## Announcements

Make sure to sign in on the google form (I send a list of which section questions are useful for which pset questions afterwards)

Pset 10 due Friday 4/21

## Cost effectiveness in global health

Significant work in economics and global health has gone into determining which healthcare interventions prevent the most suffering per dollar spent. The metric of choice for these evaluations is the DALY, a disability adjusted life year. When comparing interventions designed to save lives with those designed to improve lives, it is useful to have a metric that adjusts for the quality of life associated with various diseases and injuries, and the DALY attempts to do this. For example, "severe motor impairment with blindness due to malaria" (being blind and unable to move around or sit up without help) has a weight of 0.512. This means that avoiding a year of severe motor impairment with blindness is considered equivalent to avoiding 0.512 years of lost life.

The Tufts Medical Center Cost-Effectiveness Analysis Registry aggregates academic literature on healthcare interventions and standardizes the results to compare different interventions. This literature is notoriously variable and unstandardized, so the specific metrics for each intervention might not be accurate, but the general trends should provide some insight.
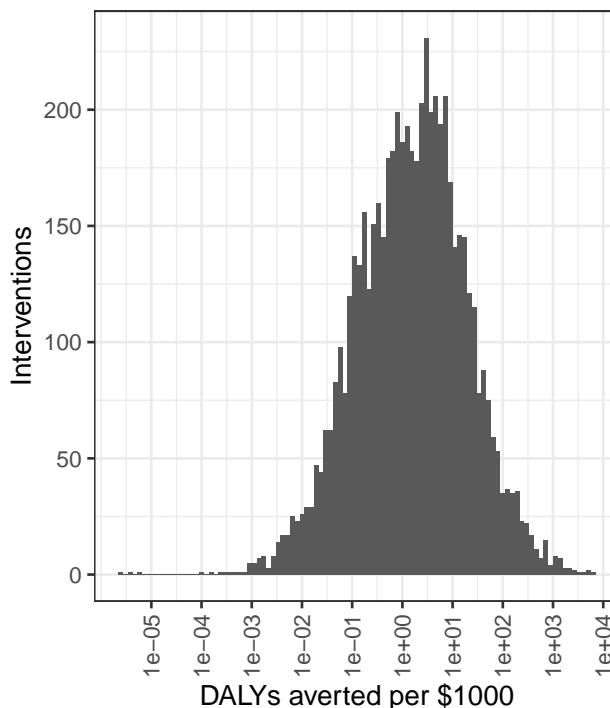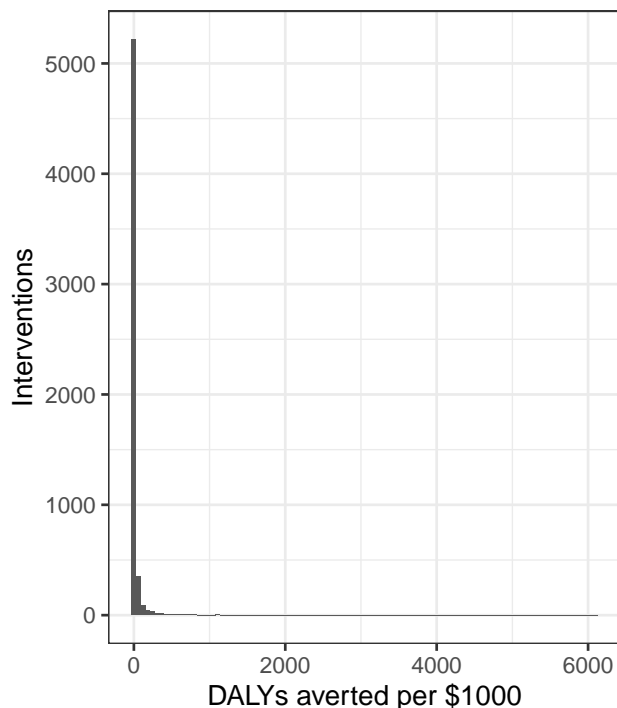


1. Let $Y_1, ..., Y_n$ be the cost-effectivenesses of the $n$ interventions. Suppose we are interested in the ratio of the cost effectiveness of the 95th percentile intervention to the median intervention (call this ratio $\theta$). Assuming the data are i.i.d. from a Log-Normal distribution with parameters $\mu$ and $\sigma^2$, find the MLE $\hat{\theta}$.

Since the $Y_i$ are Log-Normal, $X_i = \log(Y_i) \sim \mathcal{N}(\mu, \sigma^2)$. A ratio on the original scale corresponds to a difference on the log scale, and the log of a quantile is the quantile of a log, so

$$\log(\theta) = \log(Q_{Y_i}(0.95)) - \log(Q_{Y_i}(0.5)) = (\sigma Q_Z(0.95) + \mu) - (\sigma Q_Z(0.5) + \mu) = \sigma Q_Z(0.95)$$

where $Q_Z$ is the quantile function for a standard Normal. Therefore, by invariance, the MLE is

$$\hat{\theta} = \exp\left(\hat{\sigma} Q_Z(0.95)\right) \text{ with } \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\log(Y_i) - \overline{\log(Y)}\right)$$

```
exp(sqrt(var(CEA$logDalyPerThousand) * (nrow(CEA) - 1)/(nrow(CEA))) * qnorm(0.95))
```

```
## [1] 62.83007
```

In the dataset, this is 62.8, so the 95th percentile intervention is estimated to be 62.8 times as effective as the median intervention.

2. Find the asymptotic distribution for $\hat{\theta}$ and use this to approximate the standard error of the MLE. Start from the exact distribution for $\hat{\sigma}^2$ and use the representation of a $\chi_n^2$ random variable as the sum of $n$ squared standard Normals to write a CLT statement.

From the distribution of the sample variance in a Normal model, $\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-1}^2$. Since a $\chi_n^2$ can be represented as a sum of $n$ Normal random variables squared (each of which have mean 1 and variance 2 by the variance of a $\chi_1^2$ random variable), if $X_n = \sum_{i=1}^{n} Z_i^2$ for $Z_i$ i.i.d. standard Normals, the Central Limit Theorem gives

$$\sqrt{n}(X_n/n - 1) \to \mathcal{N}(0,2) \implies \sqrt{n-1}(\hat{\sigma}^2/\sigma^2 - 1) \to \mathcal{N}(0,2) \iff \sqrt{n}(\hat{\sigma}^2 - \sigma^2) \to \mathcal{N}(0, 2\sigma^4)$$

Then, with $g(x) = \exp\left(\sqrt{x}Q(0.95)\right)$ and $g'(x) = \frac{Q(0.95)}{2\sqrt{x}} \exp\left(\sqrt{x}Q(0.95)\right)$ the Delta Method gives

$$\sqrt{n}(\hat{\theta} - \theta) \to \mathcal{N}\left(0, 2\sigma^4 \left(\frac{Q(0.95)}{2\sigma} \exp\left(\sigma Q(0.95)\right)\right)^2\right)$$

As an approximation, this gives

$$\hat{\theta} \sim \mathcal{N}\left(\theta, \frac{2\sigma^4 \left(\frac{Q(0.95)}{2\sigma} \exp\left(\sigma Q(0.95)\right)\right)^2}{n}\right) \implies \text{SE}(\hat{\theta}) \approx \frac{\sqrt{2}\hat{\sigma}^2 \left(\frac{Q(0.95)}{2\hat{\sigma}} \exp\left(\hat{\sigma} Q(0.95)\right)\right)}{\sqrt{n}}$$

3. Explain how to approximate the standard error of $\hat{\theta}$ using a non-parametric bootstrap.

Take many samples of size $n$ from the original distribution with replacement. For each, compute $\exp\left(\hat{\sigma}^* Q_Z(0.95)\right)$. Return the standard deviation of these draws.

4. Explain how to approximate the standard error of $\hat{\theta}$ using a non-parametric bootstrap.

Take many samples of size $n$ from a Log-Normal distribution with parameters $\hat{\mu}$ and $\hat{\theta}$. For each, compute $\exp\left(\hat{\sigma}^* Q_Z(0.95)\right)$. Return the standard deviation of these draws.

5. Explain how to approximate the standard error of $\tilde{\theta} = \hat{Q}(0.95)/\hat{Q}(0.5)$ using a non-parametric bootstrap (the quantile is on the original scale).

Take many samples of size $n$ from the original distribution with replacement. For each, compute the sample 95th quantile and median and record the ratio. Return the standard deviation of these draws.

6. Explain how to approximate the standard error of $\tilde{\theta}$ with a parametric bootstrap.

Take many samples of size $n$ from a Log-Normal distribution with parameters $\hat{\mu}$ and $\hat{\theta}$. For each, compute the sample 95th quantile and median and record the ratio. Return the standard deviation of these draws.

7. Compare the standard errors from 2-6.

```
# First two shown as an example
n <- nrow(CEA)

# 2. Theta hat approximate SE
se_2 <- sqrt(2) * var(CEA$logDalyPerThousand) * qnorm(0.95) /
  (2 * sd(CEA$logDalyPerThousand)) *
  exp(sd(CEA$logDalyPerThousand) * qnorm(0.95)) / (sqrt(n))

# 3. Theta hat non-parametric SE
nboot <- 10^4
one_boot <- function() {
  samples <- sample(CEA$logDalyPerThousand, n, replace = T)
  return(exp(sqrt(var(samples) * (n - 1)/(n)) * qnorm(0.95)))
}
boots <- replicate(nboot, one_boot())
se_3 <- sd(boots)
```

```
##   Theta hat Theta tilde
##      62.830      50.055
```

| | |
|---|---|
| Theta hat SE approximate | 2.411 |
| Theta hat SE non-parametric bootstrapped | 2.602 |
| Theta hat SE parametric bootstrap | 2.424 |
| Theta tilde Non-parametric bootstrap | 3.616 |
| Theta tilde parametric bootstrap | 4.532 |

The standard error for $\hat{\theta}$ is lower than the standard error for $\tilde{\theta}$ in all approximations, which makes sense because it is incorporating more information (the entire variance rather than just the upper tail). The standard errors for $\tilde{\theta}$ show a considerable difference, presumably because the parametric bootstrap sometimes generates very extreme values.

8. Explain how to construct an approximate 95% bootstrap confidence interval for $\theta$ with the percentile method on $\tilde{\theta}$.

Take many samples of size $n$ from the original distribution with replacement. For each, compute the sample 95th quantile and median and record the ratio. Find the 2.5th and 97.5th percentiles of the ratios and report this as the interval.

9. Explain how to construct an approximate 95% bootstrap confidence interval for $\theta$ with a Normal approximation and a bootstrap standard error for $\tilde{\theta}$. Use the scale with better asymptotic properties and then convert the interval if necessary.

Take many samples of size $n$ from the log of the original distribution with replacement. For each, compute the sample 95th quantile and median and record the difference. Find the standard deviation of these differences. Then construct an interval by adding and subtracting $Q_{\mathcal{N}(0,1)}(0.975)$ times the standard deviation to or from $\log(\tilde{\theta})$. Finally, exponentiate both ends of the confidence interval to transform to the original scale. Performing this all on the log scale and then transforming at the end will give better asymptotic properties because the difference in quantiles will be more Normal on the log scale.

10. Compare the bootstrap intervals from 7 and 8 including both scales for the Normal approximation.

```
nboot <- 10^4
one_boot <- function() {
  samples <- sample(CEA$DalyPerThousand, n, replace = T)
  return(quantile(samples, 0.95) / quantile(samples, 0.5))
}
```

```
boots <- replicate(nboot, one_boot())
perc_method <- quantile(boots, c(0.025, 0.975))

norm_approx_1 <- c(quantile(CEA$DalyPerThousand, 0.95) /
                    quantile(CEA$DalyPerThousand, 0.5) - sd(boots) * qnorm(0.975),
  quantile(CEA$DalyPerThousand, 0.95) / quantile(CEA$DalyPerThousand, 0.5) +
    sd(boots) * qnorm(0.975))

one_boot <- function() {
  samples <- sample(CEA$logDalyPerThousand, n, replace = T)
  return(quantile(samples, 0.95) - quantile(samples, 0.5))
}
boots <- replicate(nboot, one_boot())
norm_approx_2 <- exp(c(quantile(CEA$logDalyPerThousand, 0.95) -
                        quantile(CEA$logDalyPerThousand, 0.5) - sd(boots) * qnorm(0.975),
  quantile(CEA$logDalyPerThousand, 0.95) - quantile(CEA$logDalyPerThousand, 0.5) +
    sd(boots) * qnorm(0.975)))

df <- data.frame(round(rbind(perc_method, norm_approx_1, norm_approx_2), 2))
rownames(df) <- c("Percentile method", "Normal approximation original scale",
                  "Normal approximation log scale")
colnames(df) <- c("Lower", "Upper")
knitr::kable(df)
```

|                                     | Lower | Upper |
|-------------------------------------|-------|-------|
| Percentile method                   | 42.83 | 56.91 |
| Normal approximation original scale | 42.88 | 57.23 |
| Normal approximation log scale      | 43.35 | 57.79 |

All three bootstrapped intervals are very similar.

One thing we might be interested in is how cost-effectiveness differs by country. As in week 8, we can perform a Bayesian readjustment (Normal this time) to account for the number of studies in each country. The table below shows the posterior means for intervention cost-effectiveness in the top and bottom countries.

| County | Adjusted DALYs averted per $1000 | Studies |
|--------|----------------------------------|---------|
| Zimbabwe | 46.09 | 101 |
| Guinea | 17.68 | 23 |
| Benin | 14.04 | 22 |
| Papua New Guinea | 11.71 | 11 |
| Madagascar | 11.56 | 27 |
| Burkina Faso | 10.96 | 39 |
| Cameroon | 9.55 | 35 |
| Congo, Rep. | 9.26 | 13 |
| Korea, Dem. People's Rep. | 8.51 | 14 |
| Chad | 8.08 | 22 |

| County | Adjusted DALYs averted per $1000 | Studies |
|--------|----------------------------------|---------|
| Israel | 0.01 | 15 |
| Qatar | 0.022 | 2 |

| County | Adjusted DALYs averted per $1000 | Studies |
|---|---|---|
| Denmark | 0.03 | 17 |
| Australia | 0.032 | 236 |
| Netherlands | 0.034 | 30 |
| Iceland | 0.034 | 2 |
| Norway | 0.05 | 7 |
| United States | 0.065 | 31 |
| Brunei Darussalam | 0.065 | 3 |
| Malta | 0.068 | 2 |

11. The World Bank classifies countries with gross national income per capita below $1085 as low-income and countries with GNI per capitas above $13205 as high-income. Let $\mu_1$ and $\mu_2$ be the true mean cost-effectiveness of interventions in low and high income countries respectively. Perform a two-sample $t$-test for $H_0 : \mu_1 = \mu_2$ vs $H_a : \mu_1 \neq \mu_2$.

```
income <- read.csv("data/income.csv")[,c(1,2)]
CEA_with_income <- inner_join(CEA, income, by=c("Country"))
high_income <- CEA_with_income$DalyPerThousand[CEA_with_income$Income == "High"]
low_income <- CEA_with_income$DalyPerThousand[CEA_with_income$Income == "Low"]
t.test(high_income, low_income, var.equal = F, alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  high_income and low_income
## t = -5.6572, df = 876.13, p-value = 2.084e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -50.37053 -24.42224
## sample estimates:
## mean of x mean of y
##  1.209629 38.606010
```

The p-value is much less than 0.05, so we reject the null and conclude that there is a difference in cost-effectiveness interventions in low and high income countries.

12. Perform a permutation test for whether the distributions of cost-effectiveness of interventions are different between country income groups. Use the absolute difference in sample means as the test statistic.

```
nperm <- 10^5

run_perm <- function(x, y) {
  n <- length(x)
  joined_vec <- c(x,y)
  indices <- sample(1:length(joined_vec), n)
  abs(mean(joined_vec[indices]) - mean(joined_vec[-indices]))
}

perms <- replicate(nperm,
                   run_perm(CEA_with_income$DalyPerThousand[CEA_with_income$Income == "High"],
                       CEA_with_income$DalyPerThousand[CEA_with_income$Income == "Low"]))

mean(perms > abs(mean(CEA_with_income$DalyPerThousand[CEA_with_income$Income == "High"]) -
       mean(CEA_with_income$DalyPerThousand[CEA_with_income$Income == "Low"])))
```
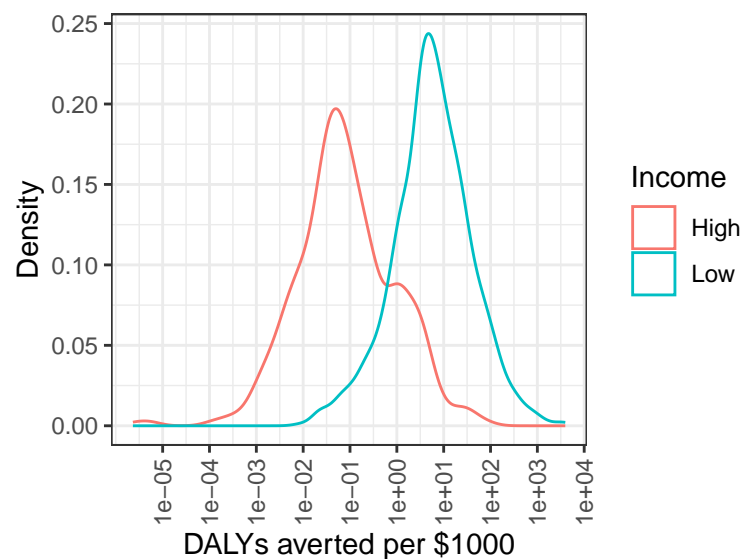
```
## [1] 0
```
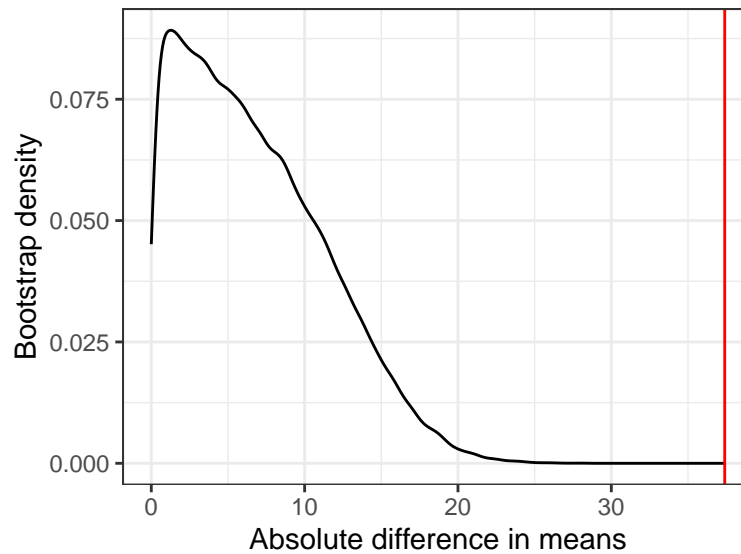


The p-value is approximately 0, and the observed difference is extremely unlikely under a null of equal distributions.

13. Compare the assumptions and conclusions for the test in 10 and the test in 11.



The assumptions of the $t$-test are that (1) the data are Normally distributed (or the sample sizes are large enough that the sample means are approximately Normal) and (2) the observations are i.i.d. within each group and are independent between groups. If using the original $t$-test, we further assume that the variances of the two groups are equal. In the permutation test, we still assume that the observations are i.i.d. within each group and independent between groups, but we do not assume the data are Normal. We can see that the two groups are approximately Normal on the log scale, so they will be extremely right skewed on the original scale.

The $t$-test null hypothesis is that the means are equal while the permutation null hypothesis is that the distributions are equal. Rejecting the $t$ null means concluding that the two groups have different means while rejecting the bootstrap null just means concluding that the distributions are not equal.

# Feedback

As part of my continual effort to improve my teaching, please take a moment to provide some feedback on section this semester.