AP® Statistics
2004 Sample Student Responses
Form B

The College Board is a not-for-profit membership association whose mission is to connect students to college success and opportunity. Founded in 1900, the association is composed of more than 4,500 schools, colleges, universities, and other educational organizations. Each year, the College Board serves over three million students and their parents, 23,000 high schools, and 3,500 colleges through major programs and services in college admissions, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT®, the PSAT/NMSQT®, and the Advanced Placement Program® (AP®). The College Board is committed to the principles of excellence and equity, and that commitment is embodied in all of its programs, services, activities, and concerns.

For further information, visit www.collegeboard.com

For the College Board's online home for AP professionals, visit AP Central at apcentral.collegeboard.com.

A1

4. The principal at Crest Middle School, which enrolls only sixth-grade students and seventh-grade students, is interested in determining how much time students at that school spend on homework each night. The table below shows the mean and standard deviation of the amount of time spent on homework each night (in minutes) for a random sample of 20 sixth-grade students and a separate random sample of 20 seventh-grade students at this school.

	Mean	Standard Deviation
Sixth-grade students	27.3	10.8
Seventh-grade students	47.0	12.4

Based on dotplots of these data, it is not unreasonable to assume that the distribution of times for each grade were approximately normally distributed.

(a) Estimate the difference in mean times spent on homework for all sixth- and seventh-grade students in this school using an interval. Be sure to interpret your interval.

Interest Statistic : $\bar{X}_1 - \bar{X}_2$

A Two - Sample T interval with a Confidence level of 95% will be constructed Since the σ of the population is unknown.

Assumptions

SRSs (Simple Random Samples) — Given

Moderate Sample Sizes & appear normal — Given

Independent Samples — Different kids in different grades

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2}, \quad df = 37.3$$

GO ON TO THE NEXT PAGE.

A2

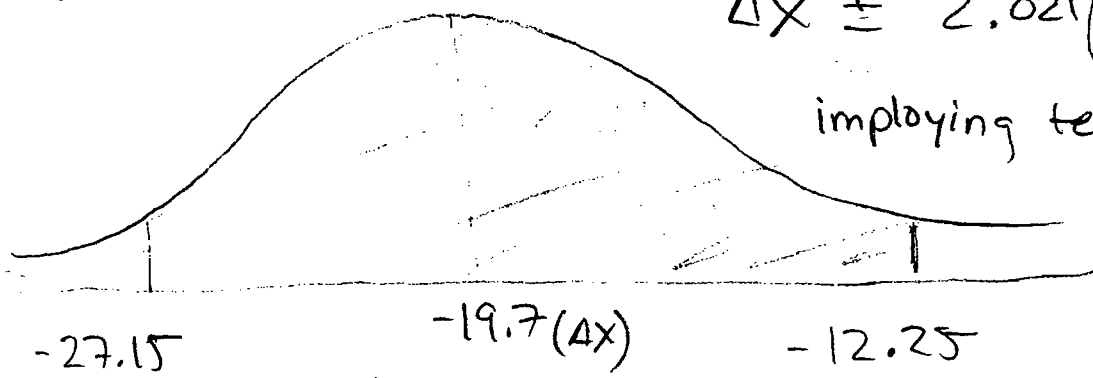
If you need more room for your work for part (a), use the space below.

$t_c = \pm 2.021$

$\Delta X = \overline{X}_1 - \overline{X}_2$ $\quad \Delta X = -19.7$

$\Delta X \pm 2.021\left(\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right)$

imploying technology: (-27.15, -12.2



We are 95% confident that the true difference in time spent on hw is between (-27.15, -12.25) for 6th graders and 7th Graders.

(b) An assistant principal reasoned that a much narrower confidence interval could be obtained if the students were paired based on their responses; for example, pairing the sixth-grade student and the seventh-grade student with the highest number of minutes spent on homework, the sixth-grade student and seventh-grade student with the next highest number of minutes spent on homework, and so on. Is the assistant principal correct in thinking that matching students in this way and then computing a matched-pairs confidence interval for the mean difference in time spent on homework is a better procedure than the one used in part (a) ? Explain why or why not.

No this is not a better method because the samples are independent. They are different kids in different Grades. One way to do matched pairs is to sample 7th Graders for the difference in hw they had ~~between~~ 6th & 7th Grade.

GO ON TO THE NEXT PAGE.

B1

4. The principal at Crest Middle School, which enrolls only sixth-grade students and seventh-grade students, is interested in determining how much time students at that school spend on homework each night. The table below shows the mean and standard deviation of the amount of time spent on homework each night (in minutes) for a random sample of 20 sixth-grade students and a separate random sample of 20 seventh-grade students at this school.

	Mean	Standard Deviation
Sixth-grade students	27.3	10.8
Seventh-grade students	47.0	12.4

Based on dotplots of these data, it is not unreasonable to assume that the distribution of times for each grade were approximately normally distributed.

(a) Estimate the difference in mean times spent on homework for all sixth- and seventh-grade students in this school using an interval. Be sure to interpret your interval.

Since the population is approximately normal, the population standard deviations are unknown, and the samples are independent and random, a two-sample t interval is appropriate.

The standard deviation of the difference in mean times is $\sqrt{\frac{10.8^2}{20}+\frac{12.4^2}{20}} \approx 3.677$. Degrees of Freedom = 37.3 (TI-89).

A 95% confidence interval for the difference in mean times would be $(6^{th} - 7^{th})$.

$(27.3-47) \pm t^* \sqrt{\frac{10.8^2}{20}+\frac{12.4^2}{20}}$

$\approx -19.7 \pm 7.45 \rightarrow \{-27.1, -12.3\}$

The 95% confidence interval for the mean difference is -27.1 to -12.3 minutes.

In other words, I am 95% confident that the population mean difference is between -27.1 and -12.3 minutes. In about 95 of 100 intervals calculated from similar samples, the population mean difference would be in the interval.

GO ON TO THE NEXT PAGE.

B2

If you need more room for your work for part (a), use the space below.

(b) An assistant principal reasoned that a much narrower confidence interval could be obtained if the students were paired based on their responses; for example, pairing the sixth-grade student and the seventh-grade student with the highest number of minutes spent on homework, the sixth-grade student and seventh-grade student with the next highest number of minutes spent on homework, and so on. Is the assistant principal correct in thinking that matching students in this way and then computing a matched-pairs confidence interval for the mean difference in time spent on homework is a better procedure than the one used in part (a) ? Explain why or why not.

The procedure in (a) was better because the samples were independent of each other. There is no reason for higher times to be matched up because this would be assuming a relationship exists. When most likely no relationship exists. Thus, the principal is wrong in his thinking.

GO ON TO THE NEXT PAGE.

C1

4. The principal at Crest Middle School, which enrolls only sixth-grade students and seventh-grade students, is interested in determining how much time students at that school spend on homework each night. The table below shows the mean and standard deviation of the amount of time spent on homework each night (in minutes) for a random sample of 20 sixth-grade students and a separate random sample of 20 seventh-grade students at this school.

	Mean	Standard Deviation
Sixth-grade students	27.3	10.8
Seventh-grade students	47.0	12.4

Based on dotplots of these data, it is not unreasonable to assume that the distribution of times for each grade were approximately normally distributed.

(a) Estimate the difference in mean times spent on homework for all sixth- and seventh-grade students in this school using an interval. Be sure to interpret your interval.

Since population st. deviations ~~are~~ ~~is~~ is not know and the size of each sample is less than 30, we should ~~use~~ construct the confidence interval for T-distribution, assuming that samples are independent.

As st. dev. for sixth-grade st. ≈ st. dev. for seventh-grade st. $10.8 \approx 12.4$, we can assume that population variances are equal and us pooled st.

95% confidence interval would be the following

$$\mu_1 - \mu_2 = \bar{X}_1 - \bar{X}_2 \pm t(n_1+n_2-2)_{0.025} \cdot S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$, where

$\bar{X}_1$ is the mean of sixth-grade students

$\bar{X}_2$ is the mean of seventh-grade students

n_1 and n_2 are sample sizes respectively

S_p is pooled st. dev.

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

S_1 and S_2 are standard deviations respectively

GO ON TO THE NEXT PAGE.

C_2

If you need more room for your work for part (a), use the space below.

$\mu_1 - \mu_2 = 27.3 - 47.0 \pm t(38)_{0.025} \cdot \sqrt{\frac{19 \cdot 10.8^2 + 19 \cdot 12.4^2}{38}}$

$\mu_1 - \mu_2 = -19.7 \pm 7.4436$

$\mu_1 - \mu_2 \in (-27.1436; -12.256)$

we are 95% confident that that true difference in means between sixth-grade st. and seventh-grade students is between -27.1436 and -12.256. ⇒

⇒ seventh-grade students spend more time on homework than sixth-grade students (because the conf. interval doesn't cover zero, it's negative)

(b) An assistant principal reasoned that a much narrower confidence interval could be obtained if the students were paired based on their responses; for example, pairing the sixth-grade student and the seventh-grade student with the highest number of minutes spent on homework, the sixth-grade student and seventh-grade student with the next highest number of minutes spent on homework, and so on. Is the assistant principal correct in thinking that matching students in this way and then computing a matched-pairs confidence interval for the mean difference in time spent on homework is a better procedure than the one used in part (a) ? Explain why or why not.

Yes it will be so, because the standard deviation of the difference will be less

GO ON TO THE NEXT PAGE.