STAT430 Homework #5: Due Friday, March 8, 2019.

Name:

---

0. We will continue with **Chapter 8** on estimation. You should have read **Sections 8.1-8.4** on point estimation; please continue with **Sections 8.5-8.10** on interval estimation and related topics.
1. Suppose $[\hat{\theta}_L, \hat{\theta}_U]$ is a $(1-\alpha)$ 100% confidence interval for the target parameter θ.

(a) Find a $(1-\alpha)$ 100% confidence interval for the new target parameter $a + b\theta$, where a is a known real number and $b > 0$ is a known real number.

Answer:

$$1 - \alpha = P(\hat{\theta}_L \leq a + b\theta \leq \hat{\theta}_U), b > 0$$
$$= P(\frac{\hat{\theta}_L - a}{b} \leq \theta \leq \frac{\hat{\theta}_U - a}{b})$$
$$[\frac{\hat{\theta}_L - a}{b}, \frac{\hat{\theta}_U - a}{b}]$$

(b) Find a $(1-\alpha)$ 100% confidence interval for the new target parameter $a + b\theta$, where a is a known real number and $b < 0$ is a known real number.

Answer:

$$1 - \alpha = P(\hat{\theta}_L \leq a + b\theta \leq \hat{\theta}_U), b < 0$$
$$= P(\frac{\hat{\theta}_L - a}{b} \geq \theta \geq \frac{\hat{\theta}_U - a}{b})$$
$$[\frac{\hat{\theta}_U - a}{b}, \frac{\hat{\theta}_L - a}{b}]$$

(c). Suppose $Y_1, \ldots, Y_{16}$ are iid $N(\mu, \sigma^2)$ and are measured in degrees Fahrenheit.
Form a 90% confidence interval for the true mean μ if the realized data are as follows:

```
Y <- c(97.91, 98.81, 97.42, 98.01, 98.25, 98.18, 98.90, 98.25, 99.59, 98.50, 98.60, 97.92, 97.48, 99.21
```

Answer:

This is the generic 90% CI for $\bar{Y}$

$$\bar{Y} \pm t_{\alpha/2,n-1}\sqrt{S^2/n}$$

```
Y <- c(97.91, 98.81, 97.42, 98.01, 98.25, 98.18, 98.90, 98.25, 99.59, 98.50, 98.60, 97.92, 97.48, 99.21

Y_bar <- mean(Y)

lower <- Y_bar + qt(.05,length(Y))*sqrt(var(Y)/length(Y))
```

```r
upper <- Y_bar - qt(.05,length(Y))*sqrt(var(Y)/length(Y))


lower
```

```
## [1] 98.15985
```

```r
upper
```

```
## [1] 98.6914
```

So a 90% CI for $\bar{Y}$ is,

$$[98.15985, 98.6914]$$

(d). Suppose we are interested in the true mean in degrees Celsius = (5/9) * (Fahrenheit - 32), instead of Fahrenheit. Use the data in part (c) to compute a 90% confidence interval for the true mean in degrees Celsius.

Answer:

```r
X <- c(97.91, 98.81, 97.42, 98.01, 98.25, 98.18, 98.90, 98.25, 99.59, 98.50, 98.60, 97.92, 97.48, 99.21

Y <- (5/9)*(X -32)

Y_bar <- mean(Y)

lower <- Y_bar + qt(.05,length(Y))*sqrt(var(Y)/length(Y))
upper <- Y_bar - qt(.05,length(Y))*sqrt(var(Y)/length(Y))


lower
```

```
## [1] 36.75547
```

```r
upper
```

```
## [1] 37.05078
```

So a 90% CI for the temperature in Celsius is,

$$[36.75547, 37.05078]$$

---

2. The code in the answer block below computes and plots exact $(1-\alpha)$ 100% two-sided confidence intervals for `nreps` replicates of the sampling experiment in which $Y_1, \ldots, Y_n$ iid $N(\theta, \sigma^2)$, with θ as our target and σ^2 unknown. The pivot used to produce the CI is

$$Q = \frac{\bar{Y} - \theta}{\sqrt{S^2/n}} \sim t_{n-1},$$

where $\bar{Y}$ is the sample mean of the observations and S^2 is the sample variance. The quantiles then come from the t distribution with $n-1$ degrees of freedom, and by isolating θ in the probability statement, we get

$$1 - \alpha = P\left(q_{\alpha/2} \leq Q \leq q_{1-\alpha/2}\right) = P\left(\bar{Y} + q_{\alpha/2}\sqrt{\frac{S^2}{n}} \leq \theta \leq \bar{Y} + q_{1-\alpha/2}\sqrt{\frac{S^2}{n}}\right),$$

so

$$\hat{\theta}_L = \bar{Y} + q_{\alpha/2}\sqrt{\frac{S^2}{n}}, \quad \hat{\theta}_U = \bar{Y} + q_{1-\alpha/2}\sqrt{\frac{S^2}{n}}.$$

(Because the t distribution is symmetric about zero, $q_{\alpha/2} = -q_{1-\alpha/2}$, and we often write $\bar{Y} \pm q_{1-\alpha/2}\sqrt{S^2/n}$ for the CI.)

Run the code below with the given settings, then increase the sample size to $n = 100$ and run again. Comment on the width and the *coverage* (empirical proportion of hits) of the CIs at the two sample sizes. What changes and what does not change?

Answer:

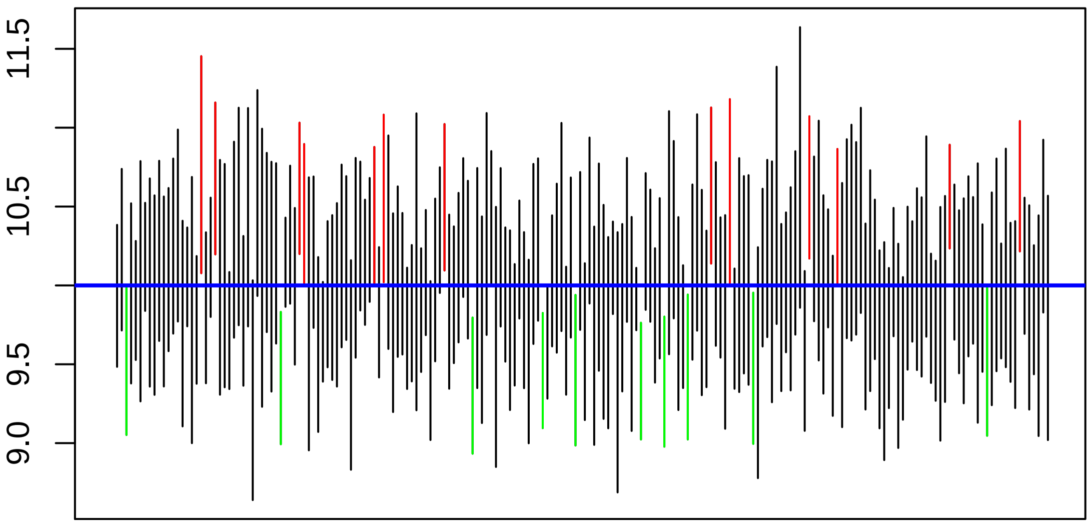
```
nreps <- 200
theta <- 10
n <- 10
alpha <- 0.1
set.seed(4302019)
Y <- rnorm(n * nreps, mean = theta, sd = 1)
YY <- matrix(Y, n, nreps)  # nreps iid samples of size n, one in each column
sample_mean <- apply(YY, MAR = 2, FUN = "mean")  # sample mean, Ybar, for each column
sample_var  <- apply(YY, MAR = 2, FUN = "var")   # sample variance, S^2, for each column
# Compute lower confidence limit, using quantile from t distribution
theta_hat_L <- sample_mean + qt(alpha / 2, df = n - 1) * sqrt(sample_var / n)
# Compute upper confidence limitk using quantile from t distribution
theta_hat_U <- sample_mean + qt(1 - alpha / 2, df = n - 1) * sqrt(sample_var / n)
too_high <- (theta < theta_hat_L)  # CI is too high and missed theta
too_low  <- (theta > theta_hat_U)  # CI is too low and missed theta
mean(too_low + too_high)      # proportion of misses should be close to alpha
```

```
## [1] 0.115
```

```
1 - mean(too_low + too_high)  # proportion of hits should be close to 1 - alpha
```

```
## [1] 0.885
```

```
# Plot the CIs.
plot(c(0,1), c(min(theta_hat_L), max(theta_hat_U)), type = "n",
     xlab = paste(nreps, "replicates of confidence intervals"), ylab = "", xaxt = "n")
x <- (1:nreps) / nreps  # index for each of the nreps CIs
segments(x, theta_hat_L, x, theta_hat_U)
segments(x[too_high], theta_hat_L[too_high], x[too_high], theta_hat_U[too_high], col = "red")
segments(x[too_low], theta_hat_L[too_low], x[too_low], theta_hat_U[too_low], col = "green")
abline(h = theta, col = "blue", lwd = 2)
```



200 replicates of confidence intervals

---

3. In class, we considered the sampling experiment in which $Y_1, \ldots, Y_n$ iid $N(\mu, \sigma^2)$, with $\theta = \sigma^2$ as our target. The pivot used to produce the CI is

$$Q = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1},$$

where S^2 is the sample variance. The quantiles then come from the χ^2 distribution with $n-1$ degrees of freedom, and by isolating θ in the probability statement, we obtained the two-sided CI,

$$\left[\frac{(n-1)S^2}{q_{1-\alpha/2}}, \frac{(n-1)S^2}{q_{\alpha/2}}\right].$$

Modify the code from problem 2 above to compute and plot `nreps=200` 95% confidence intervals for σ^2 when $n = 20$ and the true population parameters are $\mu = 10$, $\sigma^2 = 1$. Report your empirical coverage (proportion of hits).

Answer:

---

4. Consider the setting of problem 3. The true variance, σ^2, is naturally bounded below by zero. In some cases, we might only be interested in finding an upper confidence limit for σ^2, so that we have the one-sided confidence interval $(-\infty, \hat{\theta}_U]$, but we know this is actually $[0, \hat{\theta}_U]$. For this problem, the target is still σ^2 and we still use the same pivot.

(a). Show that

$$1 - \alpha = P\left(0 \le \sigma^2 \le \frac{(n-1)S^2}{q_\alpha}\right),$$

so that

$$\left[0, \frac{(n-1)S^2}{q_\alpha}\right]$$

is a valid $(1-\alpha)$ 100% confidence interval for σ^2.

Answer:

(b). Using exactly the same setting as problem 3, modify the code to compute and plot the one-sided 95% confidence intervals, and report their empirical coverage. In this case, you can replace the lower endpoint by a vector of 0's: `theta_hat_L <- rep(0, nreps)`.

Answer:

---

5. Complete **Exercise 8.40** of the text. In this problem, the variance is *known* to be 1.

Answer:

6, parts 6(a) and 6(b). Complete parts (a) and (b) of **Exercise 8.44** of the text.

Answer:

6(c). Quantiles are defined by the property that

$$F_Q(q_\alpha) = P(Q \leq q_\alpha) = \alpha,$$

where $F_Q(y)$ is the cumulative distribution function of the random variable, Q. Use this fact and the result of 6(b) to find a general expression for the αth quantile of the pivotal distribution, then use `R` to compute the quantiles $q_{0.025}$, $q_{0.05}$, $q_{0.1}$, $q_{0.9}$, $q_{0.95}$, and $q_{0.975}$.

Answer:

6(d). Use your computed quantiles and pivotal quantity to compute a 90% two-sided confidence interval for θ.

Answer:

6(e). Use your computed quantiles and pivotal quantity to compute the 90% upper confidence limit for θ (This is **Exercise 8.45(a)** of the book, so you can check your answer in the back.)

Answer:

---

7. (a), (b) and (c): Complete **Exercise 8.47** of the text. (d) Then repeat part (c) with the same setup ($\bar{y} = 4.77$, 95%), but with $n = 21$. Comment on the difference between the results in (c) and (d).

Answer: