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

   (b) [ ] [ ] [ ] [ ] [-] [0] . [9] [1] [1]
9. **nil**

1. **Problème**
   (1 point) Choisissez l'affirmation ou les affirmations correspondant à l'expression : $\Pr(\text{lundi}|\text{pluie})$.

   (a) La probabilité qu'il pleuve lundi.
   (b) La probabilité qu'il pleuve, sachant qu'on est lundi.
   (c) La probabilité qu'on soit lundi, sachant qu'il pleut.
   (d) La probabilité qu'on soit lundi et qu'il pleuve.
   (e) La probabilité qu'il pleuve et qu'on soit lundi.

   Solution

   (a) False. Ceci correspond à Pr(pluie,lundi).
   (b) False. Ceci correspond à Pr(pluie|lundi).
   (c) True.
   (d) False. Ceci correspond à Pr(lundi,pluie), équivalent à Pr(pluie,lundi).
   (e) False. Ceci correspond à Pr(pluie,lundi), équivalent à Pr(lundi,pluie).

2. **Problème**
   (2 points) Dans le modèle ci-dessous,

$$
\begin{aligned}
y_i &\sim \mathrm{Normal}(\mu, \sigma) \\
\mu_i &= \alpha + \beta \cdot x_i \\
\alpha &\sim \mathrm{Normal}(0, 10) \\
\beta &\sim \mathrm{Normal}(0, 1) \\
\sigma &\sim \mathrm{Exponential}(2)
\end{aligned}
$$

(a) La première ligne décrit les priors, la deuxième ligne la fonction de vraisemblance, et les lignes suivantes le posterior.

(b) La première ligne décrit la fonction de vraisemblance, la deuxième ligne le modèle linéaire, et les lignes suivantes les priors.

(c) La première ligne décrit le modèle linéaire, la deuxième ligne la fonction de vraisemblance, et les lignes suivantes les priors.

(d) Les deux premières lignes décrivent le modèle linéaire, et les lignes suivantes les priors.

(e) Les deux premières lignes décrivent le modèle linéaire, et les lignes suivantes le posterior.

Solution

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

3. **Problème**
(1 point) Dans le modèle ci-dessus, combien y a-t-il de paramètres dans la distribution a posteriori ?
Solution
Il y a 3 paramètres : l'intercept α, la pente β, et l'écart-type des résidus σ.

4. **Problème**
(2 points) Traduisez le modèle `brms` ci-dessous en modèle mathématique.

```
brm(
    formula = y ~ 1 + x,
    family = Beta(),
    prior = prior(Normal(0, 1), class = Intercept)
    )
```

Solution

$$
\begin{aligned}
y_i &\sim \mathrm{Beta}(\mu, \sigma) \\
\mu_i &= \alpha + \beta \cdot x_i \\
\alpha &\sim \mathrm{Normal}(0, 1)
\end{aligned}
$$

5. **Problème**
The waiting time (in minutes) at the cashier of two supermarket chains with different cashier systems is compared. The following statistical test was performed:

```
    Two Sample t-test

data:  Waiting by Supermarket
t = 1.6547, df = 123, p-value = 0.1005
alternative hypothesis: true difference in means between group Sparag and group Consumo is n
95 percent confidence interval:
```

```
 -0.2049041  2.2927772
sample estimates:
 mean in group Sparag mean in group Consumo
             7.100377             6.056441
```

Which of the following statements are correct? (Significance level 5%)

(a) The absolute value of the test statistic is larger than 1.96.
(b) A one-sided alternative was tested.
(c) The p-value is larger than 0.05.
(d) The test shows that the waiting time is longer at Sparag than at Consumo.
(e) The test shows that the waiting time is shorter at Sparag than at Consumo.

Solution

(a) False. The absolute value of the test statistic is equal to 1.655.
(b) False. The test aims at showing that the difference of means is unequal to 0.
(c) True. The p-value is equal to 0.101.
(d) False. The test result is not significant ($p \geq 0.05$).
(e) False. The test result ist not significant ($p \geq 0.05$).

6. **Problème**
The following figure shows a scatterplot. Which of the following statements are correct?

(a) The absolute value of the test statistic is larger than 1.96.
(b) A one-sided alternative was tested.
(c) The p-value is larger than 0.05.
(d) The test shows that the waiting time is longer at Sparag than at Consumo.
(e) The test shows that the waiting time is shorter at Sparag than at Consumo.

Solution

(a) False. The absolute value of the test statistic is equal to 1.655.
(b) False. The test aims at showing that the difference of means is unequal to 0.
(c) True. The p-value is equal to 0.101.
(d) False. The test result is not significant ($p \geq 0.05$).
(e) False. The test result ist not significant ($p \geq 0.05$).

7. **Problème**
What is the name of the R function for Poisson regression?

Solution
`glm` is the R function for Poisson regression. See `?glm` for the corresponding manual page.

8. **Problème**
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)
```

9. **Problème**
Consider the following regression results:

```
Call:
lm(formula = y ~ x, data = d)

Residuals:
     Min       1Q   Median       3Q      Max
-1.09342 -0.32434  0.08179  0.29419  1.19206

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.04922    0.06460   0.762     0.45
x           -0.64685    0.06020 -10.745 2.99e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4517 on 47 degrees of freedom
Multiple R-squared:  0.7107,    Adjusted R-squared:  0.7045
F-statistic: 115.5 on 1 and 47 DF,  p-value: 2.994e-14
```

Describe how the response `y` depends on the regressor `x`.

Solution
The presented results describe a linear regression.

The mean of the response `y` decreases with increasing `x`.

If `x` increases by 1 unit then a change of `y` by about -0.65 units can be expected.

Also, the effect of `x` is significant at the 5 percent level.