Final Exam Practice Problems

1. In each of the following scenarios, identify the population and the sample.

    a. 50 De Anza College students were sampled to find out the amount of sleep they get each night.
    b. 300 registered voters in a large city were asked whether they plan to vote in the next election.
    c. 75 randomly selected light bulbs made a factory were tested for defectiveness.

2. For each of the following, identify if it's an observational study or an experiment. If it's an observational study, identify the population and the sample. If it's an experiment, identify the treatment group(s), control group (if any), placebo group (if any), the explanatory variable and the response variable.

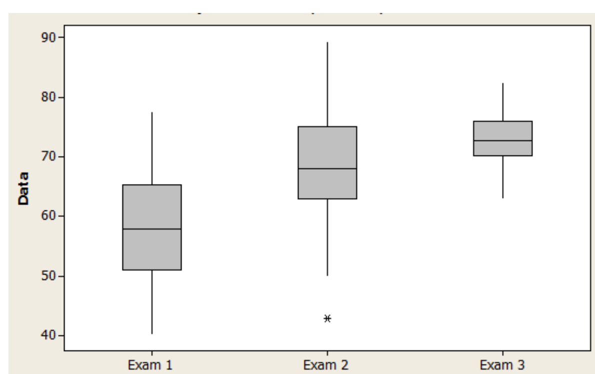
    a. Researchers wish to see if the studying with peers before a math final exam affects performance on the exam. They randomly assign 100 randomly selected students to 3 groups. The members of the first group spend 6 hours studying with peers for the final exam, the members of the second group spend 2 hours studying with peers for the final exam, while the members of the third group study alone. The final exam scores of each group are analyzed.

    b. 5 randomly selected teachers from 10 different randomly selected elementary schools in a city are interviewed. The researchers find that teachers in higher grades spend more time, on average, in preparing lessons than teachers in lower grades.

3. For each of the following, decide if it is a categorical variable or a numerical (quantitative) variable. If it's numerical, decide if it's discrete or continuous.

    a. The GPA of a student
    b. The major of a student
    c. The number of units a students is currently enrolled in
    d. Whether the student uses public transportation to get to campus
    e. The amount of money in a student's wallet
    f. How many quarters the student has been at De Anza

4. The three boxplots shown display midterm exam scores for students taking a statistics course.



    a. Which exam has the smallest IQR?

    b. If the outlier was removed from the data for Exam 2, which of the following statistics could be significantly affected? **Choose from:** IQR, Mean, Median, Range, Standard Deviation

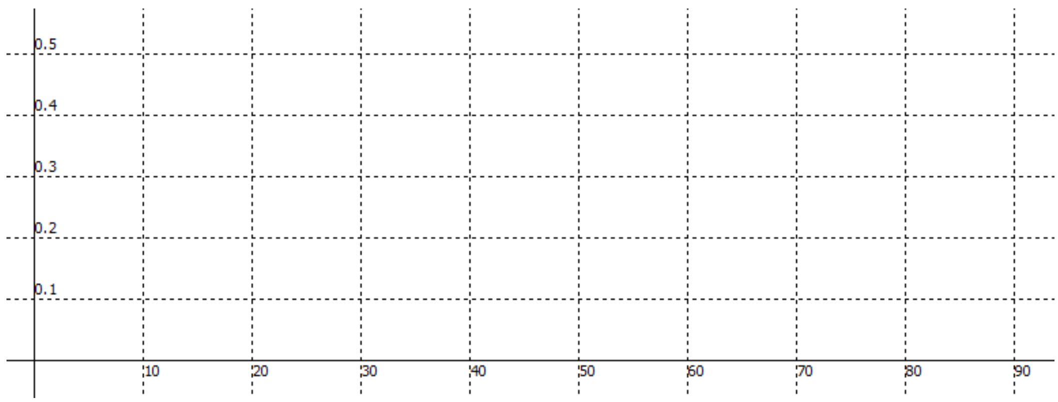
    c. Describe all of the similarities and differences between the three exam scores.

5. The following data represents the amount of time 50 students take to complete an exam:

Minutes	**Frequency**	**Relative Freq**
31-40	17	
41-50	13	
51-60	10	
61-70	8	
71-80	2	

a. Determine the relative frequencies for each bin and fill them in the column of the chart above.

b. Make a relative frequency histogram in the graph below.



c. **Without calculating**, what can you say about mean and median of this data?

6. Sketch a graph for bivariate data in which there is:
   a. Strong, positive linear correlation
   b. Weak, negative linear correlation
   c. Zero linear correlation

7. The following chart gives the gold medal times for every other Summer Olympics for the women's 100-meter freestyle swimming.

Year	1924	1932	1952	1960	1968	1976	1984	1992	2000	2008	2012
Time (in sec)	72.4	66.8	66.8	61.2	60.0	55.65	55.92	54.64	53.83	53.12	53.00

a. Find the equation of the least squares regression line, assuming 'Year' is the independent variable. Include an explanation of how you used your calculator to get this.
b. What is the correlation coefficient?
c. What is the coefficient of determination? Interpret its meaning in the context of the problem.
d. Explain the meaning of the slope of your least squares line in the context of this problem.
e. Find the predicted gold medal time for 1984 using the least squares line.
f. Compute the residual for the year 1984. Is the predicted value an overestimate or an underestimate of the actual value?
g. Would you use this model to estimate the gold medal time for women's freestyle 100-meter swimming time for 2020 Olympics? Explain.

8. A new type of flu virus has infected several people. The data was collected for the first 10 days. The number of infected people after *x* days can be modeled by the function $y = 140(1.09)^x$.

   a. In this model what does the number 140 represent?
   b. In this model what does the number 1.09 represent?
   c. Predict the number of people with the flu after 3 days.
   d. After how many days are 275 people infected?
   e. Would you use this model to predict the number of infected people a month later? Why or why not?

9. 1000 students (600 morning, 300 afternoon, 100 night) were asked how often they use the campus library. The results are summarized in the table below:

	Never uses library	*Sometimes uses library*	*Frequently uses library*	***Total***
Morning	200	240	160	**600**
Afternoon	80	150	70	**300**
Night	70	10	20	**100**
Total	**350**	**400**	**250**	**1000**

   a. Determine the percentage of students that never use the library.
   b. Determine the percentage of Night students who Never use the library.
   c. Determine the percentage of students who are Afternoon students **and**Sometimes use the library.
   d. Determine the percentage of students who are Afternoon students **or** Sometimes use the library.
   e. Are "Night Students" and "Never uses Library" independent events? Explain.
   f. Are "Morning Students" and "Sometimes uses Library" independent events? Explain.

10. Customers arrive at a restaurant in groups of 1 to 6. Based on past data, a discrete probability distribution table is shown below. The random variable x refers to the number of people in a group. The value for 6 is missing.

   a. What is the probability that the next group that arrives at the restaurant has 6 people? Write the probability statement and then find the probability.
   b. What percentage of groups have 4 people or less?
   c. Determine the expected number of people a group arriving at the restaurant (in other words, find the expected value of this random variable).

x	*P*(*x*)
1	.20
2	.30
3	.15
4	.20
5	.05
6	

11. The cooking time for a 12-pound turkey follows a normal distribution with a mean of 250 minutes and a standard deviation of 20 minutes.

    a. What is the probability that the cooking time for 12-pound turkey will be between 220 and 260 minutes? Sketch a normal curve, label it fully and shade the appropriate area. Write the probability statement and then find the probability.

    b. What is the probability that the cooking time for 12-pound turkey will be under 230 minutes? Sketch a normal curve, label it fully and shade the appropriate area. Write the probability statement and then find the probability.

    c. What is the probability that the cooking time for 12-pound turkey will be over 200 minutes? Sketch a normal curve, label it fully and shade the appropriate area. Write the probability statement and then find the probability.

    d. Determine the minimum cooking time needed for the 25% of 12-pound turkeys that need the most time to cook. Sketch a normal curve, label it fully and shade the appropriate area. Then find the percentile.

12. A person standing close to the top of a 64 feet tall building throws a baseball vertically upward. The quadratic function $s(t) = -16t^2 + 64t + 80$ models the ball's height above the ground, $s(t)$, in feet, t seconds after it was thrown.

    a. How many seconds does it take the ball to reach maximum height? Show your work.
    b. What is the maximum height the ball reaches? Show your work.
    c. How many seconds until the ball hits the ground? Show your work.

13. From a deck with 5 Black and 4 Yellow cards, you first draw one card, and then another WITHOUT REPLACEMENT.
    a. Draw the tree diagram for this situation.
    b. Using your tree diagram, find the probability that you
        i. Draw a Black card AND a Yellow card, in either order.
        ii. First draw a Yellow card and then draw a Black card.
        iii. Draw a Black card, GIVEN that you drew a Yellow card on the first draw.
        iv. Get a Black card on the second draw.

14. Given that P(E) = 0.2, P(K) = 0.6, and P(E and K) = 0.15, find P(K | E).

15. If events R and S are mutually exclusive, find P(R and S).

16. If events R and S are independent and P(R) = 0.4 and P(S) = 0.6, find P(R and S).

17. At Whatsamatta University, students rate professors on a scale of 1 to 10 with 10 being the highest. Last Spring the following ratings were collected for three different professors in three different departments:

Professor	Rating	Department	Dept. Mean $\overline{x}$	Std. Dev. s
A	10	History	8	4
B	7	Phys Ed	9	1
C	8	Math	4	2

Which of the three professors was the best professor compared to his or her department? Explain your reasons for your answer. Be complete in your answer.

18. Describe at least one problem you might have in obtaining a representative sample if you were to do a mail-in survey. In this type of survey, you would mail the surveys and each person would have to mail back their completed survey to you.

19. Suppose you are to find a sample of 100 De Anza students. Also suppose you have access to any data that the college president would have.

    a. Describe a method of finding a simple random sample.
    b. Describe a method of finding a systematic sample.
    c. Describe a method of finding a stratified random sample.
    d. Describe another method of finding a stratified random sample.

20. **Sixty** randomly selected students were asked the number of phone calls they received yesterday. The results are as follows:

# phone calls	Frequency	Rel. frequency
1	6	
2	25	
3	12	
4		
6	9	

    a. Fill in the blanks in the above table. Round to 4 decimal places.
    b. Find the sample mean, $\overline{x}$.
    c. Find the standard deviation. Is it s or σ?
    d. Find the third quartile and describe its meaning in the context of the problem.
    e. What percent of the students received at least 4 phone calls?
    f. What percent of students received less than 3 phone calls?
    g. Carefully construct a histogram for this data.
    h. Carefully construct a boxplot for this data.

21. You invest $2000 in an account that pays 2.5% annual interest compounded monthly.

    a. How much money will be in the account 3 years later?
    b. After how many years will the amount of money in the account double?

22. Iodine-131 is used to treat thyroid cancer. It has a half-life of about 8 days in the body. Suppose that a patient is given 10 mg of Iodine-131. This means that the amount of Iodine that remains in the body after t days is given by $f(t) = 10 \cdot (0.5)^{t/8} = 10 \cdot (0.917)^t$

    a. How much Iodine-131 will remain in the patient's body after 3 days?
    b. After how many days will the amount of Iodine-131 remaining in the patient's body be 2.5 mg?
    c. Fully describe the meaning of the 0.917 in the context of the problem.

23. A rare disease afflicts 0.3% of the entire population. A new test will come out positive for patients with the disease about 84% of the time. The test gives a false positive about 26% of the time. Suppose your friend takes the test and it comes out positive! Given that you're a statistics expert, he comes to you to find out what his chances of actually having the disease are. Help him out! HINT: Make a two-way table. Use a grand total of 10,000 as usual.

Partial Solutions

1. a. Population: All De Anza College students: Sample: The 50 students who the data was actually collected from.

2. a. Experiment. The treatment groups are the first and second group, the control group is the third group, and there is no placebo group. Explanatory variable: the number of hours spent studying with peers. Response variable: performance on the final exam
b. Observational study. Population: All teachers working in elementary schools in the city. Sample: the 50 teachers from whom the data was actually collected.

3. a. Numerical – continuous; b. Categorical c. Numerical – discrete

4. a. Exam 3 b. Mean, Range, Standard deviation c. All graphs are fairly symmetric (especially if we remove the outlier from Exam 2). Exams 1 and 2 have a greater spread than Exam 3. The median for Exam 1 is the lowest and that for Exam 3 is the highest. Most likely, Exam 1 was a fairly difficult exam since the entire boxplot is quite a bit lower than the other two. All students for Exam 3 were fairly evenly prepared (and therefore performed similarly) since the range of scores is so small, whereas there were many more differences in the student performances for Exam 2.

5. a.

Minutes	**Frequency**	**Relative Freq**
31-40	17	17/50 = 0.34
41-50	13	13/50 = 0.25
51-60	10	10/50 = 0.2
61-70	8	8/50 = 0.16
71-80	2	2/50 = 0.04

6. See notes

7. a. $\hat{y} = 487.2322 - 0.2168x$ b. $r = -0.9604$ c. $r^2 = 0.9223$; 92.23% of the variation in the gold medal times for women's 100-m freestyle swimming the Olympics can be explained by the variation in years.
d. Each year, the gold medal swimming time decreased by 0.2168 seconds, on average. Realistically, since the Summer Olympics occur every *four* years, you can say that the from one Summer Olympic to the next, the gold medal swimming time decreased by 0.8672 seconds. e. $\hat{y} = 487.2322 - 0.2168(1984) = 57.11$ seconds
f. residual = $y - \hat{y} = 55.92 - 57.11 = -1.19$ seconds. The predicted value is an overestimate of the actual value.
g. It could predict the gold medal swimming time for 2020, but statistically speaking, such a prediction would not be reliable since 2020 falls outside of the range of years for which the data is given.

8. a. At the start, 140 people were infected with the flu
b. The growth factor is 1.09. This means that each day, the total number of people infected with the flu is 1.09 times more than the previous day. Since 1.09 = 109% = 100% + 9%, each day, the number of people infected with the flu *increases* by 9%.
c. $y = 140(1.09)^3 \approx 181$ people infected with the flu
d. After about 8 days; To get this, set up equation $140(1.09)^x = 275$ and solve for x (you will need to use logs)

9. a. P(never use library) = 350/1000 b. P(never use library | night student) = 70/100
c. P(afternoon student AND sometimes use library) = 150/1000
d. P(afternoon student OR sometimes use library) = (300 + 400 - 150)/1000 = 550/1000 (Use Addition Rule)
e. No, since the answers to a and b are different f. No, since P(morning student and sometimes uses library) ≠ 0

10. a. Missing entry is 0.1 since the sum of the probabilities must equal 1 (or 100%).
b. $P(X \leq 4) = 0.2 + 0.3 + 0.15 + 02 = 0.85$
c. $\mu = \sum x \cdot P(x) = 1(0.2) + 2(0.3) + 2(0.15) + 4(0.2) + 5(0.05) + 6(0.1) = 2.75$ is the number of people expected in a group arriving at the restaurant.

11. a. $P(220 < X < 260) = 0.6247$ b. $P(X < 230) = 0.1587$ c. $P(X > 200) = 0.9938$
d. About 263 minutes (For a, b, c, use normalcdf; for d, use invNorm)

12. a. 2 seconds (t-value of the vertex) b. 144 feet (y-value of the vertex) c. 5 seconds (use quadratic formula)

13. (See class notes – we went over this in detail)

14. $\mathrm{P(K|E)}=\frac{\mathrm{P(K\ and\ E)}}{\mathrm{P(E)}}=\frac{0.15}{0.2}=0.75$

15. $\mathrm{P(R\ and\ S)}=0$ since R and S are mutually exclusive (they cannot both occur so their joint probability is 0)

16. $\mathrm{P(R\ and\ S)=P(R)\cdot P(S|R)=P(R)\cdot P(S)}=(0.4)(0.6)=0.24$ (Note: since R and S are independent, $\mathrm{P(S|R)=P(S)}$)

17. Compute the z-scores for each professor. The one with the best z-score is the best professor compared to his/her department. The Math professor is the best compared to his/her department since his/her z-score is 2.

18. One problem is that you will have a non-response bias.

19. a. Make a list of all students and assign each a unique number from 1 through N (total number of students). Then his randInt(1, N) until you have a list of 100 distinct students for your sample.
b. In the list above, pick a random starting point and select every N/100-th student until you have 100 distinct students for your sample.
c. Stratify the students by gender (male and female), and select proportionally from each group. For example, if there are 14,000 males and 16,000 females, choose 47 males randomly from all males and 53 females randomly from all females. (I got 47 males since 14,000/30,000 = 47%.)
d. Stratify the students by majors and select proportionally (and of course randomly from within each major until you have a sample of 100. For example, if 12% of the students are Psychology majors, choose 12%(100) = 12 Psychology majors from among all Psychology majors for your sample.

20. a.

# phone calls	Frequency	Rel. frequency
1	6	0.1
2	25	0.4167
3	12	0.2
4	**8**	0.1333
6	9	0.15

b. Put # calls in L1 and frequencies in L2. Hit 1-var Stats L1, L2 and get $\bar{x} = 2.97$
c. Since it's a sample and NOT a population, the sample standard deviation is $s = 1.53$
d. $Q3 = 4$, which means that 75% of the data values are 4 or lower.
e. Percent of students who received at least 4 phone calls: 0.1333 + 0.15 = 0.2833 = 28.33%
f. Percent of students who received less than 3 phone calls: 0.1 + 0.4167 = 0.5167 = 51.67%

21. Equation should be $A = P\left(1+\frac{r}{n}\right)^{nt} = 2000\left(1+\frac{0.025}{12}\right)^{12t} = 2000(1.00208\overline{3})^{12t}$
a. $A = 2000\left(1+\frac{0.025}{12}\right)^{12\cdot 3} = \$2155.60$
b. Solve the equation $2000(1.00208\overline{3})^{12t} = 4000$ to get about 28 years (You'll need to use logs)

22. a. 7.71 mg b. After about 16 days (you'll need to use logs) c. 0.917 = 91.7% of the Iodine-131 remains in the body the next day from the previous. This means, each day, the amount of the drug remaining reduces by 8.3%.

23.

	Tests positive	Tests negative	TOTAL
Has disease	25	5	30
Doesn't have disease	2592	7378	9970
TOTAL	2617	7383	10,000

If your friend tested positive, then his chance of having the disease is given by
$\mathrm{P(has\ disease\ |\ tests\ positive)}=\frac{25}{2617} = 0.0096 = 0.96\%$, a very small chance.