

1) Для нахождения коэффициентов возьмём из массивов final и females первые 50 значений в отдельные массивы final50 и females50, составим из них и единичного столбца матрицу Xlong для длинной регрессии и для короткой регрессии сделаем матрицу Xshort только из единичного столбца и final50, так как там мы не учитываем пол, возьмём в массив test50 первые 50 элементов массива test, затем используем формулу $\hat{\alpha} = (X^T X)^{-1} X^T y$, где X - матрица наблюдений (Xlong и Xshort), y - вектор результатов (test50), $\hat{\alpha}$ - МНК оценка весов.

```

>> final50 = [final(1, 1:50)]; >> final50 = transpose(final50);
>> females50 = [females(1, 1:50)]; >> females50 = transpose(females50);
>> test50 = [test(1, 1:50)]; >> test50 = transpose(test50);

>> Xlong = [ones(50, 1), final50, females50] >> Xshort = [ones(50, 1), final50]

Xlong =

     1     8     1
     1     4     1
     1     5     0
     1     5     0
     1     4     0
     1     4     0
     1     5     0
     1     4     0
     1     6     0
     1     2     0
     1     2     0
     1     4     0
     1     5     0
     1     4     0

Xshort =

     1     8
     1     4
     1     5
     1     5
     1     4
     1     4
     1     5
     1     4
     1     6
     1     2
     1     2
     1     4

```

По этой формуле для длинной регрессии вектор коэффициентов $\beta = [\beta_1, \beta_2, \beta_3]$ равен

```
>> coeffLong = transpose(inv(transpose(Xlong)*Xlong)*transpose(Xlong)*test50)
```

```
coeffLong =
```

```
4.4908    0.3210    1.0014
```

Для короткой регрессии вектор коэффициентов $\gamma = [\gamma_1, \gamma_2]$ равен

```
>> coeffShort = transpose(inv(transpose(Xshort)*Xshort)*transpose(Xshort)*test50)
```

```
coeffShort =
```

```
4.6410    0.3190
```

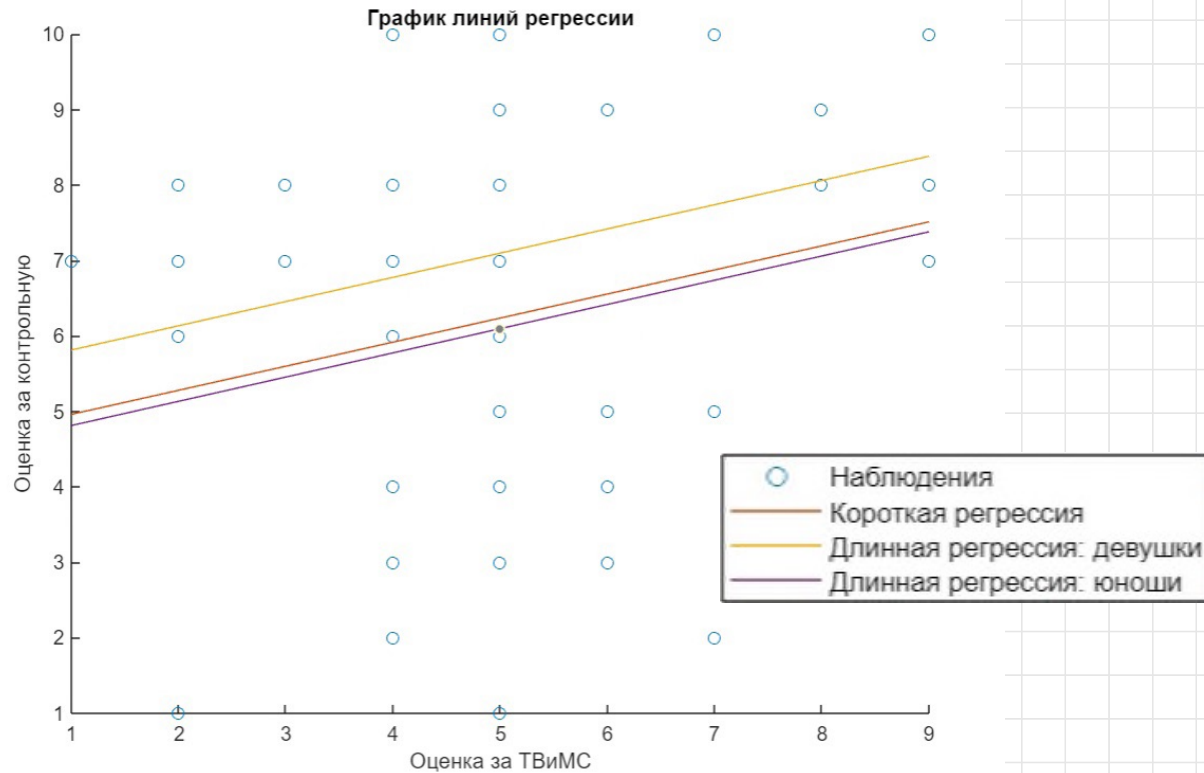
В уравнениях получается

Длинная регрессия: $\text{test50} = 4.4908 + 0.321 \cdot \text{final50} + 1.0014 \cdot \text{females50}$

Короткая регрессия: $\text{test50} = 4.641 + 0.319 \cdot \text{final50}$

Можно сказать, что веса для оценок практически не отличаются (разница ~ 0.2 и 0.002), при этом вес пола в длинной регрессии существенный - больше веса итоговой оценки более чем в 3 раза. Он положителен, так что можно предположить что в данной выборке наблюдений девушки в среднем получают оценку на 1 балл выше. Также благодаря свободному члену можно видеть, что

студенты получают оценку выше 4, даже если их оценка за ТВиМС равна 0, а с учетом положительных коэффициентов пола и итоговой оценки они скорее будут получать оценки за кр ещё выше.



2) Рассчитаем R^2 для обеих моделей

$$R^2 = ESS/TSS$$

Для этого найдём TSS и ESS.

```
TSS = Σ(y - yMean)^2
>> TSS = sum(((test50 - sum(test50)/50)).^2)

TSS =

    316.5000
```

Чтобы рассчитать ESS нам нужно найти ортогональные проекции наблюдений на пространства объясняющих векторов.

$$ESS = \sum(ortY - yMean)^2$$

$$ortLong = Xlong \cdot \beta$$

$$ortShort = Xshort \cdot \gamma$$

```
>> ortLong = Xlong*transpose(coeffLong);
>> ortShort = Xshort*transpose(coeffShort);
>> ESSshort = sum(((ortShort - sum(test50)/50)).^2)
```

```
ESSshort =
```

```
21.3762
```

```
>> ESSlong = sum(((ortLong - sum(test50)/50)).^2)
```

```
ESSlong =
```

```
27.4122
```

```
R^2 long = ESSlong/TSS = 0.0866
```

```
R^2 short = ESSshort/TSS = 0.0675
```

Ответ: R^2 для короткой модели равен 0.0675, для длинной - 0.0866.

3) Чтобы рассчитать прогноз и ошибку прогноза для значений с 51 до 65 составим матрицы XshortP и XlongP, взяв единичный столбец, столбец final15 и females15 - последние 15 значений столбцов final и females. Составим массив test15 - последние 15 значений в массиве test.

```
>> XshortP = [ones(15, 1), final15] >> XlongP = [ones(15, 1), final15, females15]
```

```
XshortP =
```

```
1 7
1 4
1 1
1 1
1 7
1 7
1 3
1 4
1 6
1 7
1 9
1 6
1 8
1 6
1 3
```

```
XlongP =
```

```
1 7 0
1 4 0
1 1 0
1 1 0
1 7 0
1 7 0
1 3 1
1 4 0
1 6 0
1 7 0
1 9 0
1 6 0
1 8 0
1 6 0
1 3 0
```

```
>> final15 = [final(1, 51:65)];
>> females15 = [females(1, 51:65)];
>> test15 = [test(1, 51:65)];
>> females15 = transpose(females15);
>> final15 = transpose(final15);
>> test15 = transpose(test15);
```

Теперь рассчитаем прогнозы с помощью короткой и длинной регрессии:

<pre>>> prognosisShort = XshortP*transpose(coeffShort)</pre>	<pre>>> prognosisLong = XlongP*transpose(coeffLong)</pre>
<pre>prognosisShort =</pre>	<pre>prognosisLong =</pre>
6.8743	6.7375
5.9171	5.7747
4.9600	4.8118
4.9600	4.8118
6.8743	6.7375
6.8743	6.7375
5.5981	6.4551
5.9171	5.7747
6.5552	6.4166
6.8743	6.7375
7.5124	7.3794
6.5552	6.4166
7.1933	7.0585
6.5552	6.4166
5.5981	5.4537

Посчитаем среднюю ошибку, насколько прогноз отличается от наблюдений

<pre>>> errorShort = sum(abs(test15 - prognosisShort))/15</pre>	<pre>>> errorLong = sum(abs(test15 - prognosisLong))/15</pre>
<pre>errorShort =</pre>	<pre>errorLong =</pre>
1.7622	1.8003

Как можно видеть, у короткой регрессии ошибка немного меньше.

3) Чтобы спрогнозировать мою оценку за контрольную, посчитаю ее с помощью модели короткой регрессии, так как она не намного, но точнее. Моя итоговая оценка за курс теории вероятности была 10, получается следующее:

<pre>>> myTest = 4.6410 + 0.3190*10</pre>
<pre>myTest =</pre>
7.8310

Вероятно, в длинной модели было бы немного больше. Но и так оценка в целом достойная.

Ответ: оценка 7.831