R University

Statistics Exam 2015-01-01 **Exam ID 00003**

Name: _____

Student ID: _____

Signature: _____

1. [ ] [ ] [ ] [ ] [2] [7] . [4] [8] [0]
2. (a) [X] (b) [ ] (c) [ ] (d) [ ] (e) [ ]
3. (a) [ ] (b) [X] (c) [X] (d) [ ] (e) [ ]
4. **glm.nb**
5. (a) (a) [ ] (b) [X] (c) [ ]

   (b) [ ] [ ] [ ] [ ] [ ] [0] . [9] [5] [9]

1. **Problem**
   What is the derivative of $f(x) = x^7 e^{3.4x}$, evaluated at $x = 0.77$?

   Solution
   Using the product rule for $f(x) = g(x) \cdot h(x)$, where $g(x) := x^7$ and $h(x) := e^{3.4x}$, we obtain

$$\begin{aligned} f'(x) &= [g(x) \cdot h(x)]' = g'(x) \cdot h(x) + g(x) \cdot h'(x) \\ &= 7x^{7-1} \cdot e^{3.4x} + x^7 \cdot e^{3.4x} \cdot 3.4 \\ &= e^{3.4x} \cdot (7x^6 + 3.4x^7) \\ &= e^{3.4x} \cdot x^6 \cdot (7 + 3.4x). \end{aligned}$$

   Evaluated at $x = 0.77$, the answer is

$$e^{3.4 \cdot 0.77} \cdot 0.77^6 \cdot (7 + 3.4 \cdot 0.77) = 27.479706.$$

   Thus, rounded to two digits we have $f'(0.77) = 27.48$.

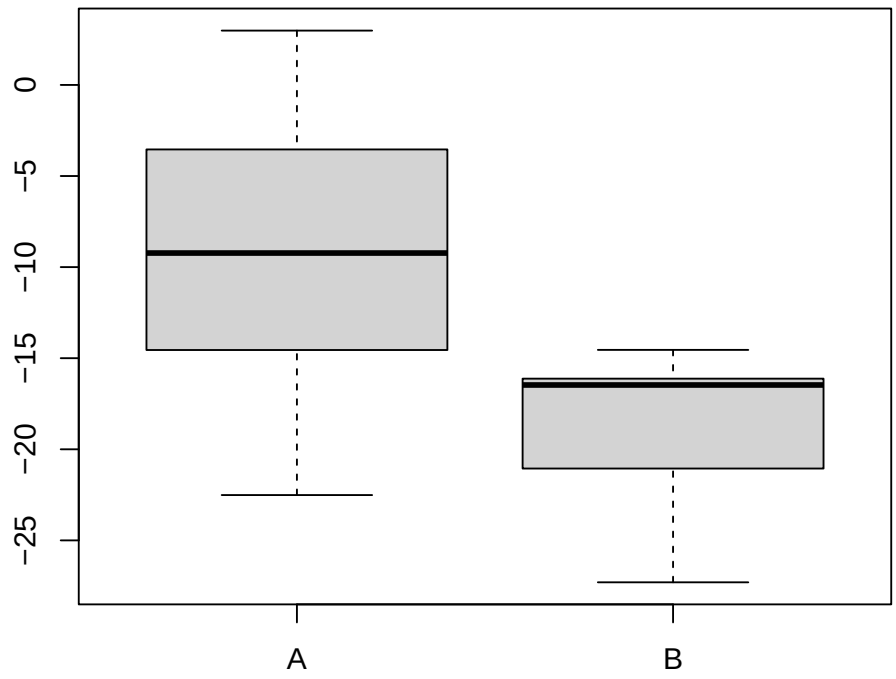
2. **Problem**
   What is the seat of the federal authorities in Switzerland (i.e., the de facto capital)?

   (a) Bern
   (b) Lausanne
   (c) St. Gallen
   (d) Zurich
   (e) Basel

   Solution
   There is no de jure capital but the de facto capital and seat of the federal authorities is Bern.

   (a) True
   (b) False
   (c) False
   (d) False
   (e) False

3. **Problem**
   In the following figure the distributions of a variable given by two samples (A and B) are represented by parallel boxplots. Which of the following statements are correct? *(Comment: The statements are either about correct or clearly wrong.)*



(a) The location of both distributions is about the same.

(b) Both distributions contain no outliers.

(c) The spread in sample A is clearly bigger than in B.

(d) The skewness of both samples is similar.

(e) Distribution B is right-skewed.

Solution

(a) False. Distribution A has on average higher values than distribution B.

(b) True. Both distributions have no observations which deviate more than 1.5 times the interquartile range from the box.

(c) True. The interquartile range in sample A is clearly bigger than in B.

(d) False. The skewness of both distributions is different. Sample A is about symmetric. Sample B is left-skewed.

(e) False. Distribution B is left-skewed.

4. **Problem**
What is the name of the R function for negative binomial regression?

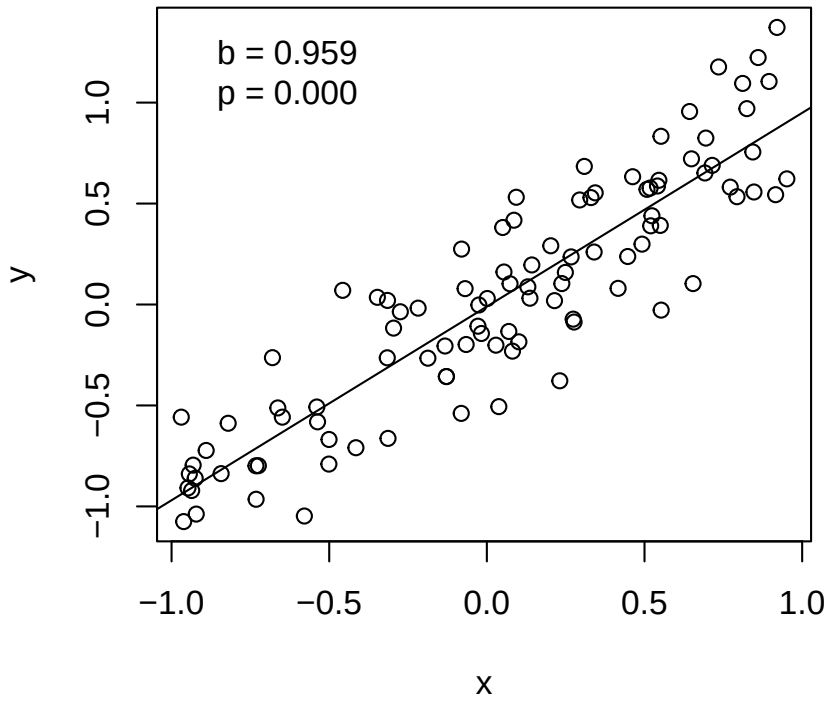
Solution
`glm.nb` is the R function for negative binomial regression. See `?glm.nb` for the corresponding manual page.

5. **Problem**
Using the data provided in `regression.csv` estimate a linear regression of `y` on `x` and answer the following questions.

(a) `x` and `y` are not significantly correlated / `y` increases significantly with `x` / `y` decreases significantly with `x`

(b) Estimated slope with respect to `x`:

Solution



To replicate the analysis in R:

```
## data
d <- read.csv("regression.csv")
## regression
m <- lm(y ~ x, data = d)
summary(m)
## visualization
plot(y ~ x, data = d)
abline(m)
```