Quiz 9.3A AP Statistics Name:

1. The Environmental Protection Agency has determined that safe drinking water should contain no more than 1.3 mg/liter of copper. You are testing water from a new source, and take 30 water samples. The mean copper content in your samples is 1.36 mg/l and the standard deviation is 0.18 mg/l. There do not appear to be any outliers in your data.

(a) Do these samples provide convincing evidence at the $\alpha = 0.05$ level that the water from this source contains unsafe levels of copper? Justify your answer.

(b) How would your conclusion change if your sample mean had been 1.355 mg/l? What point does this make about statistical significance?

2. A consumer advocacy group tests the mean vitamin C content of 50 different brands of bottled juices using, in each case, a *t*-test of significance in which the null hypothesis is the mean amount of vitamin C that is on the nutrition facts label for that brand of juice. They find that two of the 50 juice brands have statistically significantly lower Vitamin C than claimed at the $\alpha = 0.05$ level. Is this an important discovery? Explain.

3. Tai Chi is often recommended as a way to improve balance and flexibility in the elderly. Below are before-and-after flexibility ratings (on a 1 to 10 scale, 10 being most flexible) for 8 men in their 80's who took Tai Chi lessons for six months.

Subject	A	B	C	D	E	F	G	H
Flexibility rating after Tai Chi	2	4	3	3	3	4	5	10
Flexibility rating before Tai Chi	1	2	1	2	1	4	2	6

Do these paired data adequately meet the Normality condition for a *t*-procedure? Justify your answer.

4. A pharmaceutical company is testing a new drug for reducing cholesterol levels. To approve the drug for the next round of testing, they need to show that this drug reduces mean total cholesterol level by at least 50 mg/dL. They initially plan a study that involves 50 subjects and a significance level of $\alpha = 0.05$, but they discover that the power of the test against this effect size is only 0.24. What are two ways they can increase the power of the test without changing effect size?

Quiz 9.3B AP Statistics Name:

1. Sweet corn of a certain variety is known to produce individual ears of corn with a mean weight of 8 ounces. A farmer is testing a new fertilizer designed to produce larger ears of corn, as measured by their weight. He finds that 32 randomly-selected ears of corn grown with this fertilizer have a mean weight of 8.23 ounces and a standard deviation of 0.8 ounces. There are no outliers in the data.

(a) Do these samples provide convincing evidence at the $\alpha = 0.05$ level that the fertilizer had a positive impact on the weight of the corn ears? Justify your answer.

(b) How would your conclusion change if your sample mean had been 8.24 ounces? What point does this make about statistical significance?

2. Low-density lipoproteins (LDL) are a form of cholesterol in blood that has been linked to atherosclerosis, a condition of the arteries that often leads to heart attacks. Levels of LDL below 100 mg per deciliter (mg/dL) of blood are considered healthy; levels above 160 mg/dL are considered high. A pharmaceutical company tested a new drug aimed at reducing blood levels of LDL by conducting a large-scale completely-randomized experiment on 1000 individuals with LDL levels above 160 mg/dL. Subjects were randomly assigned to one of two groups: one group received the new drug and the other received the drug that is the current standard. The researchers found a statistically significant reduction of 0.6 mg/dL in the LDL levels of those in the new drug group *versus* those in the current drug group. Does this mean the new drug is an important advance in therapy? Explain.

3. The developer of a new filter for filter-tipped cigarettes claims that it leaves less nicotine in the smoke than does the current filter. Because cigarette brands differ in a number of ways, he tests each filter on one cigarette of each of nine brands and records the difference in nicotine content. His results are given in the table below.

Brand	A	B	C	D	E	F	G	H	J
Old Filter nicotine, mg	0.7	0.8	0.8	0.9	0.9	1.0	1.2	1.2	1.8
New Filter nicotine, mg	0.6	0.6	0.7	0.8	0.7	1.0	0.8	0.9	1.5

Do the differences in the paired data adequately meet the Normality condition for a *t*-procedure? Justify your answer.

4. A psychologist studying the effect of computer games on psychological well-being plans to measure changes in individuals' score on a test that measures "subjective happiness" before and after a 3-hour session of computer gaming. He has 30 volunteers, all of whom are experienced "gamers." Before starting the experiment, he determines that a two-sided test of significance using 30 subjects and a significance level of $\alpha = 0.01$ against an effect size of 1 point (on a 10-point happiness scale) only has a power of 0.29. Given that he only has these 30 subjects to work with, what two options does he have to increase the power of the test?

1. Does too much sleep impair intellectual performance? Researchers examined this commonly held belief by comparing the performance of subjects on the mornings following (a) two normal night's sleep and (b) two nights of "extended sleep." The order of these two treatments was determined randomly. In the morning they were given a number of tests of ability to think quickly and clearly. One test was for vigilance where the lower the score, the more vigilant the subject, (vigilance = alertness). The following data were collected:

	Vigilance Score									
Subject	A	B	C	D	E	F	G	H	I	J
Normal Sleep	8	9	14	4	12	11	3	26	3	8
Extended Sleep	8	9	17	2	21	16	9	38	10	0

Carry out an appropriate test to help answer the researchers' question.

2. The label on bottles of grapefruit juice say that they contain 180 ml of liquid. Your friend Jerry suspects that the true mean is less than that, so he takes an SRS of 40 bottles and performs a *t*-test on the mean volume of liquid in the bottles at a significance level of $\alpha = 0.05$.

(a) Define the parameter of interest for Jerry's test, and state the null and alternative hypotheses.

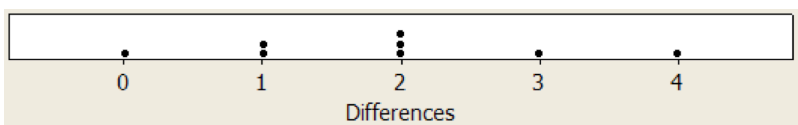
(b) Jerry wants to use a *t*-table to find the *P*-value of the test, but he discovers that there is no line for 39 degrees of freedom. He uses the line for 40 degrees of freedom instead. Explain to him what he's done wrong.

(c) Using software, Jerry determines that the power of this test against an effect size of 10 ml is only 0.42. What can Jerry do to increase power at this effect size?

(d) Eventually, Jerry finds that his *P*-value is between 0.05 and 0.10. So his test is not statistically significant at the 0.05 significance level. He concludes, "We can accept H_0: there is evidence that the volume of liquid in the bottle is indeed 180 ml." Jerry's not having a good day: what did he do wrong this time?

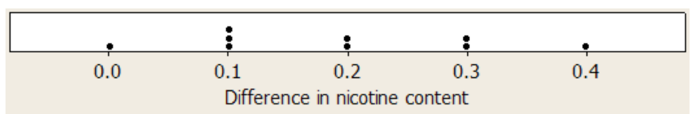
Quiz 9.3A

1. (a) State: We wish to test $H_0 : \mu = 1.3$ *versus* $H_a : \mu > 1.3$, where μ = mean copper level in all possible water samples from the new source, in mg/liter. We are using a significance level of $\alpha = 0.05$. Plan: The procedure is a one-sample *t*-test for a mean. Conditions: *Random:* We will have to assume that the 30 water samples can be viewed as an SRS of water from the source. (Since the population of all water samples is essentially infinite, we are not concerned about the 10% condition). *Normal/Large sample:* *n*=30 is large enough as long as there are no outliers in the sample. We are told there are no outliers, so it seems safe to proceed. Do: $t = \frac{1.36 - 1.3}{\frac{0.18}{\sqrt{30}}} = 1.8257$; df = 29; *P*-value = 0.0391. Conclude: A *P*-value of 0.0391 is less than $\alpha = 0.05$, so we reject H_0 and conclude that there is convincing evidence that the new water source contains unsafe levels of copper. (b) If $\bar{x} = 1.355$, then $t = \frac{1.355 - 1.3}{\frac{0.18}{\sqrt{30}}} = 1.6736$; df = 29; *P*-value = 0.0525, and we would fail to reject H_0. This points out that it may not be wise to attach to much importance to statistical significance, since a small change in mean copper level can change our statistical conclusion. **2.** Statistically significant means that the mean amount of vitamin C in a sample was far enough below the amount given on the label so that it was unlikely to be the result of sampling variability. This is not noteworthy, because at a $\alpha = 0.05$ level, we can expect to make a Type I error 5% of the time, so in 50 *t*-tests we can expect, on average, 0.05(50) = 2.5 significant results when all the null hypotheses are really true. This is approximately what we found. **3.** For paired data, we want the differences to support the assumption that the underlying population of differences is approximately Normal. A dot plot of the differences (below) shows no outliers or evidence of strong skew, so these data meet the Normality condition. **4.** Increase the sample size or increase the significance level.



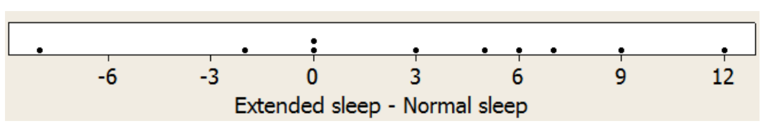
Quiz 9.3B

1. (a) State: We wish to test $H_0: \mu = 8$ *versus* $H_a: \mu > 8$, where μ = mean weight (in ounces) of all corn ears of this variety grown with the new fertilizer. We are using a significance level of α = 0.05. Plan: The procedure is a one-sample *t*-test for a mean. Conditions: *Random:* The farmer randomly selected 32 ears of corn. *10%:* It seems reasonable to assume that the farmer's crop consists of more than 280 ears of corn. *Normal/Large sample: n*=32 is large enough as long as there are no outliers or strong skew in the sample. We are told there are no outliers, so it seems safe to proceed. Do: $t = \frac{8.23 - 8}{\frac{0.8}{\sqrt{32}}} = 1.6263$; df = 31; *P*-value = 0.057. Conclude: A *P*-value of 0.057 is greater than α = 0.05, so we fail to reject H_0: we do not have convincing evidence that the corn of this variety grown with the new fertilizer weighs more than 8 oz., on average. (b) If $\bar{x} = 8.24$, then $t = \frac{8.24 - 8}{\frac{0.8}{\sqrt{32}}} = 1.697$; df = 31; *P*-value = 0.04985, and we would reject H_0. This points out that it may not be wise to attach to much importance to statistical significance, since a small change in mean corn weight can change our statistical conclusion. **2.** Statistically significant means that the 0.6 mg/dL difference in means between the two groups was large enough so that it was unlikely to arise as a result of random assignment of subjects to the two groups. A reduction of 0.6 mg/dL may be statistically significant but not practically significant, given that a 60 mg/dL change is required to bring a high LDL level down to a healthy level. **3.** For paired data, we want the differences to support the assumption that the underlying population of differences is approximately Normal. A dot plot of the differences (right) shows no outliers or evidence of strong skew, so these data meet the Normality condition. **4.** Increase the effect size or increase the significance level.



Quiz 9.3C

1. State: We wish to test $H_0: \mu_D = 0$ *versus* $H_a: \mu_D > 0$, where μ_D = the mean difference in vigilance score (extended sleep – normal sleep) for all possible subject of the experiment. Since lower values indicate more vigilance, we are looking for positive values of this difference. We will use a significance level of α = 0.05. Plan: The procedure is a one-sample *t*-test on paired data. Conditions: *Random:* The treatments (extended and normal sleep) were assigned in random order. *10%:* We aren't sampling, so there's no need to check this condition. *NormalLarge Sample:* The dotplot (right) does not indicate strong skew, and the extreme values are not outliers by the 1.5 x IQR standard. Do: $\bar{x} = 3.20$; $s = 5.87$;



$t = \frac{3.2 - 0}{\frac{5.87}{\sqrt{10}}} = 1.635$; df = 9; *P*-value = 0.0593. Conclude: A *P*-value of 0.0593 is greater than α = 0.05, so we fail to reject H_0: we do not have convincing evidence that extended sleep reduces vigilance. **2.** (a) μ = the mean volume of juice in all bottles in the population. $H_0: \mu = 180; H_a: \mu < 180$, (b) Since the appropriate *t*-distribution has only 39 degrees of freedom, Jerry would be pretending his sample is larger than it is, and the *P*-value from his test would be smaller than it actually is. We'd be more likely to reject H_0, thus increasing the likelihood of a Type I error. (c) Jerry can either increase his sample size to more than 40 or increase his significance level to increase power. (d) A test of significance can only provide evidence against the null hypothesis, not in support of it. We can't provide evidence that the mean isn't 180 ml, but that doesn't mean it is 180 ml. It could be 180.05 ml!