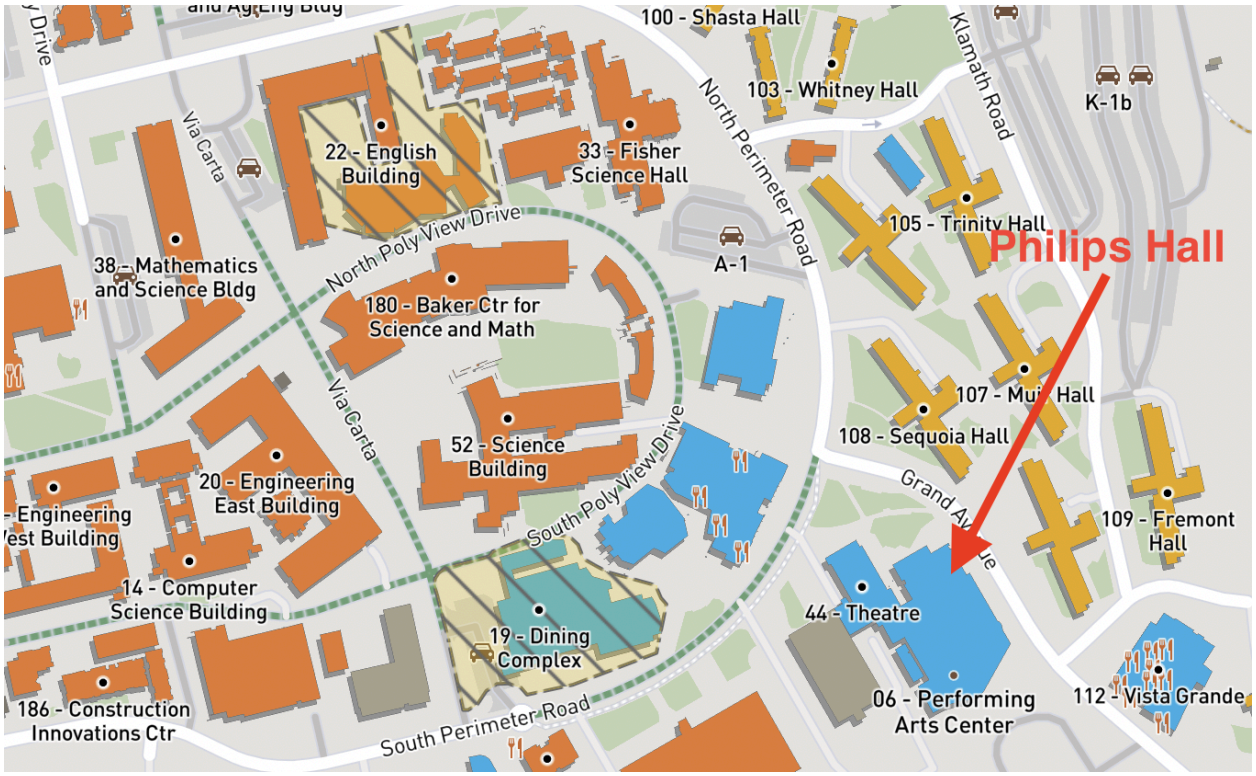
# STAT 312 Final Exam (Practice)

The final exam will take place on Saturday, March 12, from 1-4 PM, in Philips Hall, which is a large room in the back of the Performing Arts Center. See the map below.




A practice final is provided below.

Here are the instructions:

- You can print out the weekly handouts and R assignments that you have submitted to Canvas to use as a reference during the exam. They must be printed. You cannot refer to the notes on your laptop or tablet. They will be turned in at the end of the exam.
- You are **not allowed** to bring any other references. That includes in-class notes, quizzes, practice exams, etc.
- You are allowed to use a calculator.
- You may be asked to produce R code that produces the answer for any calculations that are hard to do by hand (e.g., normal calculations).
- You are expected to show all work.
- I encourage you to take the practice exam under realistic conditions. That is, you will have 170 minutes on the actual exam.

# Practice Final

1. **[10 pts]** What is the probability that you just inhaled a molecule that Julius Caesar exhaled in his dying breath?

   Suppose that there are $10^{44}$ molecules in the atmosphere, and each breath (inhaled or exhaled) contains $2.2 \times 10^{22}$ molecules. (Assume that after more than two thousand years the exhaled molecules are uniformly spread about the world.)[1]

2. **[20 pts]** The diameter $Y$ (in inches) of a randomly selected leaf on a certain species of tree is described by the PDF

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 1 \\ \frac{1}{4} & 1 < x < 4 \\ 0 & \text{otherwise} \end{cases}$$

   a. Sketch the PDF of $Y$.
   b. Calculate $P(Y > 2.3)$.
   c. Calculate $E[Y]$.
   d. Set up the integrals to show that $Var[Y] = \frac{695}{576}$.
   e. Suppose we randomly and independently select 18 leaves from this species: $Y_1, Y_2, \ldots, Y_{18}$. Let $\bar{Y}$ be the average diameter of these 18 leaves. Sketch the PDF of $\bar{Y}$ on top of the PDF of $Y$ you sketched above.
   f. Calculate $P(\bar{Y} > 2.3)$. (You can either make an approximation or provide R code.)

3. **[10 pts]** The dioxin TCDD is a toxic contaminant that was present in the chemical Agent Orange used during the Vietnam War. Researchers want to know if TCDD concentrates more in fat tissue than in blood plasma. They measured the TCDD levels in blood plasma and fat tissue of 20 Massachusetts Vietnam veterans who were possibly exposed to Agent Orange. The data is stored in R `data.frame` called `vietnam_data`, shown below:

[1] This question is taken directly from *Innumeracy*. As I mentioned, there will be one question on the exam that is like this. You don't need to have read *Innumeracy* to solve the question, but it certainly helps!

```
vietnam_data
```

A data.frame: 20 × 3

| Veteran | PlasmaTCDD | FatTCDD |
|---|---|---|
| <int> | <dbl> | <dbl> |
| 1 | 2.5 | 4.9 |
| 2 | 3.1 | 5.9 |
| 3 | 2.1 | 4.4 |
| 4 | 3.5 | 6.9 |
| 5 | 3.1 | 7.0 |
| 6 | 1.8 | 4.2 |
| 7 | 6.0 | 10.0 |
| 8 | 3.0 | 5.5 |
| 9 | 36.0 | 41.0 |
| 10 | 4.7 | 4.4 |
| 11 | 6.9 | 7.0 |
| 12 | 3.3 | 2.9 |
| 13 | 4.6 | 4.6 |
| 14 | 1.6 | 1.4 |
| 15 | 7.2 | 7.7 |
| 16 | 1.8 | 1.1 |
| 17 | 20.0 | 11.0 |
| 18 | 2.0 | 2.5 |
| 19 | 2.5 | 2.3 |
| 20 | 4.1 | 2.5 |

a. Write R code to test the hypothesis that TCDD concentrates *more* in fat than in plasma.

b. The R output is shown below. (Some parts have been redacted to avoid spoiling part a.) State your conclusion about the hypotheses.

```
        One Sample t-test

data: [redacted]
t = 1.3068, df = 19, p-value = 0.1034
alternative hypothesis: [redacted]
95 percent confidence interval:
 -0.2811607        Inf
sample estimates:
mean of x
     0.87
```

4. **[10 pts]** A mammogram is a test used to detect breast cancer. It accurately classifies about 80% of cancerous tumors and 90% of benign tumors.
    a. Using conditional probability notation, express the probability that is equal to 80%.
    b. In a study, physicians were asked what the probability of breast cancer is for a woman who was initially thought to have a 1% risk of cancer but who ended up with a positive mammogram result. Most physicians estimated the probability of cancer to be about 75%. Do you agree?
5. **[10 pts]** A city wants to know if a majority of voters favor a proposal to increase gasoline sales tax by 4% to provide badly needed revenues for street repairs. They take a simple random sample of 400 voters. They decide that if fewer than 190 of the sampled voters support the gas tax, they will conclude that the proposal does not have majority support and abandon the proposal.
    a. What is the probability of a Type I error, under the null hypothesis that 50% of voters in the city support the proposal? (You can either make an approximation or provide R code.)
    b. What is the probability of a Type II error, under the alternative hypothesis that 45% of voters in the city support the proposal? (You can either make an approximation or provide R code.)
6. **[15 pts]** The US Senate currently has 76 men and 24 women. A typical subcommittee has 10 members. If a subcommittee is randomly selected among the 100 members of the Senate, what is the probability that there are more men than women on a subcommittee?
    a. Write R code to simulate this probability.
    b. What is the exact probability? (You can calculate it or provide R code.)
7. **[10 pts]** The height of American women is normally distributed with a mean of 64.5 inches and a standard deviation of 2.5 inches. If eight American women are chosen at random, what is the probability that exactly one of them will be taller than 6 feet (= 72 inches)? (You can either calculate the probability or provide R code.)

# Answers

1. The solution can be found in *Innumeracy*, end of Chapter 1. (Remember: one question will be taken from this reading.)
2. 
    a. 
    b. .425
    c. 49/24 = 2.04167
    d. $\int_0^1 \frac{x^3}{2} dx + \int_1^4 \frac{x^2}{4} dx - (49/24)^2$
    e. It should look like a normal distribution centered around 49/24, with a smaller variance than the original distribution.
    f. About .16 (using the 68-95-99.7 rule) **or**

```
1 - pnorm(2.3, mean=49/24, sd=sqrt(695/576) / sqrt(18))
```

3. 

```
a. t.test(vietnam_data$FatTCDD - vietnam_data$PlasmaTCDD,
          mu=0,
          alternative="greater")
```

    b. Because the p-value is greater than .05, we cannot reject the null hypothesis that TCDD concentrates in fat tissue and blood plasma equally.
4. 
    a. P(mammogram is positive | cancerous) = 0.80
    b. Don't agree, the actual probability is about 7.5%.
5. .
    a. About 16%
    b. About 16% (The power is 84%).
6. .

```
a. sims <- replicate(10000,
                     sum(sample(c(rep(1, 76), rep(0, 24)),
                                size=10,
                                replace=FALSE)))
   sum(sims > 5) / 10000
b. 1 - phyper(5, 76, 24, 10)
```

7. About $\binom{8}{1}(.0015)^1(.9985)^7 = .0118$. If you want to be exact, you can do:

```
dbinom(1, 8, 1 - pnorm(72, mean=64.5, sd=2.5))
```