

STATS 205: Homework Assignment 5 (Spring 2019)

5/13/2019

Solve problems 2-7 from the textbook *HWC* available here.

Send your Rmd and PDF files to `pjeganat [at] stanford [dot] edu`.

Due on 5/20/2019 (Monday) at 1.30 p.m.

- 1) “Permutation, parametric and bootstrap tests of hypotheses by Good 2005” (Exercise 4.8, #20, page 78) link here to the textbook. Suppose the observations (X_1, \dots, X_K) are distributed in accordance with the multivariate normal probability density

$$\frac{\sqrt{|D|}}{(2\pi)^{K/2}} \exp \left[-\frac{1}{2} \sum \sum d_{ij} (x_i - \mu_i) (x_j - \mu_j) \right],$$

where the matrix $D = (d_{ij})$ is positive definite; $|D|$ denotes its determinant; $\mathbb{E}(X_j) = \mu_j$; $\mathbb{E}(x_i - \mu_i)(x_j - \mu_j) = \sigma_{ij}$; and $\sigma_{ij} = (D^{-1})_{ij}$, i, j -th element of inverse of D . If $\sigma_{ii} = \sigma^2$ when $i = j$ and $\sigma_{ij} = \sigma_{12}$, $\forall i, j$, are the observations independent? exchangeable?

- 2) **HWC** Page 409, Problem 1 (Kendall test).

The data in Table 8.3 are a subset of the data obtained by Featherston (1971). Among other things, he was interested in the relationship between the weight of tapeworms (*Taenia hydatigena*) fed to dogs and the weight of the scoleces recovered from the dogs after 20 days. (A scolex is the attachment end of a tapeworm, consisting of the head and neck.) The cysticerici used in the experiment were collected from sheep carcasses and force-fed to 10 dogs via gelatine capsules. The scoleces were recovered from each dog at autopsy, 20 days after the introduction of the tapeworms. Table 8.3 gives the mean weight of the initial cysticerici and the mean weight of the recovered worms for each of the 10 dogs in the study. Test the hypothesis of independence versus the alternative that the mean weight of introduced cysticerici is positively correlated with the mean weight of worms recovered.

```
library(NSM3)
```

```
## Loading required package: combinat
```

```
##
```

```
## Attaching package: 'combinat'
```

```
## The following object is masked from 'package:utils':
```

```
##
```

```
##      combn
```

```
## Loading required package: MASS
```

```
## Loading required package: partitions
```

```
## Loading required package: survival
```

```
## fANCOVA 0.5-1 loaded
```

```
cysticerici = c(28.9, 31.8, 12, 9.9, 15, 38, 12.5, 36.5, 8.6, 26.8)
```

```
wormsRecovered = c(1, 7.7, 7.3, 7.9, 1.1, 3.5, 18.9, 33.9, 28.6, 25)
```

```
testKendall = cor.test(cysticerici, wormsRecovered, method = "kendall", alt = "greater")
```

```
testKendall$p.value
```

```
## [1] 0.7578436
```

```
testKendall
```

```
##
## Kendall's rank correlation tau
##
## data: cysticerici and wormsRecovered
## T = 19, p-value = 0.7578
## alternative hypothesis: true tau is greater than 0
## sample estimates:
##      tau
## -0.1555556
```

We don't have enough evidence to reject that weight of tapeworms (*Taenia hydatigena*) fed to dogs and the weight of the scoleces recovered from the dogs after 20 days are independent.

3) **HWC** Page 414, Problem 20 (Kendall's sample correlation coefficient)

Estimate τ for the tapeworm data of Table 8.3.

```
testKendall$estimate
```

```
##      tau
## -0.1555556
```

4) **HWC** Page 427, Problem 34 (Kendall correlation coefficient confidence interval) do not need to compare the results with Problem 28.

For the cerebral palsy data of Table 8.4, use the bootstrap method to find a confidence interval for τ with approximate confidence coefficient .90.

```
brainWeight = c(515, 286, 469, 410, 461, 436, 479, 198, 389, 262, 536)
fiberCount = c(32500, 26800, 11410, 14850, 23640, 23820, 29840, 21830, 24650, 22500, 26000)

cerebral = data.frame(brainWeight, fiberCount)
n = dim(cerebral)[1]
bootstrap.sample = function(df){
  boot.sam = df[sample(1:n, size = n, replace = TRUE), ]
  kendall = cor(boot.sam[, 1], boot.sam[, 2], method = "kendall")
  return(kendall)
}

tau.hat.star = replicate(1000, bootstrap.sample(cerebral))
round(quantile(tau.hat.star, probs = c(.05, .95)), digits = 2)

##      5%      95%
## -0.17    0.64
```

5) **HWC** Page 440, Problem 41 (Spearman test).

Test the hypothesis of independence versus the alternative that the mean weight of introduced cysticerici is positively correlated with the mean weight of worms recovered for the tapeworm data in Table 8.3.

```
testSpearman = cor.test(cysticerici, wormsRecovered, method = "spearman", alt = "greater")
testSpearman$p.value
```

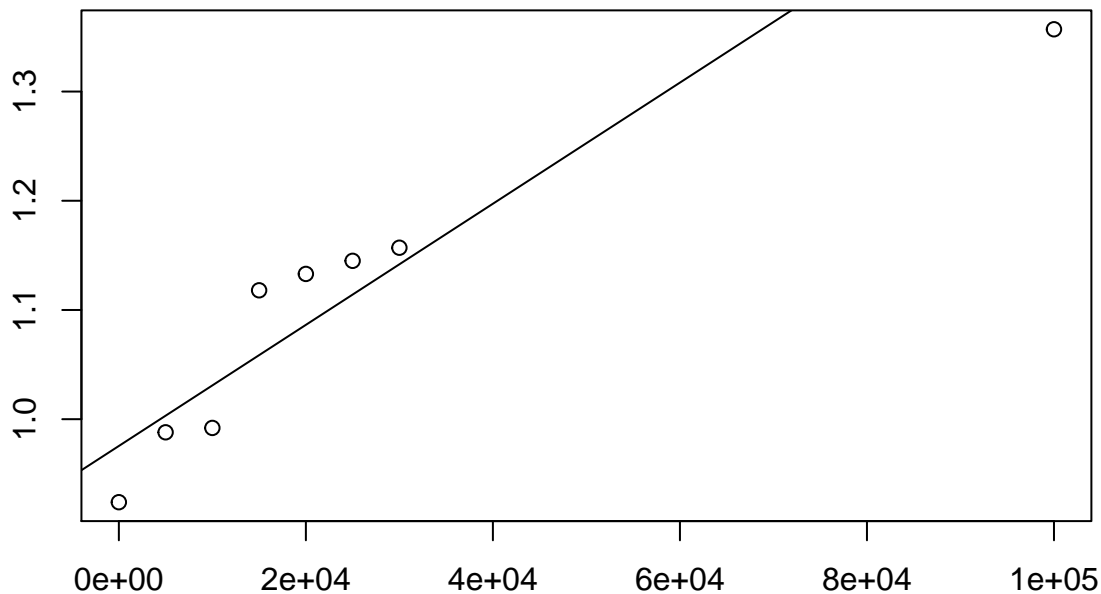
```
## [1] 0.720046
```

Based on Spearman's test, we don't have enough evidence to reject that weight of tapeworms (*Taenia hydatigena*) fed to dogs and the weight of the scoleces recovered from the dogs after 20 days are independent.

6) **HWC** Page 457, Problem 1 (test for slope).

Johnson et al. (1970) considered the behavior of a cenosphere-resin composite under hydrostatic pressure. The authors pointed out that most deep submersible vehicles utilize a buoyancy material, known as syntactic foam, that is a composite of closely packed hollow glass microspheres embedded in a resin matrix. These microspheres are relatively expensive to manufacture, and the cost of the syntactic foam is principally determined by the cost of the microspheres. The authors also noted that the ash from generating stations burning pulverized coal contains a small proportion of hollow glassy microspheres, known as cenospheres, and these have about the right size distribution for use in syntactic foam. The cenospheres can be readily collected from the ash-disposal method used in certain British generating stations. The authors were thus interested in whether the cenospheres would, in some applications, perform as well as the manufactured microspheres. In attempting to assess the usefulness of cenospheres as a component of syntactic foam, Johnson et al. investigated the effects of hydrostatic pressure (such as exists in the ocean depths) on the density of a cenosphere-resin composite. The results are given in Table 9.2. What is the P-value for a test of $H_0 : \beta = 0$ against the alternative $\beta > 0$ for these data?

```
library(NSM3)
pressure = c(0, 5000, 10000, 15000, 20000, 25000, 30000, 100000)
density = c(0.924, 0.988, 0.992, 1.118, 1.133, 1.145, 1.157, 1.357)
slopeTest = theil(pressure, density, beta.0 = 0, t = "u", r = 6)
```



```
slopeTest$p.val
```

```
## [1] 2.480159e-05
```

We conclude that the slop is significantly different from zero.

7) HWC Page 485, Problem 33 (rank-based multiple linear regression).

In heart catheterization, a 3-mm-diameter Teflon catheter (tube) is inserted into a major vein or artery at the femoral region and maneuvered up into the heart itself to assess the heart's physiology and functional ability. Heart catheterizations are sometimes performed on children with congenital heart defects. In such cases, the length of the catheter is often determined by a physician's educated guess. Rice (2007) considered a data set obtained by Weindling (1977) in a preliminary study involving 12 children. For each child, the exact catheter length required was determined by using a fluoroscope to check that the tip of the catheter had reached the pulmonary artery. The 12 catheter lengths (cm) and the heights (in) and weights (lb) for the 12 children in the study are given in Table 9.10.

```
library(Rfit)
height = c(42.8, 63.5, 37.5, 39.5, 45.5, 38.5, 43.0, 22.5, 37.0, 23.5, 33.0, 58.0)
```

```
weight = c(40.0, 93.5, 35.5, 30.0, 52.0, 17.0, 38.5, 8.5, 33.0, 9.5, 21.0, 79.0)

length = c(37.0, 49.5, 34.5, 36.0, 43.0, 28.0, 37.0, 20.0, 33.5, 30.5, 38.5, 47.0)

fit1 = rfit(length ~ height + weight)
summary(fit1)
```

```
## Call:
## rfit.default(formula = length ~ height + weight)
##
## Coefficients:
##             Estimate Std. Error t.value p.value
## (Intercept) 25.921266   7.550102  3.4332 0.00747 **
## height      0.029114   0.311060  0.0936 0.92748
## weight      0.245570   0.142480  1.7235 0.11888
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Overall Wald Test: 43.11247 p-value: 1e-05
```

```
fit2 = rfit(length ~ height)
summary(fit2)
```

```
## Call:
## rfit.default(formula = length ~ height)
##
## Coefficients:
##             Estimate Std. Error t.value   p.value
## (Intercept) 11.972952   3.990272  3.0005   0.01333 *
## height      0.594595   0.093221  6.3783 8.053e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Overall Wald Test: 40.68298 p-value: 2e-05
```

```
drop.test(fit1, fit2)
```

```
##
## Drop in Dispersion Test
## F-Statistic      p-value
##      1.55202      0.24429
```

```
fit3 = rfit(length ~ weight)
summary(fit3)
```

```
## Call:
## rfit.default(formula = length ~ weight)
##
## Coefficients:
##             Estimate Std. Error t.value   p.value
## (Intercept) 26.612065   1.769218 15.0417 3.404e-08 ***
## weight      0.258621   0.038895  6.6491 5.715e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Overall Wald Test: 44.21107 p-value: 1e-05
```

```
drop.test(fit1, fit3)
```

```
##
```

```
## Drop in Dispersion Test
## F-Statistic      p-value
##      0.014435    0.907007
```