

СПбГУ

2014-11-13

Матметоды — Множественная регрессия  
Вар.№ 01

Ф.И.О.: \_\_\_\_\_

—

1. (a) ☐ (b) ☒ (c) ☐ (d) ☐
2. (a) ☐ (b) ☐ (c) ☐ (d) ☒
3. (a) ☒ (b) ☒ (c) ☒ (d) ☒
4. (a) ☐ (b) ☐ (c) ☒ (d) ☒
5. (a) ☒ (b) ☐ (c) ☐ (d) ☐
6.      2 3     0

### 1. Задача

Во встроенном датасете `stackloss` пусть переменная `stack.loss` будет зависимой, а переменные `Air.Flow`, `Water.Temp`, `Acid.Conc` - предикторами. Проведите регрессионный анализ и определите значение коэффициента `stack.loss`

- (a) 1.2953  
(b) NA  
(c) -39.9197  
(d) -0.1521

Решение

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

### 2. Задача

Стандартизация (`z-score scaling`) шкал всех предикторов перед выполнением регрессионного анализа:

- (a) не позволяет правомерно сравнить силу эффектов разнородных предикторов  
(b) позволяет обойтись без проверки состоятельности модели

- (c) центрует значения зависимой переменной  $Y$  вокруг нуля вместо средней  
(d) позволяет правомерно сравнить силу эффектов разнородных предикторов

Решение

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

### 3. Задача

Отметьте условия применимости линейной регрессии

- (a) Нормальное распределение остатков  
(b) Гомогенность дисперсий остатков  
(c) Линейная связь  
(d) Независимость значений  $y$  друг от друга

Решение

- (a) True  
(b) True  
(c) True

(d) True

4. Задача

Во множественной регрессии, описываемой моделью  $Y = 0.01 + 1.4X_1 - 4.3X_2 + 0.8X_3$ , интерсепт это:

- (a) самый слабый предиктор
- (b) скорость изменения зависимой переменной при фиксированном значении предиктора  $X_i$
- (c) один из параметров модели
- (d) ожидаемое среднее значение зависимой переменной, когда все предикторы ( $X_1, X_2, X_3$ ) равны нулю

Решение

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

5. Задача

Поправка adjusted  $R^2$ :

(a) позволяет правомерно сравнивать модели с разным количеством предикторов

(b) всегда растет с увеличением количества предикторов

(c) всегда равна или имеет значение выше, чем  $R^2$

(d) всегда уменьшается с увеличением количества предикторов

Решение

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

6. Задача

---

ДОПОЛНИТЕЛЬНЫЙ ВОПРОС

---

Для встроенного датасета `stackloss` рассчитайте (предскажите) значение зависимой переменной `stack.loss` при условии, что предиктор `Air.Flow` принимает значение 69, а остальные независимые

переменные принимают свои средние значения. Запишите результат, округленный до третьего знака.

Решение

Сначала проведем регрессионный анализ, получим значения интерсепта и коэффициентов регрессии

```
stmod <- lm(stack.loss ~ ., data = stackloss)
```

Предсказать конкретное значение зависимой переменной можно выполнив код:

```
predict <- c(coefficients(stmod)[1] +  
coefficients(stmod)[2]*air +  
coefficients(stmod)[3]*mean(stackloss$Water.Temp) +  
coefficients(stmod)[4]*mean(stackloss$Acid.Conc.))
```

```
answer <- round(predict,3)
```

Вместо "air" подставьте нужное значение `Air.Flow`

Ф.И.О.: \_\_\_\_\_

—

1. (a) ☐ (b) ☐ (c) ☒ (d) ☐
2. (a) ☐ (b) ☐ (c) ☒ (d) ☒
3. (a) ☐ (b) ☒ (c) ☐ (d) ☐
4. (a) ☒ (b) ☐ (c) ☐ (d) ☒
5. (a) ☐ (b) ☐ (c) ☐ (d) ☒
6.      2 2 . 9 4 0

## 1. Задача

Во встроенном датасете `stackloss` пусть переменная `stack.loss` будет зависимой, а переменные `Air.Flow`, `Water.Temp`, `Acid.Conc` - предикторами. Проведите регрессионный анализ и определите значение коэффициента `stack.loss`

- (a) 0.9136  
(b) 0.7156  
(c) NA  
(d) 1.2953

Решение

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

## 2. Задача

Стандартизация (z-score scaling) шкал всех предикторов перед выполнением регрессионного анализа:

- (a) не позволяет правомерно сравнить силу эффектов разнородных предикторов  
(b) позволяет оценить ошибку предсказания модели

- (c) позволяет правомерно сравнить силу эффектов разнородных предикторов  
(d) центрует значения каждого предиктора  $X_i$  вокруг нуля вместо средней

Решение

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

## 3. Задача

Отметьте условия применимости линейной регрессии

- (a) Коэффициент детерминации  $R^2 > 0.79$   
(b) Независимость значений  $y$  друг от друга  
(c) Корреляция между независимыми переменными (мультиколлинеарность) велика  
(d) Независимые переменные стандартизованы

Решение

- (a) False

- (b) True
- (c) False
- (d) False

#### 4. Задача

Во множественной регрессии, описываемой моделью  $Y = 0.01 + 1.4X_1 - 4.3X_2 + 0.8X_3$ , интерсепт это:

- (a) один из параметров модели
- (b) угловой коэффициент регрессионной прямой
- (c) самый слабый предиктор
- (d) ожидаемое среднее значение зависимой переменной, когда все предикторы ( $X_1, X_2, X_3$ ) равны нулю

Решение

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

#### 5. Задача

Поправка adjusted  $R^2$ :

- (a) всегда равна или имеет значение выше, чем  $R^2$
- (b) всегда уменьшается с увеличением количества предикторов
- (c) всегда растет с увеличением количества предикторов
- (d) позволяет скорректировать рост  $R^2$  при добавлении каждого нового предиктора

Решение

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

#### 6. Задача

---

#### ДОПОЛНИТЕЛЬНЫЙ ВОПРОС

---

Для встроенного датасета `stackloss` рассчитайте (предскажите) значение зависимой переменной `stack.loss` при условии, что предиктор `Air.Flow` принимает

значение 68, а остальные независимые переменные принимают свои средние значения. Запишите результат, округленный до третьего знака.

Решение

Сначала проведем регрессионный анализ, получим значения интерсепта и коэффициентов регрессии

```
stmod <- lm(stack.loss ~ ., data = stackloss)
```

Предсказать конкретное значение зависимой переменной можно выполнив код:

```
predict <- c(coefficients(stmod)[1] +
coefficients(stmod)[2]*air +
coefficients(stmod)[3]*mean(stackloss$Water.Temp) +
coefficients(stmod)[4]*mean(stackloss$Acid.Conc.))
```

```
answer <- round(predict,3)
```

Вместо "air" подставьте нужное значение `Air.Flow`

СПбГУ

2014-11-13

Матметоды — Множественная регрессия  
Вар.№ 03

Ф.И.О.: \_\_\_\_\_

—

1. (a) ☐ (b) ☒ (c) ☐ (d) ☐
2. (a) ☐ (b) ☐ (c) ☒ (d) ☒
3. (a) ☐ (b) ☒ (c) ☐ (d) ☐
4. (a) ☐ (b) ☒ (c) ☐ (d) ☒
5. (a) ☐ (b) ☐ (c) ☒ (d) ☐
6. 

				2	2	9	4	0
--	--	--	--	---	---	---	---	---

### 1. Задача

Во встроенном датасете `stackloss` пусть переменная `stack.loss` будет зависимой, а переменные `Air.Flow`, `Water.Temp`, `Acid.Conc` - предикторами. Проведите регрессионный анализ и определите значение коэффициента `Water.Temp`

- (a) 0.7156
- (b) 1.2953
- (c) -39.9197
- (d) NA

Решение

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

### 2. Задача

Стандартизация (z-score scaling) шкал всех предикторов перед выполнением регрессионного анализа:

- (a) позволяет оценить ошибку предсказания модели
- (b) не позволяет правомерно сравнить силу эффектов разнородных предикторов

- (c) центрует значения каждого предиктора  $X_i$  вокруг нуля вместо средней
- (d) позволяет правомерно сравнить силу эффектов разнородных предикторов

Решение

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

### 3. Задача

Отметьте условия применимости линейной регрессии

- (a) Дисперсия остатков равна единице
- (b) Гомогенность дисперсий остатков
- (c) Коэффициент детерминации  $R^2 > 0.79$
- (d) Корреляция между независимыми переменными (мультиколлинеарность) велика

Решение

- (a) False
- (b) True

(c) False

(d) False

#### 4. Задача

Во множественной регрессии, описываемой моделью  $Y = 0.01 + 1.4X_1 - 4.3X_2 + 0.8X_3$ , интерсепт это:

(a) скорость изменения зависимой переменной при фиксированном значении предиктора  $X_i$

(b) один из параметров модели

(c) угловой коэффициент регрессионной прямой

(d) ожидаемое среднее значение зависимой переменной, когда все предикторы ( $X_1, X_2, X_3$ ) равны нулю

Решение

(a) False

(b) True

(c) False

(d) True

#### 5. Задача

Поправка adjusted  $R^2$ :

(a) всегда растёт с увеличением количества предикторов

(b) позволяет определить наиболее значимый предиктор множественной модели

(c) позволяет скорректировать рост  $R^2$  при добавлении каждого нового предиктора

(d) всегда равна или имеет значение выше, чем  $R^2$

Решение

(a) False

(b) False

(c) True

(d) False

#### 6. Задача

##### ДОПОЛНИТЕЛЬНЫЙ ВОПРОС

Для встроенного датасета `stackloss` рассчитайте (предскажите) значение зависимой переменной `stack.loss` при условии, что предиктор `Air.Flow` принимает

значение 68, а остальные независимые переменные принимают свои средние значения. Запишите результат, округленный до третьего знака.

Решение

Сначала проведем регрессионный анализ, получим значения интерсепта и коэффициентов регрессии

```
stmod <- lm(stack.loss ~ ., data = stackloss)
```

Предсказать конкретное значение зависимой переменной можно выполнив код:

```
predict <- c(coefficients(stmod)[1] +
coefficients(stmod)[2]*air +
coefficients(stmod)[3]*mean(stackloss$Water.Temp) +
coefficients(stmod)[4]*mean(stackloss$Acid.Conc.))
```

```
answer <- round(predict,3)
```

Вместо "air" подставьте нужное значение `Air.Flow`

Ф.И.О.: \_\_\_\_\_

—

1. (a) ☒ (b) ☐ (c) ☐ (d) ☐
2. (a) ☐ (b) ☐ (c) ☒ (d) ☐
3. (a) ☒ (b) ☒ (c) ☒ (d) ☒
4. (a) ☐ (b) ☒ (c) ☐ (d) ☐
5. (a) ☐ (b) ☐ (c) ☒ (d) ☐
6.      2 1 . 5 1 0

## 1. Задача

Во встроенном датасете `stackloss` пусть переменная `stack.loss` будет зависимой, а переменные `Air.Flow`, `Water.Temp`, `Acid.Conc` - предикторами. Проведите регрессионный анализ и определите значение коэффициента `Water.Temp`

- (a) 1.2953  
(b) -0.1521  
(c) -39.9197  
(d) 0.7156

Решение

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

## 2. Задача

Стандартизация (`z-score scaling`) шкал всех предикторов перед выполнением регрессионного анализа:

- (a) центрует значения зависимой переменной  $Y$  вокруг нуля вместо средней  
(b) позволяет оценить ошибку предсказания модели

- (c) центрует значения каждого предиктора  $X_i$  вокруг нуля вместо средней  
(d) позволяет обойтись без проверки на нормальность

Решение

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

## 3. Задача

Отметьте условия применимости линейной регрессии

- (a) Гомогенность дисперсий остатков  
(b) Нормальное распределение остатков  
(c) Независимость значений  $y$  друг от друга  
(d) Линейная связь

Решение

- (a) True  
(b) True  
(c) True  
(d) True

## 4. Задача

Во множественной регрессии, описываемой моделью  $Y = 0.01 + 1.4X_1 - 4.3X_2 + 0.8X_3$ , интерсепт это:

- (a) один из предикторов модели
- (b) ожидаемое среднее значение зависимой переменной, когда все предикторы ( $X_1, X_2, X_3$ ) равны нулю
- (c) угловой коэффициент регрессионной прямой
- (d) самый слабый предиктор

Решение

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

## 5. Задача

Поправка adjusted  $R^2$ :

- (a) всегда уменьшается с увеличением количества предикторов
- (b) всегда растет с увеличением количества предикторов
- (c) позволяет скорректировать рост  $R^2$  при добавлении каждого нового предиктора
- (d) всегда равна или имеет значение выше, чем  $R^2$

Решение

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

## 6. Задача

---

ДОПОЛНИТЕЛЬНЫЙ ВОПРОС

---

Для встроенного датасета stackloss рассчитайте (предскажите) значение зависимой переменной stack.loss при условии, что предиктор Air.Flow принимает

значение 66, а остальные независимые переменные принимают свои средние значения. Запишите результат, округленный до третьего знака.

Решение

Сначала проведем регрессионный анализ, получим значения интерсепта и коэффициентов регрессии

```
stmod <- lm(stack.loss ~ ., data = stackloss)
```

Предсказать конкретное значение зависимой переменной можно выполнив код:

```
predict <- c(coefficients(stmod)[1] +
coefficients(stmod)[2]*air +
coefficients(stmod)[3]*mean(stackloss$Water.Temp) +
coefficients(stmod)[4]*mean(stackloss$Acid.Conc.))
```

```
answer <- round(predict,3)
```

Вместо "air" подставьте нужное значение Air.Flow



Ф.И.О.: \_\_\_\_\_

—

1. (a) ☐ (b) ☐ (c) ☒ (d) ☐
2. (a) ☐ (b) ☐ (c) ☒ (d) ☐
3. (a) ☐ (b) ☒ (c) ☐ (d) ☐
4. (a) ☐ (b) ☐ (c) ☒ (d) ☒
5. (a) ☒ (b) ☐ (c) ☐ (d) ☐
6.      2 1 . 5 1 0

## 1. Задача

Во встроенном датасете stackloss пусть переменная stack.loss будет зависимой, а переменные Air.Flow, Water.Temp, Acid.Conc - предикторами. Проведите регрессионный анализ и определите значение коэффициента Air.Flow

- (a) -39.9197  
(b) 0.9136  
(c) 0.7156  
(d) 1.2953

Решение

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

## 2. Задача

Стандартизация (z-score scaling) шкал всех предикторов перед выполнением регрессионного анализа:

- (a) позволяет обойтись без проверки на нормальность  
(b) не позволяет правомерно сравнить силу эффектов разнородных предикторов

- (c) центрует значения каждого предиктора  $X_i$  вокруг нуля вместо средней  
(d) позволяет обойтись без проверки состоятельности модели

Решение

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

## 3. Задача

Отметьте условия применимости линейной регрессии

- (a) Дисперсия остатков равна единице  
(b) Линейная связь  
(c) Независимые переменные стандартизованы  
(d) Корреляция между независимыми переменными (мультиколлинеарность) велика

Решение

- (a) False  
(b) True  
(c) False

(d) False

#### 4. Задача

Во множественной регрессии, описываемой моделью  $Y = 0.01 + 1.4X_1 - 4.3X_2 + 0.8X_3$ , интерсепт это:

- (a) угловой коэффициент регрессионной прямой
- (b) один из предикторов модели
- (c) один из параметров модели
- (d) ожидаемое среднее значение зависимой переменной, когда все предикторы ( $X_1, X_2, X_3$ ) равны нулю

Решение

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

#### 5. Задача

Поправка adjusted  $R^2$ :

- (a) позволяет скорректировать рост  $R^2$  при добавлении каждого нового предиктора
- (b) применима только для моделей с предварительно стандартизованными предикторами
- (c) всегда уменьшается с увеличением количества предикторов
- (d) позволяет определить наиболее значимый предиктор множественной модели

Решение

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

#### 6. Задача

---

#### ДОПОЛНИТЕЛЬНЫЙ ВОПРОС

---

Для встроенного датасета `stackloss` рассчитайте (предскажите) значение зависимой переменной `stack.loss` при условии, что предиктор `Air.Flow` принимает

значение 66, а остальные независимые переменные принимают свои средние значения. Запишите результат, округленный до третьего знака.

Решение

Сначала проведем регрессионный анализ, получим значения интерсепта и коэффициентов регрессии

```
stmod <- lm(stack.loss ~ ., data = stackloss)
```

Предсказать конкретное значение зависимой переменной можно выполнив код:

```
predict <- c(coefficients(stmod)[1] +  
coefficients(stmod)[2]*air +  
coefficients(stmod)[3]*mean(stackloss$Water.Temp) +  
coefficients(stmod)[4]*mean(stackloss$Acid.Conc.))
```

```
answer <- round(predict,3)
```

Вместо "air" подставьте нужное значение `Air.Flow`

СПбГУ

2014-11-13

Матметоды — Множественная регрессия  
Вар.№ 06

Ф.И.О.: \_\_\_\_\_

—

1. (a) ☐ (b) ☐ (c) ☐ (d) ☒
2. (a) ☒ (b) ☒ (c) ☐ (d) ☐
3. (a) ☒ (b) ☒ (c) ☒ (d) ☒
4. (a) ☐ (b) ☒ (c) ☐ (d) ☒
5. (a) ☐ (b) ☒ (c) ☐ (d) ☐
6.      1 7 . 9 3 0

### 1. Задача

Во встроенном датасете stackloss пусть переменная stack.loss будет зависимой, а переменные Air.Flow, Water.Temp, Acid.Conc - предикторами. Проведите регрессионный анализ и определите значение коэффициента Acid.Conc.

- (a) -39.9197  
(b) NA  
(c) 1.2953  
(d) 0.7156

Решение

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

### 2. Задача

Стандартизация (z-score scaling) шкал всех предикторов перед выполнением регрессионного анализа:

- (a) центрует значения каждого предиктора  $X_i$  вокруг нуля вместо средней  
(b) позволяет правомерно сравнить силу эффектов разнородных предикторов

- (c) не позволяет правомерно сравнить силу эффектов разнородных предикторов  
(d) позволяет обойтись без проверки состоятельности модели

Решение

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

### 3. Задача

Отметьте условия применимости линейной регрессии

- (a) Независимость значений  $y$  друг от друга  
(b) Гомогенность дисперсий остатков  
(c) Линейная связь  
(d) Нормальное распределение остатков

Решение

- (a) True  
(b) True  
(c) True  
(d) True

#### 4. Задача

Во множественной регрессии, описываемой моделью  $Y = 0.01 + 1.4X_1 - 4.3X_2 + 0.8X_3$ , интерсепт это:

- (a) один из предикторов модели
- (b) ожидаемое среднее значение зависимой переменной, когда все предикторы ( $X_1, X_2, X_3$ ) равны нулю
- (c) скорость изменения зависимой переменной при фиксированном значении предиктора  $X_i$
- (d) один из параметров модели

Решение

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

#### 5. Задача

Поправка adjusted  $R^2$ :

- (a) всегда растет с увеличением количества предикторов
- (b) позволяет скорректировать рост  $R^2$  при добавлении каждого нового предиктора
- (c) всегда уменьшается с увеличением количества предикторов
- (d) применима только для моделей с предварительно стандартизированными предикторами

Решение

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

#### 6. Задача

---

#### ДОПОЛНИТЕЛЬНЫЙ ВОПРОС

---

Для встроенного датасета `stackloss` рассчитайте (предскажите) значение зависимой переменной `stack.loss` при условии, что предиктор `Air.Flow` принимает

значение 61, а остальные независимые переменные принимают свои средние значения. Запишите результат, округленный до третьего знака.

Решение

Сначала проведем регрессионный анализ, получим значения интерсепта и коэффициентов регрессии

```
stmod <- lm(stack.loss ~ ., data = stackloss)
```

Предсказать конкретное значение зависимой переменной можно выполнив код:

```
predict <- c(coefficients(stmod)[1] +  
coefficients(stmod)[2]*air +  
coefficients(stmod)[3]*mean(stackloss$Water.Temp) +  
coefficients(stmod)[4]*mean(stackloss$Acid.Conc.))
```

```
answer <- round(predict,3)
```

Вместо "air" подставьте нужное значение `Air.Flow`

Ф.И.О.: \_\_\_\_\_

- [illegible]

## 1. Задача

Во встроенном датасете `stackloss` пусть переменная `stack.loss` будет зависимой, а переменные `Air.Flow`, `Water.Temp`, `Acid.Conc` - предикторами. Проведите регрессионный анализ и определите значение коэффициента `stack.loss`

- (a) 0.7156  
(b) 1.2953  
(c) NA  
(d) -0.1521

Решение

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

## 2. Задача

Стандартизация (z-score scaling) шкал всех предикторов перед выполнением регрессионного анализа:

- (a) позволяет обойтись без проверки на нормальность
- (b) позволяет обойтись без проверки состоятельности модели
- (c) позволяет правомерно сравнить силу эффектов разнородных предикторов

- (d) центрует значения зависимой переменной  $Y$  вокруг нуля вместо средней

Решение

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

### 3. Задача

Отметьте условия применимости линейной регрессии

- (a) Независимость значений  $y$  друг от друга
- (b) Независимые переменные стандартизованы
- (c) Линейная связь
- (d) Дисперсия остатков равна единице

Решение

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

## 4. Задача

Во множественной регрессии, описываемой моделью  $Y = 0.01 + 1.4X_1 - 4.3X_2 + 0.8X_3$ , интерсепт это:

- (a) один из предикторов модели
- (b) один из параметров модели
- (c) ожидаемое среднее значение зависимой переменной, когда все предикторы ( $X_1, X_2, X_3$ ) равны нулю
- (d) угловой коэффициент регрессионной прямой

Решение

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

## 5. Задача

Поправка adjusted  $R^2$ :

- (a) всегда растёт с увеличением количества предикторов
- (b) всегда уменьшается с увеличением количества предикторов
- (c) применима только для моделей с предварительно стандартизированными предикторами
- (d) позволяет скорректировать рост  $R^2$  при добавлении каждого нового предиктора

Решение

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

## 6. Задача

---

### ДОПОЛНИТЕЛЬНЫЙ ВОПРОС

---

Для встроенного датасета `stackloss` рассчитайте (предскажите) значение зависимой переменной `stack.loss` при условии, что предиктор `Air.Flow` принимает

значение 65, а остальные независимые переменные принимают свои средние значения. Запишите результат, округленный до третьего знака.

Решение

Сначала проведем регрессионный анализ, получим значения интерсепта и коэффициентов регрессии

```
stmod <- lm(stack.loss ~ ., data = stackloss)
```

Предсказать конкретное значение зависимой переменной можно выполнив код:

```
predict <- c(coefficients(stmod)[1] +
coefficients(stmod)[2]*air +
coefficients(stmod)[3]*mean(stackloss$Water.Temp) +
coefficients(stmod)[4]*mean(stackloss$Acid.Conc.))
```

```
answer <- round(predict,3)
```

Вместо "air" подставьте нужное значение `Air.Flow`