1. I recently read Pete Buttigieg's book "Shortest Way Home", a beautiful biography discussing his life growing up and time as mayor of South Bend, Indiana. While most thought the book would focus on his sexuality (he was one of the first openly-gay men to run for the US presidency), instead he spends most of the book discussing the hard challenges he faced as mayor. In one section, he describes implementing a technology known as Shotspotter which listens for gun shots in communities and automatically dispatches police when it believes a shot has been fired. Naturally, the technology makes mistakes because slamming car doors and dropped objects might sound like a gun shot. Suppose you work at Shotspotter and run a bunch of tests to determine the accuracy of your technology, getting the below table.

Type 1: H_0 true, reject H_0

Type 2: H_1 true, keep H_0

	YES You fire a gun	NO You create gun-like sounds in creative ways
Shotspotter says "Bullet fired"	413	32
Shotspotter says "No bullet fired"	46	312
	459	344

Describe null and alternative hypotheses for this setup, explain what Type I and II errors would mean, and discuss the consequence of making each type of error if this technology were used in an actual city. Then, find the Type I and II error rates in this data set. Finally, discuss which type of error you *personally* think is worse *in your hometown* and explain why.

H_0: No shot fired

H_1: Shot fired

TYPE 1 ERROR: Shotspotter claims "shot fired" when in fact none were.

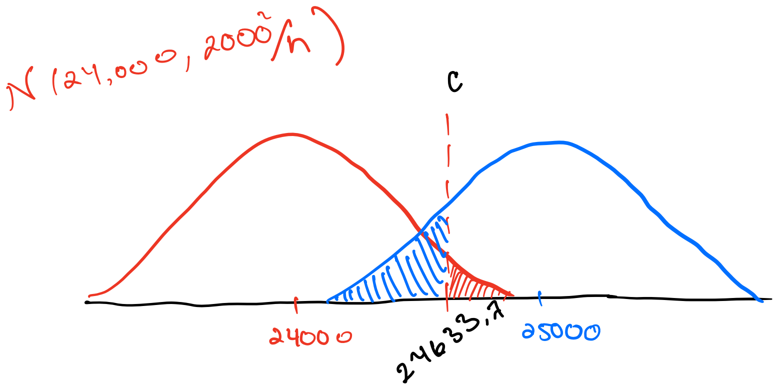
TYPE 2 ERROR: Shotspotter doesnt spot a true gun shot.

$P(\text{TYPE } 1) = P(\text{Reject } H_0, H_0 \text{ true}) = 46/459 = .10$

$P(\text{TYPE } 2) = P(\text{Keep } H_0 \mid H_1 \text{ true}) = 32/344 = .093$

If the Shotspotter is not used at all, (this would be equivalent to a failure to recognize a shot fired) guns will still be fired and someone may be injured regardless. While if the shotspotter is use and incorrectly calls the PD.... resources will be expended. → I vote Type 1 is WORSE.

2. Congratulations! You've just gotten a job at the most popular museum in America: The Air and Space Museum in Washington DC. Your first task as the resident statistician is to decide if a recent exhibit change has increased the average number of visitors per day. Before the change, the number of visitors per day was $N(24000, 2000^2)$. Your plan is to check attendance numbers on n random days in the next year, but need n to be as small as possible because it is costly and intrusive to count visitors. Your boss would be excited by a new daily average of 25000 (assume the spread is unchanged by the exhibit) and wants you to use $\alpha = 0.04$ in your test. If you demand a power of (at least) 0.85, what sample size should you use? You MUST include a picture with your answer. Feel free to use R/calculator to do some of the calculations. If you do, include your code/commands.



Red Curve: $P(\bar{x} > c) = .04$

$P\left(\frac{\bar{x} - 24000}{2000/\sqrt{n}} > \frac{c - 24000}{2000/\sqrt{n}}\right) = .04 \Rightarrow P\left(Z > \frac{c-24000}{2000/\sqrt{n}}\right) = .04$

$P(Z > z) = .04 \rightsquigarrow z = \frac{c - 24000}{2000/\sqrt{n}}$

z = qnorm(.04, lower=F)
1.751

$\boxed{1.751 = \frac{c - 24000}{2000/\sqrt{n}}} \rightarrow c = \frac{1.751(2000)}{\sqrt{n}} + 24000$

power $\geq .85 = 1 - \beta \longrightarrow \beta = .15$

Z = qnorm(.15, lower = T)
-1.036

$\boxed{-1.036 = \frac{c - 25000}{2000/\sqrt{n}}} \rightarrow c = \frac{-1.036(2000)}{\sqrt{n}} + 25000$

$\frac{1.751(2000)}{\sqrt{n}} + 24000 = \frac{-1.036(2000)}{\sqrt{n}} + 25000$

$\frac{2000}{\sqrt{n}}(1.751 + 1.036) = 1000$

$n = (2(1.751 + 1.036))^2$

$n = 31.071 \Rightarrow \boxed{n = 32}$

3. Every 4 years (or so), we get a leap day (2/29). This raises an interesting question: If you're born on leap day (a "leap-day baby") but we're in a non-leap year, would you rather celebrate your birthday on 2/28 or 3/1? On 2/29/2020, I was listening to NPR and a guest claimed that leap-day babies opt for the two options in equal proportion. Naturally, I doubted this (I wasn't sure if the preference would be for 2/28 or 3/1), and so let's think about a study you could conduct. Imagine we'll ask 500 random leap-day babies which day they prefer and record the percentage that choose 2/28. If the true percentage that opt for 2/28 is 49%, find the Type II error rate and power of our test assuming we use a significance level of 0.10. You must define a parameter, write hypotheses, and include a beautiful picture in your answer.

β $\quad$ $1-\beta$

Let p be proportion of leap-babies that opt for 2/28 celebration.

$H_0: p = 50\%$ $\quad$ $H_1: p \neq 50\%$

$$Z = \frac{\hat{p} - p}{\sqrt{pq/n}} \implies np,\ nq \geq 10$$

$\hookrightarrow 500(\tfrac{1}{2}) \geq 10$ ✓

$\alpha = .10$ $\quad$ $\alpha/2 = .05$

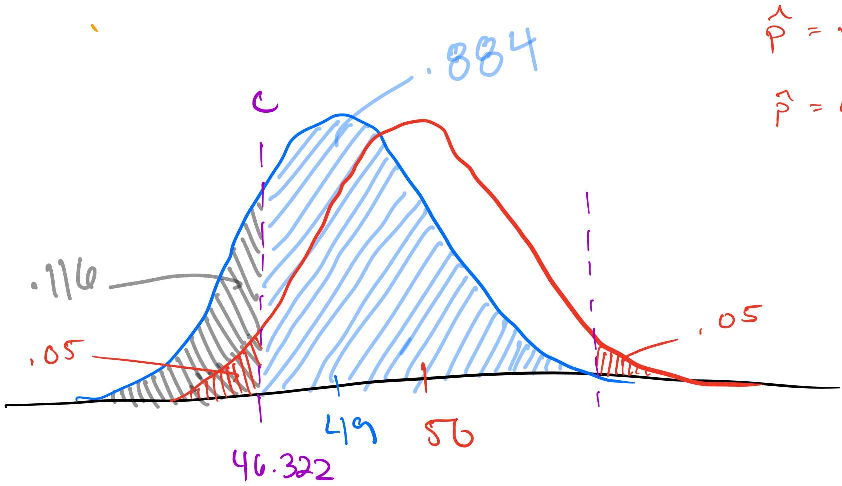
→ > qnorm(.05)

$-1.645 = $ Z-val

$$\frac{\hat{p} - p}{\sqrt{pq/n}} = -1.645$$

$$\hat{p} = \sqrt{pq/n}\,(-1.645) + p$$

$$\hat{p} = 46.322$$



$$\frac{46.322 - 49}{\sqrt{49(100-49)/n}} = -1.197$$

> pnorm(-1.197, lower = F)

.884

> pnorm(-1.197)

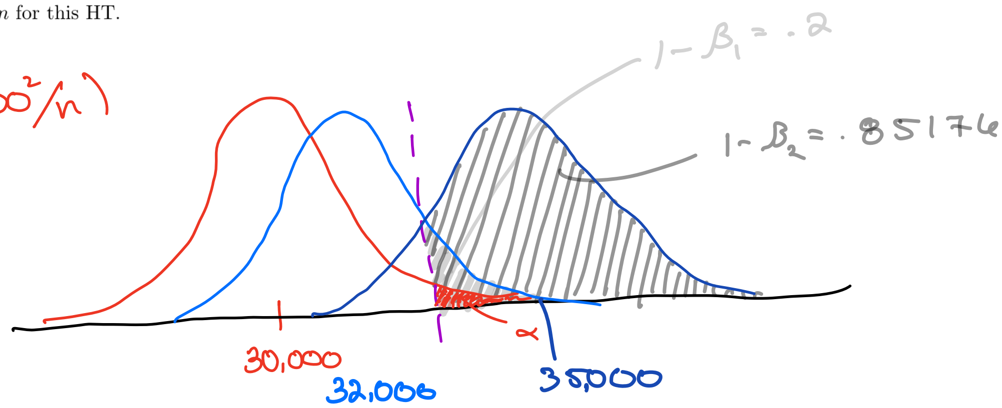
.116

$\implies \boxed{\beta \text{ (Type II error rate)} = .884}$

$\boxed{\text{power} = .116}$

4. This problem is inspired by one of my best students, who went on to get his PhD in materials science at MIT. If you look at car tires, they tend to average about 30000 miles before being replaced. In general, tire lifetimes are known to be normally distributed with SD 4500 miles. Now, my student claims to have a new manufacturing process that raises this average (he is right!) and keeps the spread the same. Let μ be the true average lifespan of tires with the new manufacturing process. When you draw a sample of size n, the power of this HT is 0.2 when $\mu = 32000$, and the power is 0.85176 when $\mu = 35000$. Find α and n for this HT.

$N(30000, 4500^2/n)$



$$z \leftarrow \underbrace{\frac{c - 30000}{4500/\sqrt{n}}}_{z} = z \Rightarrow \text{Pnorm}[z] = \alpha$$

$$\hookrightarrow c = \frac{z(4500)}{\sqrt{n}} + 30000$$

$$\frac{c - 32{,}000}{4500/\sqrt{n}} = 0.842 \qquad \beta_1 = 1 - .2 = .8 \quad > \text{qnorm}(.8)$$

$$0.842$$

$$\hookrightarrow c = \frac{0.842(4500)}{\sqrt{n}} + 32000$$

$$\frac{c - 35{,}000}{4500/\sqrt{n}} = -1.045 \qquad \beta_2 = 1 - .85176 = .148 \quad > \text{qnorm}(.148)$$

$$-1.045$$

$$\hookrightarrow c = \frac{-1.045(4500)}{\sqrt{n}} + 35000$$

$$\frac{.842(4500)}{\sqrt{n}} + 32000 = \frac{-1.045(4500)}{\sqrt{n}} + 35000 \Longrightarrow n = 8.012$$

$$c = \frac{.842(4500)}{\sqrt{8.012}} + 32000 \implies c = 33338.633$$

$$> \text{pnorm}\left(\left[\frac{33{,}338.633 - 30000}{4500/\sqrt{8.012}} \right], \text{lower} = F \right)$$

$$= .018$$

$$\boxed{\alpha = .018 \qquad n = 9}$$

5. In class, we constructed a power curve for a one-sided alternative hypothesis. Here, we construct a power curve for a two-sided alternative. We'll use question 3 as our setup, but this time, we'll vary the true percentage that support 2/28 (which was fixed at 49% in problem 3). Using R, construct a power curve for this setup. Include your code and a picture of the power curve. Your horizontal axis should go from 30 to 70, representing the most reasonable different values of the true 2/28 preference percentage. The vertical axis should be the power of the hypothesis test for each value on the horizontal axis. Label the lowest point on the curve. Also, make sure that the point (49%, your answer to question 3) is on the graph!