AP Stats Ch 24,25 - Extra Practice **Name:** Solutions **Per:** ___

1. Learning math. The Core Plus Mathematics Project (CPMP) is an innovative approach to teaching Mathematics that engages students in group investigations and mathematical modeling. After field tests in 36 high schools over a three-year period, researchers compared the performances of CPMP students with those taught using a traditional curriculum. In one test, students had to solve applied Algebra problems using calculators. Scores for 320 CPMP students were compared with those of a control group of 273 students in a traditional Math program. Computer software was used to create a confidence interval for the difference in mean scores. (*Journal for Research in Mathematics Education*, 31, no. 3 [2000])

Conf level: 95% Variable: Mu[CPMP] – Mu[Ctrl]
Interval: (5.573, 11.427)

a) What's the margin of error for this confidence interval?
b) If we had created a 98% CI, would the margin of error be larger or smaller?
c) Explain what the calculated interval means in this context.
d) Does this result suggest that students who learn Mathematics with CPMP will have significantly higher mean scores in Algebra than those in traditional programs? Explain.

(a) (5.573 ——— 11.427)

statistic = $\frac{11.427+5.573}{2} = 8.5$

statistic: 8.5 ~ margin of error

margin of error = $\frac{11.427-5.573}{2} =$ 2.927 pts

(b) higher confidence ⟷ lower precision
The margin of error would increase in size

(c) We are 95% confident that the mean for all CPMP students will be between 5.573 and 11.427 points higher than the mean for all traditional students.

(d) Yes. 0 is not within the confidence interval (all the likely values show CPMP having at least 5.573 pts higher mean).

3. CPMP, again. During the study described in Exercise 1, students in both CPMP and traditional classes took another Algebra test that did not allow them to use calculators. The table below shows the results. Are the mean scores of the two groups significantly different?

Performance on Algebraic Symbolic Manipulation Without Use of Calculators

Math Program	*n*	Mean	SD
CPMP	312	29.0	18.8
Traditional	265	38.4	16.2

a) Write an appropriate hypothesis.

b) Do you think the assumptions for inference are satisfied? Explain.

c) Here is computer output for this hypothesis test. Explain what the P-value means in this context.

2-Sample t-Test of $\mu 1 - \mu 2 \neq 0$
t-Statistic = −6.451 w/574.8761 df
P < 0.0001

d) State a conclusion about the CPMP program.

a) $\mu_{trad} = \mu_{CPMP}$ The means of both programs are the same.

$\mu_{trad} \neq \mu_{CPMP}$ The means of both programs are different.

b) ✓ — SRS (assume representative)
✓ — 320 & 273 < 10% of their populations
✓ — groups indep (different schools and students)
✓ — Nearly normal, both sample sizes are ≥ 40

c) If there were actually no difference between the programs, we would see a difference in mean scores as large as this (38.4 − 29 = 9.4 pts) or larger, with a probability < .0001, just due to chance

d) With $\alpha = .05$, $p < .0001$ is low so we reject H_0.
We do have sufficient statistical evidence to conclude that the mean score for CPMP is different from traditional.

4. CPMP and word problems. The study of the new CPMP Mathematics methodology described in Exercise 1 also tested students' abilities to solve word problems. This table shows how the CPMP and traditional groups performed. What do you conclude?

Math Program	*n*	Mean	SD
CPMP	320	57.4	32.1
Traditional	273	53.9	28.5

(Conditions already checked above)

H_0: $\mu_{trad} = \mu_{CPMP}$ The mean score for word problems is the same.

H_A: $\mu_{trad} < \mu_{CPMP}$ The mean score for word problems is lower for traditional

(I'm deciding to try a one-sided test here)

Perform a 2-Sample T-Test in T-84
using: $\bar{x}_1 = 57.4$ $\bar{x}_2 = 53.9$ $\mu_1 \neq \mu_2$
$s_{x1} = 32.1$ $s_{x2} = 28.5$ non-pooled
$n_1 = 320$ $n_2 = 273$

$t = 1.406$
$p\text{-value} = .1602$
$df = 590.0$

With $\alpha = .05$, p-value = .1602 is high so we fail to reject H_0.
We do not have sufficient statistical evidence to conclude the mean score for word problems is lower for traditional.

129. Strikes. Advertisements for an instructional video claim that the techniques will improve the ability of Little League pitchers to throw strikes, and that, after undergoing the training, players will be able to throw strikes on at least 60% of their pitches. To test this claim, we have 20 Little Leaguers throw 50 pitches each, and we record the number of strikes. After the players participate in the training program, we repeat the test. The table shows the number of strikes each player threw before and after the training.

a) Is there evidence that after training players can throw strikes more than 60% of the time?

b) Is there evidence that the training is effective in improving a player's ability to throw strikes?

Number of Strikes (out of 50)		Number of Strikes (out of 50)	
Before	After	Before	After
28	35	33	33
29	36	33	35
30	32	34	32
32	28	34	30
32	30	34	33
32	31	35	34
32	32	36	37
32	34	36	33
32	35	37	35
33	36	37	32

a) using after columns only

H_0: $\mu = .60(50) = 30$ Avg # strikes is 30.

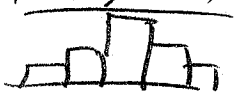
H_A: $\mu > 30$ Avg # strikes is greater than 30

conditions

SRS ✓ no but representative

$n < 10\%$ pop ✓ $20 < 10\%$ of all pitchers

Nearly Normal ✓



Manually

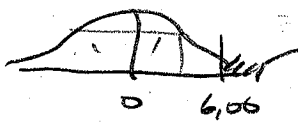
1 var stats: $\bar{x} = 33.15$, $s = 2.323$, $n = 20$, $df = 19$

$\mu_{\bar{x}} = \mu = 33.15$

$SE_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{2.323}{\sqrt{20}} = .5194$

test statistic: $t = \frac{\bar{x} - x_0}{SE_{\bar{x}}} = \frac{33.15 - 30}{.5194}$

$t = 6.06$



p-value = tcdf(6.06, 999, 19)

$= 3.95 \cdot 10^{-6}$

by calculator (checking)

TTest using after data

$\mu_0 : 30$

$\mu > 30$

$t = 6.06$

p-value = $3.923 \cdot 10^{-6}$ ✓

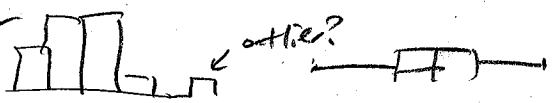
with $\alpha = .05$, p-val = $4 \cdot 10^{-6}$ is low so reject H_0

We do have sufficient evidence to conclude the average # strikes after training is >30 (more than 60% are strikes).

---

b) matched pairs, so find mean of differences ($CL_3 = \hat{L}_1 - \hat{L}_2$)

conditions matched ✓ --- NN: ✓



manually:

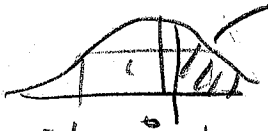
1 var stats on diffs: $\bar{x}_D = 0.1$, $s = 3.32$, $n = 20$

H_0: $\mu_D = 0$ (no improvement)

H_A $\mu_D > 0$ (training improves # strikes avg)

$SE_{\bar{x}} = \frac{3.32}{\sqrt{20}} = .74232$

$t = \frac{0.1 - 0}{.74232} = .1347$



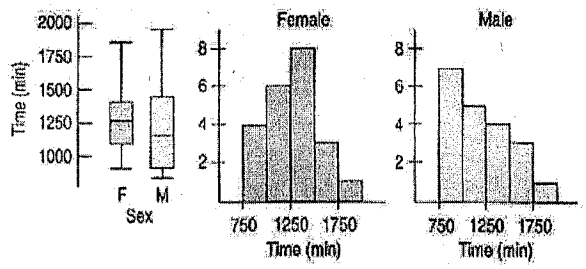
p-value = tcdf(.1347, 999, 19)

$= .447$

With $\alpha = .05$, p-value = .447 is high so we fail to reject H_0

We do not have sufficient evidence to conclude the training improves the mean number of strikes.

25. Crossing Ontario. Between 1954 and 2003, swimmers have crossed Lake Ontario 43 times. Both women and men have made the crossing. Here are some plots (we've omitted a crossing by Vikki Kieth, who swam a round trip—North to South to North—in 3390 minutes):



Summary of Time (min)			
Group	Count	Mean	StdDev
F	22	1271.59	261.111
M	20	1196.75	304.369

How much difference is there between the mean amount of time (in minutes) it would take female and male swimmers to swim the lake?

a) Construct and interpret a 95% confidence interval for the difference between female and male crossing times.

b) Comment on the assumptions and conditions.

a) $\bar{x}_M = 1196.75$ min $\quad s_M = 304.369 \quad n_M = 20$

$\bar{x}_F = 1271.59$ min $\quad s_F = 261.111 \quad n_F = 22$

$\bar{x}_{F-M} = 1271.59 - 1196.75 = 74.84$ min

$SE_{F-M} = \sqrt{\frac{261.111^2}{22} + \frac{304.369^2}{20}} = 68.286$ min

CONDITIONS

groups indep ✓ — yes

SRS ✓ — representative

Nearly Normal ✓ — $n \geq 15$ & $n \geq 65$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

$$df = \frac{\left(\frac{261.111^2}{22} + \frac{304.369^2}{20}\right)^2}{\frac{1}{21}\left(\frac{261.111^2}{22}\right)^2 + \frac{1}{19}\left(\frac{304.369^2}{20}\right)^2} = \frac{5976940.912}{457336.664 + 1129244.743} = 46.455$$ use 46 (calc says 37.67)

t^* for 95% with $df = 46$ = 2.014

CI: $74.84 \pm 2.014(68.286)$

$= (-62.688, 212.36)$ calculator gets: $(-83.21, 272.89)$

We are 95% confident the difference in mean times (women − men) is −83.2 to +272.9 min. There is no evidence of a gender difference.

b) We are assuming some things here:

1) These are individual swims, and all the swimmers are different (one swimmer doesn't appear in data more than once)

2) Neil's data is skewed. We are assuming unimodal and high n is enough that distribution of difference is approx. Normal.

7. Friday the 13th, I. In 1993 the *British Medical Journal* published an article titled, "Is Friday the 13th Bad for Your Health?" Researchers in Britain examined how Friday the 13th affects human behavior. One question was whether people tend to stay at home more on Friday the 13th. The data below are the number of cars passing Junctions 9 and 10 on the M25 motorway for consecutive Fridays (the 6th and 13th) for five different time periods.

Year	Month	6th	13th
1990	July	134012	132908
1991	September	133732	131843
1991	December	121139	118723
1992	March	124631	120249
1992	November	117584	117263

Here are summaries of two possible analyses:

Paired t-Test of mu1 = mu2 vs. mu1 > mu2
Mean of Paired Differences: 2022.4
t-Statistic = 2.9377 w/4 df
P = 0.0212

2-Sample t-Test of mu1 = mu2 vs. mu1 > mu2
Difference Between Means: 2022.4
t-Statistic = 0.4273 w/7.998 df
P = 0.3402

a) Which of the tests is appropriate for these data? Explain.
b) Using the test you selected, state your conclusion.
c) Are the assumptions and conditions for inference met?

(a) paired-t-test because the 6th & 13th in a particular month are likely to be connected (matched by year/month)

$H_0: \mu_{13} = \mu_6$ no diff. in mean # people staying home
$H_A: \mu_{13} > \mu_6$ more people (on avg) stay home on 13th

(b) With $\alpha = .05$, p-value = .0212 is low, so we reject H_0. We do have sufficient statistical evidence to conclude that more people (on average) stay home on Friday the 13th than on Friday the 6th.

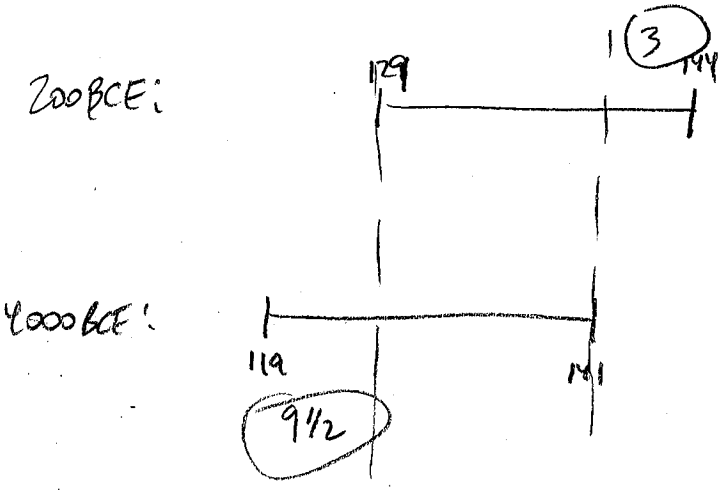
(c) ~SRS (assume these dates are representative)
✓ n < 10% pop (true even for these large n values)
✓ matched pairs (matched by year/month)
✓ Nearly normal, yes b/c n ≥ 40 (for each sample)

8. Egyptians. Some archaeologists theorize that ancient Egyptians interbred with several different immigrant populations over thousands of years. To see if there is any indication of changes in body structure that might have resulted, they measured 30 skulls of male Egyptians dated from 4000 B.C.E. and 30 others dated from 200 B.C.E. (A. Thomson and R. Randall-Maciver, *Ancient Races of the Thebaid*, Oxford: Oxford University Press, 1905.)

Perform Tukey's test for the difference in mean skull breadth between the 200 and 4000 B.C.E. data, and use your result to write a conclusion paragraph (assume conditions for inference are met).

Maximum Skull Breadth (mm)

4000 B.C.E.	4000 B.C.E.	200 B.C.E.	200 B.C.E
131	131	½ 141	131
125	135	½ 141	129 min
131	132	135	136
119 min	139	133	131
136	132	131	139
138	126	140	144 max
139	135	139	½ 141
125	134	140	130
131	128	138	133
134	130	132	138
½ 129	138	134	131
134	128	135	136
126	127	133	132
132	131	136	135
141 max	124	134	½ 141



9.5 + 3 = 12.5

Significant at $\alpha = .05$? ≥7 ✓

Significant at $\alpha = .01$? ≥10 ✓

significant at $\alpha = .001$ ≥13 ✗

p value is between .001 and .01

With $\alpha = .05$ (or .01) we can reject H_0.

We do have SSE to conclude there is a difference in mean skull breadth between 200 BCE and 4000 BCE (200 BCE appears generally wider).