Review of Statistics: Polling

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: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger. Let p denote the fraction of all likely voters who preferred the incumbent at the time of the survey, and let $\hat{p}$ be the fraction of survey respondents who preferred the incumbent.

a. Use the survey results to estimate p.

b. Use the estimator of the variance of $\hat{p}$, $\hat{p}(1-\hat{p})$, to calculate the standard error of your estimator.

c. What is the p-value for the test of $H_0 \colon p = 0.5$ vs. $H_1 \colon p \neq 0.5$?

d. What is the p-value for the test of $H_0 \colon p = 0.5$ vs. $H_1 \colon p > 0.5$?

e. Why do the results from (c) and (d) differ?

f. Did the survey contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey? Explain.

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger. Let p denote the fraction of all likely voters who preferred the incumbent at the time of the survey, and let $\hat{p}$ be the fraction of survey respondents who preferred the incumbent.

a. Use the survey results to estimate p.

b. Use the estimator of the variance of $\hat{p}$, $\hat{p}(1-\hat{p})$, to calculate the standard error of your estimator.

c. What is the p-value for the test of $H_0\colon p = 0.5$ vs. $H_1\colon p \neq 0.5$?

d. What is the p-value for the test of $H_0\colon p = 0.5$ vs. $H_1\colon p > 0.5$?

e. Why do the results from (c) and (d) differ?

f. Did the survey contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey? Explain.

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger. Let p denote the fraction of all likely voters who preferred the incumbent at the time of the survey, and let $\hat{p}$ be the fraction of survey respondents who preferred the incumbent.

a. Use the survey results to estimate p.
b. Use the estimator of the variance of $\hat{p}$, $\hat{p}(1 - \hat{p})$, to calculate the standard error of your estimator.
c. What is the p-value for the test of $H_0 \colon p = 0.5$ vs. $H_1 \colon p \neq 0.5$?
d. What is the p-value for the test of $H_0 \colon p = 0.5$ vs. $H_1 \colon p > 0.5$?
e. Why do the results from (c) and (d) differ?
f. Did the survey contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey? Explain.

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger. Let p denote the fraction of all likely voters who preferred the incumbent at the time of the survey, and let $\hat{p}$ be the fraction of survey respondents who preferred the incumbent.

a. Use the survey results to estimate p.
b. Use the estimator of the variance of $\hat{p}$, $\hat{p}(1-\hat{p})$, to calculate the standard error of your estimator.
c. What is the p-value for the test of $H_0\colon p = 0.5$ vs. $H_1\colon p \neq 0.5$?
d. What is the p-value for the test of $H_0\colon p = 0.5$ vs. $H_1\colon p > 0.5$?
e. Why do the results from (c) and (d) differ?
f. Did the survey contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey? Explain.

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger. Let p denote the fraction of all likely voters who preferred the incumbent at the time of the survey, and let $\hat{p}$ be the fraction of survey respondents who preferred the incumbent.

a. Use the survey results to estimate p.
b. Use the estimator of the variance of $\hat{p}$, $\hat{p}(1-\hat{p})$, to calculate the standard error of your estimator.
c. What is the p-value for the test of $H_0\colon p = 0.5$ vs. $H_1\colon p \neq 0.5$?
d. What is the p-value for the test of $H_0\colon p = 0.5$ vs. $H_1\colon p > 0.5$?
e. Why do the results from (c) and (d) differ?
f. Did the survey contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey? Explain.

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger. Let p denote the fraction of all likely voters who preferred the incumbent at the time of the survey, and let $\hat{p}$ be the fraction of survey respondents who preferred the incumbent.

a. Use the survey results to estimate p.
b. Use the estimator of the variance of $\hat{p}$, $\hat{p}(1-\hat{p})$, to calculate the standard error of your estimator.
c. What is the p-value for the test of $H_0\colon p = 0.5$ vs. $H_1\colon p \neq 0.5$?
d. What is the p-value for the test of $H_0\colon p = 0.5$ vs. $H_1\colon p > 0.5$?
e. Why do the results from (c) and (d) differ?
f. Did the survey contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey? Explain.

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger.

a. Use the survey results to estimate p.

$$\hat{p} = \frac{215}{400} = 0.5375$$

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger.

a. Use the survey results to estimate p.

$$\hat{p} = \frac{215}{400} = 0.5375$$

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger.

b. Use the estimator of the variance of $\hat{p}$, $\hat{p}(1-\hat{p})$, to calculate the standard error of your estimator.

We have $\hat{p} = 0.5375$ and $n = 400$,

$$\begin{aligned} \text{SE}(\hat{p}) &= \sqrt{\frac{\text{var}(\hat{p})}{n}}, \quad \text{where } \text{var}(\hat{p}) = \hat{p}(1-\hat{p}) \\ &= \sqrt{\frac{0.5375(1-0.5375)}{400}} \\ &\approx 0.0250 \end{aligned}$$

R code: `sqrt(0.5375*(1-0.5375)/400)`

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger.

b. Use the estimator of the variance of $\hat{p}$, $\hat{p}(1-\hat{p})$, to calculate the standard error of your estimator.

We have $\hat{p} = 0.5375$ and $n = 400$,

$$\begin{aligned}
\text{SE}(\hat{p}) &= \sqrt{\frac{\text{var}(\hat{p})}{n}}, \quad \text{where } \text{var}(\hat{p}) = \hat{p}(1-\hat{p}) \\
&= \sqrt{\frac{0.5375(1-0.5375)}{400}} \\
&\approx 0.0250
\end{aligned}$$

R code: `sqrt(0.5375*(1-0.5375)/400)`

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger.

c. What is the p-value for the test of $H_0\colon p = 0.5$ vs. $H_1\colon p \neq 0.5$?

We have $p_0 = 0.5$ and $\text{SE}(\hat{p}) \approx 0.0250$,

$$\begin{aligned} p\text{-value} &= \mathrm{P}\left[-z < \frac{\hat{p} - p_0}{\text{SE}(\hat{p})} < z\right] \\ &= 2 \cdot \mathrm{P}\left[z > \frac{\hat{p} - p_0}{\text{SE}(\hat{p})}\right] \\ &\approx 2 \cdot \mathrm{P}\left[z > \frac{0.5375 - 0.5}{0.0250}\right] \\ &\approx 2 \cdot (1 - \mathrm{P}[z < 1.5]) \\ &\approx .0133 \end{aligned}$$

R code:
```
z = (0.5375-0.5)/sqrt(0.5375*(1-0.5375)/400)
2*(1-pnorm(z))
```

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger.

c. What is the p-value for the test of $H_0\colon p = 0.5$ vs. $H_1\colon p \neq 0.5$?

We have $p_0 = 0.5$ and $\text{SE}(\hat{p}) \approx 0.0250$,

$$
\begin{aligned}
p\text{-value} &= \mathrm{P}\left[-z < \frac{\hat{p} - p_0}{\text{SE}(\hat{p})} < z\right] \\
&= 2 \cdot \mathrm{P}\left[z > \frac{\hat{p} - p_0}{\text{SE}(\hat{p})}\right] \\
&\approx 2 \cdot \mathrm{P}\left[z > \frac{0.5375 - 0.5}{0.0250}\right] \\
&\approx 2 \cdot (1 - \mathrm{P}[z < 1.5]) \\
&\approx .0133
\end{aligned}
$$

R code:
```
z = (0.5375-0.5)/sqrt(0.5375*(1-0.5375)/400)
2*(1-pnorm(z))
```

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger.

d. What is the p-value for the test of $H_0\colon p = 0.5$ vs. $H_1\colon p > 0.5$?

We have $p_0 = 0.5$ and $\text{SE}(\hat{p}) \approx 0.0250$,

$$\begin{aligned} p\text{-value} &= \mathrm{P}\left[z > \frac{\hat{p} - p_0}{\text{SE}(\hat{p})}\right] \\ &\approx \mathrm{P}\left[z > \frac{0.5375 - 0.5}{0.0250}\right] \\ &\approx 1 - \mathrm{P}[z < 1.5] \\ &\approx .066 \end{aligned}$$

R code:
```
z = (0.5375-0.5)/sqrt(0.5375*(1-0.5375)/400)
1-pnorm(z)
```

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger.

d. What is the p-value for the test of $H_0\colon p = 0.5$ vs. $H_1\colon p > 0.5$?

We have $p_0 = 0.5$ and $\text{SE}(\hat{p}) \approx 0.0250$,

$$
\begin{aligned}
p\text{-value} &= \mathrm{P}\left[z > \frac{\hat{p} - p_0}{\text{SE}(\hat{p})}\right] \\
&\approx \mathrm{P}\left[z > \frac{0.5375 - 0.5}{0.0250}\right] \\
&\approx 1 - \mathrm{P}[z < 1.5] \\
&\approx .066
\end{aligned}
$$

R code:
```
z = (0.5375-0.5)/sqrt(0.5375*(1-0.5375)/400)
1-pnorm(z)
```

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger.

e. Why do the results from (c) and (d) differ?

The first test is a two-sided test, the second test is a one-sided test. The one-sided test assumes the probability of an extreme sample observation falls into the right tail of the probability distribution, ruling out entirely that it could occur in the left tail. It is a stronger test (assuming it is reasonable to make that assumption) and leads to a larger p-value.

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger.

e. Why do the results from (c) and (d) differ?

   The first test is a two-sided test, the second test is a one-sided test. The one-sided test assumes the probability of an extreme sample observation falls into the right tail of the probability distribution, ruling out entirely that it could occur in the left tail. It is a stronger test (assuming it is reasonable to make that assumption) and leads to a larger p-value.

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger.

f. Did the survey contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey? Explain.

   This is a matter of judgment. We are not given information about the preferred significance level. The one-sided p-value is larger than the common benchmark of 0.05, but the two-sided p-value is smaller. I see no reason to conduct a one-sided test and a p-value of 0.01 feels small to me, so I would say that, yes, the survey evidence suggests that the incumbent may have been ahead of the challenger at the time of the survey. (I am aware that a Google search leads to answers that suggest the opposite ...)

Problems and Applications: Polling

In a survey of 400 likely voters, 215 responded that they would vote for the incumbent, and 185 responded that they would vote for the challenger.

f. Did the survey contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey? Explain.

   This is a matter of judgment. We are not given information about the preferred significance level. The one-sided p-value is larger than the common benchmark of 0.05, but the two-sided p-value is smaller. I see no reason to conduct a one-sided test and a p-value of 0.01 feels small to me, so I would say that, yes, the survey evidence suggests that the incumbent may have been ahead of the challenger at the time of the survey. (I am aware that a Google search leads to answers that suggest the opposite ...)