concordance=TRUE

Statistics 2 Pi

Team 8

April 25, 2018

Contents

1 Task 95

a) Fitting Weibull distribution

Using the same approach as in Task 36, we get the following method of percentiles (based on quartiles) estimations for the 2-parameter Weibull distribution in the form:

$$F_{\lambda,\beta}(x) = \begin{cases} 1 - e^{-(\frac{x}{\lambda})^{\beta}}\,,\; x \geq 0 \\ 0\,,\; x < 0 \end{cases}$$

with $\beta > 0$ is the shape and $\lambda > 0$ is the scale parameter:

```
vandal <- c(38, 56, 77, 110, 112, 138, 152, 168, 188, 210, 228, 241, 252, 273, 283, 288,
lenv <- length(vandal)
g1_W <- 0.25
g2_W <- 0.75
x1_v <- quantile(vandal, probs = g1_W, na.rm = FALSE,  names = FALSE)
x2_v <- quantile(vandal, probs = g2_W, na.rm = FALSE,  names = FALSE)
shape_v_QME <- log(-log(1-g1_W)/-log(1-g2_W))/log(x1_v/x2_v)
shape_v_QME

## [1] 2.072234
```

which is γ in CDF in the form that is required in the task: $F(x|c, \gamma) = 1 - exp(-cx^{\gamma})$

```
scale_v_QME <- x2_v/(-log(1-g2_W))^(1/shape_v_QME)
scale_v_QME

## [1] 315.6161
```

Let's calculate c for CDF in the form that is required in the task: $F(x|c, \gamma) = 1 - exp(-cx^{\gamma})$:

```
c_v_QME <- scale_v_QME ^ (-shape_v_QME)
c_v_QME

## [1] 6.624574e-06
```

Let's compare with $fitdistrplus$:

```
require(fitdistrplus)

## Loading required package:  fitdistrplus
## Warning:  package 'fitdistrplus' was built under R version 3.4.3
## Loading required package:  MASS
## Loading required package:  survival

fit.weibull <- fitdist(vandal, "weibull", method = "qme", probs = c(0.25, 0.75))
summary(fit.weibull)

## Fitting of the distribution ' weibull ' by matching quantiles
## Parameters :
##          estimate
## shape    2.072232
## scale  315.614506
## Loglikelihood:  -193.7117   AIC:  391.4233   BIC:  394.2257
```

c for CDF in the form that is required in the task:

```
unname(summary(fit.weibull)$estimate[2]^(-summary(fit.weibull)$estimate[1]))

## [1] 6.624711e-06
```

Our results are almost the same.

b) Chi squared goodness-of-fit test

In order to perform the test, first let's calculate the actual number of observations in each interval:

```
border_v <- c(0, 145, 225, 310, 420)
observed_v <- c()
for (i in 1:length(border_v)) {
  observed_v[i] <- length(vandal[vandal < border_v[i + 1]]) - length(vandal[vandal < bo
}
observed_v

## [1] 6 4 9 4 7
```

Then let's calculate the number of expected observations in each interval:

```
cumul_probs_W <- pweibull(border_v, shape_v_QME, scale_v_QME, lower.tail = TRUE, log.
dist_probs_W <- c()
for (i in 2:length(border_v)) {dist_probs_W[i-1] <- cumul_probs_W[i]-cumul_probs_W[i-
estimated_W <- lenv * c(dist_probs_W, 1 - sum(dist_probs_W))
estimated_W

## [1] 5.426639 6.303436 6.823023 6.526331 4.920571
```

And finally obtain the value of χ^2 test statistic:

```
chi_W <- sum(((estimated_W - observed_v)^2)/estimated_W)
chi_W

## [1] 3.453611
```

and it's p-value:

```
pchisq(chi_W, length(border_v) - 1 - 2, ncp = 0, lower.tail = FALSE, log.p = FALSE)

## [1] 0.1778517
```

Since p-value is higher than 0.05, the null hypothesis that the data set is Weibull distibuted couldn't be rejected.