

STA 2101/442 Assignment 1 (Mostly Review)¹

The questions on this assignment are practice for the quiz on Friday September 15th, and are not to be handed in. For the linear algebra part starting with Question 10, 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. For each of the following distributions, derive a general expression for the Maximum Likelihood Estimator (MLE); don't bother with the second derivative test. Then use the data to calculate a numerical estimate; you should bring a calculator to the quiz in case you have to do something like this.

(a) $p(x) = \theta(1 - \theta)^x$ for $x = 0, 1, \dots$, where $0 < \theta < 1$. Data: 4, 0, 1, 0, 1, 3, 2, 16, 3, 0, 4, 3, 6, 16, 0, 0, 1, 1, 6, 10. Answer: 0.2061856

(b) $f(x) = \frac{\alpha}{x^{\alpha+1}}$ for $x > 1$, where $\alpha > 0$. Data: 1.37, 2.89, 1.52, 1.77, 1.04, 2.71, 1.19, 1.13, 15.66, 1.43 Answer: 1.469102

(c) $f(x) = \frac{\tau}{\sqrt{2\pi}} e^{-\frac{\tau^2 x^2}{2}}$, for x real, where $\tau > 0$. Data: 1.45, 0.47, -3.33, 0.82, -1.59, -0.37, -1.56, -0.20 Answer: 0.6451059

(d) $f(x) = \frac{1}{\theta} e^{-x/\theta}$ for $x > 0$, where $\theta > 0$. Data: 0.28, 1.72, 0.08, 1.22, 1.86, 0.62, 2.44, 2.48, 2.96 Answer: 1.517778

2. In the *centered* linear regression model, sample means are subtracted from the explanatory variables, so that values above average are positive and values below average are negative. Here is a version with one explanatory variable. Independently for $i = 1, \dots, n$, let $Y_i = \beta_0 + \beta_1(x_i - \bar{x}) + \epsilon_i$, where

β_0 and β_1 are unknown constants (parameters).

x_i are known, observed constants.

$\epsilon_1, \dots, \epsilon_n$ are random variables with $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma^2$ and $Cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$.

σ^2 is an unknown constant (parameter).

Y_1, \dots, Y_n are observable random variables.

- (a) What is $E(Y_i)$? $Var(Y_i)$?
- (b) Prove that $Cov(Y_i, Y_j) = 0$. Use the definition $Cov(U, V) = E\{(U - E(U))(V - E(V))\}$.

¹This 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](https://creativecommons.org/licenses/by-sa/3.0/). Use any part of it as you like and share the result freely. The L^AT_EX source code is available from the course website: <http://www.utstat.toronto.edu/~brunner/oldclass/appliedf17>

- (c) If ϵ_i and ϵ_j are independent (not just uncorrelated), then so are Y_i and Y_j , because functions of independent random variables are independent. Proving this in full generality requires advanced definitions, but in this case the functions are so simple that we can get away with an elementary definition. Let X_1 and X_2 be independent random variables, meaning $P\{X_1 \leq x_1, X_2 \leq x_2\} = P\{X_1 \leq x_1\}P\{X_2 \leq x_2\}$ for all real x_1 and x_2 . Let $Y_1 = X_1 + a$ and $Y_2 = X_2 + b$, where a and b are constants. Prove that Y_1 and Y_2 are independent.
- (d) In *least squares estimation*, we observe random variables Y_1, \dots, Y_n whose distributions depend on a parameter θ , which could be a vector. To estimate θ , write the expected value of Y_i as a function of θ , say $E_\theta(Y_i)$, and then estimate θ by the value that gets the observed data values as close as possible to their expected values. To do this, minimize

$$Q = \sum_{i=1}^n (Y_i - E_\theta(Y_i))^2.$$

The value of θ that makes Q as small as possible is the least squares estimate.

Using this framework, find the least squares estimates of β_0 and β_1 for the centered regression model. The answer is a pair of formulas. Show your work.

- (e) Because of the centering, it is possible to verify that the solution actually *minimizes* the sum of squares Q , using only single-variable second derivative tests. Do this part too.
- (f) How about a least squares estimate of σ^2 ?
- (g) You know that the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ must be unbiased, but show it by calculating their expected values for this particular case.
- (h) Calculate $\hat{\beta}_0$ and $\hat{\beta}_1$ for the following data. Your answer is a pair of numbers.

| | | | | | |
|-----|---|----|---|---|---|
| x | 8 | 7 | 7 | 9 | 4 |
| y | 9 | 13 | 9 | 8 | 6 |

I get $\hat{\beta}_1 = \frac{1}{2}$.

- (i) Going back to the general setting (not just the numerical example with $n = 5$), Suppose the ϵ_i are normally distributed.
- i. What is the distribution of Y_i ?
 - ii. Write the log likelihood function.
 - iii. Obtain the maximum likelihood estimates of β_0 and β_1 ; don't bother with σ^2 . The answer is a pair of formulas. *Don't do more work than you have to!* As soon as you realize that you have already solved this problem, stop and write down the answer.
3. For the general uncentered regression model $Y_i = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_{p-1} x_{i,p-1} + \epsilon_i$ (you fill in details as necessary), show that the maximum likelihood and least squares estimates of the β_j parameters are the same. You are most definitely *not* being asked to an explicit formula; that's too much work. Just show that (assuming existence) they are the same.

4. Let $x_1, \dots, x_{n_1} \stackrel{i.i.d.}{\sim} N(\mu_1, \sigma^2)$, and $y_1, \dots, y_{n_2} \stackrel{i.i.d.}{\sim} N(\mu_2, \sigma^2)$. These two random samples are independent, meaning all the x variables are independent of all of the y variables.

Every elementary Statistics text tells you that

$$t = \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2),$$

where

$$s_p^2 = \frac{\sum_{i=1}^{n_1} (x_i - \bar{x})^2 + \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{n_1 + n_2 - 2}$$

This is the basis of tests and confidence intervals for $\mu_1 - \mu_2$.

- Using (not proving) the distribution fact above, derive a $(1 - \alpha)100\%$ confidence interval for $\mu_1 - \mu_2$. “Derive” means show all the High School algebra. Use $t_{\alpha/2}$ to denote the point cutting off the upper $\alpha/2$ of the t distribution with $n_1 + n_2 - 2$ degrees of freedom. The core of the answer is a formula for the lower confidence limit and another formula for the upper confidence limit.
- Suppose you wanted to test $H_0 : \mu_1 = \mu_2$. Give a formula for the test statistic.
- Prove that $H_0 : \mu_1 = \mu_2$ is rejected at significance level α if and only if the $(1 - \alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ does not include zero. It is easiest to start with the event that the null hypothesis is *not* rejected.
- Here are some x_i and y_i values, together with a set of $t_{0.025}$ critical values produced by R.

| | | | | |
|-----|---|----|---|-----|
| x | 6 | 10 | 8 | 12 |
| y | 7 | 4 | 8 | 7 9 |

```
> df = 1:10; critvalue = qt(0.975,df)
> round(rbind(df,critvalue),3) # Round to 3 digits
```

```
df          1.000 2.000 3.000 4.000 5.000 6.000 7.000 8.000 9.000 10.000
critvalue 12.706 4.303 3.182 2.776 2.571 2.447 2.365 2.306 2.262 2.228
```

- Calculate s_p^2 . My answer is $34/7$.
- Give a 95% confidence interval for $\mu_1 - \mu_2$. I get $(-1.496, 5.496)$, give or take a little rounding error.
- What is the numerical value of the t statistic for testing $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$? I get $t = 1.353$. Do you reject H_0 at $\alpha = 0.05$? Are you able to conclude that there is a real difference between means? Answer Yes or No.

5. A medical researcher conducts a study using twenty-seven litters of cancer-prone mice. Two members are randomly selected from each litter, and all mice are subjected to daily doses of cigarette smoke. For each pair of mice, one is randomly assigned to Drug A and one to Drug B. Time (in weeks) until the first clinical sign of cancer is recorded.
- State a reasonable model for these data. Remember, a statistical model is a set of assertions that partly specify the probability distribution of the observable data. For simplicity, you may assume that the study continues until all the mice get cancer, and that log time until cancer has a normal distribution.
 - What is the parameter space for your model?
6. Suppose that volunteer patients undergoing elective surgery at a large hospital are randomly assigned to one of three different pain killing drugs, and one week after surgery they rate the amount of pain they have experienced on a scale from zero (no pain) to 100 (extreme pain).
- State a reasonable model for these data. For simplicity, you may assume normality.
 - What is the parameter space?
7. A fast food chain is considering a change in the blend of coffee beans they use to make their coffee. To determine whether their customers prefer the new blend, the company selects a random sample of $n = 100$ coffee-drinking customers and asks them to taste coffee made with the new blend and with the old blend. Customers indicate their preference, Old or New. State a reasonable model for these data. What is the parameter space?
8. 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.
- _____ The p -value is the probability that the null hypothesis is true.
 - _____ The p -value is the probability that the null hypothesis is false.
 - _____ In a study comparing a new drug to the current standard treatment, the null hypothesis is rejected. We conclude that the new drug is ineffective.
 - _____ If $p > .05$ we reject the null hypothesis at the .05 level.
 - _____ If $p < .05$ we reject the null hypothesis at the .05 level.
 - _____ The greater the p -value, the stronger the evidence against the null hypothesis.
 - _____ 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.
 - _____ The 95% confidence interval for β_3 is from -0.26 to 3.12 . This means $P\{-0.26 < \beta_3 < 3.12\} = 0.95$.

9. Suppose that several studies have tested the same hypothesis or similar hypotheses. For example, six studies could all be testing the effect of a treatment for arthritis. Two studies rejected the null hypothesis, one was almost significant, and the other three were clearly non-significant. What should be concluded? Ideally, one would pool the data from the six studies, but in practice the raw data are not available. All we have are the published test statistics. How do we combine the information we have and come to an overall conclusion? That is the main task of *meta-analysis*. In this question you will develop some simple, standard tools for meta-analysis.
- (a) Let the test statistic T be continuous, with pdf $f(t)$ and strictly increasing cdf $F(t)$ under the null hypothesis. The null hypothesis is rejected if $T > c$. Show that if H_0 is true, the distribution of the p -value is $U(0, 1)$. Derive the density. Start with the cumulative distribution function of the p -value: $Pr\{P \leq x\} = \dots$
- (b) Suppose H_0 is false. Would you expect the distribution of the p -value to still be uniform? Pick one of the alternatives below. You are not asked to derive anything for now.
- The distribution should still be uniform.
 - We would expect more small p -values.
 - We would expect more large p -values.
- (c) Let $P_i \sim U(0, 1)$. Show that $Y_i = -2 \ln(P_i)$ has a χ^2 distribution. What are the degrees of freedom? Remember, a chi-squared with ν degrees of freedom is $\text{Gamma}(\alpha = \nu/2, \beta = 2)$.
- (d) Let P_1, \dots, P_n be a random sample of p -values with the null hypotheses all true, and let $Y = \sum_{i=1}^n -2 \ln(P_i)$. What is the distribution of Y ? Only derive it (using moment-generating functions) if you don't know the answer.
- (e) Suppose we observe the following random sample of p -values: 0.016 0.188 0.638 0.148 0.917 0.124 0.695. For the test statistic of Question 9d,
- What is the critical value at $\alpha = 0.05$? The answer is a number. It doesn't matter how you find it. If you have to do something like this on a quiz or the final, several numbers will be supplied and you will have to choose the right one. My answer is 23.68.
 - What is the value of the test statistic? The answer is a number. My answer is 21.41.
 - What null hypothesis are you testing? I find it easier to state in words, rather than in symbols.
 - Do you reject the null hypothesis at $\alpha = 0.05$? Answer Yes or No.
 - What if anything do you conclude?

10. 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
11. Which statement is true?
- (a) $a(\mathbf{B} + \mathbf{C}) = a\mathbf{B} + a\mathbf{C}$
 - (b) $a(\mathbf{B} + \mathbf{C}) = \mathbf{Ba} + \mathbf{Ca}$
 - (c) Both a and b
 - (d) Neither a nor b
12. 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
13. 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
14. 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
15. 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

16. 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

17. 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

18. 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.

19. 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.

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

21. Suppose that the square matrices \mathbf{A} and \mathbf{B} both have inverses. Using the definition of an inverse, prove that $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$. Because you are using the definition, you have two things to show.

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

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

24. Using Question 23, prove that if the inverse of a symmetric matrix exists, it is also symmetric.

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

26. The $p \times p$ matrix Σ is said to be *positive definite* if $\mathbf{a}^\top \Sigma \mathbf{a} > 0$ for all $p \times 1$ 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: $\Sigma \mathbf{v} = \lambda \mathbf{v}$. Eigenvectors are typically scaled to have length one, so you may assume $\mathbf{v}^\top \mathbf{v} = 1$.
27. Recall the *spectral decomposition* of a symmetric matrix (for example, a variance-covariance matrix). Any such matrix Σ can be written as $\Sigma = \mathbf{P} \Lambda \mathbf{P}^\top$, where \mathbf{P} is a matrix whose columns are the (orthonormal) eigenvectors of Σ , Λ is a diagonal matrix of the corresponding eigenvalues, and $\mathbf{P}^\top \mathbf{P} = \mathbf{P} \mathbf{P}^\top = \mathbf{I}$. If Σ is real, the eigenvalues are real as well.
- (a) Let Σ be a square symmetric matrix with eigenvalues that are all strictly positive.
- What is Λ^{-1} ?
 - Show $\Sigma^{-1} = \mathbf{P} \Lambda^{-1} \mathbf{P}^\top$
- (b) Let Σ be a square symmetric matrix, and this time the eigenvalues are non-negative.
- What do you think $\Lambda^{1/2}$ might be?
 - Define $\Sigma^{1/2}$ as $\mathbf{P} \Lambda^{1/2} \mathbf{P}^\top$. Show $\Sigma^{1/2}$ is symmetric.
 - Show $\Sigma^{1/2} \Sigma^{1/2} = \Sigma$, justifying the notation.
- (c) Now return to the situation where the eigenvalues of the square symmetric matrix Σ are all strictly positive. Define $\Sigma^{-1/2}$ as $\mathbf{P} \Lambda^{-1/2} \mathbf{P}^\top$, where the elements of the diagonal matrix $\Lambda^{-1/2}$ are the reciprocals of the corresponding elements of $\Lambda^{1/2}$.
- Show that the inverse of $\Sigma^{1/2}$ is $\Sigma^{-1/2}$, justifying the notation.
 - Show $\Sigma^{-1/2} \Sigma^{-1/2} = \Sigma^{-1}$.
- (d) Let Σ be a symmetric, positive definite matrix. How do you know that Σ^{-1} exists?
28. 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 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 $\mathbf{A} \mathbf{v} = \mathbf{0}$. If $\mathbf{A} \mathbf{v} = \mathbf{0}$ 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.
- (b) Show that if $\mathbf{X}^\top \mathbf{X}$ is positive definite then $(\mathbf{X}^\top \mathbf{X})^{-1}$ exists.
- (c) Show that if $(\mathbf{X}^\top \mathbf{X})^{-1}$ exists then the columns of \mathbf{X} are linearly independent.

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