Advanced Statistical Methods Assignment 1

Dahye Kim

2021 9 20

1.2

Since most of bootstrap replications in Figure 1.3 show a similar pattern to Figure1.2, one might suggest that the flat spot between ages 25 and 35 is genuine.

1.3

Under the assumption, I think the largest observed t value might be 3.826058. Because in all 7128 cases t exactly follows a student-t distribution with 70 degrees, the smallest p-value of t might be $\frac{1}{7128} = 0.00014$, which makes t value the largest. And since $t_{70}(3.826058) = 0.00014$, I think the largest observed t value might be 3.826058.

```
qt(1/7128, 70)
```

```
## [1] -3.825437
```

2.3

The definition of frequentism in terms of probabilistic accuracy attribute for an observed estimate $\hat{\theta} = t(x)$ the probabilistic properties of $\hat{\Theta} = t(X)$ as an estimator of $\hat{\theta}$. That is, $\hat{\theta}$ is just a single number but $\hat{\Theta}$ takes on a range of values whose spread can define measures of accuracy. Similarly, the definition of frequentism in terms of an infinite sequence of future trials attribute for $\hat{\theta}$ the accuracy properties of the ensemble of $\hat{\Theta}$ values.

2.4

In (2.15),

$$\hat{\theta} = \bar{x}_2 - \bar{x}_1 \sim N(0, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right))$$

under H_0, which depends on σ. And if we plug in the unbiased estimate of σ^2,

$$\hat{\sigma}^2 = \left[\sum_1^{n_1}(x_{1i} - \bar{x}_1)^2 + \sum_1^{n_2}(x_{2i} - \bar{x}_2)^2\right]/(n_1 + n_2 - 2)$$

in (2.15), then the approximate 0.95 confidence interval for the difference $\mu_2 - \mu_1$ is

$$\bar{x}_2 - \bar{x}_1 \pm 1.96 \cdot \widehat{sd}$$

However, when we use student-t hypothesis test,

$$t = \frac{\bar{x}_2 - \bar{x}_1}{\widehat{sd}} \sim t_{n_1+n_2-2}$$

under H_0, where $\widehat{sd} = \hat{\sigma}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}$. t is pivotal, having the same distribution, no matter what the value of the “nuisance parameter” σ. So we do not have to “plug in”. And the exact 0.95 confidence interval for the difference $\mu_2 - \mu_1$ is

$$\bar{x}_2 - \bar{x}_1 \pm 1.99 \cdot \widehat{sd}.$$

, which is larger than that of approximate 95% normal theory hypothesis test, covering the true value in 95% of repetitions of probability model (2.12)-(2.13). This is because the tail of student-t distribution is usually fatter than that of normal distribution.

Therefore, student-t hypothesis test performs better.