AP Statistics
Inference Worksheet #1

Name [illegible] Pd___

For each problem state: the **type** of problem, the **hypotheses** if appropriate, and the **formula** needed.
For #1, 2, 7, and 11, **fill in the values** for the formula. **Please write all solutions on separate paper.**

1. There seems to be no end to how large the signing bonuses professional athletes can obtain when they start their careers (Peyton Manning - $11.6 million). Suppose a sample of 18 new NFL players report their signing bonuses at the start of the 1998 season, and the results show a mean of $3.81 million and a standard deviation of $1.7 million. Estimate with 90% confidence the mean signing bonus for that season. $\bar{x} = 3.81$ $s_x = 1.7$ $n = 18$

1-sample t-interval

$$\bar{x} \pm t^*_{17} \cdot \frac{s_x}{\sqrt{n}} \qquad 3.81 \pm 1.740 \cdot 1.7/\sqrt{18}$$

2. The student council is thinking about discontinuing the student poetry magazine because only 20% of the students read it. A vote was taken at the school and it was decided to continue the magazine if more than 20% of the students are known to read it. A random sample of 256 students showed that 77 of them had read the last issue. Use a level of significance of 0.05 to determine if the magazine should be continued.

1-proportion Z-test

$H_0: p = 0.20$

$H_A: p > 0.20$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{\frac{77}{256} - 0.20}{\sqrt{\frac{0.20(0.80)}{256}}}$$

3. Consumer Reports gave the following information about annual automobile insurance premiums for similar coverage for randomly selected consumers in California and Florida.

California Premiums (in dollars)	727	672	417	651	500	536	486	537	738	936
Florida Premiums (in dollars)	654	828	494	512	757	712	584	746	770	

Is there a significant difference between mean premiums in California and Florida?

Two separate, independent groups.

2-sample t-test (or interval?)

$H_0: \mu_c = \mu_F$ $H_A: \mu_c \neq \mu_F$

$$t = \frac{(\bar{x}_c - \bar{x}_F) - 0}{\sqrt{\frac{s_c^2}{n_c} + \frac{s_F^2}{n_F}}} \qquad SE(\bar{x}_c - \bar{x}_F)$$

4. The data below give the mercury content of randomly selected fish in the lakes of Maine.

1 Behind Dam	1.05	.23	.77	.91	.25	.13	.41	.36	.24	.25	.4	.45	1.12	.6	.68	.22	.47	.37	.43
2 Natural Lake	.1	.29	.21	.94	1.22	.9	.34	.32	.37	.54	.86	.77	.67	.29	.16	.49			

Is there sufficient evidence to suggest that there is a difference in the average mercury content between lakes formed behind a dam and natural-formed lakes?

2-sample t-test

$H_0: \mu_1 - \mu_2 = 0$

$H_A: \mu_1 - \mu_2 \neq 0$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad SE(\bar{x}_1 - \bar{x}_2)$$

5. *Consumer Reports* gave the following information about annual premiums (in dollars) for 18 randomly selected renewable life insurance policies with similar benefits:

300	345	328	426	660	388	410	563	303
395	278	455	577	470	455	373	365	360

Find a 90% confidence interval for the average of all annual premiums for such life insurance policies.

1-sample t-interval

$$\bar{x} \pm t^*_{17} \cdot s_x/\sqrt{n}$$

6. A random sample of students from a high school were chosen to determine if their sitting pulse rate was lower than their standing pulse rate. Each student's pulse rate was measured in both positions.

Sitting	62	74	82	88	82	66	64	84	72	82	80	72	64	62
Standing	68	78	80	92	58	96	72	100	82	76	92	74	60	58

Can we conclude the sitting pulse rate is lower?

Since pulse rates will vary by student, we will use a paired t-test (matched pairs)

μ_d = difference of sitting − standing

$H_0: \mu_d = 0$

$H_A: \mu_d < 0$

$$t = \frac{\bar{x}_d - 0}{(s_d/\sqrt{n})}$$

7. Fatty acids present in fish oil may be useful for treating some psychiatric disorders. An article reported September 3, 1998 at the Yahoo Health website described a randomized experiment done by a Harvard Medical School researcher in which 14 bipolar patients received fish oil daily, while 16 other patients received a placebo daily. After four months, 9 of the 14 patients who received fish oil had responded favorably, but only 3 or the 16 placebo patients had done so.

$\hat{p}_1 = \frac{9}{14}$ $\hat{p}_2 = \frac{3}{16}$

Calculate and interpret a 90% confidence interval for the difference between the proportions showing favorable response in the two groups.

2-proportion z-interval

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

☆ But actually... this data FAILS the success/failure condition... :(

$$\left(\frac{9}{14} - \frac{3}{16}\right) \pm 1.645 \sqrt{\frac{\frac{9}{14}\left(1-\frac{9}{14}\right)}{14} + \frac{\frac{3}{16}\left(1-\frac{3}{16}\right)}{16}}$$

8. Two types of instruments for measuring the amount of sulfur monoxide in the atmosphere are being compared in an air-pollution experiment. The following readings were recorded for the two instruments on each of the nine randomly selected air samples.

Instrument A	0.86	0.82	0.75	0.61	0.89	0.64	0.81	0.68	0.65
Instrument B	0.87	0.74	0.63	0.55	0.76	0.70	0.69	0.57	0.53

Are the instruments measuring the same amount of sulfur monoxide?

μ_d = difference of A − B

If each pair is measuring the same air sample, then paired t-test (or interval).

$H_0: \mu_d = 0$

$H_A: \mu_d \neq 0$

$$t = \frac{\bar{x}_d - 0}{(s_d/\sqrt{n})}$$

(if we are measuring 18 different samples, then 2-sample t-test)

9. A major car manufacturer wants to test a new engine to determine whether it meets new air-pollution standards. The mean emission of all engines of this type must be less than 20 parts per million of carbon. Ten engines are manufactured for testing purposes, and the emission level of each is determined. The data (in parts per million) are listed:

!!

15.6 16.2 22.5 20.5 16.4 19.4 16.6 17.9 12.7 13.9

Do the data supply sufficient evidence to allow the manufacturer to conclude that this type of engine meets the pollution standard? $\alpha = 0.01$

1-sample t-test

these are somewhat tricky...

$H_0: \mu = 20$

$H_A: \mu < 20$

$$t = \frac{\bar{x} - \mu}{(s_x/\sqrt{n})}$$

10. Suppose you wish to compare a new method of teaching reading to “slow learners” to the current standard method. You decide to base this comparison on the results of a reading test given at the end of a learning period of 6 months. Of a random sample of 20 slow learners, 8 are taught by the new method and 12 are taught by the standard method. All 20 children are taught by qualified instructors under similar conditions for a 6-month period. The results of the reading test at the end of this period are given below.

Reading Scores for Slow Learners

New Method	80	80	79	81	76	66	79	76				
Standard	79	62	70	68	73	76	86	73	72	68	75	66

a. Use a 90% confidence interval, and interpret the interval for the true mean difference between the test scores for the new method and the standard method.
b. Test to see if the new method’s test scores are higher. Let $\alpha = 0.10$

a) 2-sample t-interval $\quad (\bar{x}_N - \bar{x}_S) \pm t^* \sqrt{\frac{s_N^2}{n_N} + \frac{s_S^2}{n_S}}$

b) 2-sample t-test

$H_0: \mu_N - \mu_S = 0$

$H_A: \mu_N - \mu_S > 0$

$$t = \frac{(\bar{x}_N - \bar{x}_S) - 0}{\sqrt{\frac{s_N^2}{n_N} + \frac{s_S^2}{n_S}}}$$

11. A dietician has developed a diet that is low in fats, carbohydrates, and cholesterol. Although the diet was initially intended to be used by people with heart disease, the dietician wishes to examine the effect this diet has on the weights of obese people. Two random samples of 100 obese people each are selected, and one group of 100 is placed on the low-fat diet. The other 100 are placed on a diet that contains approximately the same quantity of food but is not as low in fats, carbohydrates and cholesterol. Each person, the amount of weight lost (or gained) in a 3-week period is recorded. The results are below.

	Low-Fat Diet (1)	Other Diet (2)
Mean weight loss	9.3 lbs	7.4 lbs
Sample variance	22.4	16.3

← these are already s^2

a. Form and interpret a 95% confidence interval for the difference between the population mean weight losses for the two diets.
b. Test to see if there is a difference in weight loss for the two diets.

a) 2-sample t-interval

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(9.3 - 7.4) \pm t^* \sqrt{\frac{22.4}{100} + \frac{16.3}{100}}$$

---

b) 2-sample t-test

$H_0: \mu_1 - \mu_2 = 0$

$H_A: \mu_1 - \mu_2 \neq 0$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{9.3 - 7.4}{\sqrt{\frac{22.4}{100} + \frac{16.3}{100}}}$$

Booyah.