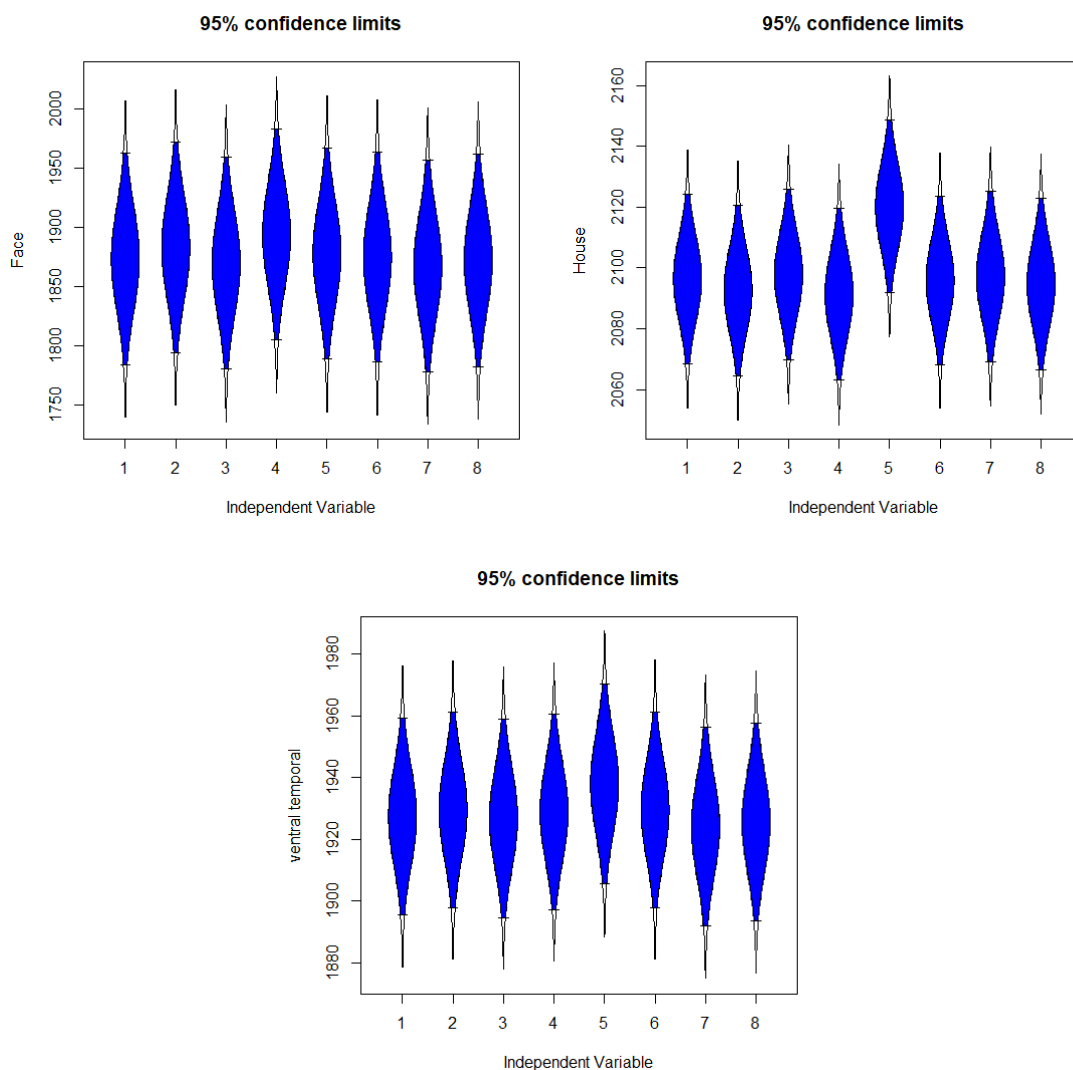
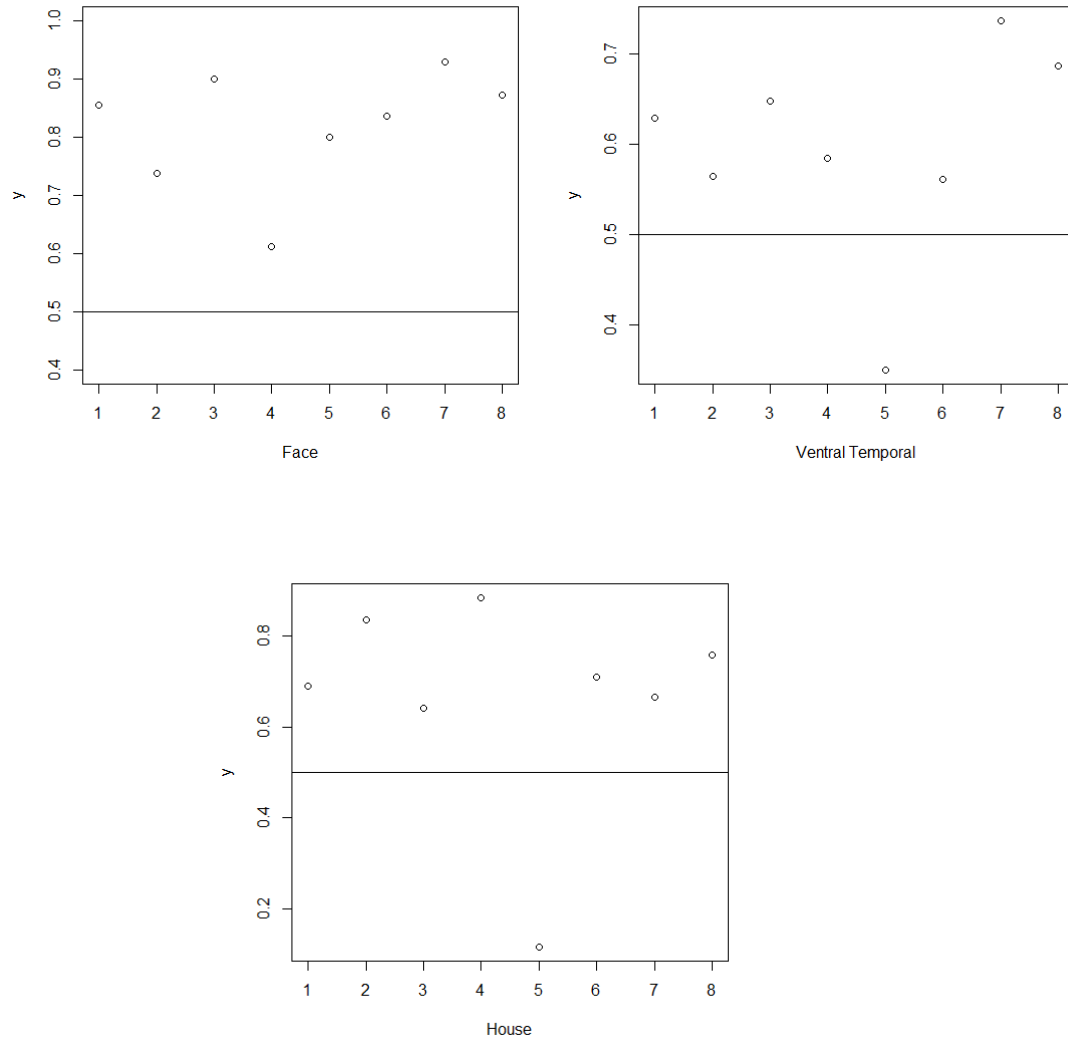


سوال 1 :

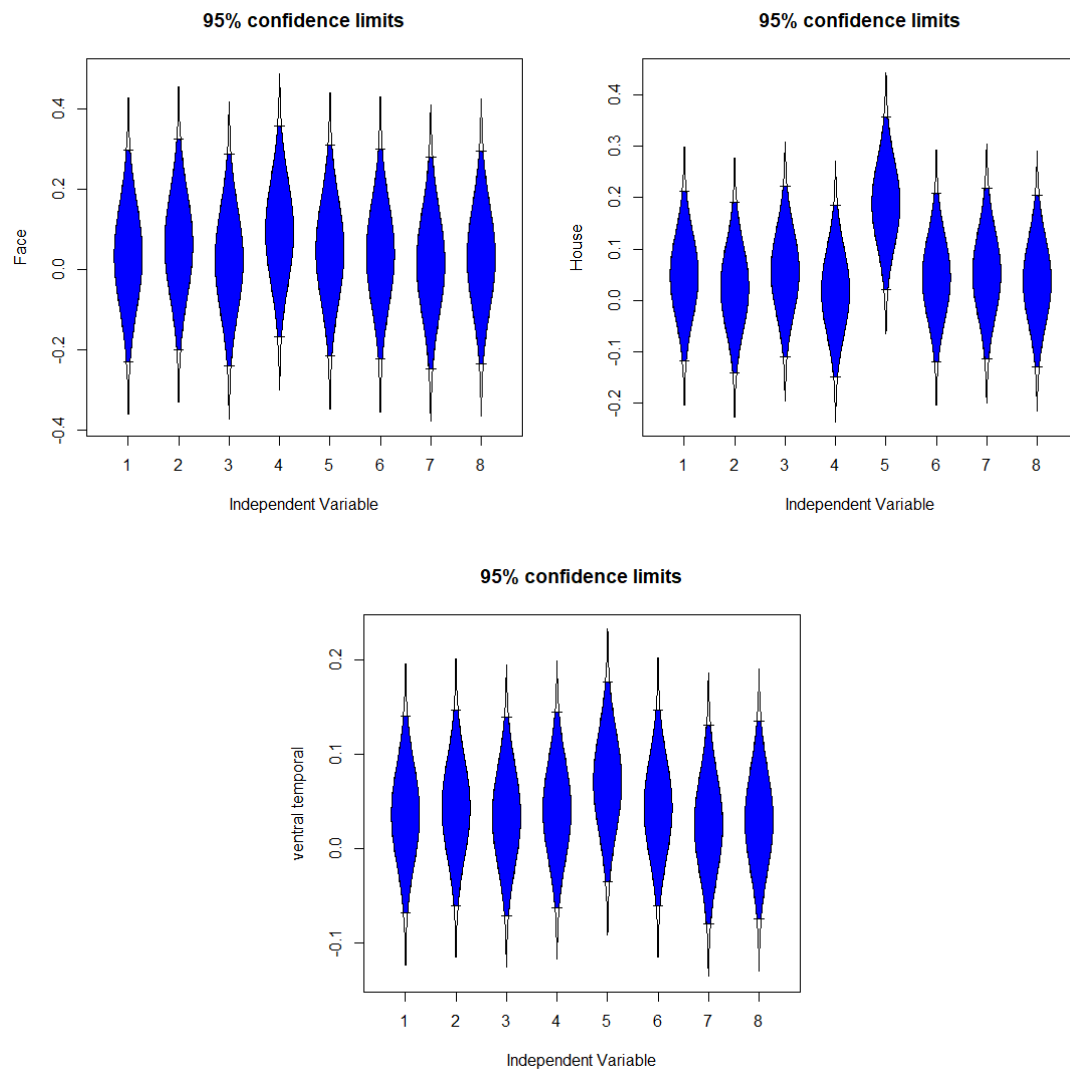
1- با انجام Pre-Processing مورد نیاز و انجام محاسبات ضروری خروجی ها به این شکل به دست آمدند:



2- برای این بخش از t-test استفاده کردم، مقایسه خروجی p-value ها با مقدار 0.05 به صورت گرافیکی در زیر آمده است. سایر پارامترهای خروجی تست در کد موجود است. ANOVA به نظر تست خوبی نمی آید، زیرا هدف بررسی تفاوت میانگین پاسخ به هر تصویر با مقدار حالت استراحت نیست و تک تک پاسخ ها برای ما ارزشمند است.



3- این بخش با نرمال کردن داده ها و سپس رسم منحنی های خطا گزارش می شود:



4- تمام خروجی ها در کد موجود است. یک نمونه به شرح زیر است:

```

[1] "House"
> t.test(mat3[,1], mat3[,6],

      Welch Two Sample t-test

data:  mat3[, 1] and mat3[, 6]
t = 0.030018, df = 273.96, p-value = 0.9761
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -38.50540  39.69783
sample estimates:
mean of x mean of y
 2096.454  2095.858

> wilcox.test(mat3[,1], mat3[,6])

      Wilcoxon rank sum test with continuity correction

data:  mat3[, 1] and mat3[, 6]
W = 9554, p-value = 0.9621
alternative hypothesis: true location shift is not equal to 0

```

با توجه به مقادیر p-value برای هم تست پارامتری و هم غیرپارامتری، تفاوت آماری قابل اثباتی وجود ندارد.

5- مشابه قسمت قبل، یک مورد را به عنوان نمونه آورده ام.

```

> wilcox.test(mat1[,4] , rest1, paired = TRUE)

      Wilcoxon signed rank test with continuity correction

data:  mat1[, 4] and rest1
V = 53920, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> t.test(mat1[,4] , rest1, paired=TRUE)

      Paired t-test

data:  mat1[, 4] and rest1
t = 14.555, df = 347, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 10.83483 14.22069
sample estimates:
mean of the differences
      12.52776

```

خروجی هر دو تست موید تفاوت بین این دو دسته داده است.

6- خروجی مقایسه برای دو روش pearson و spearman به این شکل است:

```
> cor.test(mat3[,4], mat3[,5], method="pearson")

Pearson's product-moment correlation

data:  mat3[, 4] and mat3[, 5]
t = 119.52, df = 136, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9933833 0.9966247
sample estimates:
      cor 
0.9952736

> cor.test(mat3[,4], mat3[,5], method="spearman")

Spearman's rank correlation rho

data:  mat3[, 4] and mat3[, 5]
S = 2816, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho 
0.9935706
```

7- برای بررسی این قسمت از تست Permutation استفاده شد.

```
[[1]]
[1] "Unpaired two-sample permutation test was performed."

[[2]]
[1] "p-value was estimated based on 20000 simulations."

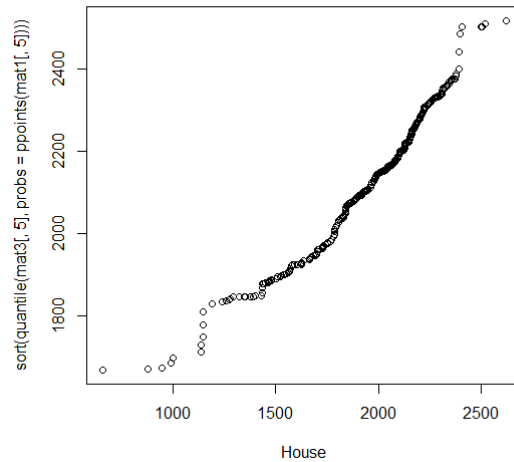
$alternative
[1] "two.sided"

$mu
[1] 0

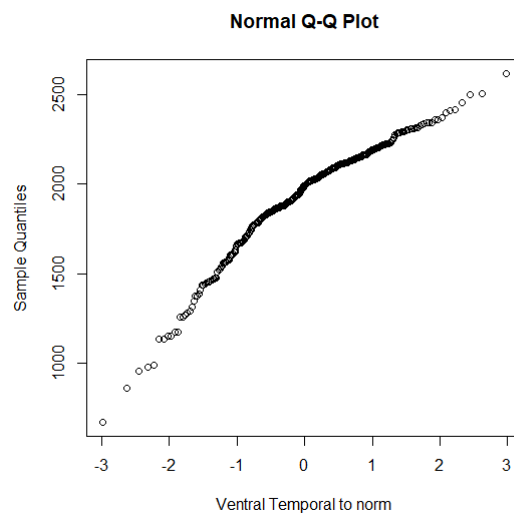
$p.value
[1] 0.15595
```

9- آزمون kolmogorov-smirnov و QQ-plot برای بررسی تشابه توزیع ها استفاده شد ، خروجی ها به این شکلند:

به عنوان مثال برای تصویر خانه شباهت پاسخ نواحی را داریم:



و مقایسه ی پاسخ ناحیه ventral temporal با نرمال به شکل زیر است:



یک نمونه خروجی آزمون Kolmogorov-smirnov هم آمده است. تمام حالت های کامل در کد موجود است:

```
Two-sample Kolmogorov-Smirnov test

data:  mat1[, 4] and mat3[, 4]
D = 0.2565, p-value = 4.511e-06
alternative hypothesis: two-sided

Warning message:
In ks.test(mat1[, 4], mat3[, 4]) :
  p-value will be approximate in the presence of ties
```

سوال 2 :

داده ی سنگین مرکز آمار لود شده و پیش پردازش مناسب انجام شد.

1- خروجی جدول به صورت منطبق بر صورت سوال گزارش شده :

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	0.9208838	28	30.14519	0.9373990	0.22928994	83545
[2,]	0.8955868	35	37.34298	0.9034948	0.25899375	94368
[3,]	0.9130192	42	44.31709	0.7883780	0.15977704	58217
[4,]	0.9531819	48	50.23199	0.6056435	0.11146271	40613
[5,]	0.9705413	53	54.97989	0.4319031	0.07861095	28643
[6,]	0.9943684	56	57.77669	0.3074910	0.06188043	22547
[7,]	1.0201291	59	60.08580	0.2181577	0.04344008	15828
[8,]	1.0481999	60	61.12057	0.1644393	0.02833979	10326
[9,]	0.9614072	61	43.24127	0.7126857	1.00000000	364364

به عنوان مثال میانه سن در مادران با تعداد 5 بچه مقدار 53 سال است.

2- دو مدل مختلف بر داده ها منطبق کردم، اولی را دقیقاً بر جدول فوق رسم کردم، یعنی سواد را باینری ندیدم که مشخصات مدل به شرح زیر شد:

```
Call:
lm(formula = ChildNo ~ Literacy)

Coefficients:
(Intercept)      Literacy
      8.773         -7.845
```

مدل دوم با باینری دیدن سواد یا بی سواد بود که فکر کنم مد نظر سوال همین مورد باشد :

```
Call:
lm(formula = o ~ Literacy)
```

```
Coefficients:
(Intercept)      Literacy
    -0.7386         1.5014
```

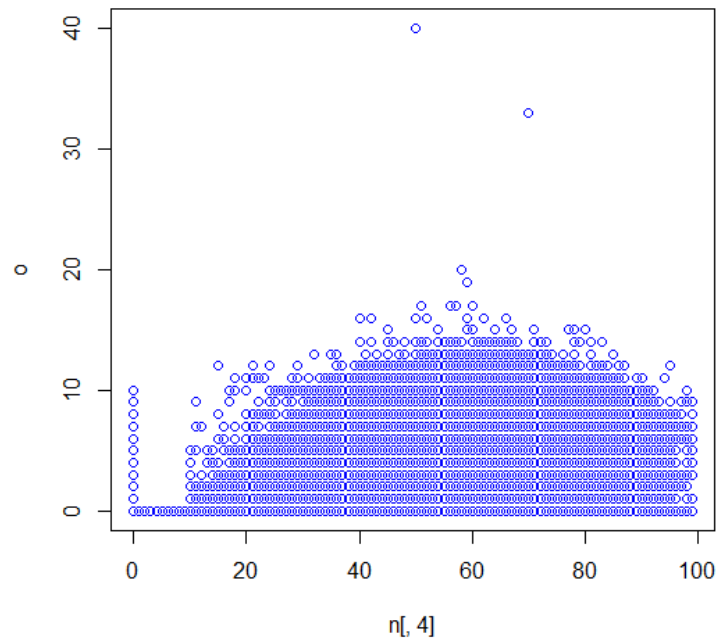
به هر حال هر دو مدل بر کاهش تعداد فرزند با افزایش سواد اتفاق نظر داشتند.

3-4 خطای هر دو مدل مشابه سوالات تمرین طول ترم محاسبه شد، یعنی با انحراف معیار مانده ها. سپس با افزودن وضعیت کار مادران (که به نظر می تواند پارامتر خوبی برای توصیف بهتر تعداد فرزند باشد) خطا مجددا بررسی شد:

```
> sd(residuals_vec2)
[1] 1.62216
> sd(residuals_vec3)
[1] 1.58811
```

دیده می شود که مطابق انتظار خطا مقداری کمتر شد.

5- برای فهم بهتر سوال ابتدا نمایشی از داده ها گرفتیم :



صرف نظر از داده های outlier ، داده ها به طور کلی مدلی از سن و تعداد فرزند ارائه می دهند. با استفاده از آزمون correlation این همبستگی را بررسی کردم:

Pearson's product-moment correlation

```
data: (m[, 18] + m[, 19]) and m[, 13]
t = -51.71, df = 304410, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.09683452 -0.08979167
sample estimates:
 cor
-0.09331426
```

مورد بعدی مقایسه نیز در کد موجود است. برخلاف همبستگی کم بین تعداد فرزند و وضع شغلی (که احتمالاً به دلیل نیاز به پارامتربندی دوباره ی داده های شغلی است)، سن همبستگی خوبی را نشان می دهد.

6- این سوال را با آزمون واریانسی روی تعامل استفاده از اینترنت با تعداد فرزندان + وضعیت امادگی برای کار با تعداد فرزندان بررسی کردم. خروجی تست به این شرح است. روال منطقی کار در کد موجود است.

```

F test to compare two variances

data:  lm(m[, 11] ~ (m[, 18] + m[, 19])) and lm(m[, 5] ~ (m[, 18] + m[, 19]))
F = 0.35799, num df = 304410, denom df = 304410, p-value < 2.2e-16
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3554568 0.3605437
sample estimates:
ratio of variances
 0.3579912

```

تست نشان می دهد واریانس ها برابر نیستند. با توجه به تست و confidence interval با احتمال خیلی بالا عدد یک در بازه نیست. پس واریانس ها برابر نیستند