stats100b_hw6

Takao

5/23/2022

1

$$H_0 : p = 0.2$$
$$H_1 : p > 0.2$$

```
p <- 24/100
e.s.e <- sqrt(0.2*(1-0.2)/100)
z_stats <- (p-0.2) / e.s.e
cat("p-value = ", pnorm(z_stats, 0, 1, lower.tail = FALSE))
```

```
## p-value =  0.1586553
```

Since the p-value is greater than 0.05 we do not have significant evidence to reject the null hypothesis

For the second part of the question,

$$P(\hat{p} > C|p = 0.2) = \alpha$$

$$P(Z > \frac{C - 0.2}{e.s.e}) = 0.05$$

```
c <- qnorm(0.05, 0, 1, lower.tail = FALSE)*e.s.e+0.2
cat("C  = ",c)
```

```
## C  =  0.2657941
```

$$P(\hat{p} > C|0.3) = 1 - \beta$$
$$P(Z > \frac{C - 0.3}{e.s.e}) = 1 - \beta$$

```
e.s.e <- sqrt(0.3*(1-0.3)/100)
cat("beta = ", 1-pnorm((c - 0.3) / e.s.e, 0, 1, lower.tail = FALSE))
```

```
## beta =  0.227703
```

2

We state the hypotheses

$$H_0 : \mu = 12000$$
$$H_1 : \mu > 12000$$

```r
z_stats <- (12500 - 12000) / (2000/sqrt(30))
cat("p-value = ", pnorm(z_stats, 0, 1, lower.tail = FALSE))
```

```
## p-value =  0.08545176
```

Since the p-value is greater than 0.05, we fail to reject the null hypothesis

3

Ratio lambda is found through

$$\frac{0.1}{0.2} = \frac{1}{2}, \frac{0.4}{0.3} = \frac{4}{3}, \frac{0.1}{0.3} = \frac{1}{3}, \frac{0.1}{0.1} = 1, \frac{0.3}{0.1} = 3$$

The rejection priority by ranking from the magnitude above is is

$$4, 2, 5, 3, 1$$

a

The most powerful test at alpha = 0.1 rejects the null when lambda = 0.3/0.1

$$\alpha = P(x_5|H_0) = 0.1$$

The power of this test is

$$P(x_5|H_1) = 0.3$$

b

The most powerful test at alpha = 0.4 rejects the null when lambda is greater than or equal to 0.4/0.3

$$\alpha = P(x_5|H_0) + P(x_2|H_0) = 0.1 + 0.3 = 0.4$$

The power of this test is

$$P(x_5|H_1) + P(x_2|H_1) = 0.3 + 0.4 = 0.7$$
$$\beta = 1 - 0.7 = 0.3$$

4

$$\Lambda = \frac{\Pi_i f(x_i|H_1)}{\Pi_i f(x_i|H_0)} = \frac{\lambda_1^n \exp(\lambda_1 \sum_i |x_i|)}{\lambda_0^n \exp(-\lambda_0 \sum_i |x_i|)} = (\frac{\lambda_1}{\lambda_0})^n \exp[(\lambda_0 - \lambda_1) \sum_i |x_i|]$$

$$log(\Lambda) = nlog(\frac{\lambda_1}{\lambda_0}) + (\lambda_0 - \lambda_1) \sum_i |x_i|$$

Since lambda1 is greater than lambda0, we reject the null hypothesis when

$$-\sum_i |x_i| > c$$

Since lambda0 and lambda1 are arbitrary, and the test is most powerful for any lambda1 greater than lambda0, it is uniformly most powerful for testing

$$H_0 : \lambda = \lambda_0$$

$$H_1 : \lambda > \lambda_1$$

5

The test with the maximum power at a certain significance level is the likelihood ratio test.

$$\Lambda = \frac{f(x|H_1)}{f(x|H_0)} = \frac{2x}{1} = 2x$$

The definition of a significance level alpha:

$$\alpha = P(X > c|H_0) = \int_c^1 f_0(x)dx = x|_c^1 = 1 - c$$

$$1 - c = 0.1$$

$$c = 0.1$$

The definition of power 1-beta is

$$1 - \beta = P(X > c|H_1) = \int_{0.9}^1 f_1(x)dx = x^2|_{0.9}^1 = 1 - 0.9^2 = 0.19$$

6

a

Denote lambda hat as the MLE of the exponential distribution

$$\Lambda = \frac{\Pi_i f(x_i|H_1)}{\Pi_i f(x_i|H_0)} = \frac{\hat{\lambda}^n \exp(\lambda_1 \sum_i |x_i|)}{\lambda_0^n \exp(-\lambda_0 \sum_i |x_i|)} = (\frac{\hat{\lambda}}{\lambda_0})^n \exp[(\lambda_0 - \hat{\lambda}) \sum_i |x_i|]$$

b

Taking the log and simplifying,

$$2log(\Lambda) = 2nlog(\hat{\lambda}) - 2nlog(\lambda_0) + 2(\lambda_0 - \hat{\lambda}) \sum_i x_i$$

We can reject the null hypothesis if

$$2log(\Lambda) \geq \chi^2_{1-\alpha}$$

with df = 1

7

a

Making substition for lambda as 1/x

$$log(\Lambda) = nlog(1) - nlog(\bar{x}) - nlog(\lambda_1) + (\lambda_0 - \frac{1}{\bar{x}})n\bar{x}$$
$$log(\Lambda) = -nlog(\bar{x}) - nlog(\lambda_0) + n\lambda_0\bar{x} - n$$

$$\frac{1}{n}log(\Lambda) = -log(\bar{x}) - log(\lambda_0) + \lambda_0\bar{x} - 1$$

Removing the known constant, we have that the rejection region is

$$\lambda_0\bar{x} - log(\bar{x}) \geq C$$

b

$$\frac{1}{n}log(\Lambda) = -log(\lambda_0\bar{x}) + \lambda_0\bar{x} - 1$$

Letting t = lambda0 multiplied by xbar

$$g(t) = -log(t) + t - 1$$

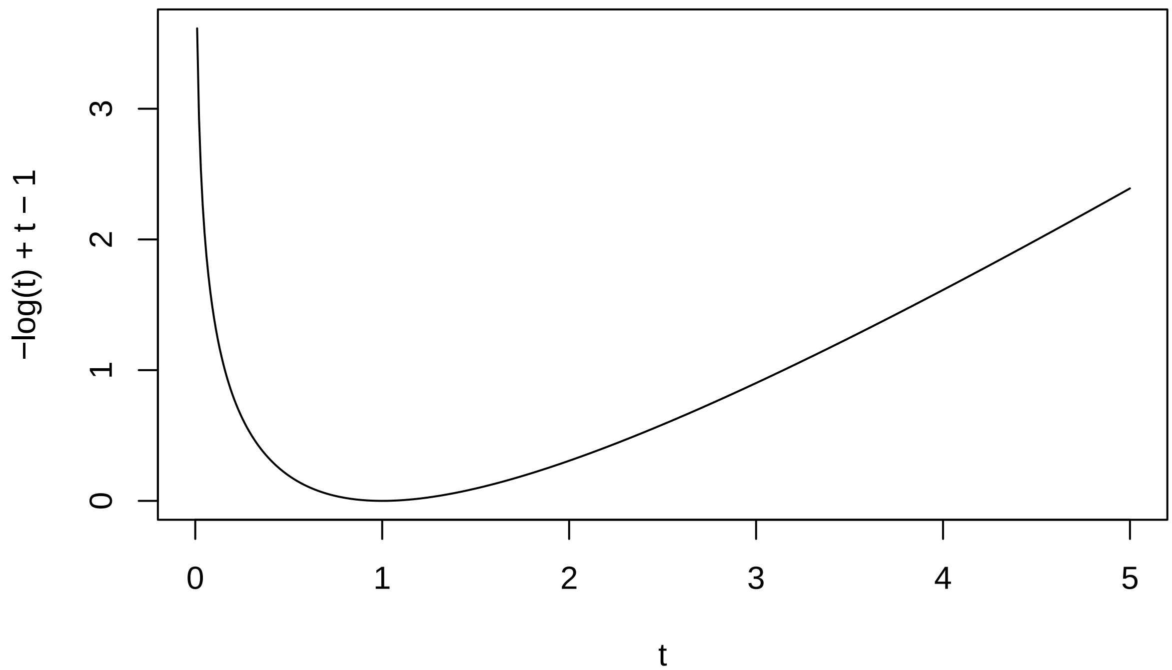
First, we will find the minimum by

$$\frac{d}{dt}g(t) = -\frac{1}{t} + 1 = 0$$

Solving the equation above, we have that

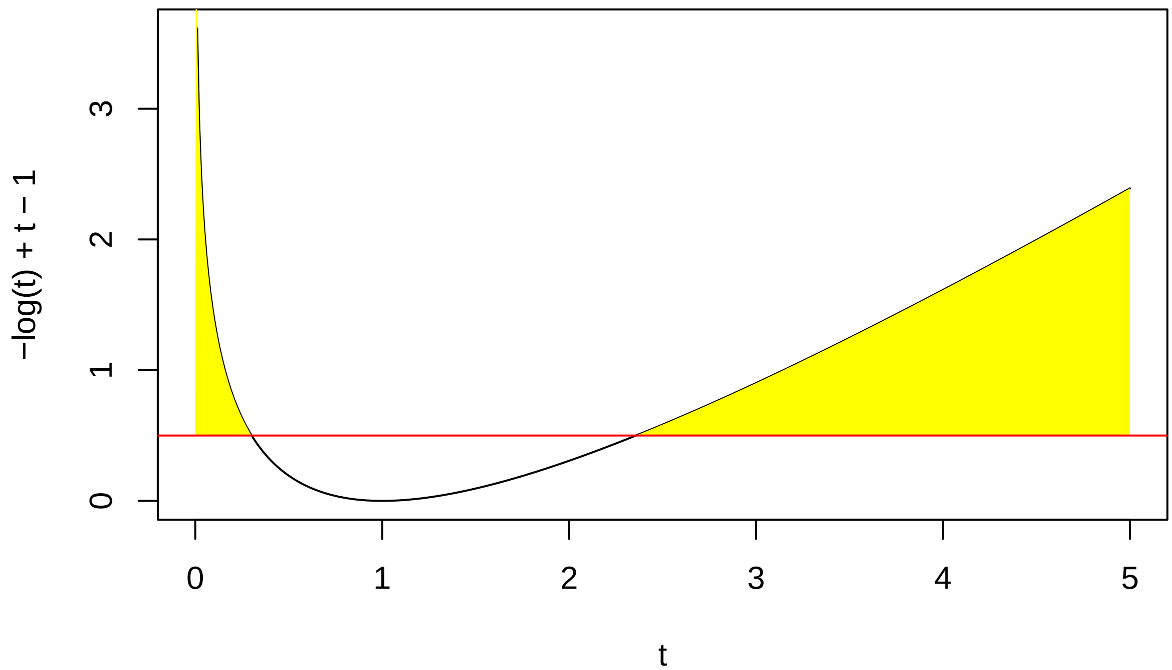
$$t = 1$$

```
t <- seq(0,5,0.01)
plot(t,-log(t)+t-1,type = "l")
```



c

```
t <- seq(0,5,0.01)
x <- seq(0,0.301, length.out = 100);y <- -log(x)+x-1
x1 <- c(0.301, x, 0); y1 <- c(0.5,y,0.5)
x <- seq(2.357, 5, length.out = 100);y <- -log(x)+x -1
x2 <- c(2.357, x, 5); y2 <- c(0.5,y,0.5)
plot(t, -log(t) + t - 1, type = "l")
polygon(x1,y1, col = "yellow", border = NA); polygon(x2, y2, col = "yellow", border = NA)
abline(h = 0.5, col = "red")
```



d

There are two particular value of t as threshold

$$\lambda_0 \bar{x} \geq c_1 and \lambda_0 \bar{x} \leq c_2$$
$$\sum_i x_i \geq n \frac{c1}{\lambda_0} and \sum_i x_i \leq n \frac{c2}{\lambda_0}$$

The sum of n independent variables of exponential (special case of Gamma) is Gamma (alpha = n, lambda)

8

a

To determine if there is a difference in mean

$$H_0 : \mu_x - \mu_y = 0$$
$$H_1 : \mu_x - \mu_y \neq 0$$

Assuming equal variance,

$$e.s.e = S_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

Where the pooled variance is

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{m+n-2}$$

```
x <- c(3.03, 5.6,9.3,12.51,15.21,16.84)
y <- c(3.19,4.47,4.53,4.69,6.79,12.75)
Sp <- sqrt( ((6-1)*var(x)+(6-1)*var(y))/(6+6-2))
e.s.e <- Sp*sqrt(1/6 + 1/6)
t_score <- (mean(x) - mean(y) - 0) / e.s.e
cat("P-vale = ", 2*pt(t_score,6+6-2, lower.tail = FALSE))
```

```
## P-vale =  0.1298458
```

Now, assuming unequal variance, we have that

$$e.s.e = \sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}$$

$$df = \frac{(\frac{S_x^2}{n} + \frac{S_y^2}{n})^2}{(\frac{S_x^2}{n})^2/(n-1) + (\frac{S_y^2}{m})^2/(m-1)}$$

```
e.s.e <- sqrt(var(x) / 6 + var(y)/6)
df <- (var(x)/6 + var(y)/6)^2 / ((var(x) / 6)^2/5 + (var(y)/6)^2/5)
t_score <- (mean(x) - mean(y) - 0)/e.s.e
cat("P-value = ", 2*pt(t_score, df, lower.tail = FALSE))
```

```
## P-value =  0.1352171
```

Since the p-value is greater than 0.05, we fail to reject the null hypothesis

b

Testing to see if there is a difference in variance,

$$H_0 : \frac{\sigma_y}{\sigma_x} = 1$$
$$H_1 : \frac{\sigma_y}{\sigma_x} \neq 1$$

```
2*pf(var(y)/var(x), df1 = 5, df2 = 5)
```

```
## [1] 0.3479363
```

Since the p-value is greater than 0.05, we fail to reject the null hypothesis

9

```r
data = c(3.03, 5.6,9.3,12.51,15.21,16.84,3.19,4.47,4.53,4.69,6.79,12.75)
data_rank <- rank(data)
R <- sum(data_rank[1:6])
R_prime <- 6*(6+6+1) - R
R_star <- min(R, R_prime); cat("R* = ", R_star)
```

```
## R* =  31
```

Since $31 > 30$, the p-vale is greater than 0.2, We do not reject the null hypothesis that there is no difference between the two types.

10

$$H_0 : \text{There is no significant difference among types}$$

$$H_1 : \text{There is significant difference among types}$$

```r
data <- c(1.7,6.1,12.5,25.1,42.1,13.6,25.2,46.2,13.4,29.7,46.9)
label <- c(rep("type 1", 5), rep("type 2", 3), rep("type 3", 3))
anova(lm(data~label))
```

```
## Analysis of Variance Table
##
## Response: data
##           Df  Sum Sq Mean Sq F value Pr(>F)
## label      2  375.38  187.69  0.6904 0.5289
## Residuals  8 2174.89  271.86
```

Since the p-value is greater than 0.05, we fail to reject the null hypothesis