STAT S4201 001, Homework 2

Brian Weinstein (bmw2148)

Feb 10, 2016

Code is attached here and also posted at https://github.com/BrianWeinstein/advanced-data-analysis. Where relevant, code snippets and output are are included in-line.

Problem 1:

Suppose that $\{X_1, X_2, \ldots, X_n\}$ is a random sample from $N(\mu, \sigma^2)$. Construct a 95% confidence interval for σ^2 under the following scenarios: (a) μ is known to be 0. (b) μ is unknown.

(a) asdf

Problem 1a

(b) asdf

Type up Problem 1b

Fix $n = 10$ and $\sigma = 1$. Run a Monte Carlo simulation to confirm that the confidence interval you constructed under the scenario (a) produces a coverage of 95%. Report how many random samples were drawn in your simulation and how close your coverage was to 95%.

See attached code. In my simulation I drew 1000 random samples, with a 95% confidence interval capturing the true variance in 948 of the trials (94.8% coverage).

Problem 2: Ramsey 3.22

```
# Data input
time26 <- c(5.79, 1579.52, 2323.70)
time28 <- c(68.8, 108.29, 110.29, 426.07, 1067.60)
```

(a) *Form two new variables by taking the logarithms of the breakdown times.*

```
> Y1 <- log(time26) ; Y1
[1] 1.756132 7.364876 7.750916
> Y2 <- log(time28) ; Y2
[1] 4.231204 4.684813 4.703113 6.054604 6.973168
```

(b) $\overline{Y}_1 - \overline{Y}_2 = 5.6240 - 5.3294 = 0.2946$

(c) $\exp\left(\overline{Y}_1 - \overline{Y}_2\right) = \exp(0.2946) = 1.3426$, where 1.3426 is the multiplicative treatment effect, indicating that the breakdown time at 26 kV is estimated to be 1.3426 times larger than the breakdown time at 28 kV.

(d) *Compute a 95% confidence interval for the difference in mean log breakdown times. Take the antilogaritms of the endpoints and express the result in a sentence.*

The pooled standard deviation is given by

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}} = \sqrt{\frac{(3-1)(11.257) + (5-1)(1.310)}{(3+5-2)}} = 2.1508,$$

where $n_1 = 3$ and $n_2 = 5$ are the sample sizes and $s_1^2 = 11.2574$ and $s_2^2 = 1.3104$ are the sample variances of the log-transformed measurements.
The standard error of $\left(\overline{Y}_1 - \overline{Y}_2\right)$ is given by

$$\text{SE}\left(\overline{Y}_1 - \overline{Y}_2\right) = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (2.1508)\sqrt{\frac{1}{3} + \frac{1}{5}} = 1.5707.$$

Therefore a 95% confidence interval for the difference in mean log breakdown times is

$$\begin{gathered}
\left(\overline{Y}_1 - \overline{Y}_2\right) \pm t_{(3+5-2)}\left(1 - \frac{\alpha}{2}\right)\text{SE}\left(\overline{Y}_1 - \overline{Y}_2\right) \\
0.2946 \pm t_6\left(1 - \frac{0.05}{2}\right)1.5707 \\
0.2946 \pm (2.4469)(1.5707) \\
0.2946 \pm 3.8435 \\
\Longrightarrow -3.5489 \le \left(\overline{Y}_1 - \overline{Y}_2\right) \le 4.1381.
\end{gathered}$$

Taking the anilogarithms of the confidence interval endpoints,

$$\begin{aligned}
\text{Lower confidence limit} &= e^{-3.5489} = 0.0288 \\
\text{Upper confidence limit} &= e^{4.1381} = 62.682,
\end{aligned}$$

we find that the breakdown time at 26 kV is estimated to be 1.3426 (from part c) times longer than the breakdown time at 28 kV, (95% confidence: 0.0288 to 62.682 times).

Problem 3: Ramsey 3.25

Problem 4: Ramsey 3.28

Problem 5: Ramsey 3.32

Problem 6: Ramsey 4.19

Todo list

Problem 1a 1
Type up Problem 1b 1