1. Suppose we have two SRSs from two distinct populations and the samples are independent. We measure the same variable for both samples. Suppose both populations of the values of these variables are Normally distributed but the means and standard deviations are unknown. For purposes of comparing the two means, we use
   (a) Two-sample t procedures
   (b) Matched pairs t procedures
   (c) Two-proportion z procedures
   (d) The least-squares regression line
   (e) None of the above. The answer is ______________________.

2. An SRS of size 100 is taken from a population having proportion 0.8 successes. An independent SRS of size 400 is taken from a population having proportion 0.5 successes. The sampling distribution of the difference in sample proportions has what mean?
   (a) 0.3
   (b) 0.15
   (c) The smaller of 0.8 and 0.5
   (d) The mean cannot be determined without the sampling results.
   (e) None of the above. The answer is ______________________.

3. A study was conducted to investigate the effectiveness of a new drug for treating Stage 4 AIDS patients. A group of AIDS patients was randomly divided into two groups. One group received the new drug; the other group received a placebo. The difference in mean subsequent survival (those with drugs – those without drugs) was found to be 1.04 years, and a 95% confidence interval was found to be 1.04 ± 2.37 years. Based upon this information, we can conclude that
   (a) the drug was effective since those taking the drug lived, on average, 1.04 years longer.
   (b) the drug was ineffective since those taking the drug lived, on average, 1.04 years less.
   (c) there is no evidence the drug was effective since the 95% confidence interval covers zero.
   (d) there is evidence the drug was effective since the 95% confidence interval does not cover zero.
   (e) we can make no conclusions since we do not know the sample size or the actual mean survival of each group.

The next two questions refer to this scenario.
Different varieties of fruits and vegetables have different amounts of nutrients. These differences are important when these products are used to make baby food. We wish to compare the carbohydrate content of two varieties of peaches. The data were analyzed with SAS, and the following output was obtained:

VARIETY	N	MEAN	STD DEV	STD ERROR	MIN	MAX
1	5	33.6	3.781	1.691	29.000	38.000
2	7	25.0	10.392	3.927	2.000	33.000

VARIANCES	T	DF	PROB > |T|
UNEQUAL	2.0110	8.0	0.0791
EQUAL	1.7490	10.0	0.1109

4. We wish to test if the two varieties are significantly different in their mean carbohydrate content. The null and alternative hypotheses are
   (a) $H_0: \mu_1 = \mu_2$; $H_a: \mu_1 < \mu_2$
   (b) $H_0: \mu_1 = \mu_2$; $H_a: \mu_1 > \mu_2$
   (c) $H_0: \mu_1 = \mu_2$; $H_a: \mu_1 \neq \mu_2$
   (d) $H_0: \bar{x}_1 = \bar{x}_2$; $H_a: \bar{x}_1 < \bar{x}_2$
   (e) $H_0: \bar{x}_1 = \bar{x}_2$; $H_a: \bar{x}_1 \neq \bar{x}_2$

5. The test statistic and P-value are
   (a) 1.7490; 0.0318
   (b) 1.7490; 0.0554
   (c) 2.0110; 0.1582
   (d) 2.0110; 0.0791
   (e) 2.0110; 0.0396

For this we use STD dev. for s,

$$t = \frac{33.6 - 25 - 0}{\sqrt{\frac{3.781^2}{5} + \frac{10.392}{7}}}$$

6. Two types of tennis balls were tested to determine which one goes faster on a serve. Eight different players served one of each type of ball and the results were recorded.

Server	Raphael	Roger	Serena	Venus	Andy	Justine	LLeyton	Maria
Type A	120	125	119	110	118	82	115	105
Type B	115	122	114	114	115	91	110	106

Assuming that the speeds are approximately normally distributed, how many degrees of freedom will there be in the appropriate t-test used to determine which type of tennis ball travels faster?

(a) 6 (b) 7 (c) 16 (d) 15 (e) 14

7. Sometimes, the order in which a question is asked makes a difference in how it is answered. For example, if you ask people if they prefer chocolate or strawberry ice cream, you might get different answers than if you asked them if they prefer strawberry or chocolate. Seventy-five randomly selected people were asked, "Do you prefer chocolate or strawberry?" and 75 different randomly selected people were asked, "Do you prefer strawberry or chocolate?" The results are given in the following table.

	Response: Prefer Chocolate	Response: Prefer Strawberry
Question: Do you prefer chocolate or strawberry?	42	33
Question: Do you prefer strawberry or chocolate?	34	41

A two-proportion z-test was performed on these data to see if the order of the question made a difference. What is the approximate P-value of the test?

(a) 0.453
(b) 0.093
(c) 0.560
(d) 0.055
(e) 0.191

The next three questions refer to the following scenario.
Sixty-eight people from a random sample of 128 residents of Uppsala, Sweden, had blue eyes. 45 people from a random sample of 110 people from Preston, England, had blue eyes. Let p_1 represent the proportion of people in Uppsala with blue eyes, and let p_2 represent the proportion of people in Preston with blue eyes.

8. If researchers suspected that the distribution of eye color is different in these two countries before collecting the data, which of the following pairs of hypotheses would be appropriate to test?

(a) $H_0: p_1 = 0.53, p_2 = 0.41; H_a: p_1 \neq 0.53, p_2 \neq 0.41$

(b) $H_0: p_1 = p_2 = 0.47; H_a: p_1 \neq p_2 \neq 0.47$

(c) $H_0: p_1 = p_2; H_a: p_1 > p_2$

(d) $H_0: p_1 = p_2; H_a: p_1 \neq p_2$

(e) $H_0: p_1 = p_2; H_a: p_1 < p_2$

9. Which of the following represents the correct conclusion for the significance test described in Question 8?

(a) Reject H_0 at the 5% significance level since the *P*-value is 0.06.

(b) Fail to reject H_0 at the 5% significance level since the *P*-value is 0.06.

(c) Reject H_0 at the 5% significance level since the 95% confidence interval for $p_1 - p_2$ is $(-0.004, 0.248)$.

(d) Reject H_0 at the 1% significance level.

(e) Fail to reject H_0 at the 10% significance level.

10. Which of conditions (a)-(c) is *not* necessary in order to perform the test in Question 8?

(a) There must be at least 1280 people in Uppsala, Sweden and at least 1100 people in Preston, England.
(b) *np* and *n*(1 - *p*) must be large enough for Normal calculations to be reasonably accurate.
(c) Two independent random samples must be taken.
(d) None of conditions (a)-(c) is necessary.
(e) All of conditions (a)-(c) are necessary.

11. Popular wisdom is that eating presweetened cereal tends to increase the number of dental caries (cavities) in children. A sample of children was (with parental consent) entered into a study and followed for several years. Each child was classified as a sweetened-cereal lover or a non-sweetened cereal lover. At the end of the study, the amount of tooth damage was measured. Here are the summary data:

Group	n	Mean	Std. Dev.
Sugar bombed	10	6.41	5.0
No sugar	15	5.20	15.0

An approximate 95% confidence interval for the difference in the mean tooth damage is

(a) $(6.41-5.20)\pm 2.26\sqrt{\frac{5}{10}+\frac{15}{15}}$

(b) $(6.41-5.20)\pm 2.26\sqrt{\frac{25}{10}+\frac{225}{15}}$

(c) $(6.41-5.20)\pm 1.96\sqrt{\frac{25}{10}+\frac{225}{15}}$

(d) $(6.41-5.20)\pm 2.26\sqrt{\frac{25}{100}+\frac{225}{225}}$

Free Response Questions

12. Just before the presidential election in November 2008, a local newspaper conducted a poll of residents of a medium-sized city and found that 120 out of a simple random sample of 250 men intended to vote for Barack Obama and 132 out of an SRS of 240 women intended to vote for Obama.

(a) Is this convincing evidence that there was a gender difference in Obama's support in this city? Support your conclusion with a test of significance, using $\alpha = 0.05$.

2 prop - z test

$\hat{p}_1 = \frac{120}{250}$ (men)

$\hat{p}_2 = \frac{132}{240}$ (women)

$H_0: p_1 - p_2 = 0$ $\alpha = .05$

$H_a: p_1 - p_2 \neq 0$

z-statistic = -1.55

p-value = .121

Conditions

2 SRS from 2 populations - stated.

Assume pop men > 2500

and pop women > 2400

250(.48) = ~~120~~ > 10

250 (.52) = 130 > 10

240(.55) = 132 > 10

240 (.45) = 108 > 10

At the 5% level there is not suff. evidence to reject Ho b/c p-value > α. There may be no difference in opinion by gender.

(b) Construct and interpret a 95% confidence interval for the difference in proportion of women and men who supported Obama in this city.

(-.1583, .0183)

with 95% confidence the difference in support for Obama by gender is that men support Obama at a rate of 16% less to 2% more than women.

13. Jordan's cat —Fern|| is a finicky eater. Jordan is trying to determine which of two brands of canned cat food Fern prefers, Tab-a-Cat or Chow Lion. For two months, she flips a coin each day to decide which of the two foods to feed Fern, and weighs how much Fern eats in grams. Here is the data:

	n	$\bar{x}$	s
Tab-a-Cat	31	85.2	3.45
Chow Lion	30	82.1	4.62

(a) Construct an appropriate significance test at a 0.05 alpha level to determine if there is a significant difference in Fern's weight based on the brand of cat food Fern is eating.

2 sample t-test for difference in means.

$H_0: \mu_1 - \mu_2 = 0$ (or $\mu_1 = \mu_2$)

$H_a: \mu_1 - \mu_2 \neq 0$ (or $\mu_1 \neq \mu_2$)

$\bar{x}_1 \to$ Tab-a cat

$\bar{x}_2 \to$ Chow Lion.

Conditions

2 SRS → stated (flipping coin)

Assume Pop (days) > 310

Since $n \geq 30$ CLT applies

dist. of $(\bar{x}_1 - \bar{x}_2)$ is Approx normal.

$\alpha = .05$

t stat = 2.96

p-value = .004

At The 5% level there is ~~no~~ sufficient evidence to reject Ho, b/c p-value < α. The type of food may matter.

(b) Construct and interpret a 99% confidence interval for the difference in mean amount of food Fern eats when she is offered Tab-a-Cat and when she is offered Chow Lion.

All Conditions the same

$$3.1 \pm t^* \sqrt{\frac{3.45^2}{31} + \frac{4.62^2}{30}}$$

(leave t^*)

(1.0013, 5.1987)

At The 95% confidence level the average grams of Tab-a-cat eaten is between 1.0013g and 5.1987g. more than Chow Lion.