



**Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Computação Científica
Departamento de Estatística**

Relatório Trabalho de ME610

**Eliane Ramos de Siqueira RA:155233
Guilherme Pazian RA:160323**

Professor: Maurício Zevallos

Campinas-SP, 28 de Abril de 2017

1. Introdução

As consequências do hábito de fumar são um dos assuntos mais importantes para pesquisa médica hoje em dia. Em particular, pesquisas anteriores estão apontando uma possível relação decrescente entre este hábito em mulheres durante a gravidez e o peso da criança ao nascer, ou seja, com a presença do hábito, o peso da criança é menor.

Como relatado no Surgeon General's Report, 1989:

“O tabagismo parece ser um determinante mais significativo do peso ao nascer do que a altura, o peso, o número de fetos, a renda anual da mãe, o histórico de resultados de gestações anteriores ou o sexo do bebê. A redução do peso ao nascer associada ao hábito de fumar parece ser um efeito direto do tabagismo sobre o crescimento fetal.”

O objetivo deste trabalho é verificar o quão válida é esta opinião através da análise do peso da criança ao nascer.

2. Descrição

Os dados utilizados foram coletados através de entrevistas com as mães durante a gravidez e correspondem a um ano de estudo. Eles incluem 1236 nascimentos onde a criança sobreviveu pelo menos 28 dias. Juntamente com o peso das crianças, foram coletadas informações de outras variáveis que influenciam no peso das crianças, como sexo do bebê, número de gestações anteriores, hábito de fumar, número de cigarros que fuma, se parou de fumar, quanto tempo faz, além de raça, altura, peso, nível educacional e idade do pai e da mãe, estado civil da mãe e rendimento anual da família.

Para este estudo as variáveis disponíveis no banco de dados são:

- Peso: Peso da criança ao nascer (em onça)
- Sexo: Sexo da criança
- Data_nasc: Data de nascimento (1996= 1 de janeiro de 1961)
- Vivo: Nascido vivo ou não (1= Nasceu vivo e sobreviveu por pelo menos 28 dias)
- Qtd_feto: Quantidade de fetos (5= feto único)
- Tempo_gestacao: Tempo da gestação (em dias)
- Nmero_gestacoes: Número de gestações anteriores (incluindo morte fetal)
- Altura_pai: Altura do pai (em polegadas)
- Peso_pai: Peso do pai (em libras)
- Altura_mae: Altura da mãe (em polegadas)
- Peso_mae: Peso da mãe (em libras)
- Tempo_sem_fumar: Tempo sem fumar (em anos)
- Estado_civil: estado civil da mãe (5 níveis)
- Rendimento_anual: Rendimento anual da família ()
- Número_cigarros: Número de cigarros fumados por dia (5 níveis)
- Fuma: Hábito de fumar (5 níveis)
- Educacao_mae: Nível educacional da mãe (5 níveis)
- Cor_mae: Raça da mãe (5 níveis)
- Idade_mae: Idade da mãe (5 níveis)

- Educacao_pai: Nivel educacional do pai (5 nÍveis)
- Cor_pai: Raça do pai (5 nÍveis)
- Idade_pai: Idade do pai (5 nÍveis)

Afim de avaliar a influencia das demais variáveis sobre o peso do bebê, foi considerado o peso da criança ao nascer como variável resposta e as demais variáveis como explicativas.

3. Metodologia

Em uma primeira análise do banco de dados disponível, notou-se a presença de crianças apenas do sexo masculino que sobreviveram pelo menos 28 dias e eram feto único, tal característica levou a desconsideração das variáveis correspondentes, Sexo,Vivo e Qtd_feto. As variáveis correspondentes à Peso e Altura do pai foram desconsideradas devido ao elevado número de observações faltantes, 499 e 492 respectivamente. As variáveis Peso_pai e Altura_pai foram desconsideradas devido a quantidade elevada de observações desconhecidas. Para as demais variáveis, foram consideradas nas análises, apenas as observações com informações completas de todas as variáveis envolvidas. Além disso, no banco de dados original, as variáveis correspondentes ao peso da criança,peso e altura da mãe e do pai estavam em unidades de medidas não usuais, onça, libras e polegadas respectivamente, então foram feitas as devidas conversões para kg no caso dos pesos e centímetro no caso das alturas.

4. Objetivo

5. Resultados

Escreva os resultados aqui

6. Discussão

Escreva a discussão aqui

7. Anexos

Coloque os anexos aqui

Escreva as referencias aqui

Limpando os dados

Excluimos as váriaveis: Sexo (apenas masculino), Qtd_feto (apenas 5) e Vivo (apenas 1).

Na variavel Cor_mae, temos uma observação com valor desconhecido (99)

Na variavel Idade_mae temos duas observações desconhecidas (99)

Na variavel Educacao_mae temos uma observação desconhecida (9)

Na variavel Altura_mae temos 22 observações desconhecidas (99)

Na variavel Peso_mae temos 36 observações desconhecidas (999)

Na variavel Cor_pai, temos 5 observações com valor desconhecido (99)
Na variavel Idade_pai, temos 7 observações com valor desconhecido (99)
Na variavel Educacao_pai, temos 13 observações com valor desconhecido (9)
Na variavel Altura_pai temos 492 observações desconhecidas (99)
Na variavel Peso_pai temos 499 observações desconhecidas (999)
Na variavel Rendimento_anual temos 124 observações desconhecidas (98)
Na variavel Fuma temos 10 observações desconhecidas (9)
Na variavel Tempo_sem_fumar temos 9 observações desconhecidas (98) e 1 não perguntado(99)
Na variavel Numero_cigarros temos 10 observações desconhecidas (98)

Tiramos as variaveis Peso_pai e Altura_pai devido a quantidade elevada de observações desconhecidas.

Análise Descritiva

Com base na vizualização de análises descritivas (gráficos de dispersão e boxplots), observou-se as relações entre as variáveis explicativas com a variável resposta (peso do bebê ao nascer).Notou-se que existe pouca relação entre as variáveis explicativas “Altura_mae_cm”, “Idade_pai”, “Idade_mae” e “Data_nasc” em relação ao peso do bebê, ou seja, o peso do bebê parece não estar correlacionado com estas variáveis. Portanto optou-se em não incluir estas variáveis explicativas mesmo num modelo inicial. Em relação às demais variáveis explicativas, todas apresentaram indícios de uma possível corelação com o peso dos bebês, de maneira que a distribuição dos dados referêntes à variável resposta muda conforme os valores observados nas variáveis explicativas. Nenhuma variável contínua/discreta aparentou ter uma relação não-linear com o peso dos bebês.

Grafico de Dispersão

MOderado

FRACA relaLinear

FORTE

Pouca relação linear

Pouca relação linear

Pouca relação linear

Boxplot para as variáveis categóricas e Dispersão para continua

Note que as variáveis “Fuma” e “Tempo sem Fumar” contém a mesma informação, ou seja, as mães não fumantes (classificação em “Fuma”), são classificadas como “Nunca fumou” (classificação em “Tempo_sem_fumar”) assim como as mães fumantes (classificação em “Fuma”) são classificadas como “Ainda Fuma” (classificação em “Tempo_sem_fumar”), as mães do grupo “fumava” (classificação em “Fuma”) foram classificadas em relação ao tempo em que fumaram na variável “Tempo sem fumar”. A partir dessa premissa, vamos ajustar um primeiro modelo excluindo-se a variável “Tempo_sem_fumar” já que esta possui a mesma informação que a variável “Fuma” e é menos específica (não temos interesse específico na quantidade diária de cigarros consumida por indivíduo).

```
## [1] 0.3156834
```

```
## [1] 0.2887799
```

```
##
```

```
## t test of coefficients:
```

```
##
```

	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	3.3279884	0.1565832	21.2538	< 2.2e-16 ***
## FumaFuma	-0.2397740	0.0311682	-7.6929	3.460e-14 ***
## FumaFumava	-0.0165962	0.1000682	-0.1658	0.8683101
## Tempo_sem_fumar1-2	0.1897359	0.1234835	1.5365	0.1247279
## Tempo_sem_fumar10+	-0.2267149	0.1969476	-1.1511	0.2499507
## Tempo_sem_fumar2-3	-0.1517817	0.1747509	-0.8686	0.3852978
## Tempo_sem_fumar3-4	-0.0642165	0.1972027	-0.3256	0.7447675
## Tempo_sem_fumar5-9	-0.0903418	0.1426614	-0.6333	0.5267096
## Tempo_sem_fumarDurante_a_g.	0.0267920	0.1091561	0.2454	0.8061612
## Tempo_sem_fumarNão Sabe	0.2038489	0.1246025	1.6360	0.1021580
## Rendimento_anual>15000	-0.0729141	0.1446775	-0.5040	0.6143898
## Rendimento_anual12500-15000	0.0638234	0.1134508	0.5626	0.5738585
## Rendimento_anual2500-4999	0.1039730	0.0817864	1.2713	0.2039287
## Rendimento_anual5000-12499	0.1001895	0.0765230	1.3093	0.1907451
## Estado_civil.L	0.0669886	0.0767611	0.8727	0.3830435
## Educacao_pai.L	0.0036359	0.0642319	0.0566	0.9548710
## Educacao_pai.Q	0.0015179	0.0510487	0.0297	0.9762851
## Educacao_pai.C	0.0332485	0.0389288	0.8541	0.3932633
## Educacao_pai^4	0.0204529	0.0310009	0.6598	0.5095663
## Educacao_mae.L	0.0071564	0.0506006	0.1414	0.8875590
## Educacao_mae.Q	-0.0293142	0.0352200	-0.8323	0.4054300
## Educacao_mae.C	0.0337548	0.0440751	0.7658	0.4439494
## Educacao_mae^4	0.0058333	0.0357381	0.1632	0.8703765
## Cor_maeBlack	-0.2082667	0.1155002	-1.8032	0.0716646 .

```

## Cor_maeMex          0.0093341  0.1210044  0.0771  0.9385290
## Cor_maeMixed        -0.1025473  0.1221244 -0.8397  0.4012813
## Cor_maeNão Definido -0.0172877  0.2104531 -0.0821  0.9345480
## Cor_maeWhite        -0.1184129  0.0663011 -1.7860  0.0744067 .
## Cor_paiBlack         0.2070021  0.1238913  1.6708  0.0950693 .
## Cor_paiMex           0.3626010  0.1423767  2.5468  0.0110225 *
## Cor_paiMixed         0.3136446  0.1096976  2.8592  0.0043364 **
## Cor_paiNão Definido  0.0513837  0.2448111  0.2099  0.8337956
## Cor_paiWhite         0.3034963  0.0781270  3.8847  0.0001093 ***
## Peso_mae_kg          0.0433447  0.0187230  2.3150  0.0208137 *
## Tempo_gestacao       0.1951718  0.0181476 10.7547 < 2.2e-16 ***
## Data_nasc            0.0276228  0.0137833  2.0041  0.0453344 *
## Numero_gestacoes     0.0611437  0.0210702  2.9019  0.0037911 **
## Idade_mae            -0.0046516  0.0050863 -0.9145  0.3606626
## Altura_mae_cm        0.0733644  0.0178144  4.1183  4.135e-05 ***
## Idade_pai            0.0021838  0.0038371  0.5691  0.5693908
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Nenhuma estimativa relacionada com as Variáveis “Tempo sem Fumar”, “Rendimento Anual”, “Estado Civil”, “Educação Pai”, “Educação Mãe”, “idade da mãe” e “idade do pai” foram significativos, portanto, será ajustado um novo modelo sem estas variáveis:

```
## [1] 0.3019706
```

```
## [1] 0.2902679
```

```
##
```

```
## t test of coefficients:
```

```
##
```

	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	3.30299041	0.07531261	43.8571	< 2.2e-16 ***
## FumaFuma	-0.23805979	0.03036035	-7.8411	1.126e-14 ***
## FumaFumava	-0.00071938	0.03923311	-0.0183	0.985374
## Cor_paiBlack	0.21210270	0.11657772	1.8194	0.069144 .
## Cor_paiMex	0.35117010	0.13754185	2.5532	0.010820 *
## Cor_paiMixed	0.32611518	0.10418980	3.1300	0.001798 **
## Cor_paiNão Definido	0.06034492	0.23696474	0.2547	0.799039
## Cor_paiWhite	0.29638458	0.07101076	4.1738	3.253e-05 ***
## Cor_maeBlack	-0.19410874	0.10782201	-1.8003	0.072115 .
## Cor_maeMex	0.04760597	0.11487530	0.4144	0.678658
## Cor_maeMixed	-0.09638131	0.11406428	-0.8450	0.398325

```
## Cor_maeNão Definido -0.00659248 0.20104102 -0.0328 0.973847
## Cor_maeWhite -0.09617289 0.05745861 -1.6738 0.094483 .
## Peso_mae_kg 0.04225092 0.01790505 2.3597 0.018477 *
## Tempo_gestacao 0.19786235 0.01832006 10.8003 < 2.2e-16 ***
## Data_nasc 0.02783670 0.01333976 2.0867 0.037160 *
## Numero_gestacoes 0.04988123 0.01660001 3.0049 0.002722 **
## Altura_mae_cm 0.07193353 0.01714494 4.1956 2.960e-05 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

No novo modelo, note que o coeficiente referente ao fator “Fumava” da variável Fuma é não significativo, ou seja, não é diferente do grupo de referência (“Nunca”), portanto, no próximo modelo vamos ajustar esta variável com só dois fatores (“Fuma” e “Nunca/Fumava”).

Note também que os coeficientes, referentes as cores do pai, “Não definido” é estatisticamente não significativo, portanto, este é bem similire ao grupo de cor “Asian”, então vamos ajustar o novo modelo com as cores “Asian” e “Não definido” pertencendo a um mesmo grupo de cor (“Asian/Não definido”).

O mesmo acontece para a variável “Cor mãe” nos fatores “Mex”, “Mixed” e “Não Definido”, portanto os novos fatores no proximo modelo são “Asian/Mex/Mixed/Não Definido”, “Black” e “White”

```
## [1] 0.3013758
## [1] 0.2931486
##
## t test of coefficients:
##
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.309448 0.069118 47.8812 < 2.2e-16 ***
## FumaFuma -0.237660 0.028301 -8.3977 < 2.2e-16 ***
## Cor_paiBlack 0.162732 0.112315 1.4489 0.1476778
## Cor_paiMex 0.373616 0.110853 3.3704 0.0007788 ***
## Cor_paiMixed 0.267769 0.099389 2.6941 0.0071730 **
## Cor_paiWhite 0.271125 0.088283 3.0711 0.0021892 **
## Cor_maeBlack -0.151854 0.091433 -1.6608 0.0970569 .
## Cor_maeWhite -0.077158 0.063392 -1.2172 0.2238228
## Peso_mae_kg 0.042687 0.017867 2.3892 0.0170679 *
## Tempo_gestacao 0.197533 0.018246 10.8261 < 2.2e-16 ***
## Data_nasc 0.027418 0.013260 2.0677 0.0389238 *
## Numero_gestacoes 0.049916 0.016619 3.0035 0.0027345 **
## Altura_mae_cm 0.071030 0.017067 4.1619 3.423e-05 ***
## ---
```

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

dados Cor do pai “Black” é igual à referência. dados Cor da mãe “White” é igual à referência.

```
## [1] 0.2998204
```

```
## [1] 0.2929626
```

```
##
```

```
## t test of coefficients:
```

```
##
```

	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	3.336745	0.056216	59.3555	< 2.2e-16 ***
## FumaFuma	-0.238720	0.028016	-8.5207	< 2.2e-16 ***
## Cor_paiMex	0.330556	0.102732	3.2177	0.001333 **
## Cor_paiMixed	0.178070	0.079487	2.2402	0.025289 *
## Cor_paiWhite	0.169355	0.057006	2.9708	0.003039 **
## Cor_maeBlack	-0.019746	0.063508	-0.3109	0.755929
## Peso_mae_kg	0.043372	0.017666	2.4551	0.014250 *
## Tempo_gestacao	0.196892	0.018183	10.8281	< 2.2e-16 ***
## Data_nasc	0.028219	0.013241	2.1312	0.033308 *
## Numero_gestacoes	0.049858	0.016673	2.9903	0.002854 **
## Altura_mae_cm	0.070550	0.017001	4.1497	3.607e-05 ***

```
## ---
```

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

dados Cor da mãe “Black” é igual à referência. Portanto, todas as cores da mãe são equivalentes quanto ao peso do bebê.

O novo modelo não contemplará a variável Cor da mãe.

```
## [1] 0.2997577
```

```
## [1] 0.2935912
```

```
##
```

```
## t test of coefficients:
```

```
##
```

	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	3.321503	0.032831	101.1694	< 2.2e-16 ***
## FumaFuma	-0.239034	0.028046	-8.5228	< 2.2e-16 ***
## Cor_paiMex	0.344686	0.094448	3.6495	0.000276 ***
## Cor_paiMixed	0.188571	0.072892	2.5870	0.009819 **
## Cor_paiWhite	0.184669	0.035866	5.1489	3.143e-07 ***
## Peso_mae_kg	0.042842	0.017594	2.4350	0.015062 *
## Tempo_gestacao	0.197025	0.018155	10.8527	< 2.2e-16 ***


```
## Data_nasc      0.028168    0.013237    2.1279    0.033584 *
## Numero_gestacoes 0.049505    0.016557    2.9900    0.002856 **
## Altura_mae_cm   0.070133    0.016945    4.1389 3.777e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

9 - Educacao_pai 136 - Cor_pai 1043 - Idade_pai

8. Anexos

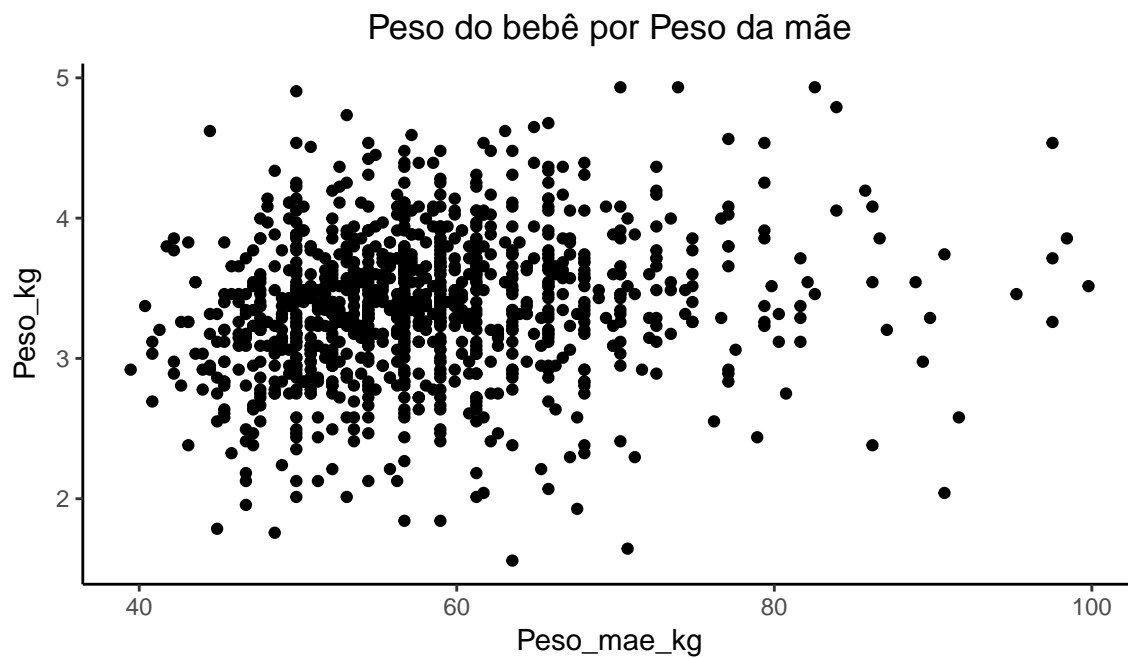


Figura 1: Gráfico de dispersão entre o peso do bebe e peso da mãe

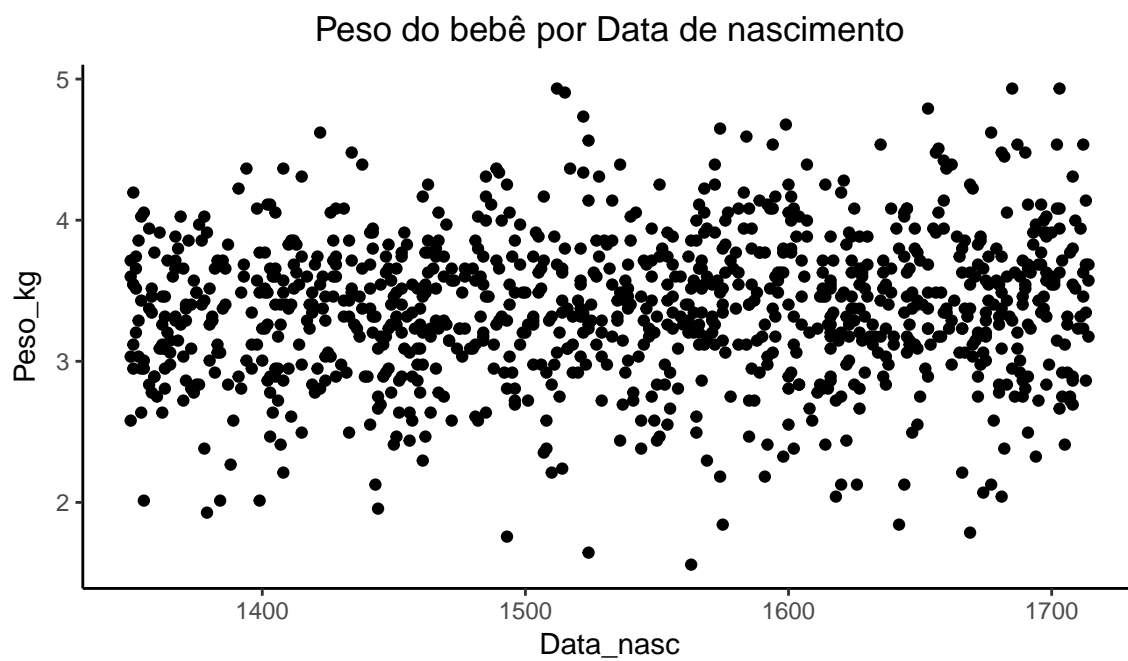


Figura 2: Gráfico de dispersão entre o peso do bebe e a data de nascimento dele

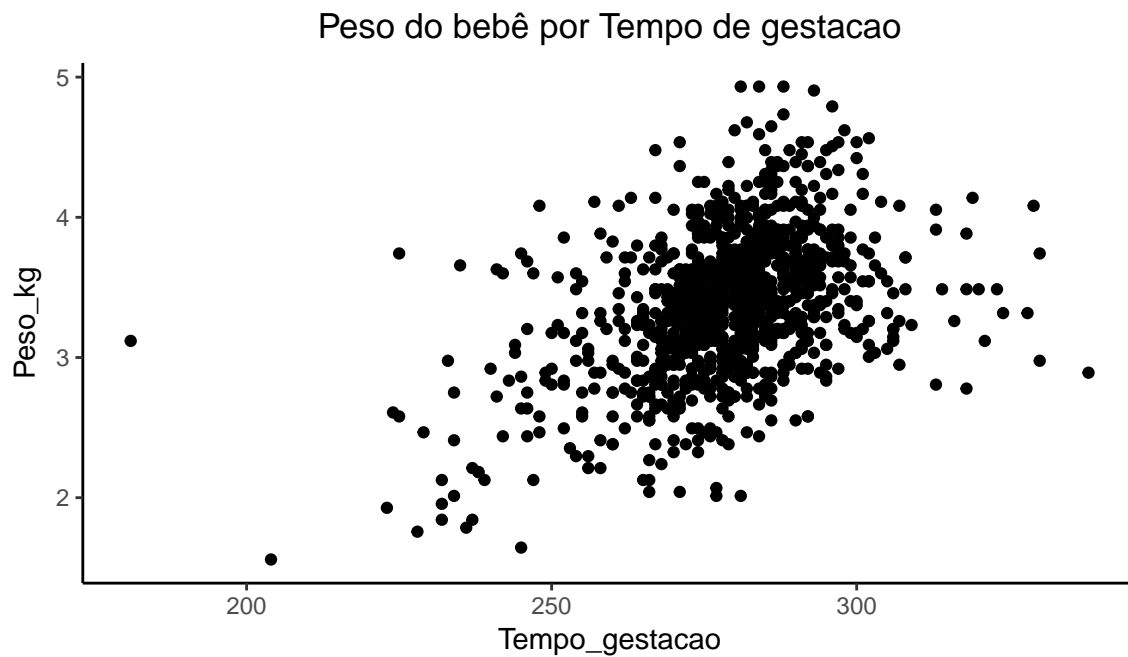


Figura 3: Gráfico de dispersão entre o peso do bebe e e o tempo de gestação

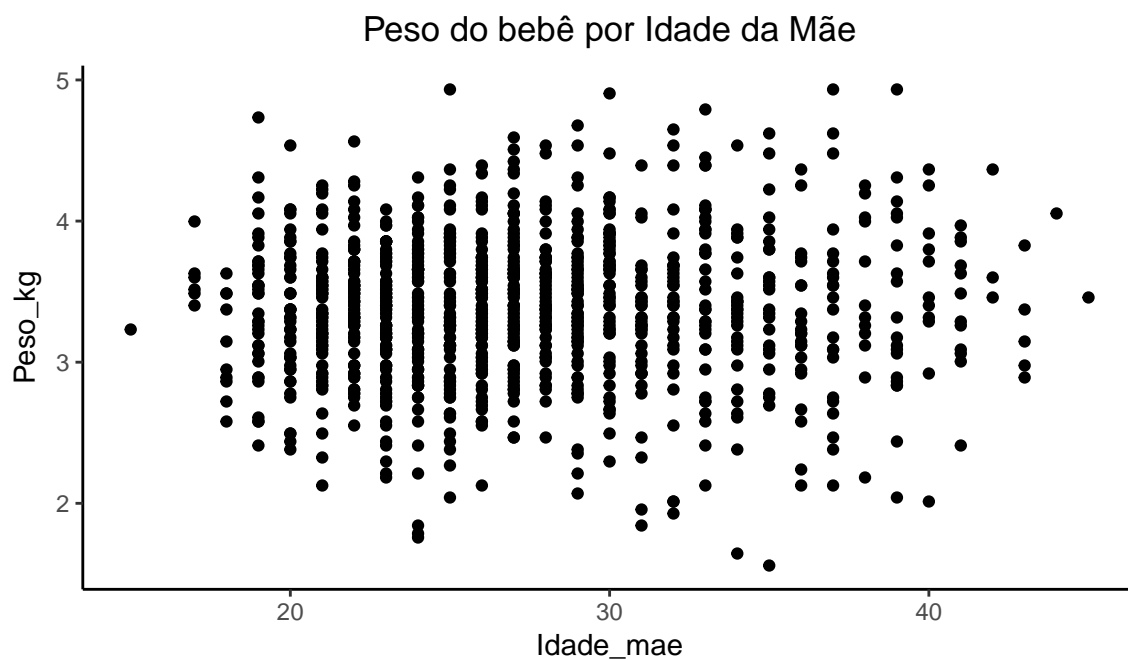


Figura 4: Gráfico de dispersão entre o peso do bebe e idade da mãe

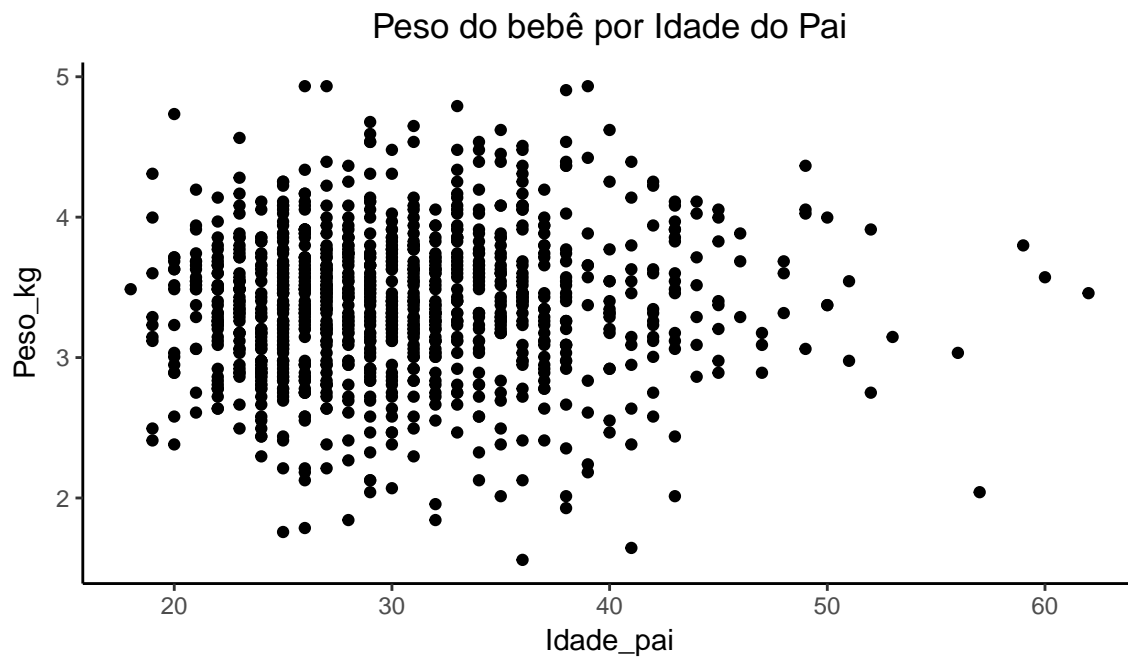


Figura 5: Gráfico de dispersão entre o peso do bebe e idade do pai

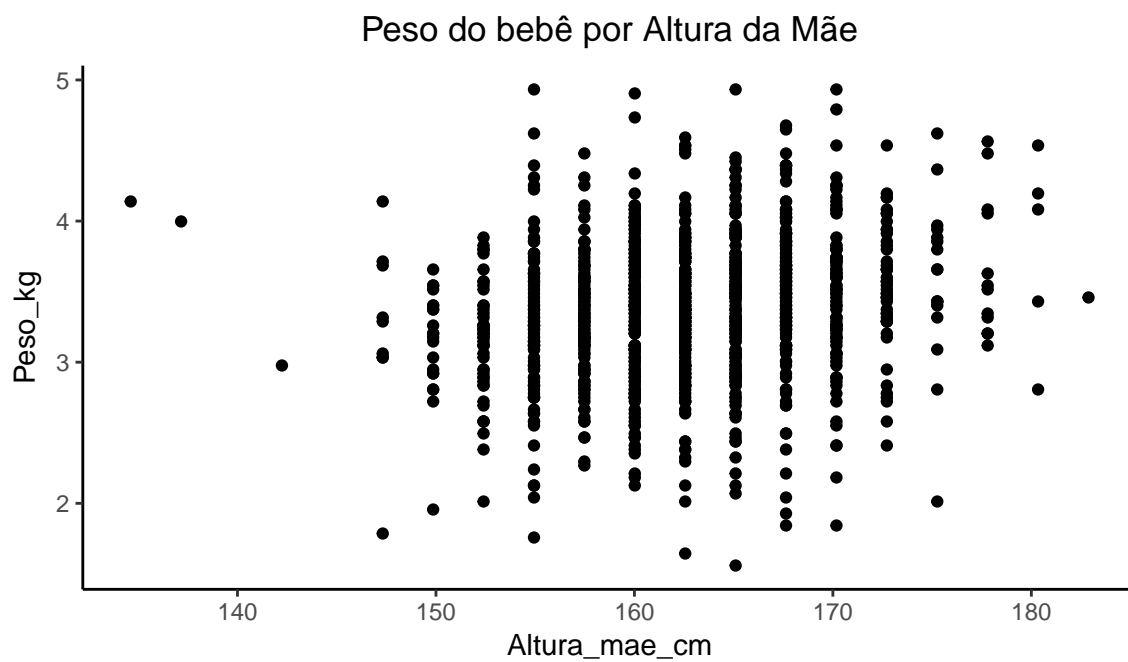


Figura 6: Gráfico de dispersão entre o peso do bebe e a altura da mãe

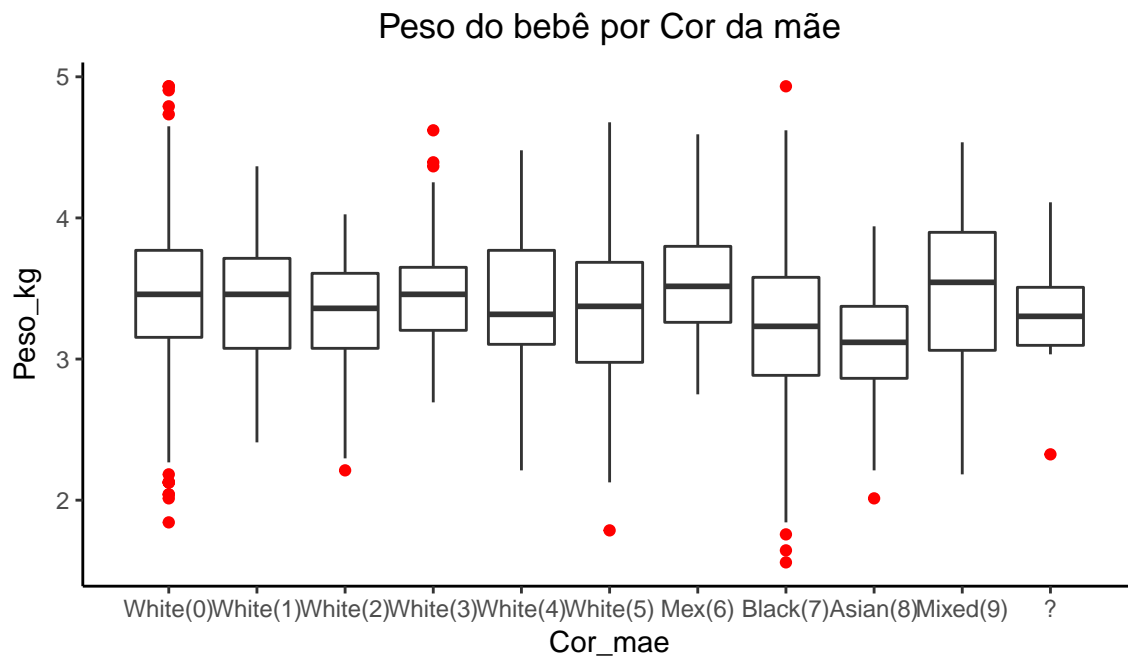


Figura 7: Boxplot entre o peso do bebe e raça da mãe

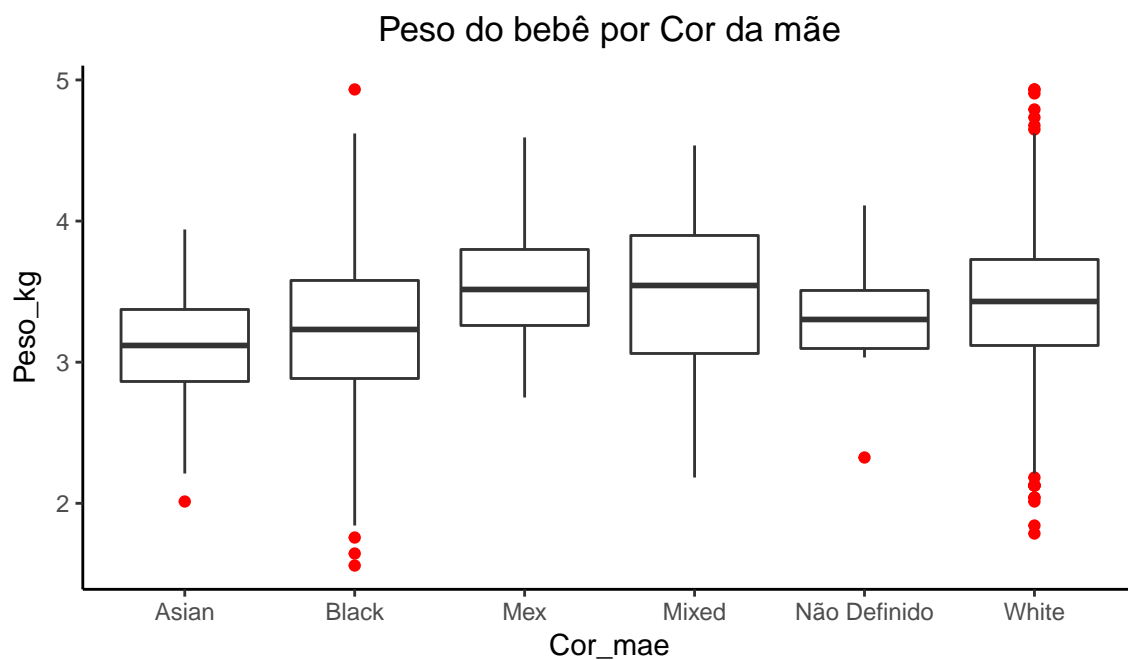


Figura 8: Boxplot entre o peso do bebe e raça da mãe - mudanças nos níveis

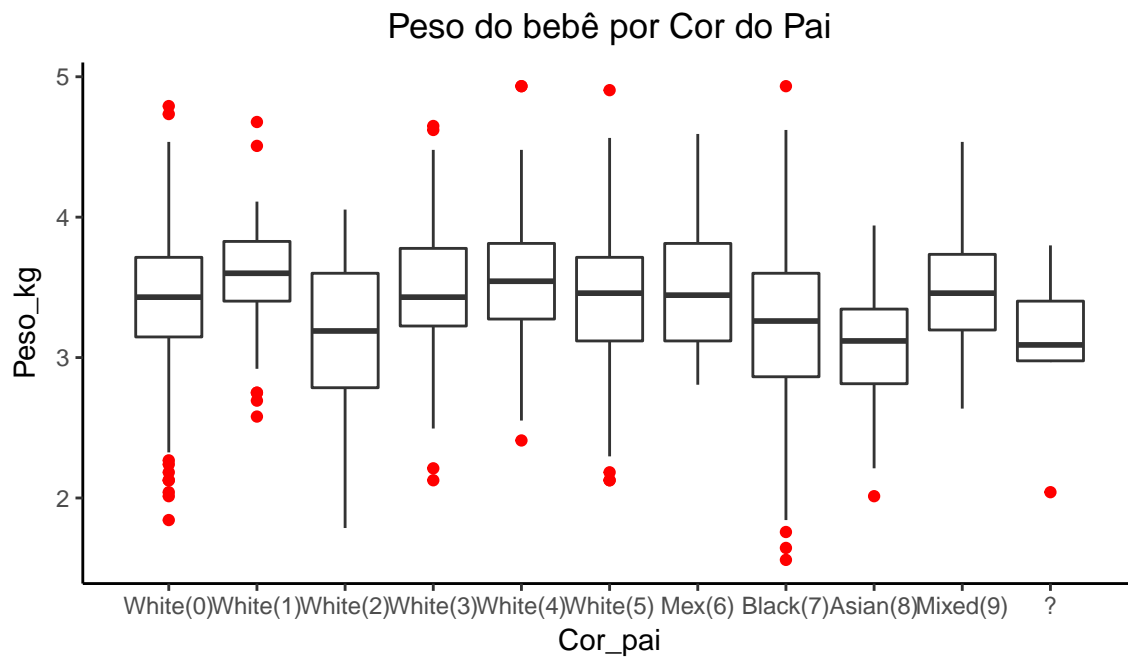


Figura 9: Boxplot entre o peso do bebe e raça o pai

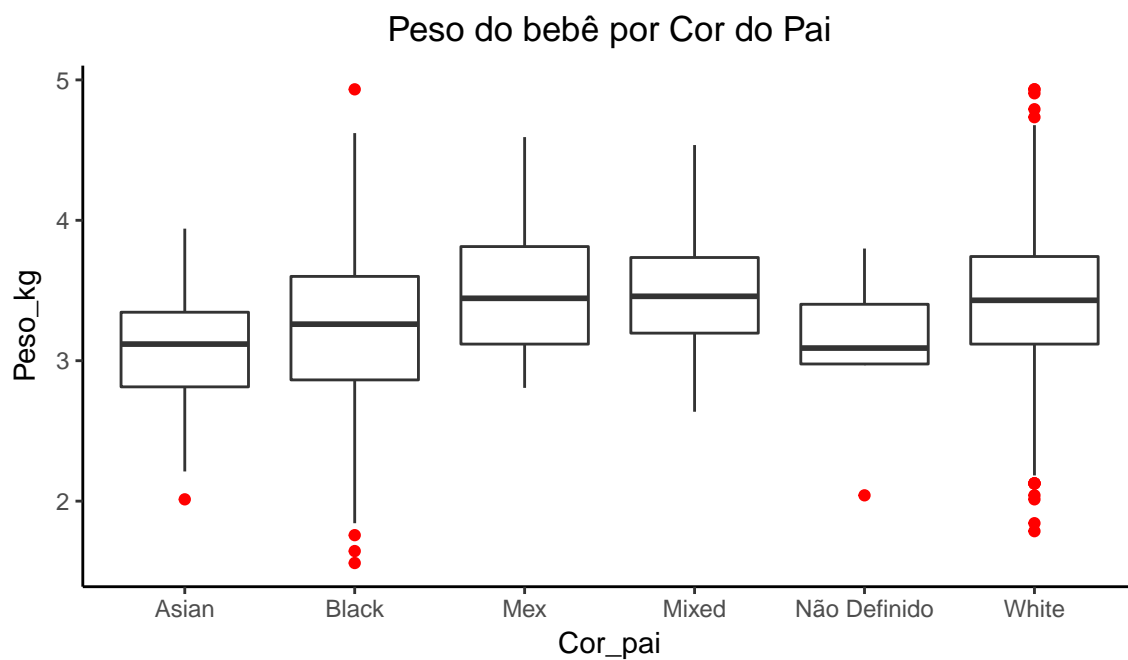


Figura 10: Boxplot entre o peso do bebe e raça do pai - mudanças nos níveis

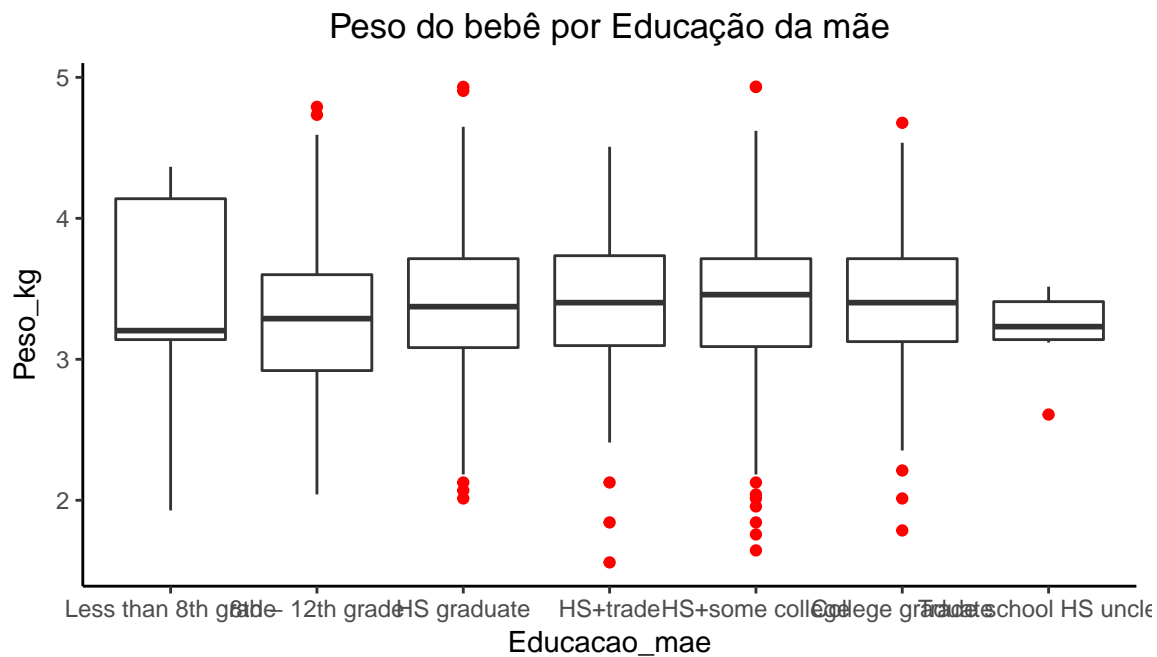


Figura 11: Boxplot entre o peso do bebe e o nível educacional da mãe

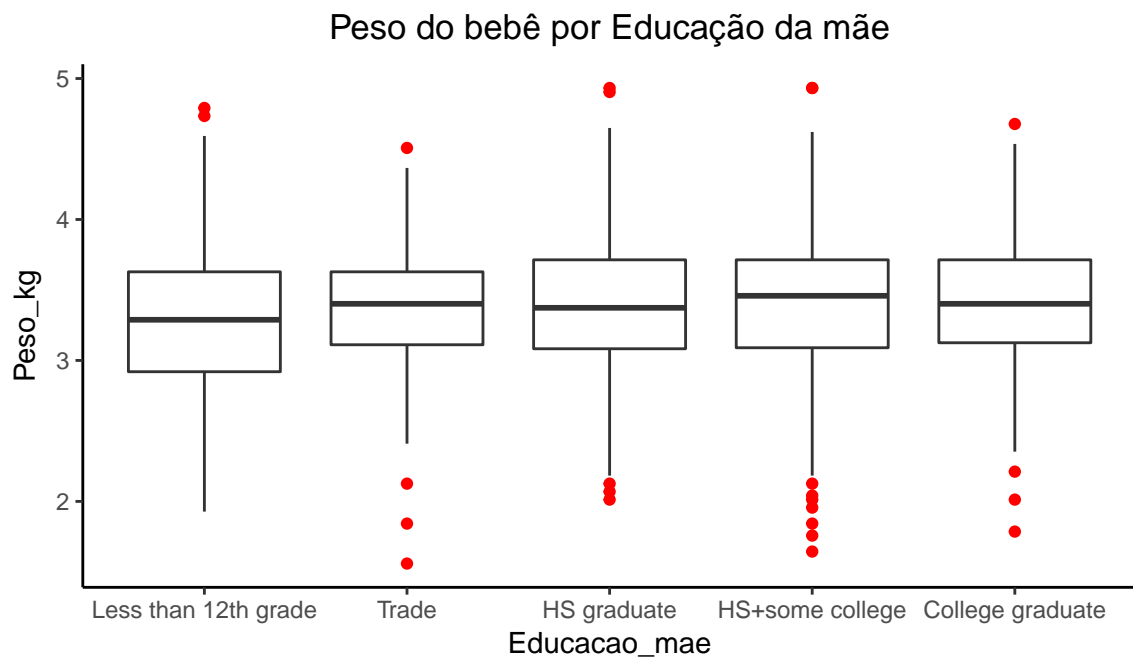


Figura 12: Boxplot entre o peso do bebe e o nível educacional da mãe - mudanças nos níveis

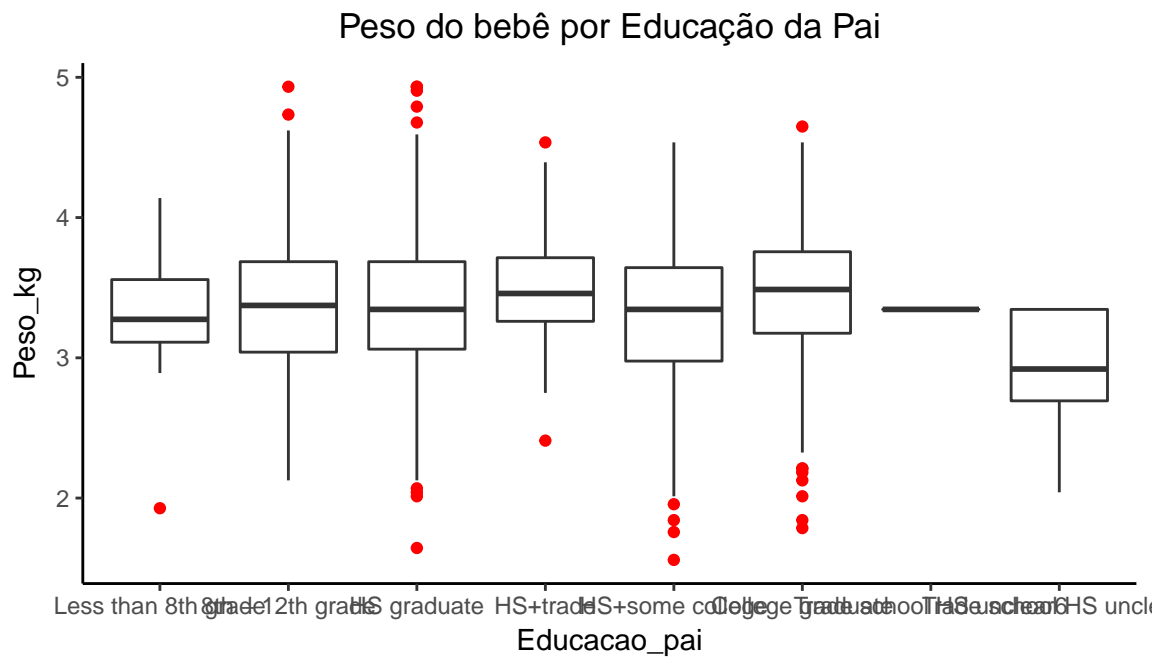


Figura 13: Boxplot entre o peso do bebe e o nível educacional do pai

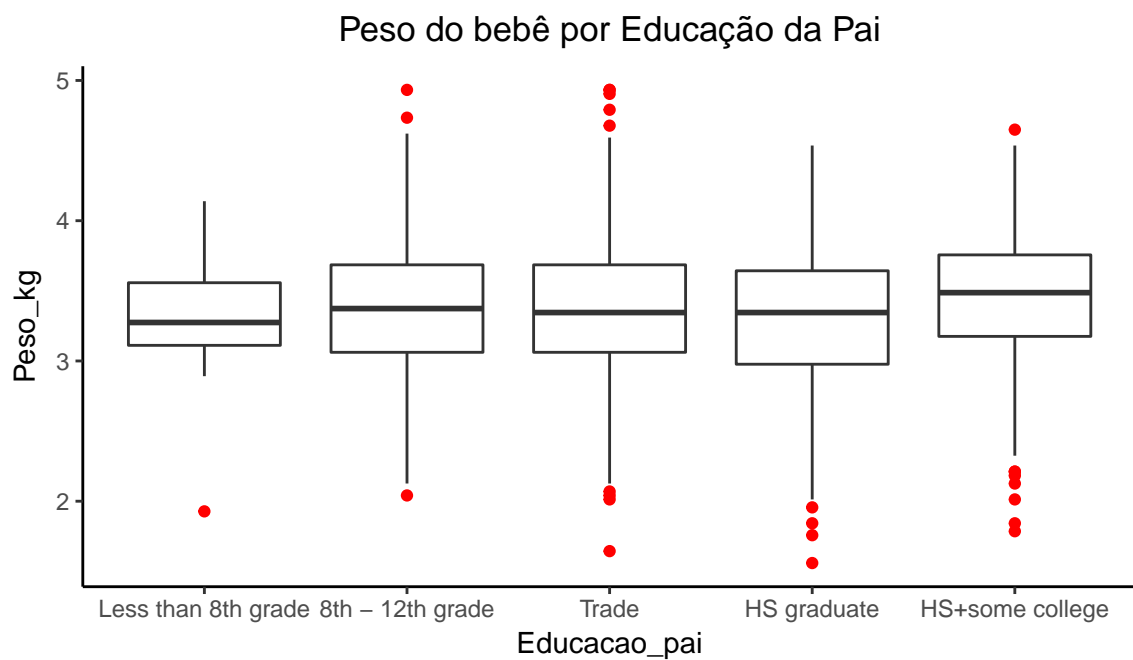


Figura 14: Boxplot entre o peso do bebe e o nível educacional do pai - mudanças nos níveis

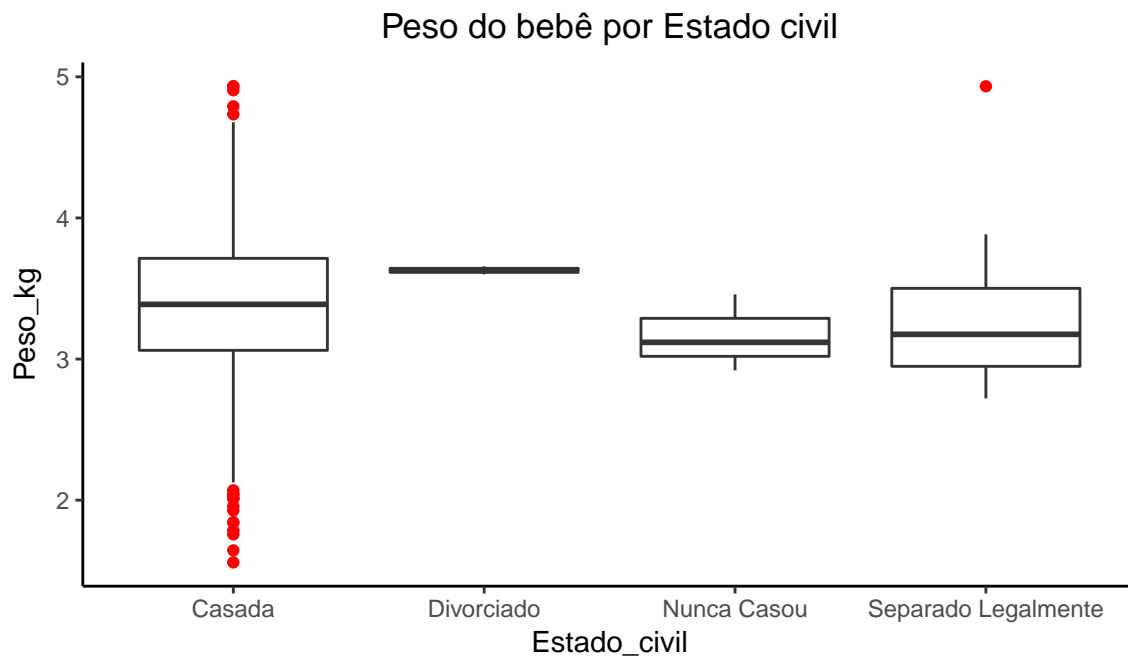


Figura 15: Boxplot entre o peso do bebe e o estado civil da mãe

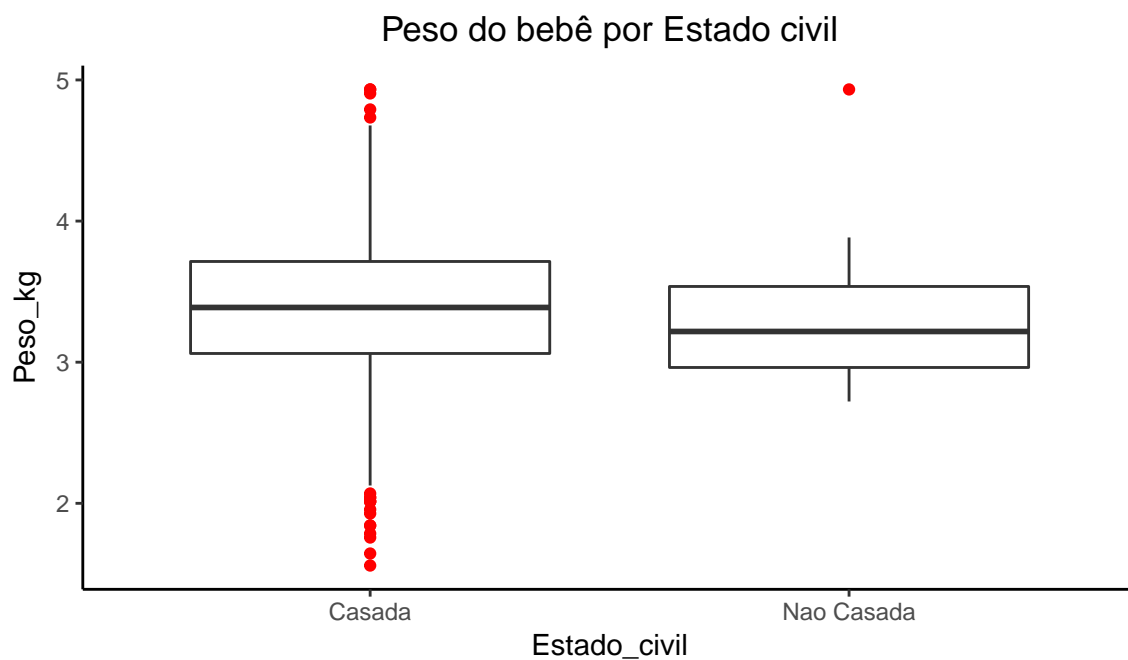


Figura 16: Boxplot entre o peso do bebe e o estado civil da mãe - mudanças nos níveis

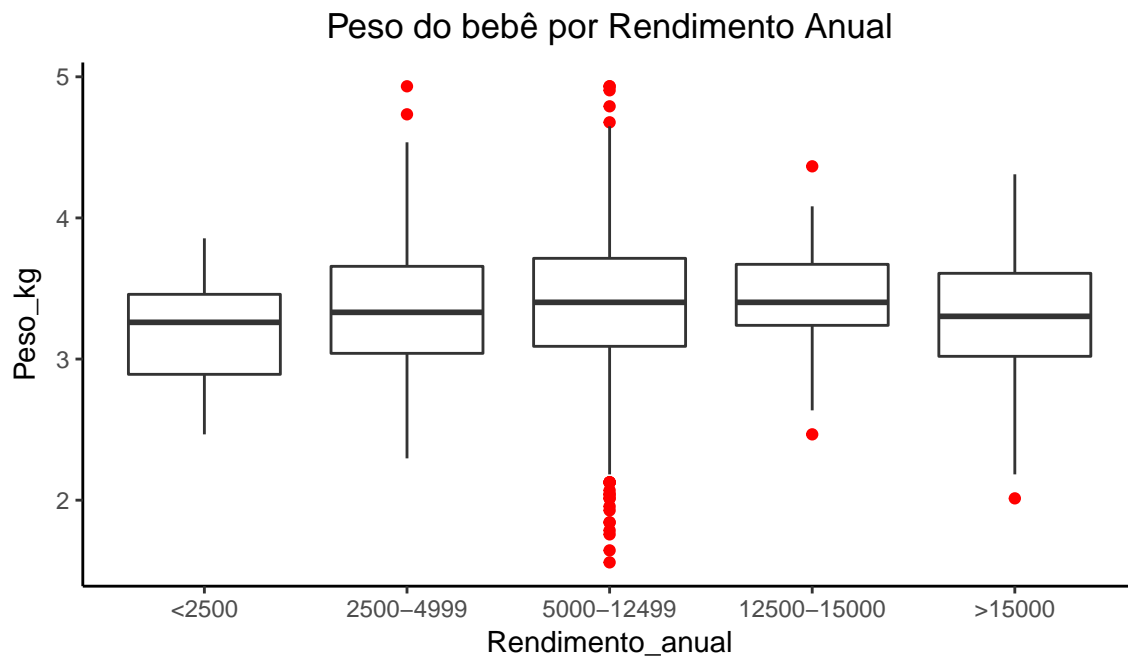


Figura 17: Boxplot entre o peso do bebe e o renda anual da família

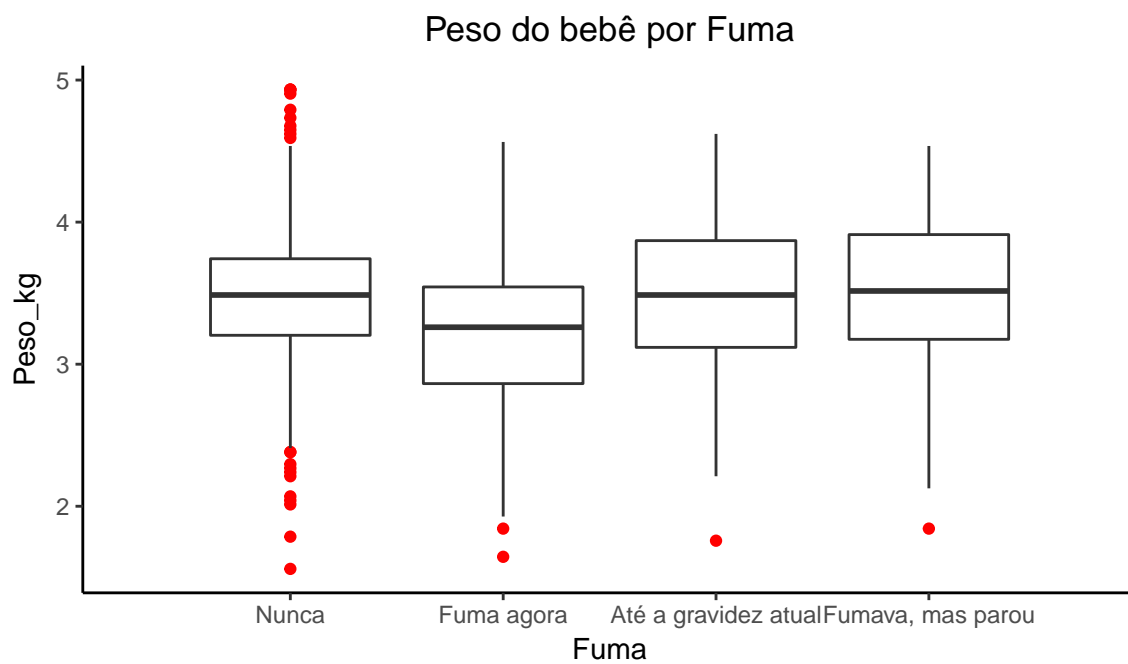


Figura 18: Boxplot entre o peso do bebe e o hábito de fumar

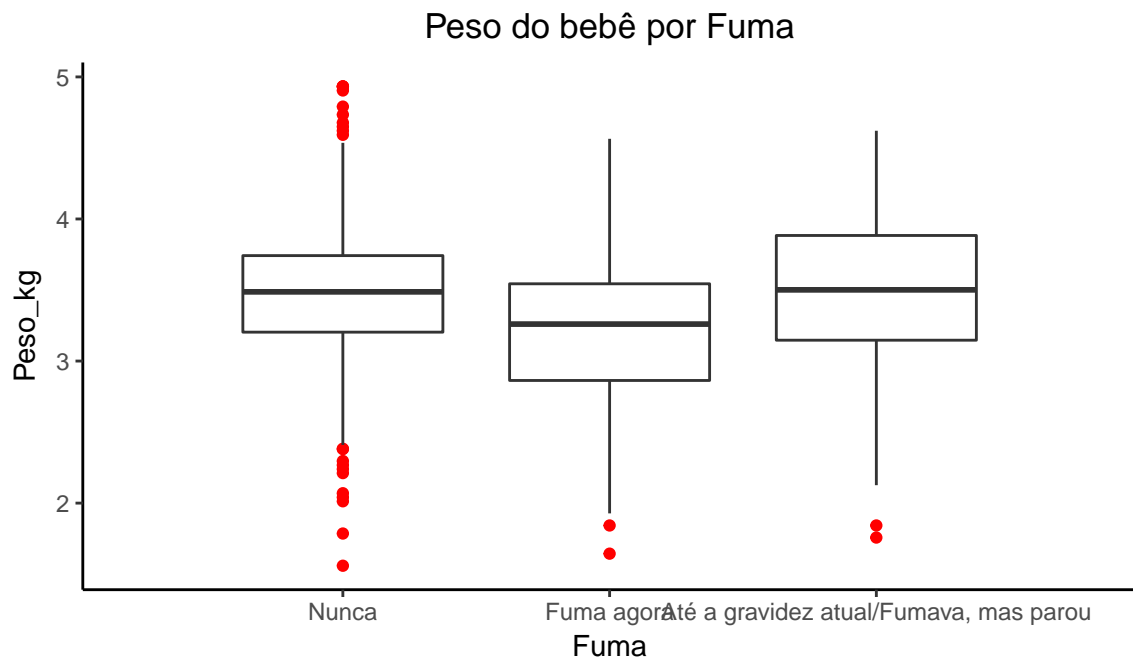


Figura 19: Boxplot entre o peso do bebe e o hábito de fumar -mudanças de níveis

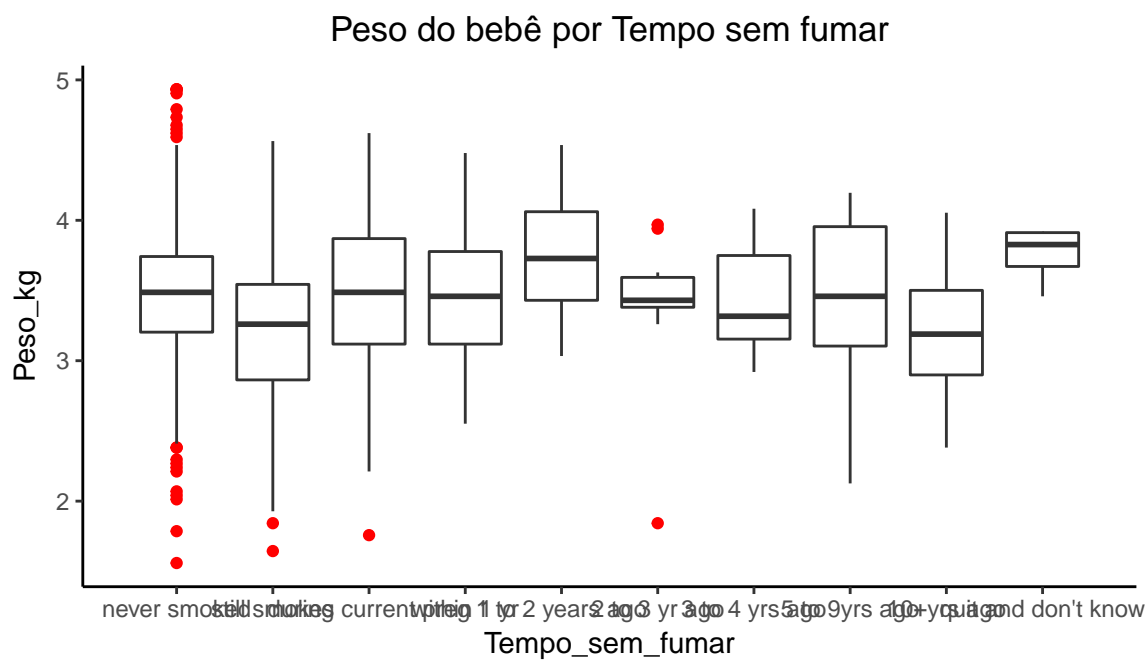


Figura 20: Boxplot entre o peso do bebe e o tempo sem fumar

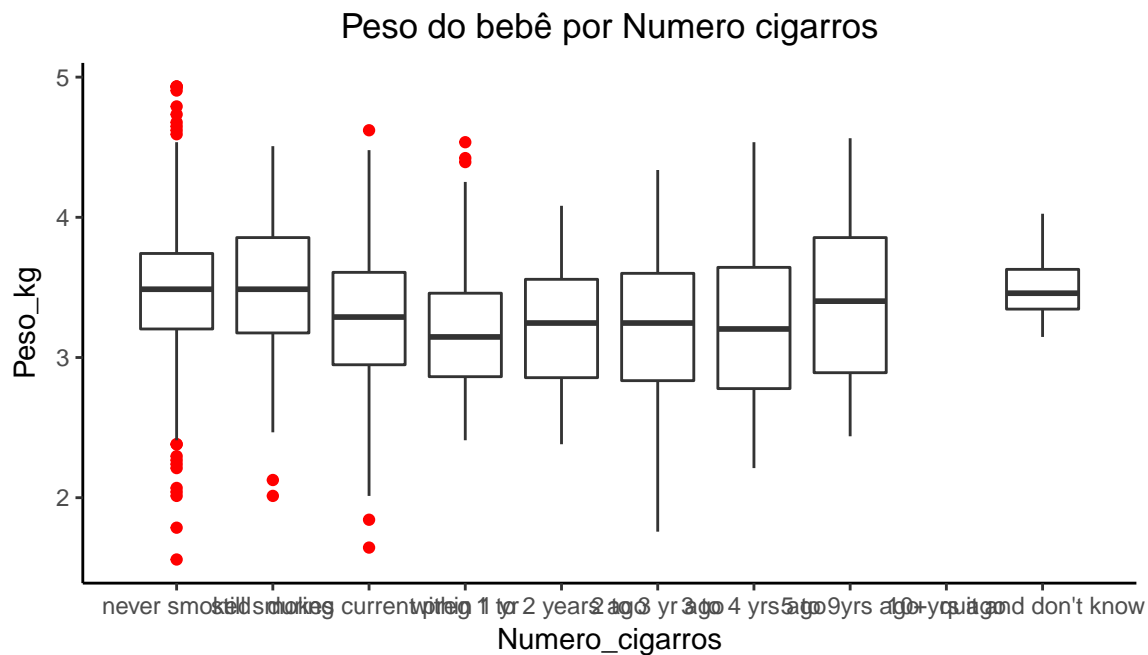


Figura 21: Boxplot entre o peso do bebe e o número de cigarros fumados por dia

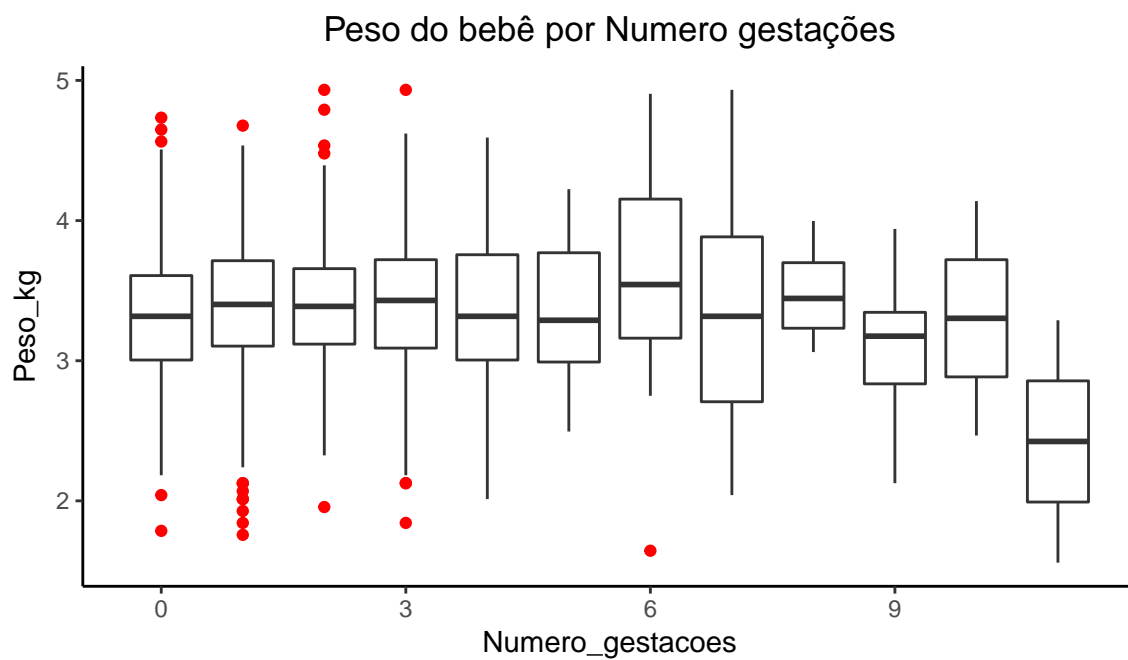


Figura 22: Boxplot entre o peso do bebe e o número de gestações anteriores da mãe

9. Referências