Assignment 9a

Satvik Saha

2024-10-29

Answer 1

(a) The new standard error with a sample size of n will be around $se = 4.0\sqrt{100/n}$. We want $2.8se = 5.0 - 1.4$, hence $n \approx (2.8 \cdot 4.0/(5.0 - 1.4))^2 \cdot 100 \approx 968$.

(b) This time, we want $5.0 - 1.4$ to be

```
qnorm(0.8) + qnorm(0.95)
```

```
## [1] 2.486475
```

standard errors, so $n \approx (2.5 \cdot 4.0/(5.0 - 1.4))^2 \cdot 100 \approx 772$.

Answer 2

(a) We equate $0.03 = 0.5/\sqrt{n}$, whence $n \approx (0.5/0.03)^2 \approx 278$.

(b) We make the approximation that 14% of our sample will be Latinos, hence we ask for $n \approx (0.5/0.03)^2/0.14 \approx 1985$.

(c) We approximate $0.01 = \sqrt{0.14 \cdot (1 - 0.14)/n}$, whence $n \approx 0.14 \cdot (1 - 0.14)/0.01^2 \approx 1204$.

Research homework assignment

Suppose that we are in the setup $y_i \sim N(g(x_i), \sigma^2)$, and x_i take values in $0, 0.5, 1$ with frequencies $(1 - w)n/2, wn, (1 - w)n/2$. Then, for $x \in \{0, 0.5, 1\}$, we have the means within each group $\bar{y}_x \sim N(g(x), \sigma^2/n_x)$. In order to estimate the full effect $\theta_i = g(1) - g(0)$ and the half-effect $\theta_{0.5} = g(0.5) - g(0)$, suppose that we use $\hat{\theta}_i = \bar{y}_1 - \bar{y}_0$ and $\hat{\theta}_{0.5} = \bar{y}_{0.5} - \bar{y}_0$. We are also interested in the relative non-linearity $\delta = (0.5\theta_1 - \theta_{0.5})/\theta_1$.

Check that

$$\mathrm{MSE}(\hat{\theta}_1) = \frac{4}{1 - w}\,\frac{\sigma^2}{n}, \qquad \mathrm{MSE}(\hat{\theta}_{0.5}) = \frac{1 + w}{w(1 - w)}\,\frac{\sigma^2}{n}.$$

When $w = 0$, we have

$$\mathrm{MSE}(\hat{\theta}_1^0) = 4\,\frac{\sigma^2}{n}.$$

When $w = 1/3$, we have

$$\mathrm{MSE}(\hat{\theta}_1^{1/3}) = 6\,\frac{\sigma^2}{n}, \qquad \mathrm{MSE}(\hat{\theta}_{0.5}^{1/3}) = 6\,\frac{\sigma^2}{n}.$$

We would like to measure the deviation from linearity $\Delta = 0.5\theta_1 - \theta_{0.5}$, estimated by $\hat{\Delta} = 0.5\hat{\theta}_1 - \hat{\theta}_{0.5}$, which has

$$\mathrm{MSE}(\hat{\Delta}^{1/3}) = 7.5\,\frac{\sigma^2}{n}.$$

Thus, if we are satisfied with the estimate of the full effect θ_1 with the scheme $w = 0$ for some sample size n, say with our MSE $4\sigma^2/n < \epsilon$, then we would need a sample size of around $2n$ with the scheme $w = 1/3$ for our MSE for the estimated deviation from linearity Δ to be similarly small, since $7.5\sigma^2/2n < \epsilon$.