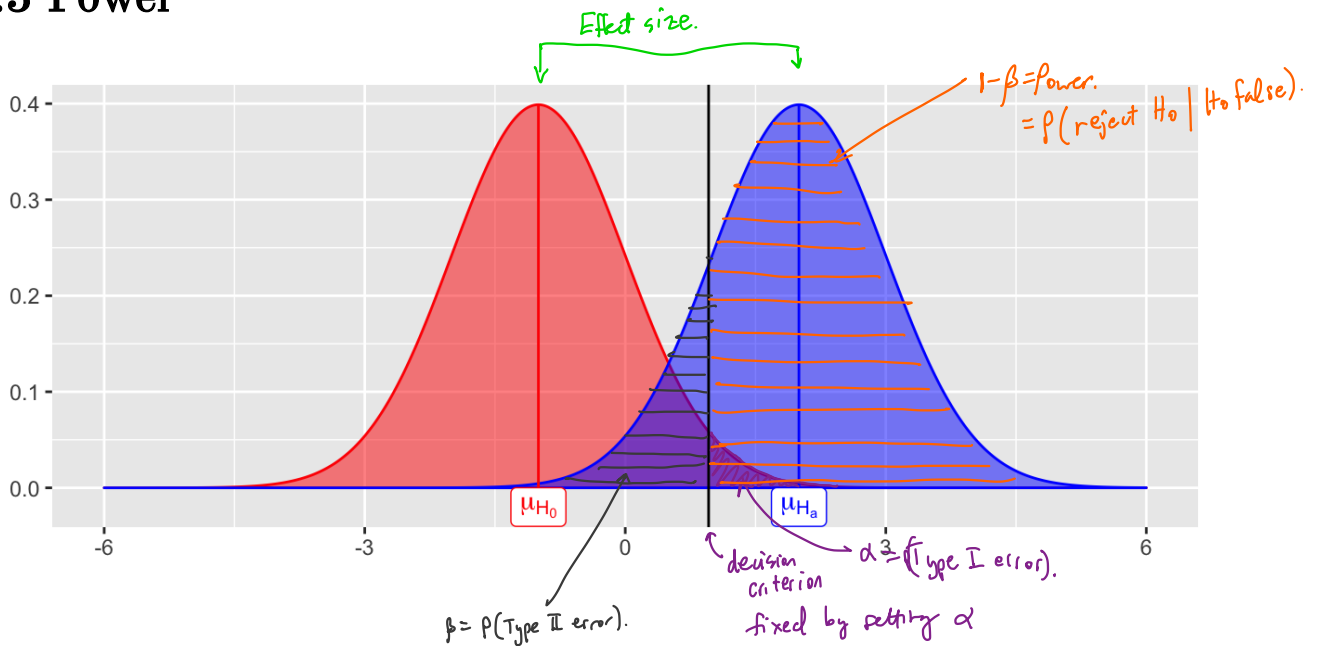


## 2.3 Power

$$P(\text{Type I error}) = P(\text{reject } H_0 \mid H_0 \text{ true}) = \alpha$$

$$P(\text{Type II error}) = P(\text{Fail To reject } H_0 \mid H_0 \text{ false}) = \beta.$$



Consider a hypothesis test about the parameter  $\theta$ :

$$H_0 : \theta = \theta_0$$

$$H_a : \theta > \theta_0$$

We let  $\beta = P(\text{fail to reject } H_0 \mid H_0 \text{ is false}) = P(\text{Type II error})$ , then Power =  $P(\text{reject } H_0 \mid H_0 \text{ is false}) = 1 - \beta$ .

Power depends on the distance between the hypothesized value of the parameter  $\theta_0$  and the actual value  $\theta_1$ , so we can write  $1 - \beta(\theta_1)$ .

Why is power important?  
 $\uparrow$  effect size.

1. If we have multiple statistical testing methods for the same hypothesis, choose test that has highest power.
2. If you are going to spend time/money to do an experiment, need to check beforehand that your study will be powerful enough to detect an effect.

For a few simple cases, you can derive a closed form expression of power.

All others: use Monte Carlo methods to estimate power.

**Example 2.4** Consider a one-sample z-test. Sample  $X_1, \dots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$ .  
 $\uparrow$  unknown       $\leftarrow$  known

$H_0: \mu = \mu_0$  vs.  $H_a: \mu > \mu_0$ .

Using statistic  $Z^* = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

We reject  $H_0$  if  $Z^* > Z_{1-\alpha}$   $\swarrow$  crit. value

If  $\mu_0 = 5$  (hypothesized value) but true mean  $\mu_1 = 6$ .

What is probability of correctly rejecting  $H_0: \mu = 5$ ? This is Power!

Effect size:  $\mu_1 - \mu_0 = 6 - 5 = 1$ . If effect size is 10, our test would have more power!  
 easier to detect the truth!

For z-test we can analytically derive power (Chihara & Hestenberg p. 229-230).

$$1 - \beta = P(\text{reject } H_0 \mid H_0 \text{ is false}) \\ = P\left(Z^* > Z_{1-\alpha} - \frac{(\mu_1 - \mu_0)}{\sigma/\sqrt{n}}\right)$$

So power is a function of smallest z-value where you will reject  $H_0$

1. Significance level: as  $\alpha \uparrow$ , power  $\uparrow$  (trade-off btw/ Type I and Type II error).
2. Effect size:  $\mu_1 - \mu_0$  as effect size  $\uparrow$ , power  $\uparrow$
3. Sample size: as  $n \uparrow$ , power  $\uparrow$
4. Variance: as variance  $\uparrow$ , power  $\downarrow$  (no control over this in practice).

Notes: ① as power  $= 1 - \beta \uparrow$ ,  $P(\text{type I error}) = \alpha \uparrow$ . For fixed  $n, \sigma^2, \mu_1 - \mu_0$  the only way to increase power is to  $\uparrow \alpha$ .

② Only way to simultaneously  $\uparrow$  power &  $\downarrow \alpha$  is  $\uparrow n$ .

## 2.4 MC Estimator of $1 - \beta$

Assume  $X_1, \dots, X_n \sim F(\theta_0)$  (i.e., assume  $H_0$  is true).

Then, we have the following hypothesis test –

$$H_0 : \theta = \theta_0$$

$$H_a : \theta > \theta_0$$

and the statistics  $T^*$ , which is a test statistic computed from data. Then we **reject**  $H_0$  if  $T^* >$  the critical value from the distribution of the test statistic.

This leads to the following algorithm to estimate the power of the test ( $1 - \beta$ )

- ① Select model, setup hypothesis test.
- ② Select value of alternative  $\theta_1$
- ③ Set  $n$ , other param values (e.g.  $\sigma^2$ ), and  $\alpha$
- ④ For each  $j = 1, \dots, M$ 
  - a) Sample  $X_1^{(j)}, \dots, X_n^{(j)}$  from model under alternative hypothesis  $\theta = \theta_1$
  - b) Compute  $T^{*(j)}$  based on data from (a)
  - c) Compute  $y_j = \mathbb{I} \{ \text{reject } H_0 \text{ based on } T^{*(j)} \}$ .
- ⑤ Compute  $1 - \hat{\beta} = \frac{1}{M} \sum_{j=1}^M y_j$  (i.e. count # of correct answers).

## Your Turn

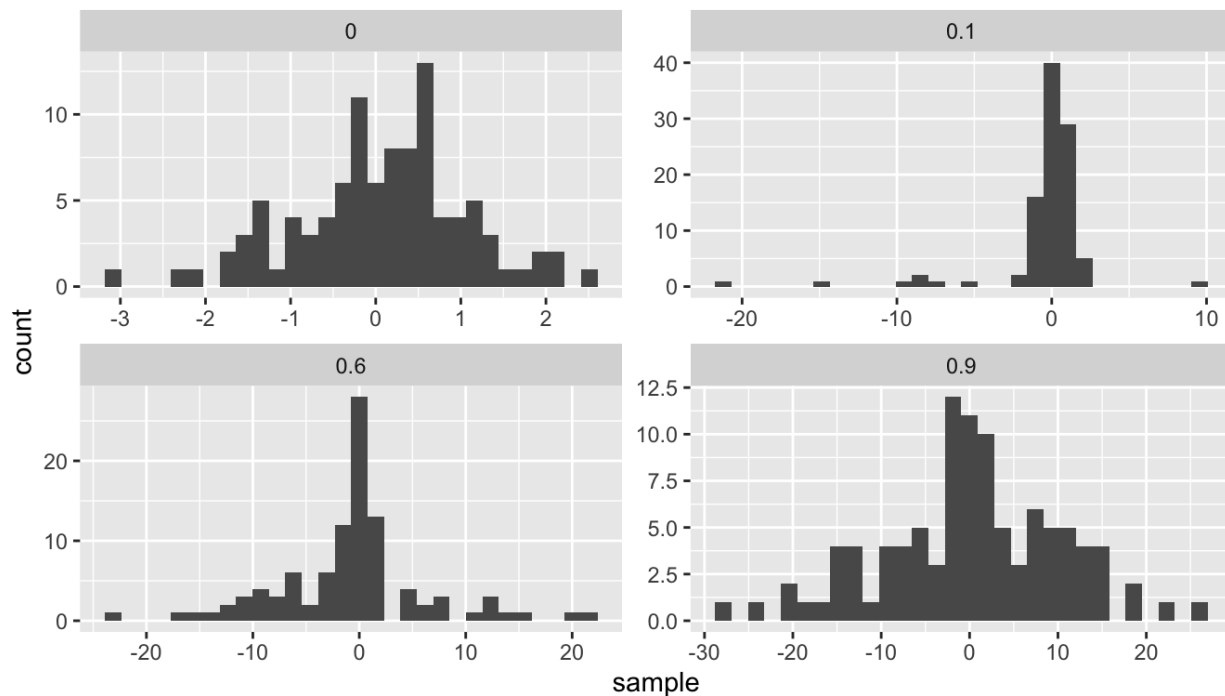
Consider data generated from the following mixture distribution:

$$f(x) = (1 - \epsilon)f_1(x) + \epsilon f_2(x), \quad x \in \mathbb{R}$$

where  $f_1$  is the pdf of a  $N(0, 1)$  distribution,  $f_2$  is the pdf of a  $N(0, 100)$  distribution, and  $\epsilon \in [0, 1]$ .

```
r_noisy_normal <- function(n, epsilon) {
  z <- rbinom(n, 1, 1 - epsilon)
  z*rnorm(n, 0, 1) + (1 - z)*rnorm(n, 0, 10)
}

n <- 100
data.frame(e = 0, sample = r_noisy_normal(n, 0)) %>%
  rbind(data.frame(e = 0.1, sample = r_noisy_normal(n, 0.1))) %>%
  rbind(data.frame(e = 0.6, sample = r_noisy_normal(n, 0.6))) %>%
  rbind(data.frame(e = 0.9, sample = r_noisy_normal(n, 0.9))) %>%
  ggplot() +
  geom_histogram(aes(sample)) +
  facet_wrap(~e, scales = "free")
```



We will compare the power of various tests of normality. Let  $F_X$  be the distribution of a random variable  $X$ . We will consider the following hypothesis test,

$$H_0 : F_x \in N \quad \text{vs.} \quad H_a : F_x \notin N,$$

*i.e.  $H_0$  says  $X$  is Normally distributed*

*$H_a$  says it isn't.*

where  $N$  denotes the family of univariate Normal distributions.

Recall Pearson's moment coefficient of skewness (See Example 2.2).

*and corresponding skewness test*

*$H_0 : \sqrt{\beta_1} = 0$   
 $H_a : \sqrt{\beta_1} \neq 0.$*

We will compare Monte Carlo estimates of power for different levels of contamination ( $0 \leq \epsilon \leq 1$ ). We will use  $\alpha = 0.1$ ,  $n = 100$ , and  $m = 100$ .

```
# skewness statistic function
skew <- function(x) {
  xbar <- mean(x)
  num <- mean((x - xbar)^3)
  denom <- mean((x - xbar)^2)
  num/denom^1.5
}

# setup for MC
alpha <- .1
n <- 100
m <- 100
epsilon <- seq(0, 1, length.out = 200)
var_sqrt_b1 <- 6*(n - 2)/((n + 1)*(n + 3)) # adjusted variance for
  skewness test
crit_val <- qnorm(1 - alpha/2, 0, sqrt(var_sqrt_b1)) #crit value for
  the test
empirical_pwr <- rep(NA, length(epsilon)) #storage

# estimate power for each value of epsilon
for(j in 1:length(epsilon)) {
  # perform MC to estimate empirical power
  ## Your turn

}

## store empirical se
empirical_se <- "Your Turn: fill this in"

## plot results --
## x axis = epsilon values
## y axis = empirical power
## use lines + add band of estimate +/- se
```

Compare the power with  $n = 100$  to the power with  $n = 10$ . Make a plot to compare the two for many values of  $\epsilon$ .