/\*Importando as bases\*/

/\*Moradores\*/

**PROC** **IMPORT** OUT= mor DATAFILE= "/media/bi\_modelagem/\_pld/Risco\_de\_Credito/Analise\_Alfredo/job/pdad.xlsx"

DBMS=xlsx REPLACE;

GETNAMES=YES;

**RUN**;

/\*Exercício 1.4\*/

/\*Item i)\*/

/\*Observação: a variável escolaridade foi tratada no R\*/

/\*Seleciona apenas as pessoas com 25 anos ou mais do Plano Piloto e da RA Samambaia\*/

**proc** **sql**;

create table mor2 as

select

case when A01ra=**1** then 'Plano Piloto' else 'Samambaia' end as 'Região Administrativa'n,

Escolaridade,

case when E03=**1** then 'Masculino' else 'Feminino' end as 'Sexo'n,

case when Escolaridade='Sem escolaridade' then **1**

when Escolaridade='Fundamental incompleto' then **2**

when Escolaridade='Fundamental completo' then **3**

when Escolaridade='Médio incompleto' then **4**

when Escolaridade='Médio completo' then **5**

when Escolaridade='Superior incompleto' then **6**

when Escolaridade='Superior completo' then **7**

else **8** end as escol,

E03

from mor

where A01ra in (**1**,**12**) and idade\_calculada >=**25**;

;**quit**;

**proc** **sort** data=mor2;

by 'Região Administrativa'n; **run**;

/\*Crie tabela com frequências nas linhas, colunas e total\*/

**proc** **freq** data=mor2;

by 'Região Administrativa'n;

tables Escolaridade\*'Sexo'n / out=FreqCount nopercent nocol norow ;

**run**;



/\*Coeficiente de Pearson\*/

**proc** **corr** data=mor2;

var escol/\*Escolaridade\*/ E03/\*Sexo\*/;

**run**;

/\* Resposta: -0.03252\*/

/\*Não há diferença significativa da escolaridade entre homens e mulheres dentro da Região, entretanto, é possível

verificar uma diferença entre o grau de escolaridade da Região Samambaia para o Plano Piloto.

Observa-se que as pessoas do Plano Piloto, em sua maioria, possuem ensino superior completo ou incompleto.

O coeficiente de correlação de Pearson, comprova que não há relação entre as variáveis escolaridade e sexo\*/

/\*Item ii)\*/

/\*Cria variável da região de origem e trata dados\*/

**proc** **sql**;

create table mor3 as

select

E142 as regiao\_nascimento,

case when A01ra=**1** then 'Plano Piloto' else 'Samambaia' end as 'Região Administrativa'n,

case when G16 in (**77777**,**88888**,**99999**) then **.** else G16 end as Renda,

case when E142 in (**33**,**35**,**31**,**32**,**41**,**42**,**43**) then 'Sul/Sudeste'

when E142 in (**50**,**51**,**52**) then 'Centro Oeste'

when E142 = **99** then 'DF'

else 'Norte/Nordeste' end as 'Região Nascimento'n

from mor

where A01ra in (**1**,**12**);

**quit**;

/\*Média e Variância por região de nascimento e Sambambaia e Plano Piloto\*/

**proc** **sql**;

create table med\_var as

select

'Região Administrativa'n,

'Região Nascimento'n,

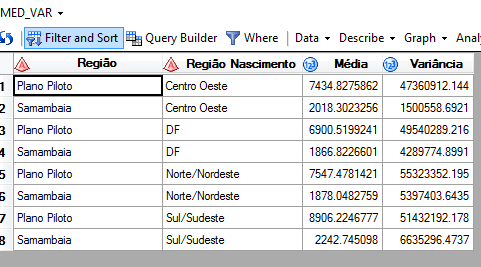
mean(Renda) as 'Média'n,

var(Renda) as 'Variância'n

from mor3

group by 'Região Nascimento'n, 'Região Administrativa'n

;**quit**;



/\*Boxplot\*/

/\*Renda por Local de Nascimento\*/

**proc** **sort** data=mor3;

by 'Região Nascimento'n 'Região Administrativa'n;

**run**;

options nogstyle;

title 'Renda por Local de Nascimento';

symbol value=dot color=salmon;

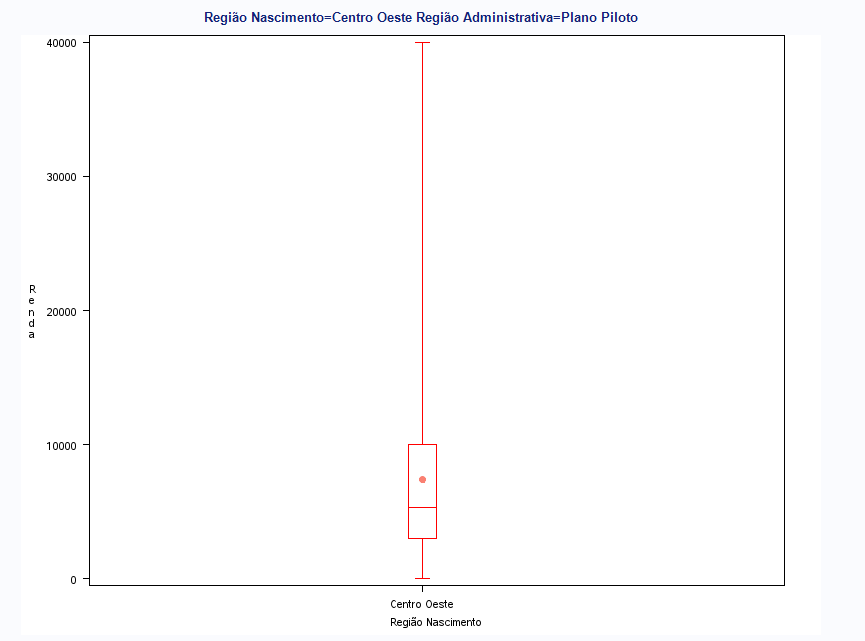
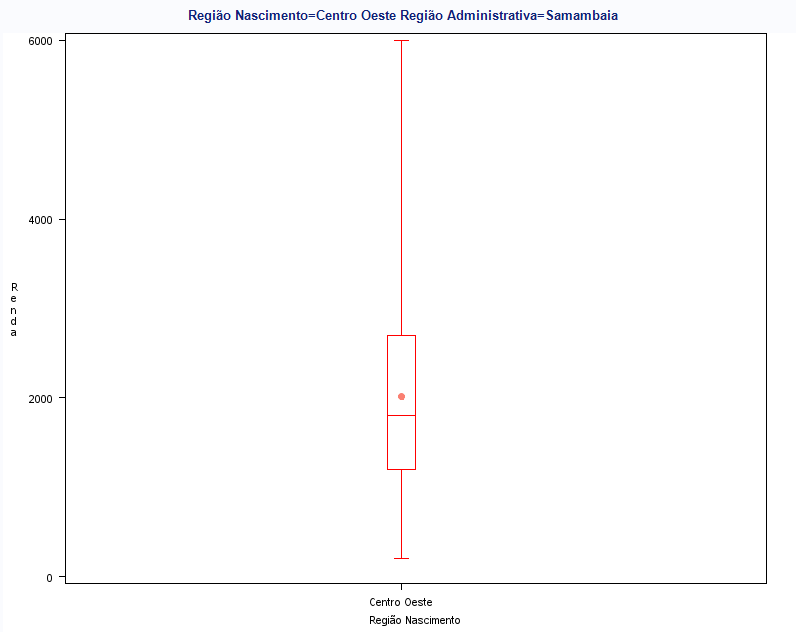
**proc** **boxplot** data=mor3;

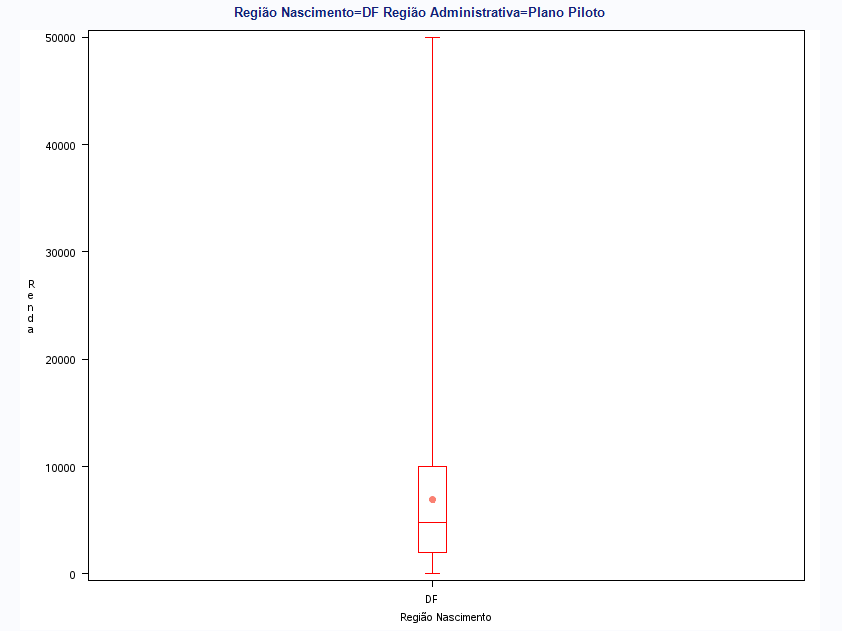
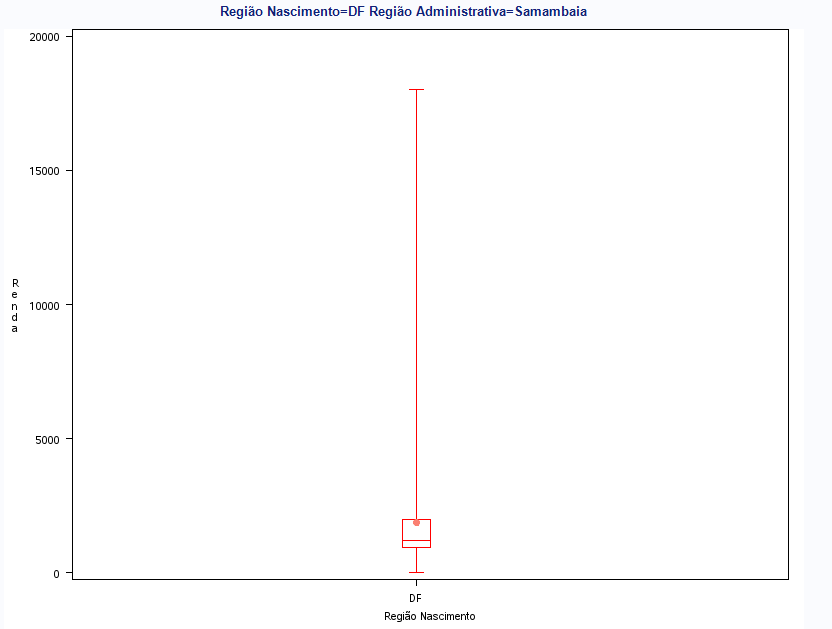
ID Renda;

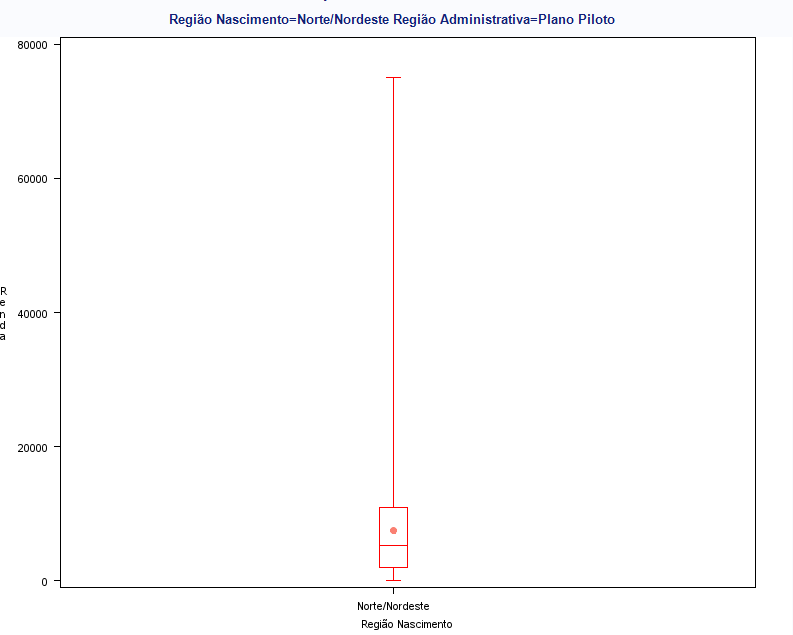
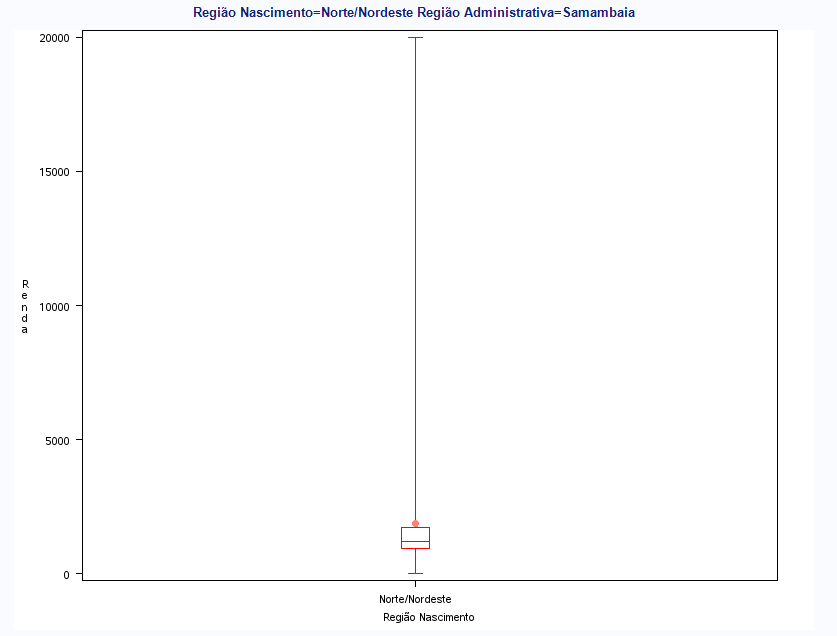
by 'Região Nascimento'n 'Região Administrativa'n;

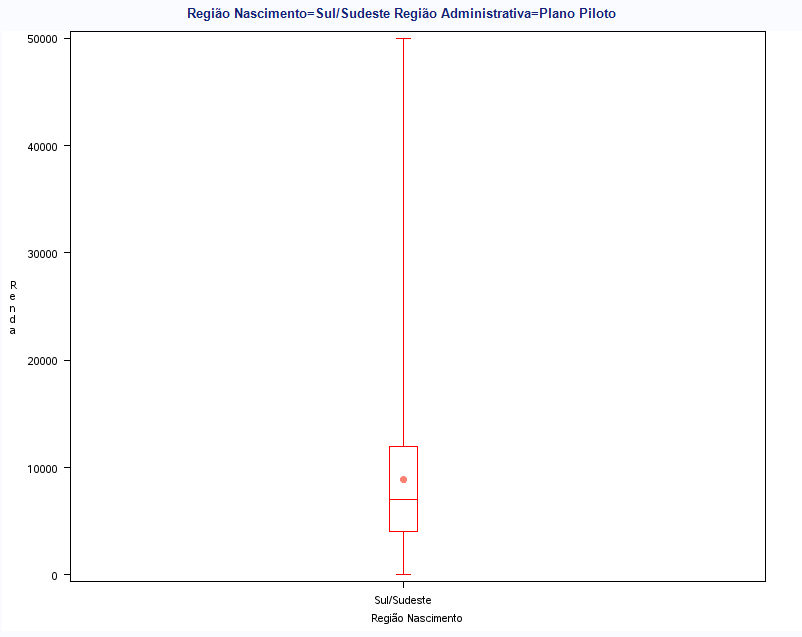
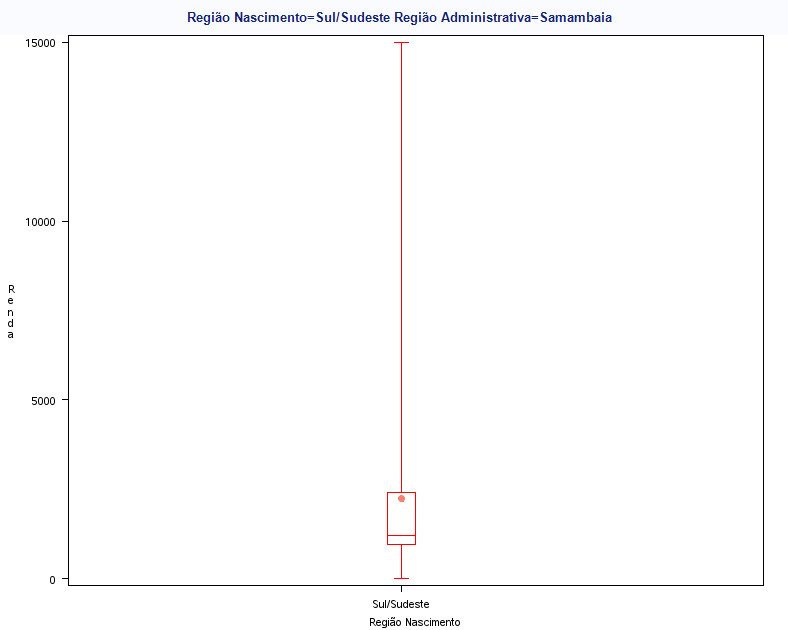
plot Renda\*'Região Nascimento'n;

**run**;

/\*R²- Considerando a Região Administrativa - Plano Piloto\*/

**proc** **reg** data=mor3 outest=outest plots=none;

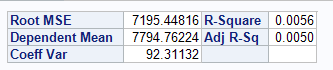
model Renda = regiao\_nascimento/ rsquare press sse adjrsq;

where 'Região Administrativa'n='Plano Piloto';

**run**;

**quit**;

/\*R²: 0.0056\*/



/\*R²- Considerando a Região Administrativa - Samambaia\*/

**proc** **reg** data=mor3 outest=outest plots=none;

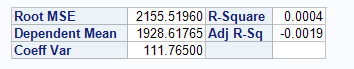
model Renda = regiao\_nascimento/ rsquare press sse adjrsq;

where 'Região Administrativa'n='Samambaia';

**run**;

**quit**;

/\*R²: 0.0004/



/\*R²\*/

**proc** **reg** data=mor3 outest=outest plots=none;

model Renda = regiao\_nascimento/ rsquare press sse adjrsq;

**run**;

**quit**;

/\*R²: 0.0085/



/\*Dentro das regiões administrativas é possível ver que, em média, as pessoas do DF e do Nordeste ganham menos

quando comparada com pessoas naturais das outras regiões.

Entretanto, há diferença expressiva da renda média é entre as regiões administrativas do Plano Piloto para Samambaia,

O valor do R² é baixo, o que indica que só a região de nascimento não é suficiente para explicar o valor da renda.

Agrupar a renda poderia melhorar o ajuste do modelo ou imputar os valores missings e/ou a inclusão de outras variáveis.\*/

/\*Item iii)\*/

/\*Trata dados\*/

**proc** **sql**;

create table mor4 as

select

E03 as sexo,

case when A01ra=**1** then 'Plano Piloto' else 'Samambaia' end as 'Região Administrativa'n,

case when G18 in (**88888**,**99999**) then **.** else G18 end as Horas,

case when E03=**1** then 'Masculino' else 'Feminino' end as 'Sexo'n

from mor

where A01ra in (**1**,**12**);

**quit**;

/\*Média e Variância por horas trabalhadas e sexo das regiões Sambambaia e Plano Piloto\*/

**proc** **sql**;

create table med\_var1 as

select

'Região Administrativa'n,

'Sexo'n,

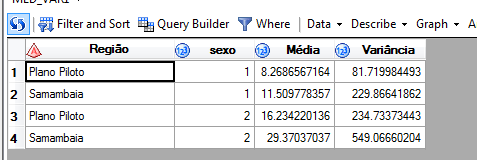
mean(Horas) as 'Média'n,

var(Horas) as 'Variância'n

from mor4

group by 'Sexo'n, 'Região Administrativa'n

;**quit**;



/\*Boxplot\*/

/\*Renda por Local de Nascimento\*/

**proc** **sort** data=mor4;

by 'Sexo'n 'Região Administrativa'n;

**run**;

options nogstyle;

title 'Horas Gastas Semanalmente com Tarefa Doméstica por Sexo';

symbol value=dot color=salmon;

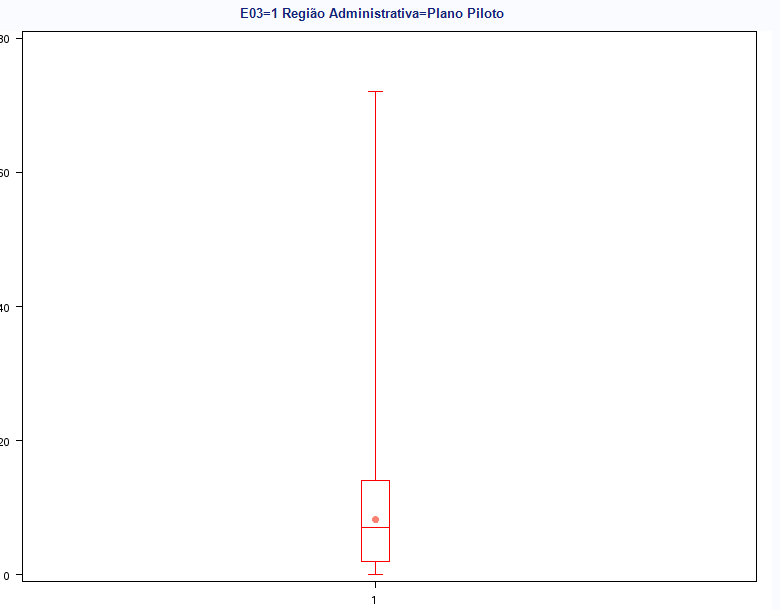
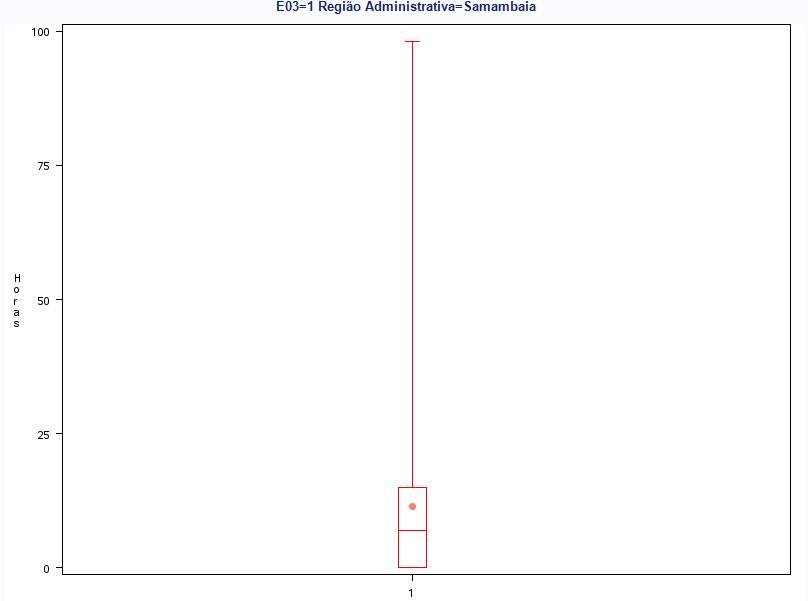
**proc** **boxplot** data=mor4;

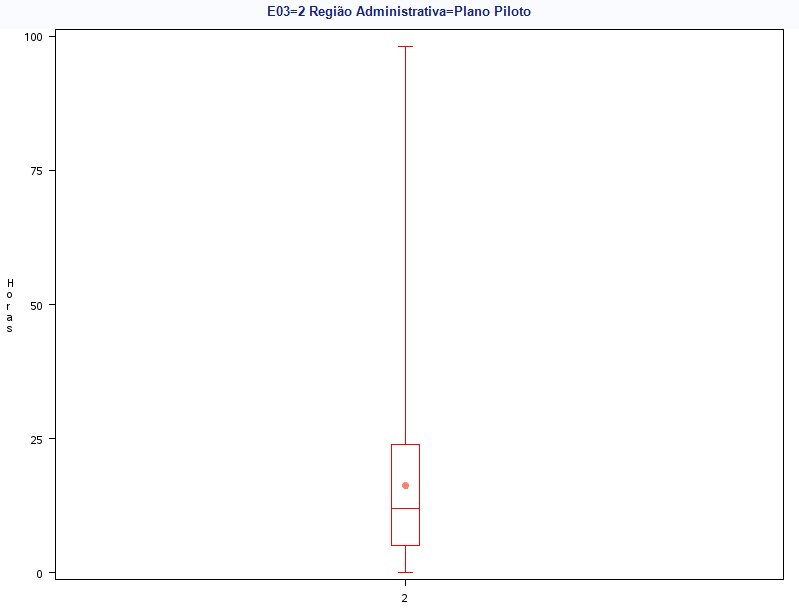
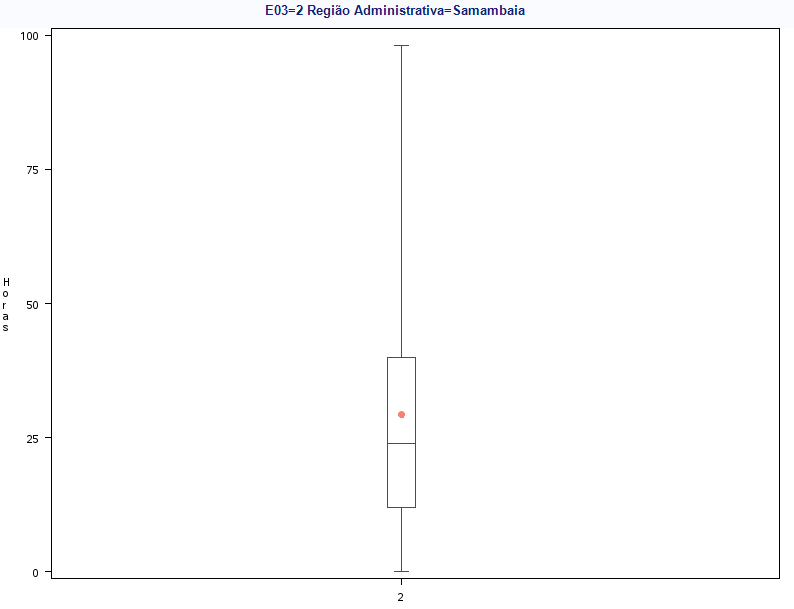
ID Horas;

by 'Sexo'n 'Região Administrativa'n;

plot Horas \*'Sexo'n;

**run**;

/\*R²- Considerando a Região Administrativa - Plano Piloto\*/

**proc** **reg** data=mor4 outest=outest plots=none;

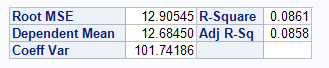
model horas = sexo/ rsquare press sse adjrsq;

where 'Região Administrativa'n='Plano Piloto';

**run**;

**quit**;

/\*R²: 0.0861\*/



/\*R²- Considerando a Região Administrativa - Samambaia\*/

**proc** **reg** data=mor4 outest=outest plots=none;

model horas = sexo/ rsquare press sse adjrsq;

where 'Região Administrativa'n='Samambaia';

**run**;

**quit**;

/\*R²: 0.1640\*/



/\*R²\*/

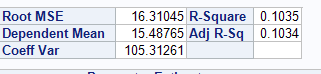
**proc** **reg** data=mor4 outest=outest plots=none;

model horas = sexo/ rsquare press sse adjrsq;

**run**;

**quit**;

/\*R²: 0.1035/



/\*É possível observar que em ambas as regiões as mulheres se dedicam mais as tarefas domésticas que os homens,

sendo em média. Entretanto, as mulheres de Samambaia, trabalham em média 13h a mais que as do Plano Piloto.

O valor do R² já um pouco mais elevado o que indica há indícios que o sexo (homem/mulher) é uma variável

que consegue explicar a quantidade de horas gastas com tarefas domésticas.\*/