Vanderbilt University
Political Science Department
Fall 2024

Stats 1
(PSCI 8356)
Professor Jim Bisbee

PRACTICE MIDTERM EXAMINATION:

These are meant to be examples of the types of questions that might appear on the exam. You will need either statistical tables like those appearing in the back of the textbook or access to a statistical software program (i.e., `R`) that can provide similar information.

Be sure to show all your work and to use complete sentences to provide explanations.

1. Show that:

    (a) You can derive Bayes' Theorem using the conditional probability and multiplicative law of probability definitions.

    (b) For a Bernoulli distribution $E[Y] = \pi$ and $V[Y] = \pi(1 - \pi)$

    (c) $E[c] = c$

    (d) $E[aY] = aE[Y]$

    (e) $V[X + Y] = V[X] + 2Cov[X, Y] + V[Y]$.

2. Imagine the Democratic Party organizes a convention to reinvent itself and invites 19 former presidential candidates: Bennet, Biden, Booker, Bullock, Buttigieg, Castro, Delaney, Gabbard, Harris, Klobuchar, Messam, O'Rourke, Ryan, Sanders, Sestak, Steyer, Warren, Williamson, and Yang.

    Seeing this, a renowned streaming company decides to create a special that features a conversation between two of these members to capture the different perspectives at the convention. To ensure the pairing is equally likely, they will randomly select two people from the list.

    (a) How many events are in the sample space?

    (b) Define event C to be: select at least one woman (the women are Gabbard, Harris, Klobuchar, Warren, and Williamson). What is the probability of this event?

    (c) Define event D to be: select two women. What is the probability of this event?

    (d) Are C and D mutually exclusive?

    (e) Are C and D independent?

3. Suppose there is a new international cyberthreat a malicious computer virus that is difficult to detect. Suppose the CIA has determined that 3% of computers in the US have it, so the best estimate of the baseline probability of having the virus is .03. The CIA has also developed a diagnostic test that detects it in a computer according to the following probabilities.

	Has Virus (3 %)	**Doesn't Have Virus (97 %)**
Test Positive	85%	11%
Test Negative	15%	89%

Table 1: Test Results vs Virus Presence

    (a) Given that your computer tests positive for the virus, what is the probability that it actually has it?

    (b) Consider the following scenario: You are given a new diagnostic test. Every time you ask this test whether your computer has a virus, it always tell you that your computer is virus-free, regardless of the actual state. This test would correctly classify all true negatives—every time the computer does not have a virus, this new test would be correct. Would this be a better test overall? Why or why not?

(c) Now suppose the CIA rolls out an improved version of its test. In the improved version, the false positives are reduced from 11% to 4%. What is the new probability that your computer actually has the virus when the test says it does?

(d) What if the improved version of the CIA's test instead leaves the false positive rate at 11% but improves the true positive rate from 85% to 92%. Now what is the probability that your computer has the virus when the test says it does?

(e) The two different, improved CIA tests were both an improvement of 7%. Why did they not have the same effect on the probability that your computer has the virus given that the test says it does?

4. Suppose $\hat{\theta}$ is an estimator for θ and $E(\hat{\theta}) = a\theta + b$ for some nonzero constants a and b.

   (a) In terms of a, b, and θ, what is $B(\hat{\theta})$?

   (b) Find a function of $\hat{\theta}$ – say, $\hat{\theta}^*$ – that is an unbiased estimator for θ.

   (c) Express $MSE(\hat{\theta}^*)$ as a function of $VAR(\hat{\theta})$.

   (d) Give an example of a value for a for which $MSE(\hat{\theta}^*) < MSE(\hat{\theta})$.

   (e) Give an example of values for a and b for which $MSE(\hat{\theta}^*) > MSE(\hat{\theta})$

5. If Y has a binomial distribution with parameters n and p, then $\hat{p}_1 = Y/n$ is an unbiased estimator of p. Another estimator of p is $\hat{p}_2 = (Y+1)/(n+2)$.

   (a) Derive the bias of $\hat{p}_2$.

   (b) Derive $MSE(\hat{p}_1)$ and $MSE(\hat{p}_2)$.

   (c) For what values of p is $MSE(\hat{p}_1) < MSE(\hat{p}_2)$?