Exam 2

Vocabulary

1. Null hypothesis: A conjecture about a population or populations, that certain parameter(s) are equal to each other or to some specified value

2. Type II error: Failure to reject the null hypothesis when it's false

3. Sampling distribution: The probability distribution for some statistic (i.e., for the values it will take on) across repeated samples from a given population

4. Critical region: The set of values for a test statistic that would lead to rejection of the null hypothesis

5. Point estimate: A single number that is the best guess for some population parameter

Conceptual questions

1. A study on sex differences in reading ability measures the reading speeds of a sample of first-graders. The researchers do a two-tailed t-test comparing the averages of the boys and of the girls. Write the null and alternative hypotheses as sentences and as equations.

Null (sentence): Mean reading speed is the same for boys and girls.

Null (equation): $\mu_{boy} = \mu_{girl}$

Alternative (sentence): Mean reading speed differs between boys and girls.

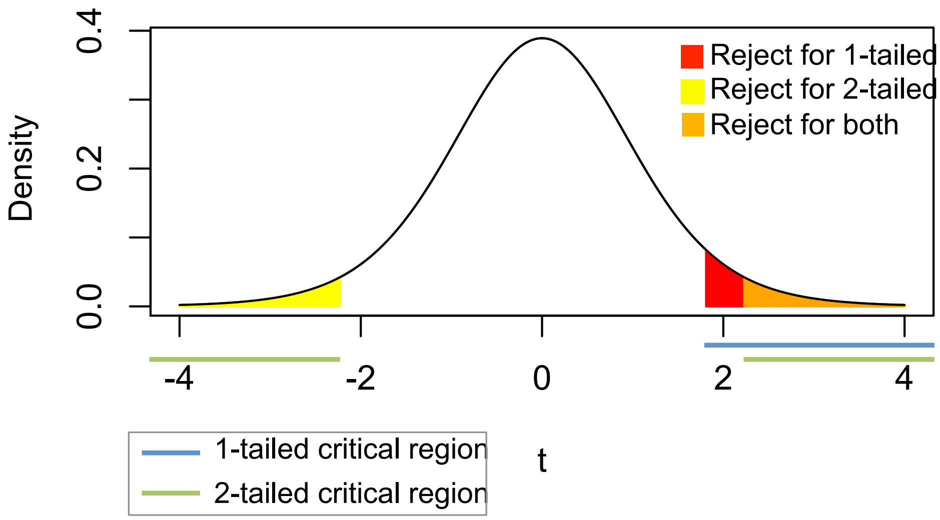
Alternative (equation): $\mu_{boy} \neq \mu_{girl}$

2. A comparison of two samples gives a 99% confidence interval of [1.3, 3.4] for the difference between population means ($\mu_A - \mu_B$). The critical value is $t_{crit} = 3.05$.
Because the confidence interval doesn't include zero, we reject the null hypothesis at $\alpha = 1 - .99 = .01$.

What can you say about the value of t for testing whether $\mu_A = \mu_B$? $t > 3.05$

What can you say about the p-value? $p < .01$

3. Imagine a researcher plans to run a t-test using $\alpha = 5\%$, and the null hypothesis is true (of course she doesn't know that). (a) Draw a picture of the probability distribution for the *t* statistic that will result from this study. (b) Indicate in the picture the critical region for a one-tailed test, and the critical region(s) for a two-tailed test (the two tests will partially overlap). (c) Based on your picture, what's the probability the study will yield a result that would be significant using a one-tailed test but not using a two-tailed test? (d) What's the probability the study will yield a result that would be significant using a two-tailed test but not a one-tailed test?



The yellow and orange regions are each equal to $\alpha/2 = 2.5\%$. The red and orange regions together are equal to 5%, so the red region equals 2.5%. (c) Rejecting the null for 1-tailed but not 2-tailed happens in the red region: 2.5%. (d) Rejecting the null for 2-tailed but not 1-tailed happens in the yellow region: 2.5%.

4. Instead of testing whether two populations have the same mean, there's a test for whether they have the same variance. You haven't seen this kind of hypothesis test, but you should be able to figure out how it works. The null hypothesis is that the two variances are equal. The alternative hypothesis is that they differ. To decide between the hypotheses, you calculate the variances of the two samples and divide the larger one by the smaller one. This ratio is your test statistic, and we call it F. Notice that a bigger value of F means that the sample variances are further apart. Next you find the sampling distribution or likelihood function for F according to the null hypothesis, which tells you the probabilities for what F should be if that hypothesis is correct. You use this to determine the critical value for F, which tells you how big F needs to be before you reject the null hypothesis.

5. A new drug is discovered that is thought to alleviate depression. Eager to be the first to demonstrate its effectiveness, 100 different labs run an experiment comparing people on and off the drug. Each lab gets its own sample of 40 subjects and does a paired t-test. Because it's important to be sure the drug works before marketing it, they all use a Type I error rate of $\alpha = .01$. Of the 100 experiments, 85 conclude the drug has an effect. Assuming the drug really does have an effect, what is the power of this experiment?

85/100 = 85%

Math questions

1. Bullfrogs can jump an average of 84 inches, with a standard deviation of 6 inches. Imagine you measured jumps from 50 frogs and then calculated their mean.

(A) What is the expected value of your result? That is, what should you expect to get, on average?

$E(M) = \mu =$ 84 inches

(B) How far from the expected value is your result likely to be? That is, what's your standard error?

$\sigma_M = \sigma/\sqrt{n} = 6/\sqrt{50} \approx$ .85 inches

2. Blindsight is a condition wherein a person believes s/he is blind but still has implicit knowledge of objects in the visual field. One way to test for blindsight is to flash a spot of light in one of several locations and ask the person to guess where it is. If they do better than they should by answering randomly (the null hypothesis), we conclude they have blindsight (the alternative hypothesis). Imagine we run this test on a patient using four locations (up, down, left, right), and out of 10 trials she's right 7 times. (a) Based on the binomial distribution below, what's the (one-tailed) p-value of our test? (b) What do we conclude? (Assume we had chosen $\alpha = 5\%$.)

Frequency:	0	1	2	3	4	5	6	7	8	9	10
Probability:	.0563	.1877	.2816	.2503	.1460	.0584	.0162	.0031	.0004	.0000	.0000

(a) $p = .0031 + .0004 + .0000 + .0000 = \underline{.0035}$

(b) We conclude the patient has blindsight.

3. In the Stroop task, subjects see color words printed in different colors. Sometimes the word and the color are the same (e.g. GREEN printed in green) and sometimes they're different (e.g., GREEN printed in red). The subject has to say the color of each word as quickly as possible. The data below show the average reaction times for five subjects, for words printed in the same color and for words printed in a different color. Is there a reliable difference between mean reaction time in the two conditions? Do a two-tailed test with $\alpha = .05$, using the critical value $t_{\text{crit}} = 2.78$.

	Time (milliseconds)	
Subject	Same	Different
A	413	442
B	349	380
C	419	440
D	390	414
E	470	495

$$X_{\text{diff}} = \{29, 31, 21, 24, 25\}$$

$$M_{\text{diff}} = 26$$

$$s_{\text{diff}} = \sqrt{\frac{\sum (X_{\text{diff}} - M_{\text{diff}})^2}{n-1}} = \sqrt{\frac{3^2 + 5^2 + 5^2 + 2^2 + 1^2}{4}} = \sqrt{16} \approx 4$$

$$\sigma_{M_{\text{diff}}} = \frac{s_{\text{diff}}}{\sqrt{n}} = \frac{4}{\sqrt{5}} \approx 1.79$$

$$t = \frac{M_{\text{diff}}}{\sigma_{M_{\text{diff}}}} \approx \frac{26}{1.79} \approx 14.53$$

$t > t_{\text{crit}}$, so we conclude there is a reliable difference

4. Find the standardized effect size (d) for Question 3.

$$d = \frac{M_{\text{diff}}}{s_{\text{diff}}} \approx \frac{26}{4} = \underline{6.5}$$

5. A two-tailed independent-samples t-test results in $M_A = 25$, $M_B = 21$, $t = 3.7$, and $t_{crit} = 2.31$. Find the confidence interval for the difference between means, $\mu_A - \mu_B$. (Hint: First figure out the standard error.)

$$t = \frac{M_A - M_B}{\sigma_{M_A - M_B}}$$

$$3.7 = \frac{25 - 21}{\sigma_{M_A - M_B}}$$

$$\sigma_{M_A - M_B} = \frac{25 - 21}{3.7} = 1.08$$

$$CI = (M_A - M_B) \pm t_{crit} \cdot \sigma_{M_A - M_B} = 4 \pm 2.31 \cdot 1.08 \approx \underline{4 \pm 2.49} \text{ or } \underline{[1.51, 6.49]}$$

R questions

Consider the following command and output for Questions 1-4.

```
> t.test(X,mu=7)

data:  X
t = -3.3663, df = 7, p-value = 0.01198
alternative hypothesis: true mean is not equal to 7
95 percent confidence interval:
 2.956695 6.293305
sample estimates:
mean of x
    4.625
```

1. What kind of hypothesis test is being run?
Single-sample t-test

2. Which hypothesis is supported by the results, using $\alpha = .05$?
Alternative, because $p < .05$ and because 7 (μ_0) is not in the confidence interval

3. If α had been .01, which hypothesis would the data support? Why?
Null, because $p > .01$

4. If the null hypothesis had been $\mu = 6$, which hypothesis would the data support (using $\alpha = .05$)? Why?
Null, because 6 is inside the confidence interval

5. Fill in the letters A-E next to the commands on the left, to show what each command computes (one will be left blank)

B	`qt(alpha/2,df,lower.tail=FALSE)`	**A:** Critical value for a one-tailed t-test
A	`qt(alpha,df,lower.tail=FALSE)`	**B:** Critical value for a two-tailed t-test
D	`pt(t,df,lower.tail=FALSE)`	**C:** Critical value for a one-tailed binomial test
E	`2*pt(abs(t),df,lower.tail=FALSE)`	**D:** p-value for a one-tailed t-test
C	`qbinom(alpha,n,q,lower.tail=FALSE)`	**E:** p-value for a two-tailed t-test
–	`pbinom(f,n,q,lower.tail=FALSE)`	