Lognormal Mean Testing

Ryan Lafferty

2/23/2020

Let $Y_1, ..., Y_n$ be a sample from a normal distribution $N(\mu, \sigma^2)$ and let $\psi(\mu, \sigma) = a\mu + b\sigma^2$ for constants $a, b \in \mathbb{R}$. We are interested in testing hypotheses of the form $H_0 : \psi(\mu, \sigma) = \psi_0$ vs $H_1 : \psi(\mu, \sigma) \neq \psi_0$.

Consider the estimator $\hat{\psi} = a\overline{Y} + bS_y^2$, where $\overline{Y}$ is the sample mean and S_y^2 is the (unbiased) sample variance. Then

$$E(\hat{\psi}) = aE(\overline{Y}) + bE(S_y^2) = a\mu + b\sigma^2,$$

so $\hat{\psi}$ is an unbiased estimator of ψ.

We can compute the variance of $\hat{\psi}$ as follows:

$$\begin{aligned} V(\hat{\psi}) &= V(a\overline{Y} + bS_y^2) \\ &= a^2 V(\overline{Y}) + b^2 V(S_y^2) \\ &= a^2 \frac{\sigma^2}{n} + b^2 V\left(\frac{\sigma^2}{n-1}\chi^2_{n-1}\right) \\ &= a^2 \frac{\sigma^2}{n} + 2b^2 \frac{\sigma^4}{(n-1)} \end{aligned}$$

Next we look for an unbiased estimator $\hat{V}$ of the variance $V = V(\hat{\psi})$. We already know that the variance of S_y^2 is $2\frac{\sigma^4}{(n-1)}$ and its mean is σ^2. Thus we can write

$$E(S_y^4) = [E(S_y^2)]^2 + V(S_y^2) = \sigma^4 + 2\frac{\sigma^4}{(n-1)}.$$

It is sensible to try a linear combination of S_y^2 and S_y^4: Let $\hat{V} = pS_y^2 + qS_y^4$. Then we have

$$\begin{aligned} E(\hat{V}) &= p\sigma^2 + q\left[1 + \frac{2}{n-1}\right]\sigma^4 \\ &= p\sigma^2 + q\left[\frac{n+1}{n-1}\right]\sigma^4 \\ &= a^2\frac{\sigma^2}{n} + 2b^2\frac{\sigma^4}{(n-1)} \end{aligned}$$

This results in $p = \frac{a^2}{n}$ and $q = \frac{2b^2}{n+1}$. Thus we see that

$$\hat{V} = \frac{a^2}{n}S_y^2 + \frac{2b^2}{n+1}S_y^4$$

is an unbiased estimator of V.

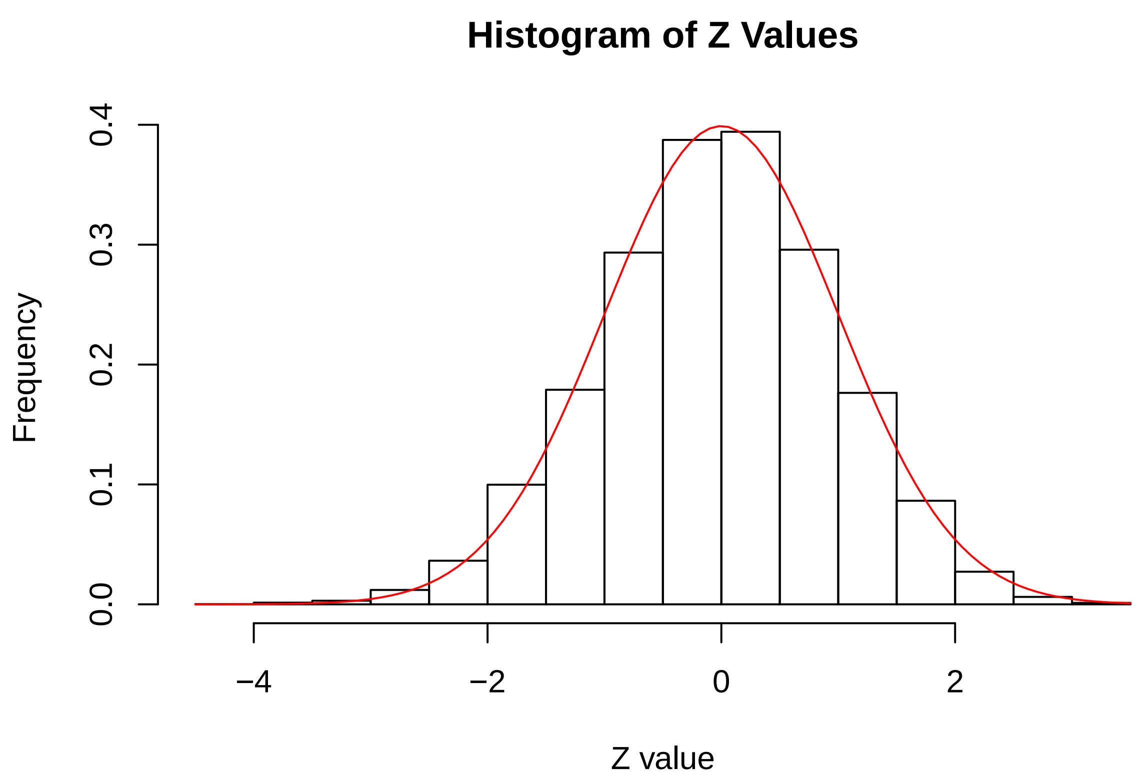
Now, for testing $H_0 : \psi = \psi_0$, we consider the test statistic Z given by:

$$Z := \frac{\hat{\psi} - \psi_0}{\sqrt{\hat{V}}}.$$

We can implement this in R as follows:

```
getZ <- function(data,psi0 = 0,a = 1, b=1/2){
  ybar <- mean(data)
  sy2 <- var(data)
  n <- length(data)
  psihat <- a * ybar + b * sy2
  vhat <- (a^2/n)* sy2 + (2*b^2/(n+1))*sy2^2
  Z <- (psihat - psi0)/sqrt(vhat)
  return(Z)
}
```

Suppose $\mu = 1$ and $\sigma^2 = 4$. We can generate some values of Z and plot a histogram to observe the distribution of Z.



A standard normal curve (in red) has been overlaid to show that the Z values closely follow a normal distribution. The upper and lower 5% cutoff values for the standard normal distribution are given by

$$Z_{.05/2} = \text{-}1.96$$

and

$$Z_{1-.05/2} = 1.96.$$

Using these we can construct an $\alpha = 5\%$ level confidence interval for ψ, whose bounds are given by:

$$\hat{\psi} \pm 1.96\sqrt{\hat{V}}.$$

In our example this confidence interval is computed as follows:

```r
#Generate the data
data <- rnorm(n,mean=1,sd=2)

a <- 1
b <- 1/2
ybar <- mean(data)
sy2 <- var(data)
n <- length(data)
psihat <- a * ybar + b * sy2
vhat <- (a^2/n)* sy2 + (2*b^2/(n+1))*sy2^2

#Compute the confidence interval
CI <- psihat + c(-1,1)* 1.96 * sqrt(vhat)
print(paste("Confidence Interval = [",CI[1],",",CI[2],"]"),quote=FALSE)
```

```
## [1] Confidence Interval = [ 2.84567205025805 , 3.26828021630341 ]
```

Indeed, this interval contains the value $\psi_0 = a(1) + b\left(\frac{1}{2}\right) = 3$.