**Examining Type I Error Rates and Statistical Power**
In the lab today we will be examining the type I error rates and power for a hypothesis test.
Consider the GPA example in your class notes. The population is characterized by the following probability distribution:



Recall the population mean and variance are μ = 2.8571 and σ2 = 0.4082.

We will estimate the type I error rate of hypothesis tests when the population can be characterized by this distribution. Remember

* A type I error occurs when $H\_{0}$ is true, but the test we are using rejects $H\_{0}. $
* $α=$the significance level = the probability of a type 1 error
* If $H\_{0} $is true and the hypothesis testing procedure is repeated 1000 times (take a new sample and perform a new hypothesis test), we would expect approximately 1000×α of the hypothesis tests to incorrectly reject $H\_{0}$ (we expect $1000×α$ type 1 errors).

We will first simulate 1000 samples of size 20 from the given distribution for Y, and then perform hypothesis testing ($H\_{0}$: μ = 2.8571 vs.$H\_{a}$: μ ≠ 2.8571) for each sample using the confidence interval method. We will use α = 0.05.

> mu <- 2.8571

> n <- 20

> set.seed(1234)

> set1 <- 4\*matrix(data=rbeta(n = 1000\*n, shape1 = 5, shape2 = 2), nrow = 1000, ncol = n)

> means <- apply(X = set1, MARGIN = 1, FUN = mean)

> SDs <- apply(X = set1, MARGIN = 1, FUN = sd)

> lower <- means - qt(p=.975, df = 20 - 1) \* SDs/sqrt(n)

> upper <- means + qt(p=.975, df = 20 - 1) \* SDs/sqrt(n)

> count <- sum(lower < mu & upper > mu)

> count

[1] 946

> count/1000

[1] 0.946

> (1000 - count)/1000

[1] 0.054

* 1000 – 946 = 54 intervals do not contain μ = 2.8571, leading to rejection of $H\_{0}$.
* The estimated type I error rate is then 54/1000 = 0.054.
* Notice the code that we used above is essentially the same as when finding estimated confidence level in the previous lab.

**Lab Questions Part 1**

1. What is the estimated type one error rate for a sample size of 10?
2. What sample size less than 10 results in an estimated type I error rate of much higher than 0.05?

Next, we would like to use a much more skewed distribution and investigate what would happen to the type I error rate when the normal distribution assumption is violated for small samples. The probability distribution for the population is given as



The distribution is heavily skewed to the left and has population mean and variance: μ = 3.2787 and σ2 = 0.3331. You can use the following code for simulating samples from this distribution.

> mu <- 3.2787

> n <- 20

> N <- 10000

> set.seed(1111)

> set1 <- 4\*matrix(data = rbeta(n = N\*n, shape1 = 5, shape2 = 1.1), nrow = N, ncol =

 n)

**Lab Questions Part 2**

1. Calculate the estimated type I error rates for the test statistic based hypothesis test for n = 20 and 10. Use α = 0.05 for Ho: μ = 3.2787 vs. Ha: μ ≠ 3.2787.

**Recall** the test statistic is calculated as $t=\frac{ \overbar{y}-μ\_{0}}{s/√n}$ and H0 is rejected when $|t|>|t\_{α/2, n-1}|$

Now we will examine the power of a hypothesis test.

* In hypothesis testing, we assume that Ho is true and then try to prove it to be incorrect using evidence gathered in the sample.
* It is important to define P(Reject $H\_{0} $ | $H\_{a}$ is **TRUE**)
* This is called the power of the test.

Consider the general form of a left-tail hypothesis test

Ho: μ ≥ $µ\_{0}$
Ha: μ < $µ\_{0}$

 We reject Ho if t < -tα, n-1.

* Power analysis examines P(T < -tα,n-1) assuming that Ha is true.
* Ho is no longer true, therefore, μ0 is NOT the population mean and

$$T=\frac{\bar{Y}-μ\_{0}}{S/\sqrt{n}}$$

*no longer* has a t distribution with n – 1 degrees of freedom.

* Now, μa is the actual population mean.
* This leads to T having a non-central t distribution with n – 1 degrees of freedom and a “non-centrality” parameter of $\frac{\sqrt{n}(μ\_{a}-μ\_{0})}{σ}$.

Using α = 0.01 and sample size of 30, we will calculate the probability of rejecting the null hypothesis of Ho: μ ≥ $2.4$ if $µ\_{a}=2$, $µ\_{0}=2.4$, and σ2 = 0.8. (The $µ\_{a}$ and σ2 are based on mean and variance for a beta probability distribution with parameters a = 2 and b = 2)

> # sample size

> n <- 30

> # actual population mean

> # according to the given hypothesis this has to be less than

> # mu\_a

> pop\_mean <- 2

> # actual population standard deviation

> pop\_sd <- sqrt(0.8)

> # value of mu\_0 for the null hypothesis

> mu\_0 <- 2.4

> # effect size

> effect.size <- abs(mu\_0 – pop\_mean)

> # difference in means

> diff.in.means <- pop\_mean – mu\_0

> # Type 1 error rate

> alpha <- 0.01

> # calculate power using power.t.test

> power.t.test(n = n, d = effect.size, sd = pop\_sd, sig.level = alpha, type =

 “one.sample”, alternative = “one.sided”)

 One-sample t test power calculation

 n = 30

 delta = 0.4

 sd = 0.8944272

 sig.level = 0.01

 power = 0.5035215

 alternative = one.sided

> # critical value

> crit\_Val <- qt(p = 0.99, df = n – 1)

> # power is P(T < -t\_(n-1,alpha))

> 1 – pt(q = crit\_Val, df = n-1, ncp = sqrt(n) \*effect.size / pop\_sd)

[1] 0.5035215

The probability of rejecting the null hypothesis is 0.5. What do you think of this result?

* Often, power analysis is performed when planning a study to find the ideal sample size
* Suppose you have been asked to find a sample size for the above setting that will result in a power of at least 0.8.

**Lab Questions Part 3**

1. What is the sample size that results in a power of at least 0.8?
2. Can you give an intuitive reason as to why the power increased for the sample size you found?

Now we will look at how to verify the power that we got above using simulations. We obtained a power of 0.5, this means that if the sampling process were repeated 10000 times, we would expect the null hypothesis to be rejected approximately 5000 times.

> # simulate N samples of size n each

> # number of simulated samples

> N <- 10000

> # set a seed for reproducibility

> set.seed(4167)

> set1 <- 4\*matrix(data = rbeta(n = N\*n, shape1 = 2, shape2 = 2), nrow = N, ncol =

 n)

> # get the mean of each sample

> means <- apply(X = set1, MARGIN = 1, FUN = mean)

> # get the SD of each sample

> SDs <- apply(X = set1, MARGIN = 1, FUN = sd)

> # get the test statistic for each sample

> tstat <- ((means - mu\_0)/(SDs/sqrt(n)))

> # sum how many times we got an abs value of test statistic less than critical value

> sum(abs(tstat) > abs(crit\_Val))

[1] 4909

# divide this by the number of simulations to get the power

> sum(abs(tstat) > abs(crit\_Val))/N

[1] 0.4909

The estimated power is 0.49 which is very close to the power calculated earlier.