Andrew J. Spieker, PhD
BIOS 6342 - Contemporary Statistical Inference
Lab #6: Size and power by simulation

Standard hypothesis testing: The Neyman-Pearson lemma and the Karlin-Rubin theorem inform us on how to construct certain kinds of optimal tests in specific settings. In the real world, the problems we encounter are often a bit more complex. Simulation methods can come in handy when seeking to evaluate the size and power of a procedure. This skill is useful when you're developing a medical study and want to evaluate the real-world operating characteristics of a proposed procedure under a hypothetical scenario. This lab is in some ways the flip side of the previous lab in that it is not intended to be theoretically rigorous; instead, it is to make sure we don't lose sight of real-world challenges and practical methods to address them.

Setup: Suppose we're preparing for a study of $n = 40$ independent observations in which we plan to measure some outcome, Y, and a binary exposure, X. We plan to model the association using a linear regression model:

$$Y = \beta_0 + \beta_1 X + \epsilon.$$

Here are things we anticipate *a priori*:

- We expect exactly half of the sample to be exposed ($X = 1$).
- We expect the error, ϵ, to be right-skewed with a standard deviation of 10.

We intend to use the following rule to decide whether to reject $H_0 : \beta_1 = 0$ in favor of $H_1 : \beta_1 \neq 0$:

$$\text{Reject } H_0 \text{ if } \left| \frac{\widehat{\beta}_1}{\widehat{\text{SE}}[\widehat{\beta}_1]} \right| > t_{0.975,40-2} = 2.024394.$$

If we have not already covered it in class, this is called a *Wald* test. It is asymptotically justified, but good performance in finite samples depends upon a number of criteria (such as...?).

<u>Exercise</u>: Consider generating $\epsilon \sim \text{Gamma}(\alpha, \beta) - \alpha/\beta$. This distribution has mean zero, variance α/β^2, and skewness $s = 2/\sqrt{\alpha}$. Examine the code below and make sense of what it's doing.

```r
## Function to generate errors accordingly
gen.epsilon <- function(skew, sd = 10, n = 40) {
  alpha <- 4/(skew^2)
  beta <- 2/(sd*skew)
  eps <- rgamma(n, alpha, beta) - alpha/beta
  return(eps)
}
```

You should be able to solve for α and β given the skewness and variance. You could even conceptualize reparameterizing this and naming it a "Gamma0(s, σ)" distribution!

Exercise: Fiddle around with the code below to see how the skewness parameter controls the shape of the error distribution.

```r
## Set seed for reproducibility
set.seed(6342)

## Generate error for illustration
eps <- gen.epsilon(skew = 1)
hist(eps)
```

Exercise: Investigate the code below and understand why it captures the salient parts of the data generating mechanism. Why do I effectively enforce $\beta_0 = 0$ without loss of generality? Your answer can be heuristic.

```r
gen.data <- function(n = 40, beta1, skew) {
  X <- c(rep(0, n/2), rep(1, n/2))
  epsilon <- gen.epsilon(skew)
  Y <- beta1 * X + epsilon
  dat <- data.frame(cbind(X, Y))
  return(dat)
}
```

Exercise: Investigate the code below and try to follow along with each step.

```r
simulation <- function(seed = 6342, M = 10000, n = 40, beta1, skew) {
  set.seed(seed)
  critical.value <- qt(0.975, df = 38)
  res <- matrix(0, nrow = M, ncol = 1)
  for (m in 1:M){
    dat <- gen.data(n, beta1, skew)
    zz <- lm(Y ~ X, data = dat)
    t.stat <- summary(zz)$coef[2,3]
    res[m,1] <- as.numeric(abs(t.stat) > critical.value)
    if (round(m/2500) == (m/2500)) {print(paste(m, "simulations complete!"))}
  }
  power <- mean(res)
  return(power)
}
```

Exercise: What is the size of this test under various choices for s?

Exercise: Let $\beta_1 = 6$ and $s = 0.001$. What is the power of this test?

Exercise: How does the power of this test vary with the skewness parameter?

Exercise: Let $s = 2$. What effect will you have 80% power to detect?

Exercise: Let $s = 0.001$? How do the size and power change if X is actually randomly assigned rather than known in advance?

Note: In practice, we care about additional properties such as bias, variability, and coverage. This set of exercises focuses on a narrow range of things we could reasonably care about.