mathbf{A}^\top)^{-1}$.

22. Using Question 21, prove that the if the inverse of a symmetric matrix exists, it is also symmetric.

23. Let $\mathbf{a}$ be an $n \times 1$ matrix of real constants. How do you know $\mathbf{a}^\top\mathbf{a} \geq 0$?

24. Recall the *spectral decomposition* of a symmetric matrix (for example, a variance-covariance matrix). Any such matrix $\mathbf{\Sigma}$ can be written as $\mathbf{\Sigma} = \mathbf{P\Lambda P}^\top$, where $\mathbf{P}$ is a matrix whose columns are the (orthonormal) eigenvectors of $\mathbf{\Sigma}$, $\mathbf{\Lambda}$ is a diagonal matrix of the corresponding eigenvalues, and $\mathbf{P}^\top\mathbf{P} = \mathbf{PP}^\top = \mathbf{I}$. If $\mathbf{\Sigma}$ is real, the eigenvalues are real as well.

   (a) Let $\mathbf{\Sigma}$ be a square symmetric matrix with eigenvalues that are all strictly positive.

      i. What is $\mathbf{\Lambda}^{-1}$?

      ii. Show $\mathbf{\Sigma}^{-1} = \mathbf{P\Lambda}^{-1}\mathbf{P}^\top$

   (b) Let $\mathbf{\Sigma}$ be a square symmetric matrix, and this time the eigenvalues are non-negative.

      i. What do you think $\mathbf{\Lambda}^{1/2}$ might be?

      ii. Define $\mathbf{\Sigma}^{1/2}$ as $\mathbf{P\Lambda}^{1/2}\mathbf{P}^\top$. Show $\mathbf{\Sigma}^{1/2}$ is symmetric.

      iii. Show $\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2} = \mathbf{\Sigma}$, justifying the notation.

   (c) Now return to the situation where the eigenvalues of the square symmetric matrix $\mathbf{\Sigma}$ are all strictly positive. Define $\mathbf{\Sigma}^{-1/2}$ as $\mathbf{P\Lambda}^{-1/2}\mathbf{P}^\top$, where the elements of the diagonal matrix $\mathbf{\Lambda}^{-1/2}$ are the reciprocals of the corresponding elements of $\mathbf{\Lambda}^{1/2}$.

      i. Show that the inverse of $\mathbf{\Sigma}^{1/2}$ is $\mathbf{\Sigma}^{-1/2}$, justifying the notation.

      ii. Show $\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{-1/2} = \mathbf{\Sigma}^{-1}$.

   (d) The (square) matrix $\mathbf{\Sigma}$ is said to be *positive definite* if $\mathbf{a}^\top\mathbf{\Sigma a} > 0$ for all vectors $\mathbf{a} \neq \mathbf{0}$. Show that the eigenvalues of a positive definite matrix are all strictly positive. Hint: start with the definition of an eigenvalue and the corresponding eigenvalue: $\mathbf{\Sigma v} = \lambda\mathbf{v}$. Eigenvectors are typically scaled to have length one, so you may assume $\mathbf{v}^\top\mathbf{v} = 1$.

   (e) Let $\mathbf{\Sigma}$ be a symmetric, positive definite matrix. Putting together a couple of results you have proved above, establish that $\mathbf{\Sigma}^{-1}$ exists.

25. Let $\mathbf{X}$ be an $n \times p$ matrix of constants. The idea is that $\mathbf{X}$ is the "design matrix" in the linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, so this problem is really about linear regression.

   (a) Recall that $\mathbf{A}$ symmetric means $\mathbf{A} = \mathbf{A}^\top$. Let $\mathbf{X}$ be an n by p matrix. Show that $\mathbf{X}^\top\mathbf{X}$ is symmetric.

   (b) Recall the definition of linear independence. The columns of $\mathbf{A}$ are said to be *linearly dependent* if there exists a column vector $\mathbf{v} \neq \mathbf{0}$ with $\mathbf{Av} = \mathbf{0}$. If $\mathbf{Av} = \mathbf{0}$ implies $\mathbf{v} = \mathbf{0}$, the columns of $\mathbf{A}$ are said to be linearly *independent*. Show that if the columns of $\mathbf{X}$ are linearly independent, then $\mathbf{X}^\top\mathbf{X}$ is positive definite.

   (c) Show that if $\mathbf{X}^\top\mathbf{X}$ is positive definite then $(\mathbf{X}^\top\mathbf{X})^{-1}$ exists.

   (d) Show that if $(\mathbf{X}^\top\mathbf{X})^{-1}$ exists then the columns of $\mathbf{X}^\top\mathbf{X}$ are linearly independent.

   (e) Show that if the columns of $\mathbf{X}^\top\mathbf{X}$ are linearly independent, then the columns of $\mathbf{X}$ are linearly independent.

   This is a good problem because it establishes that the least squares estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$ exists if and only if the columns of $\mathbf{X}$ are linearly independent.