Review of Statistics: California Test Scores

Dr. Patrick Toche

Textbook:

James H. Stock and Mark W. Watson, *Introduction to Econometrics*, 4th Edition, Pearson.

Other references:

Joshua D. Angrist and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, 1st Edition, Princeton University Press.
Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 7th Edition, Cengage Learning.

The textbook comes with online resources and study guides. Other references will be given from time to time.

Problems and Applications: California Test Scores

Data on fifth-grade test scores (reading and mathematics) for 420 school districts in California yield average score $\overline{Y} = 654.2$ and standard deviation $s_Y = 19.1$.

a. Construct a 95% confidence interval for the mean test score in the population.

b. When the districts were divided into those with small classes (< 20 students per teacher) and those with large classes (≥ 20 students per teacher), the following results were found:

Class Size	Average Score ($\overline{Y}$)	Standard Deviation (s_Y)	n
Small	657.4	19.4	238
Large	650.0	17.9	182

Is there statistically significant evidence that the districts with smaller classes have higher average test scores? Explain.

Problems and Applications: California Test Scores

Data on fifth-grade test scores (reading and mathematics) for 420 school districts in California yield average score $\overline{Y} = 654.2$ and standard deviation $s_Y = 19.1$.

a. Construct a 95% confidence interval for the mean test score in the population.

b. When the districts were divided into those with small classes (< 20 students per teacher) and those with large classes (≥ 20 students per teacher), the following results were found:

Class Size	Average Score ($\overline{Y}$)	Standard Deviation (s_Y)	n
Small	657.4	19.4	238
Large	650.0	17.9	182

Is there statistically significant evidence that the districts with smaller classes have higher average test scores? Explain.

Problems and Applications: California Test Scores

Data on fifth-grade test scores (reading and mathematics) for 420 school districts in California yield average score $\overline{Y} = 654.2$ and standard deviation $s_Y = 19.1$.

a. Construct a 95% confidence interval for the mean test score in the population.

b. When the districts were divided into those with small classes (< 20 students per teacher) and those with large classes (≥ 20 students per teacher), the following results were found:

Class Size	Average Score ($\overline{Y}$)	Standard Deviation (s_Y)	n
Small	657.4	19.4	238
Large	650.0	17.9	182

Is there statistically significant evidence that the districts with smaller classes have higher average test scores? Explain.

Problems and Applications: California Test Scores

$n = 420, \overline{Y} = 654.2, s_Y = 19.1.$

a. Construct a 95% confidence interval for the mean test score in the population.

As the population standard deviation is unknown, the asymptotic distribution of the test statistic follows a Student-t distribution. The significance level is $1 - 0.95 = 0.05$. The corresponding critical t-statistic for $420 - 1 = 419$ degrees of freedom is about 1.965642. This compares with about 1.959964 for the standard normal distribution.

The standard error is:

$$\text{SE} = \frac{s_Y}{\sqrt{n}} = \frac{19.1}{\sqrt{420}} \approx 0.93$$

A two-sided confidence interval for mean test score in the population is:

$$\begin{aligned} \overline{Y} \pm t_{1-\alpha/2} \times \text{SE} &= 654.2 \pm 1.96 \times 0.93 \\ &= (652.37, 656.03) \end{aligned}$$

The margin of error is about $1.83.$

Problems and Applications: California Test Scores

$n = 420, \overline{Y} = 654.2, s_Y = 19.1.$

a. Construct a 95% confidence interval for the mean test score in the population.

As the population standard deviation is unknown, the asymptotic distribution of the test statistic follows a Student-t distribution. The significance level is $1 - 0.95 = 0.05$. The corresponding critical t-statistic for $420 - 1 = 419$ degrees of freedom is about 1.965642. This compares with about 1.959964 for the standard normal distribution.

The standard error is:

$$\text{SE} = \frac{s_Y}{\sqrt{n}} = \frac{19.1}{\sqrt{420}} \approx 0.93$$

A two-sided confidence interval for mean test score in the population is:

$$\begin{aligned} \overline{Y} \pm t_{1-\alpha/2} \times \text{SE} &= 654.2 \pm 1.96 \times 0.93 \\ &= (652.37, 656.03) \end{aligned}$$

The margin of error is about 1.83.

Problems and Applications: California Test Scores

b. Is there statistically significant evidence that the districts with smaller classes have higher average test scores? Explain.

A two-sided test:

$$H_0: \mu_{\text{large}} - \mu_{\text{small}} = 0$$
$$H_1: \mu_{\text{large}} - \mu_{\text{small}} \neq 0$$

The Student-t statistic for this test:

$$t = \frac{(657.4 - 650.0) - 0}{\sqrt{\frac{(19.4)^2}{238} + \frac{(17.9)^2}{182}}} \approx \frac{7.4}{1.828} \approx 4.05$$

There are enough degrees of freedom that we can refer to the standard normal distribution for an accurate critical value. Since $|4.05| > 1.96$, we reject the null hypothesis of no effect of class sizes.

Problems and Applications: California Test Scores

b. Is there statistically significant evidence that the districts with smaller classes have higher average test scores? Explain.

A two-sided test:

$$H_0 : \mu_{\text{large}} - \mu_{\text{small}} = 0$$
$$H_1 : \mu_{\text{large}} - \mu_{\text{small}} \neq 0$$

The Student-t statistic for this test:

$$t = \frac{(657.4 - 650.0) - 0}{\sqrt{\frac{(19.4)^2}{238} + \frac{(17.9)^2}{182}}} \approx \frac{7.4}{1.828} \approx 4.05$$

There are enough degrees of freedom that we can refer to the standard normal distribution for an accurate critical value. Since $|4.05| > 1.96$, we reject the null hypothesis of no effect of class sizes.