

Ф.И.О.: _____

1. (a) ☐ X (b) ☐ X (c) ☐ X (d) ☐

2. (a) ☐ (b) ☐ (c) ☐ (d) ☐ X

3.

| | | | | | |
|--|--|--|--|---|---|
| | | | | 2 | 6 |
|--|--|--|--|---|---|

 .

| | | |
|---|---|---|
| 1 | 0 | 0 |
|---|---|---|

4.

| | | | | | |
|--|--|--|--|---|---|
| | | | | 2 | 6 |
|--|--|--|--|---|---|

 .

| | | |
|---|---|---|
| 0 | 0 | 0 |
|---|---|---|

5. (a) ☐ X (b) ☐ X (c) ☐ (d) ☐

6. (a) ☐ X (b) ☐ (c) ☐ X (d) ☐ X

1. Задача

Какие из этих утверждений являются условиями применимости дисперсионного анализа

- (a) Гомогенность дисперсий остатков
- (b) Нормальное распределение остатков
- (c) Случайность и независимость групп и наблюдений внутри групп
- (d) Значимое влияние факторов

Решение

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

2. Задача

Ошибка I рода — это...

- (a) отказ от ложной нулевой гипотезы
- (b) различий не обнаружено там, где они есть
- (c) неправильный результат теста
- (d) различия считаются достоверными, когда нулевая гипотеза верна

Решение

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

3. Задача

Загрузите датасет `InsectSprays`, выполнив код

```
data(InsectSprays)
```

В этом датасете собраны данные об числе насекомых на участках, обработанных различными инсектицидами. Проверьте, различается ли эффективность инсектицидов. В качестве зависимой переменной используйте квадратный корень из числа насекомых (функция `sqrt()`).

Каково значение суммы квадратов, оценивающей изменчивость оставшуюся необъясненной? Округлите до десятых

Решение

Посчитать дисперсионный анализ можно, выполнив код

```
data(InsectSprays)
fit <- lm(sqrt(count) ~ spray, data = InsectSprays)
library(car)
Anova(fit)
```

Anova Table (Type II tests)

```
Response: sqrt(count)
      Sum Sq Df F value    Pr(>F)
spray   88.438  5  44.799 < 2.2e-16 ***
Residuals 26.058 66
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Значение суммы квадратов, оценивающей изменчивость оставшуюся необъясненной
26.1.

4. Задача

Австралийские фермеры изучали число семян в помидорах разных сортов. Всего было три сорта, или три градации дискретного предиктора: желтый (Yellow, X_{level1}), красный (Red, X_{level2}) и черный (Black, X_{level3}). Были построены две модели, связывающие зависимую переменную с предиктором:

| | Names | Base_Yellow | Base_Red |
|---|-----------|-------------|----------|
| 1 | Intercept | 19 | 20 |
| 2 | yellow | NA | -1 |
| 3 | red | 1 | NA |
| 4 | black | 7 | 6 |

Рассчитайте (предскажите) значение интерсепта для модели, в которой в качестве базового уровня выбраны черные помидоры.

Решение

Интерсептом в модели, где в качестве базового уровня выбраны черные помидоры будет предсказанное среднее зависимой переменной при таком уровне предиктора. То есть

```
> Base_Yellow[1]+Base_Yellow[4]
```

```
[1] 26
```

или

```
> Base_Red[1]+Base_Red[4]
```

```
[1] 26
```

5. Задача

Дополните предложение. Дисперсионный анализ:

- (a) проверяет, есть ли различия между средними значениями зависимой переменной при разных уровнях предиктора
- (b) тестирует гипотезу о том, что все групповые средние равны
- (c) тестирует гипотезу о равенстве дисперсий в группах
- (d) делает поправку для снижения вероятности ошибки II рода

Решение

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

6. Задача

ДОПОЛНИТЕЛЬНЫЙ ВОПРОС

Датафрейм WBC содержит информацию о выбросах CO2 на душу населения в разных странах.

Предикторы: POP - непрерывный, численность населения, income - дискретный, ВВП на душу населения (градации rich и poor), demographic - дискретный, место проживания большинства населения (градации rural и urban).

Сравните три модели: Модель 1

`lm(formula = CO2 ~ POP + demographic + income, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|------------------|----------|------------|---------|----------|
| (Intercept) | 137.719 | 114.292 | 1.205 | 0.232 |
| POP | 421.102 | 70.491 | 5.974 | 0.000 |
| demographicUrban | 21.155 | 143.136 | 0.148 | 0.883 |
| incomeRich | 541.192 | 164.077 | 3.298 | 0.002 |

R^2 Adjusted R^2

[1] 0.395 0.369

Модель 2

`lm(formula = CO2 ~ POP + income, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 150.036 | 77.675 | 1.932 | 0.057 |
| POP | 418.778 | 68.247 | 6.136 | 0.000 |
| incomeRich | 535.703 | 158.730 | 3.375 | 0.001 |

R^2 Adjusted R^2

[1] 0.395 0.378

Модель 3

`lm(formula = CO2 ~ POP + demographic, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|------------------|----------|------------|---------|----------|
| (Intercept) | 323.171 | 106.113 | 3.046 | 0.003 |
| POP | 397.403 | 74.780 | 5.314 | 0.000 |
| demographicUrban | -85.704 | 148.679 | -0.576 | 0.566 |

R^2 Adjusted R^2

[1] 0.302 0.283

Определите, правомерно или нет исключение факторов, и почему. Выберите правильные утверждения.

- (a) можно исключать предиктор `demographic`, без него модель лучше объясняет изменчивость
- (b) нельзя исключать ни одного из предикторов, они все значимы
- (c) нельзя исключать предиктор `POP`, он значим
- (d) нельзя исключать предиктор `income`, без него модель хуже объясняет изменчивость

Решение

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

Ф.И.О.: _____

1. (a) ☐ X (b) ☐ (c) ☐ X (d) ☐ X

2. (a) ☐ X (b) ☐ X (c) ☐ (d) ☐ X

3.

| | | | | | |
|--|--|--|--|---|---|
| | | | | 2 | 6 |
|--|--|--|--|---|---|

 .

| | | |
|---|---|---|
| 1 | 0 | 0 |
|---|---|---|

4.

| | | | | | |
|--|--|--|--|---|---|
| | | | | 2 | 4 |
|--|--|--|--|---|---|

 .

| | | |
|---|---|---|
| 0 | 0 | 0 |
|---|---|---|

5. (a) ☐ X (b) ☐ (c) ☐ (d) ☐

6. (a) ☐ X (b) ☐ X (c) ☐ (d) ☐ X

1. Задача

Какие из этих утверждений являются условиями применимости дисперсионного анализа

- (a) Случайность и независимость групп и наблюдений внутри групп
- (b) Значимое влияние факторов
- (c) Гомогенность дисперсий остатков
- (d) Нормальное распределение остатков

Решение

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

2. Задача

Ошибка II рода — это...

- (a) невозможность отказа от ложной нулевой гипотезы
- (b) достоверных различий не найдено когда альтернативная гипотеза верна
- (c) невозможность отказа от истинной нулевой гипотезы
- (d) различий не обнаружено там, где они есть

Решение

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

3. Задача

Загрузите датасет `InsectSprays`, выполнив код

```
data(InsectSprays)
```

В этом датасете собраны данные об числе насекомых на участках, обработанных различными инсектицидами. Проверьте, различается ли эффективность инсектицидов. В качестве зависимой переменной используйте квадратный корень из числа насекомых (функция `sqrt()`).

Каково значение суммы квадратов, оценивающей изменчивость оставшуюся необъясненной? Округлите до десятых

Решение

Посчитать дисперсионный анализ можно, выполнив код

```
data(InsectSprays)
fit <- lm(sqrt(count) ~ spray, data = InsectSprays)
library(car)
Anova(fit)
```

Anova Table (Type II tests)

```
Response: sqrt(count)
      Sum Sq Df F value    Pr(>F)
spray   88.438  5  44.799 < 2.2e-16 ***
Residuals 26.058 66
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Значение суммы квадратов, оценивающей изменчивость оставшуюся необъясненной
26.1.

4. Задача

Австралийские фермеры изучали число семян в помидорах разных сортов. Всего было три сорта, или три градации дискретного предиктора: желтый (Yellow, X_{level1}), красный (Red, X_{level2}) и черный (Black, X_{level3}). Были построены две модели, связывающие зависимую переменную с предиктором:

| | Names | Base_Yellow | Base_Red |
|---|-----------|-------------|----------|
| 1 | Intercept | 17 | 18 |
| 2 | yellow | NA | -1 |
| 3 | red | 1 | NA |
| 4 | black | 7 | 6 |

Рассчитайте (предскажите) значение интерсепта для модели, в которой в качестве базового уровня выбраны черные помидоры.

Решение

Интерсептом в модели, где в качестве базового уровня выбраны черные помидоры будет предсказанное среднее зависимой переменной при таком уровне предиктора. То есть

```
> Base_Yellow[1]+Base_Yellow[4]
```

```
[1] 24
```

или

```
> Base_Red[1]+Base_Red[4]
```

```
[1] 24
```

5. Задача

Дополните предложение. Дисперсионный анализ:

- (a) проверяет нулевую гипотезу о том, что все группы принадлежат к одной генеральной совокупности
- (b) тестирует альтернативную гипотезу об отсутствии различий между средними значениями в группах по фактору
- (c) ничем не отличается от множественных сравнений

- (d) имеет альтернативную гипотезу о том, что все групповые средние отличаются друг от друга

Решение

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

6. Задача

ДОПОЛНИТЕЛЬНЫЙ ВОПРОС

Датафрейм WBC содержит информацию о выбросах CO₂ на душу населения в разных странах.

Предикторы: POP - непрерывный, численность населения, income - дискретный, ВВП на душу населения (градации rich и poor), demographic - дискретный, место проживания большинства населения (градации rural и urban).

Сравните три модели: Модель 1

`lm(formula = CO2 ~ POP + demographic + income, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|------------------|----------|------------|---------|----------|
| (Intercept) | 122.983 | 102.398 | 1.201 | 0.234 |
| POP | 413.547 | 69.829 | 5.922 | 0.000 |
| demographicUrban | 57.448 | 140.556 | 0.409 | 0.684 |
| incomeRich | 523.957 | 162.222 | 3.230 | 0.002 |

R^2 Adjusted R^2

[1] 0.396 0.371

Модель 2

`lm(formula = CO2 ~ POP + income, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 150.036 | 77.675 | 1.932 | 0.057 |
| POP | 418.778 | 68.247 | 6.136 | 0.000 |
| incomeRich | 535.703 | 158.730 | 3.375 | 0.001 |

R^2 Adjusted R^2

[1] 0.395 0.378

Модель 3

`lm(formula = CO2 ~ POP + demographic, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|------------------|----------|------------|---------|----------|
| (Intercept) | 206.912 | 105.334 | 1.964 | 0.053 |
| POP | 394.400 | 73.994 | 5.330 | 0.000 |
| demographicUrban | 137.871 | 147.115 | 0.937 | 0.352 |

R^2 Adjusted R^2

[1] 0.307 0.288

Определите, правомерно или нет исключение факторов, и почему. Выберите правильные утверждения.

- (a) нельзя исключать предиктор `income`, без него модель хуже объясняет изменчивость
- (b) можно исключать предиктор `demographic`, без него модель лучше объясняет изменчивость
- (c) нельзя исключать ни одного из предикторов, они все значимы
- (d) нельзя исключать предиктор `POP`, он значим

Решение

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

Ф.И.О.: _____

1. (a) ☐ (b) ☐ (c) ☐ (d) ☐

2. (a) ☐ (b) ☐ (c) ☐ (d) ☐

3.

| | | | | | |
|---|---|---|---|---|---|
| 2 | 3 | 1 | 1 | 2 | 9 |
|---|---|---|---|---|---|

 .

| | | |
|---|---|---|
| 2 | 0 | 0 |
|---|---|---|

4.

| | | | | | |
|--|--|--|--|---|---|
| | | | | 3 | 1 |
|--|--|--|--|---|---|

 .

| | | |
|---|---|---|
| 0 | 0 | 0 |
|---|---|---|

5. (a) ☐ (b) ☐ (c) ☐ (d) ☐

6. (a) ☐ (b) ☐ (c) ☐ (d) ☐

1. Задача

Какие из этих утверждений являются условиями применимости дисперсионного анализа

- (a) Корреляция между факторами
- (b) Коэффициент детерминации $R^2 > 0.8$
- (c) Равные размеры выборок в группах по фактору
- (d) Гомогенность дисперсий остатков

Решение

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

2. Задача

Ошибка II рода — это...

- (a) различий не обнаружено там, где они есть
- (b) неправильный результат теста
- (c) невозможность отказа от истинной нулевой гипотезы
- (d) достоверных различий не найдено когда альтернативная гипотеза верна

Решение

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

3. Задача

Загрузите датасет `chickwts`, выполнив код

```
data(chickwts)
```

В этом датасете собраны данные об влиянии различных пищевых добавок на вес цыплят. Проверьте, различается ли эффективность пищевых добавок.

Каково значение суммы квадратов по фактору? Округлите до десятых

Решение

Посчитать дисперсионный анализ можно, выполнив код

```
data(chickwts)
fit <- lm(weight ~ feed, data = chickwts)
library(car)
Anova(fit)
```

Anova Table (Type II tests)

Response: weight

| | Sum Sq | Df | F value | Pr(>F) |
|-----------|--------|----|---------|---------------|
| feed | 231129 | 5 | 15.365 | 5.936e-10 *** |
| Residuals | 195556 | 65 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Значение суммы квадратов по фактору **231129.2**.

4. Задача

Австралийские фермеры изучали число семян в помидорах разных сортов. Всего было три сорта, или три градации дискретного предиктора: желтый (Yellow, X_{level1}), красный (Red, X_{level2}) и черный (Black, X_{level3}). Были построены две модели, связывающие зависимую переменную с предиктором:

| | Names | Base_Yellow | Base_Red |
|---|-----------|-------------|----------|
| 1 | Intercept | 24 | 25 |
| 2 | yellow | NA | -1 |
| 3 | red | 1 | NA |
| 4 | black | 7 | 6 |

Рассчитайте (предскажите) значение интерсепта для модели, в которой в качестве базового уровня выбраны черные помидоры.

Решение

Интерсептом в модели, где в качестве базового уровня выбраны черные помидоры будет предсказанное среднее зависимой переменной при таком уровне предиктора. То есть

```
> Base_Yellow[1]+Base_Yellow[4]
```

```
[1] 31
```

или

```
> Base_Red[1]+Base_Red[4]
```

```
[1] 31
```

5. Задача

Дополните предложение. Дисперсионный анализ:

- (a) имеет альтернативную гипотезу о том, что все групповые средние отличаются друг от друга
- (b) тестирует гипотезу о равенстве дисперсий в группах
- (c) делает поправку для снижения вероятности ошибки II рода
- (d) проверяет, есть ли различия между средними значениями зависимой переменной при разных уровнях предиктора

Решение

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

6. Задача

ДОПОЛНИТЕЛЬНЫЙ ВОПРОС

Датафрейм WBC содержит информацию о выбросах CO2 на душу населения в разных странах.

Предикторы: POP - непрерывный, численность населения, income - дискретный, ВВП на душу населения (градации rich и poor), demographic - дискретный, место проживания большинства населения (градации rural и urban).

Сравните три модели: Модель 1

`lm(formula = CO2 ~ POP + demographic + income, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|------------------|----------|------------|---------|----------|
| (Intercept) | 49.993 | 102.253 | 0.489 | 0.626 |
| POP | 432.534 | 68.308 | 6.332 | 0.000 |
| demographicUrban | 202.812 | 136.330 | 1.488 | 0.141 |
| incomeRich | 513.123 | 158.139 | 3.245 | 0.002 |

R^2 Adjusted R^2

[1] 0.413 0.388

Модель 2

`lm(formula = CO2 ~ POP + income, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 150.036 | 77.675 | 1.932 | 0.057 |
| POP | 418.778 | 68.247 | 6.136 | 0.000 |
| incomeRich | 535.703 | 158.730 | 3.375 | 0.001 |

R^2 Adjusted R^2

[1] 0.395 0.378

Модель 3

`lm(formula = CO2 ~ POP + demographic, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|------------------|----------|------------|---------|----------|
| (Intercept) | 151.065 | 103.638 | 1.458 | 0.149 |
| POP | 423.569 | 72.628 | 5.832 | 0.000 |
| demographicUrban | 245.270 | 144.401 | 1.699 | 0.094 |

R^2 Adjusted R^2

[1] 0.326 0.307

Определите, правомерно или нет исключение факторов, и почему. Выберите правильные утверждения.

- (a) нельзя исключать предиктор income, без него модель хуже объясняет изменчивость
- (b) нельзя исключать предиктор POP, он значим
- (c) нельзя исключать предиктор demographic, без него модель хуже объясняет изменчивость
- (d) можно исключать предиктор demographic, без него модель лучше объясняет изменчивость

Решение

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

Ф.И.О.: _____

1. (a) ☐ (b) ☐ X (c) ☐ X (d) ☐ X

2. (a) ☐ X (b) ☐ X (c) ☐ (d) ☐ X

3.

| | | | | | |
|---|---|---|---|---|---|
| 1 | 9 | 5 | 5 | 5 | 6 |
|---|---|---|---|---|---|

 .

| | | |
|---|---|---|
| 0 | 0 | 0 |
|---|---|---|

4.

| | | | | | |
|--|--|--|--|---|---|
| | | | | 2 | 9 |
|--|--|--|--|---|---|

 .

| | | |
|---|---|---|
| 0 | 0 | 0 |
|---|---|---|

5. (a) ☐ X (b) ☐ (c) ☐ X (d) ☐

6. (a) ☐ (b) ☐ X (c) ☐ (d) ☐

1. Задача

Какие из этих утверждений являются условиями применимости дисперсионного анализа

- (a) Значимое влияние факторов
- (b) Случайность и независимость групп и наблюдений внутри групп
- (c) Гомогенность дисперсий остатков
- (d) Нормальное распределение остатков

Решение

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

2. Задача

Ошибка II рода — это...

- (a) невозможность отказа от ложной нулевой гипотезы
- (b) различий не обнаружено там, где они есть
- (c) отказ от истинной нулевой гипотезы
- (d) достоверных различий не найдено когда альтернативная гипотеза верна

Решение

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

3. Задача

Загрузите датасет `chickwts`, выполнив код

```
data(chickwts)
```

В этом датасете собраны данные об влиянии различных пищевых добавок на вес цыплят. Проверьте, различается ли эффективность пищевых добавок.

Каково значение суммы квадратов, оценивающей изменчивость оставшуюся необъясненной? Округлите до десятых

Решение

Посчитать дисперсионный анализ можно, выполнив код

```
data(chickwts)
fit <- lm(weight ~ feed, data = chickwts)
library(car)
Anova(fit)
```

Anova Table (Type II tests)

Response: weight

| | Sum Sq | Df | F value | Pr(>F) |
|-----------|--------|----|---------|---------------|
| feed | 231129 | 5 | 15.365 | 5.936e-10 *** |
| Residuals | 195556 | 65 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Значение суммы квадратов, оценивающей изменчивость оставшуюся необъясненной **195556**.

4. Задача

Австралийские фермеры изучали число семян в помидорах разных сортов. Всего было три сорта, или три градации дискретного предиктора: желтый (Yellow, X_{level1}), красный (Red, X_{level2}) и черный (Black, X_{level3}). Были построены две модели, связывающие зависимую переменную с предиктором:

| | Names | Base_Yellow | Base_Red |
|---|-----------|-------------|----------|
| 1 | Intercept | 22 | 23 |
| 2 | yellow | NA | -1 |
| 3 | red | 1 | NA |
| 4 | black | 7 | 6 |

Рассчитайте (предскажите) значение интерсепта для модели, в которой в качестве базового уровня выбраны черные помидоры.

Решение

Интерсептом в модели, где в качестве базового уровня выбраны черные помидоры будет предсказанное среднее зависимой переменной при таком уровне предиктора. То есть

```
> Base_Yellow[1]+Base_Yellow[4]
```

```
[1] 29
```

или

```
> Base_Red[1]+Base_Red[4]
```

```
[1] 29
```

5. Задача

Дополните предложение. Дисперсионный анализ:

- (a) проверяет нулевую гипотезу о том, что все группы принадлежат к одной генеральной совокупности
- (b) тестирует альтернативную гипотезу об отсутствии различий между средними значениями в группах по фактору
- (c) тестирует гипотезу о том, что все групповые средние равны

(d) делает поправку для снижения вероятности ошибки II рода

Решение

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

6. Задача

ДОПОЛНИТЕЛЬНЫЙ ВОПРОС

Датафрейм WBC содержит информацию о выбросах CO2 на душу населения в разных странах.

Предикторы: POP - непрерывный, численность населения, income - дискретный, ВВП на душу населения (градации rich и poor), demographic - дискретный, место проживания большинства населения (градации rural и urban).

Сравните три модели: Модель 1

`lm(formula = CO2 ~ POP + demographic + income, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|------------------|----------|------------|---------|----------|
| (Intercept) | 210.499 | 107.751 | 1.954 | 0.055 |
| POP | 425.997 | 68.985 | 6.175 | 0.000 |
| demographicUrban | -111.574 | 137.453 | -0.812 | 0.420 |
| incomeRich | 525.521 | 159.601 | 3.293 | 0.002 |

R^2 Adjusted R^2

[1] 0.400 0.375

Модель 2

`lm(formula = CO2 ~ POP + income, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 150.036 | 77.675 | 1.932 | 0.057 |
| POP | 418.778 | 68.247 | 6.136 | 0.000 |
| incomeRich | 535.703 | 158.730 | 3.375 | 0.001 |

R^2 Adjusted R^2

[1] 0.395 0.378

Модель 3

`lm(formula = CO2 ~ POP + demographic, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|------------------|----------|------------|---------|----------|
| (Intercept) | 355.121 | 104.902 | 3.385 | 0.001 |
| POP | 416.131 | 73.480 | 5.663 | 0.000 |
| demographicUrban | -147.147 | 146.094 | -1.007 | 0.317 |

R^2 Adjusted R^2

[1] 0.309 0.289

Определите, правомерно или нет исключение факторов, и почему. Выберите правильные утверждения.

- (a) нельзя исключать ни одного из предикторов, они все значимы
- (b) нельзя исключать предиктор `income`, без него модель хуже объясняет изменчивость
- (c) нельзя исключать предиктор `demographic`, без него модель хуже объясняет изменчивость
- (d) можно исключать предиктор `income`, он не значим

Решение

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

Ф.И.О.: _____

1. (a) ☐ (b) ☐ (c) ☐ (d) ☐

2. (a) ☐ (b) ☐ (c) ☐ (d) ☐

3.

| | | | | | |
|---|---|---|---|---|---|
| 2 | 3 | 1 | 1 | 2 | 9 |
|---|---|---|---|---|---|

 .

| | | |
|---|---|---|
| 2 | 0 | 0 |
|---|---|---|

4.

| | | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|---|---|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | 2 | 3 |
|--------------------------|--------------------------|--------------------------|--------------------------|---|---|

 .

| | | |
|---|---|---|
| 0 | 0 | 0 |
|---|---|---|

5. (a) ☐ (b) ☐ (c) ☐ (d) ☐

6. (a) ☐ (b) ☐ (c) ☐ (d) ☐

1. Задача

Какие из этих утверждений являются условиями применимости дисперсионного анализа

- (a) Значимое влияние факторов
- (b) Равные размеры выборок в группах по фактору
- (c) Дисперсия остатков равна единице
- (d) Случайность и независимость групп и наблюдений внутри групп

Решение

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

2. Задача

Ошибка II рода — это...

- (a) различия считаются достоверными, когда альтернативная гипотеза верна
- (b) различий не обнаружено там, где они есть
- (c) достоверных различий не найдено когда альтернативная гипотеза верна
- (d) невозможность отказа от ложной нулевой гипотезы

Решение

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

3. Задача

Загрузите датасет `chickwts`, выполнив код

```
data(chickwts)
```

В этом датасете собраны данные об влиянии различных пищевых добавок на вес цыплят. Проверьте, различается ли эффективность пищевых добавок.

Каково значение суммы квадратов по фактору? Округлите до десятых

Решение

Посчитать дисперсионный анализ можно, выполнив код

```
data(chickwts)
fit <- lm(weight ~ feed, data = chickwts)
library(car)
Anova(fit)
```

Anova Table (Type II tests)

Response: weight

| | Sum Sq | Df | F value | Pr(>F) |
|-----------|--------|----|---------|---------------|
| feed | 231129 | 5 | 15.365 | 5.936e-10 *** |
| Residuals | 195556 | 65 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Значение суммы квадратов по фактору **231129.2**.

4. Задача

Австралийские фермеры изучали число семян в помидорах разных сортов. Всего было три сорта, или три градации дискретного предиктора: желтый (Yellow, X_{level1}), красный (Red, X_{level2}) и черный (Black, X_{level3}). Были построены две модели, связывающие зависимую переменную с предиктором:

| | Names | Base_Yellow | Base_Red |
|---|-----------|-------------|----------|
| 1 | Intercept | 16 | 17 |
| 2 | yellow | NA | -1 |
| 3 | red | 1 | NA |
| 4 | black | 7 | 6 |

Рассчитайте (предскажите) значение интерсепта для модели, в которой в качестве базового уровня выбраны черные помидоры.

Решение

Интерсептом в модели, где в качестве базового уровня выбраны черные помидоры будет предсказанное среднее зависимой переменной при таком уровне предиктора. То есть

```
> Base_Yellow[1]+Base_Yellow[4]
```

```
[1] 23
```

или

```
> Base_Red[1]+Base_Red[4]
```

```
[1] 23
```

5. Задача

Дополните предложение. Дисперсионный анализ:

- (a) тестирует гипотезу о равенстве дисперсий в группах
- (b) ничем не отличается от множественных сравнений
- (c) имеет альтернативную гипотезу о том, что хотя бы одна из групповых средних отличается от прочих
- (d) тестирует альтернативную гипотезу об отсутствии различий между средними значениями в группах по фактору

Решение

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

6. Задача

ДОПОЛНИТЕЛЬНЫЙ ВОПРОС

Датафрейм WBC содержит информацию о выбросах CO2 на душу населения в разных странах.

Предикторы: POP - непрерывный, численность населения, income - дискретный, ВВП на душу населения (градации rich и poor), demographic - дискретный, место проживания большинства населения (градации rural и urban).

Сравните три модели: Модель 1

`lm(formula = CO2 ~ POP + demographic + income, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|------------------|----------|------------|---------|----------|
| (Intercept) | 19.617 | 108.748 | 0.180 | 0.857 |
| POP | 436.019 | 68.149 | 6.398 | 0.000 |
| demographicUrban | 231.381 | 136.791 | 1.692 | 0.095 |
| incomeRich | 577.792 | 158.680 | 3.641 | 0.001 |

R^2 Adjusted R^2

[1] 0.418 0.393

Модель 2

`lm(formula = CO2 ~ POP + income, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 150.036 | 77.675 | 1.932 | 0.057 |
| POP | 418.778 | 68.247 | 6.136 | 0.000 |
| incomeRich | 535.703 | 158.730 | 3.375 | 0.001 |

R^2 Adjusted R^2

[1] 0.395 0.378

Модель 3

`lm(formula = CO2 ~ POP + demographic, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|------------------|----------|------------|---------|----------|
| (Intercept) | 198.902 | 104.893 | 1.896 | 0.062 |
| POP | 417.069 | 73.507 | 5.674 | 0.000 |
| demographicUrban | 153.276 | 146.147 | 1.049 | 0.298 |

R^2 Adjusted R^2

[1] 0.309 0.290

Определите, правомерно или нет исключение факторов, и почему. Выберите правильные утверждения.

- (a) нельзя исключать предиктор income, он значим
- (b) можно исключать предиктор POP, он не значим
- (c) можно исключать предиктор demographic, без него модель лучше объясняет изменчивость
- (d) можно исключать предиктор income, он значим, но без него модель лучше объясняет изменчивость

Решение

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

Ф.И.О.: _____

1. (a) ☐ (b) ☐ X (c) ☐ X (d) ☐ X

2. (a) ☐ (b) ☐ X (c) ☐ X (d) ☐

3.

| | | | | | |
|--|--|--|--|---|---|
| | | | | 2 | 6 |
|--|--|--|--|---|---|

 .

| | | |
|---|---|---|
| 1 | 0 | 0 |
|---|---|---|

4.

| | | | | | |
|--|--|--|--|---|---|
| | | | | 2 | 7 |
|--|--|--|--|---|---|

 .

| | | |
|---|---|---|
| 0 | 0 | 0 |
|---|---|---|

5. (a) ☐ X (b) ☐ X (c) ☐ X (d) ☐ X

6. (a) ☐ (b) ☐ X (c) ☐ (d) ☐ X

1. Задача

Какие из этих утверждений являются условиями применимости дисперсионного анализа

- (a) Дисперсия остатков равна единице
- (b) Нормальное распределение остатков
- (c) Случайность и независимость групп и наблюдений внутри групп
- (d) Гомогенность дисперсий остатков

Решение

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

2. Задача

Ошибка I рода — это...

- (a) отказ от ложной нулевой гипотезы
- (b) отказ от истинной нулевой гипотезы
- (c) различия найдены там, где их нет
- (d) достоверных различий не найдено, когда нулевая гипотеза верна

Решение

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

3. Задача

Загрузите датасет `InsectSprays`, выполнив код

```
data(InsectSprays)
```

В этом датасете собраны данные об числе насекомых на участках, обработанных различными инсектицидами. Проверьте, различается ли эффективность инсектицидов. В качестве зависимой переменной используйте квадратный корень из числа насекомых (функция `sqrt()`).

Каково значение суммы квадратов, оценивающей изменчивость оставшуюся необъясненной? Округлите до десятых

Решение

Посчитать дисперсионный анализ можно, выполнив код

```
data(InsectSprays)
fit <- lm(sqrt(count) ~ spray, data = InsectSprays)
library(car)
Anova(fit)
```

Anova Table (Type II tests)

```
Response: sqrt(count)
      Sum Sq Df F value    Pr(>F)
spray   88.438  5  44.799 < 2.2e-16 ***
Residuals 26.058 66
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Значение суммы квадратов, оценивающей изменчивость оставшуюся необъясненной
26.1.

4. Задача

Австралийские фермеры изучали число семян в помидорах разных сортов. Всего было три сорта, или три градации дискретного предиктора: желтый (Yellow, X_{level1}), красный (Red, X_{level2}) и черный (Black, X_{level3}). Были построены две модели, связывающие зависимую переменную с предиктором:

| | Names | Base_Yellow | Base_Red |
|---|-----------|-------------|----------|
| 1 | Intercept | 20 | 21 |
| 2 | yellow | NA | -1 |
| 3 | red | 1 | NA |
| 4 | black | 7 | 6 |

Рассчитайте (предскажите) значение интерсепта для модели, в которой в качестве базового уровня выбраны черные помидоры.

Решение

Интерсептом в модели, где в качестве базового уровня выбраны черные помидоры будет предсказанное среднее зависимой переменной при таком уровне предиктора. То есть

```
> Base_Yellow[1]+Base_Yellow[4]
```

```
[1] 27
```

или

```
> Base_Red[1]+Base_Red[4]
```

```
[1] 27
```

5. Задача

Дополните предложение. Дисперсионный анализ:

- (a) имеет альтернативную гипотезу о том, что хотя бы одна из групповых средних отличается от прочих
- (b) тестирует гипотезу о том, что все групповые средние равны
- (c) проверяет, есть ли различия между средними значениями зависимой переменной при разных уровнях предиктора

- (d) проверяет нулевую гипотезу о том, что все группы принадлежат к одной генеральной совокупности

Решение

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

6. Задача

ДОПОЛНИТЕЛЬНЫЙ ВОПРОС

Датафрейм WBC содержит информацию о выбросах CO2 на душу населения в разных странах.

Предикторы: POP - непрерывный, численность населения, income - дискретный, ВВП на душу населения (градации rich и poor), demographic - дискретный, место проживания большинства населения (градации rural и urban).

Сравните три модели: Модель 1

`lm(formula = CO2 ~ POP + demographic + income, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|------------------|----------|------------|---------|----------|
| (Intercept) | 202.160 | 104.193 | 1.940 | 0.056 |
| POP | 419.163 | 68.455 | 6.123 | 0.000 |
| demographicUrban | -102.473 | 136.012 | -0.753 | 0.454 |
| incomeRich | 540.546 | 159.338 | 3.392 | 0.001 |

R^2 Adjusted R^2

[1] 0.399 0.374

Модель 2

`lm(formula = CO2 ~ POP + income, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 150.036 | 77.675 | 1.932 | 0.057 |
| POP | 418.778 | 68.247 | 6.136 | 0.000 |
| incomeRich | 535.703 | 158.730 | 3.375 | 0.001 |

R^2 Adjusted R^2

[1] 0.395 0.378

Модель 3

`lm(formula = CO2 ~ POP + demographic, data = WBC)`

| | Estimate | Std. Error | t value | Pr(> t) |
|------------------|----------|------------|---------|----------|
| (Intercept) | 322.212 | 104.908 | 3.071 | 0.003 |
| POP | 406.520 | 73.172 | 5.556 | 0.000 |
| demographicUrban | -83.859 | 145.482 | -0.576 | 0.566 |

R^2 Adjusted R^2

[1] 0.302 0.283

Определите, правомерно или нет исключение факторов, и почему. Выберите правильные утверждения.

- (a) нельзя исключать ни одного из предикторов, они все значимы
- (b) нельзя исключать предиктор POP, он значим
- (c) нельзя исключать предиктор demographic, без него модель хуже объясняет изменчивость
- (d) нельзя исключать предиктор income, он значим

Решение

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