

# hw9-report

Jack McShane

2022-04-23

**1. Let  $Y_1, Y_2, \dots, Y_n \sim_{iid} N(\theta, \sigma^2)$  where  $\sigma^2$  is known. Consider  $H_0 : \theta \leq \theta_0$  vs.  $H_1 : \theta > \theta_0$ .**

**(a) What would be Type I Error? What would be Type II Error?**

Type I error involves selecting or accepting the wrong hypothesis when the evidence does not support it. An example of this kind of error would be a company pushing the idea that their drug is better than another even though the evidence supports the opposite conclusion.

Type II error involves selecting the wrong hypothesis due to a lack of data (or if the data is not representative). If there is not enough data to support the new hypothesis even though it is the correct one, our test procedure could lead one to select the incorrect hypothesis.

**(b) In HW8, we show that the likelihood ratio test procedure is to reject  $H_0$  if**

$$z = \frac{\bar{y} - \theta_0}{\sigma/\sqrt{n}} \geq k_1 = \sqrt{-2\log(k)}$$

**The power function of this test is:**

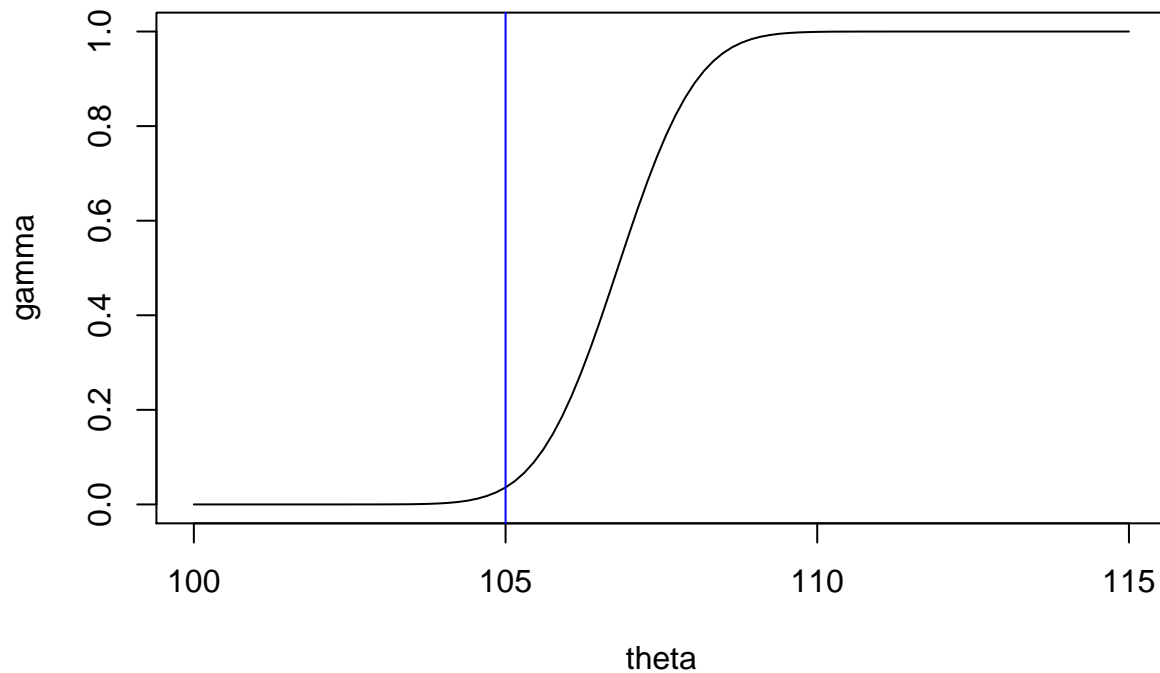
$$\gamma(\theta) = P(N(0,1) \geq k_1 + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}})$$

**Let  $\theta_0 = 105$ ,  $\sigma = 10$ ,  $n = 100$  and  $k_1 = 1.8$ , plot this function, and comment on your plot.**

```
theta0 <- 105
sigma <- 10
n <- 100
k1 <- 1.8
ntheta <- 100

theta <- seq(100, 115, length=ntheta)
gamma <- 1 - pnorm((theta0 - theta)/(sigma/sqrt(n)) + k1)
```

```
plot(theta, gamma, type='l')
abline(v=105, col='blue')
```



The curve takes the general shape of a power function curve. The significance level that was used in generating the graph has allowed us to guarantee a rather low probability of type I error when selecting our hypothesis (of course this leads to a higher probability of type II error when rejecting  $H_0$  on the other side of  $\theta_0$ ). We have a relatively high number of samples which gives the curve a rather steep increase in power value following the  $\theta_0$  value.

**(c) For this test, what is the probability of Type I Error when  $\theta = 105$ ?**

```
typeI <- 1 - pnorm((theta0 - 105)/(sigma/sqrt(n)) + k1)
```

```
## [1] "Probability of Type I Error:"
```

```
## [1] 0.03593032
```

(d) For this test, what is the probability of Type II Error when  $\theta = 110$ ? What is the power of rejecting  $H_0$  when  $\theta = 110$ ?

```
gamma <- 1 - pnorm((theta0 - 110)/(sigma/sqrt(n)) + k1)
typeII <- 1 - gamma
```

```
## [1] "Probability of Type II Error:"
```

```
## [1] 0.0006871379
```

```
## [1] "Power of rejecting Ho:"
```

```
## [1] 0.9993129
```

(e) If we set the significance level  $\alpha = 0.05$ , what is  $k_1$ ?

```
alpha <- .05
k1 <- qnorm(1 - alpha)
```

```
## [1] "Value of k1:"
```

```
## [1] 1.644854
```

(f) For this test procedure with  $\alpha = 0.05$ , what sample size  $n$  is necessary to ensure that the power of rejecting  $H_0$  at  $\theta = 108$  is at least 80%?

$$\begin{aligned} .80 &= \gamma(108) = P(N(0, 1) \geq \frac{105 - 108}{10/\sqrt{n}} + k_1) \\ .80 &= P(N(0, 1) \geq -3\sqrt{n}/10 + k_1) \\ qnorm(1 - .80) &= -3\sqrt{n}/10 + k_1 \\ \Rightarrow n &\geq \left(\frac{10}{3}(qnorm(1 - .80) - k_1)\right)^2 \\ \Rightarrow n &\geq \left(\frac{10}{3}(qnorm(1 - .80) - qnorm(1 - \alpha))\right)^2 \end{aligned}$$

```
alpha <- .05
pwr <- .80
n <- ((qnorm(1 - pwr) - qnorm(1 - alpha)) * 10 / 3)^2
```

```
## [1] "Minimum sample size:"
```

```
## [1] 68.69508
```

We would therefore need at least 69 samples for this assurance.

## 2. Consider the carprice example from ‘Notes 7’.

(a) What is the  $\hat{\beta}_1$ ? How do you interpret this number?

$\hat{\beta}_1$  is an estimator for the true  $\beta_1$ .  $\beta_1$  is the estimated parameter which determines the contribution of our observed feature (age) to the car price prediction, and which minimizes the least-squared error calculation.  $\hat{\beta}_1 \approx -9.5$ . This means that for every year the car gets older, its value drops approximately \$950 as long as the mileage on the car does not change.

(b) To test  $H_0 : \beta_1 = 0$  vs  $H_1 : \beta_1 \neq 0$ , what is the P-value? What is your conclusion?

From the output, we can see that the p-value for  $\hat{\beta}_1$  is .0397. Assuming  $\alpha = .05$ , we can conclude that the  $\hat{\beta}_1$  coefficient is significant in the prediction.

(c) Based on this output, what is the prediction of the average car price for a 3-year-old car with mileage of 25,000?

```
pred <- 183.0352 - 9.5043*3 - .8215*25
pred
```

```
## [1] 133.9848
```

## 3. Consider the Default example from ‘Notes 8’.

(a) What is the  $\hat{\beta}_1$ ? How do you interpret this number?

$\hat{\beta}_1$  is the coefficient associated with whether or not a particular person is a student. It is an estimate for the true  $\beta_1$  which scales the contribution of student status to the probability of default. Using likelihood ratios, we can determine the effect that a change in the input of whether or not the person is a student has on said person defaulting. The ratio evaluates to  $\approx .5$ . In combination with the negative value of the coefficient itself, we can determine that the student status has a negative correlation with the probability of default and that the transition from non-student to student decreases the probability of default by roughly half.

(b) What is your prediction of Default for someone who is a student, with balance of 800 and income of 15,000?

```
pred <- 1 / ( 1 + exp( -(-1.087e1 -6.468e-1 + (5.737e-3)*(800) + (3.033e-6)*(15000))))
```

```
## [1] "Predicted value:"
```

```
## [1] 0.00102534
```

The person is unlikely to default because the probability is below 50%.

(c) To test  $H_0 : \beta_1 = 0$  vs  $H_1 : \beta_1 \neq 0$ , what is the p-value? What is your conclusion?

From the output, we can see that the p-value for  $\hat{\beta}_1$  is .00619. Assuming  $\alpha = .05$ , we can conclude that the  $\hat{\beta}_1$  coefficient is significant in the prediction.