Control 6: Investigación de Operaciones

Nombre: Rodrigo Cayazaya M.

Correo: rodrigo.cayazaya@sansano.usm.cl

Rol: 201773538-4

Código:

```
library(ggplot2)
library(lmtest)
library(fBasics)
library(chemCal)

#pregunta A
data = read.csv("io/control 6/FormaD.csv")
model = lm(Puntaje ~ Horas, data=data)
model
ggplot(data, aes(x=Horas, y=Puntaje)) +
  geom_point(color = "#298089", size=3) +
  geom_smooth(method= lm, formula= y~x, color= "#000000") +
  labs(title= "Puntaje vs Horas") +
  theme_test(base_size = 20)

#pregunta B
summary(data)
```

```
#pregunta C
subset(data, Horas == 3.5) #no hay valores
#Para un nuevo valor (X=3.5), observamos el IC para la respuesta de Y
new.dat <- data.frame(Horas=3.5)
predict(model, newdata = new.dat, interval = 'prediction', level = 0.95)

#pregunta D
modelSummary <- summary(model)
C <- 10.0 #La hip sería que Beta_1 es 10
modelCoeffs <- modelSummary$coefficients
beta.estimate <- modelCoeffs["Horas", "Estimate"]
std.error <- modelCoeffs["Horas", "Std. Error"]
t_value <- (beta.estimate - C)/std.error
abs(t_value)
nrow(data)

#pregunta E
fm = aov(model)
dwtest(model)
ksnormTest(fm$residuals)
1.35810/sqrt(200)
plot(fm$residuals)
```

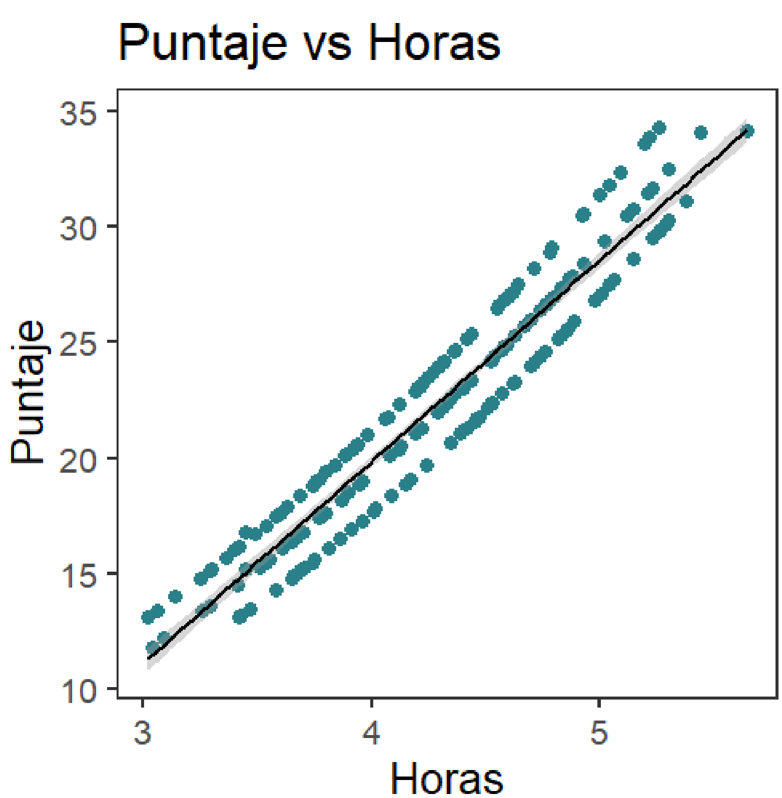
La edad promedio en que un adolescente adquiere un teléfono móvil ha ido disminuyendo con el pasar de los años. Lamentablemente, su utilización ha traído dificultades en ciertas actividades de su diario vivir. El sueño es una de esas actividades y su calidad se puede ver comprometida considerando las horas que se utiliza un celular al día. Se ha visto que en promedio, los adolescentes utilizan 3 horas su teléfono durante el día. Sin embargo, se han observado situaciones donde sobrepasan las 5 horas diarias. Preocupados por esta situación, el Ministro de Salud ha decidido encarar esta situación realizando un estudio en el Colegio Pumahue de la ciudad de Puerto Montt. Los adolescentes encuestados tienen en promedio 16 años. El estudio implica realizar un cuestionario que considera el Índice de Calidad de Sueño de Pittisburgh (PSQI)[1] para evaluar la calidad del sueño del estudiantado. Aquí, mientras más grande el puntaje obtenido en el cuestionario, peor es su calidad de sueño.
Según el dataset asignado disponible en Aula (ver siguiente página FormaX.csv), parámetros asignados y utilizando R, conteste las siguientes preguntas:

(a) (10 puntos) Presente un modelo de regresión lineal entre la cantidad de horas que utilizan el celular y el puntaje obtenido en el PSQI.

```
Call:
lm(formula = Puntaje ~ Horas, data = data)

Coefficients:
(Intercept)        Horas
    -15.015        8.717
```

El modelo de regresión lineal es:

Puntaje = 8.717 Horas – 15.015



(b) (10 puntos) Sobre la variación en el puntaje PSQI, ¿Qué tanto no podría ser explicado por la cantidad de horas de uso de celular?

```
> summary(data)
    Puntaje           Horas      
 Min.   :11.80   Min.   :3.020  
 1st Qu.:17.60   1st Qu.:3.777  
 Median :22.30   Median :4.295  
 Mean   :22.25   Mean   :4.275  
 3rd Qu.:25.93   3rd Qu.:4.740  
 Max.   :34.30   Max.   :5.650  
```

La cantidad de datos que existe sólo abarca entre 3.02 hasta 5.65 horas, por lo que no se pueden conocer los valores fuera de este rango.

(c) (25 puntos) Con un 95 % de certeza, describa la posible respuesta esperada del puntaje PSQI cuando se usa un celular 3.5 horas.

```
> new.dat <- data.frame(Horas=3.5)
> predict(model, newdata = new.dat, interval = 'prediction', level = 0.95)
       fit      lwr      upr
1 15.49542 12.49673 18.49411
```

Los valores aceptados se encuentran entre 12.49673 y 18.49411, centrado en 15.49542.

(d) (25 puntos) Una autoridad indica en un medio nacional que "existe una relación importante entre ambas variables. En general, el puntaje obtenido crece unas **C** veces la cantidad de horas de uso de los móviles". Verifique dicha aseveración.

```
> #t-test de Hipótesis para verificar "La producción aumenta al doble respecto a la temperatura"
> modelSummary <- summary(model)
> C <- 10.0 #La hip sería que Beta_1 es 10
> modelCoeffs <- modelSummary$coefficients
> beta.estimate <- modelCoeffs["Horas", "Estimate"]
> std.error <- modelCoeffs["Horas", "Std. Error"]
> t_value <- (beta.estimate - C)/std.error
> abs(t_value)
[1] 7.089056
> nrow(data)
[1] 200
```

Se obtiene un valor de 7.089. Ahora se busca el valor en la tabla con n-2 grados de libertad, sin embargo, no se encontró con 198 grados de libertad, por lo que se utilizó el valor de la tabla con 200 grados (el cual no debería variar mucho). Así se obtiene un valor de 1.972 (ver imagen de tabla adjunta).

Por lo tanto, se rechaza la hipótesis de que beta_1 es igual a 10. En otras palabras, el puntaje no crece C=10 veces la cantidad de horas de uso de los móviles.

Degrees of freedom	*Two-tailed test:* / *One-tailed test:*	10% / 5%	5% / 2.5%	2% / 1%	1% / 0.5%	0.2% / 0.1%	0.1% / 0.05%
1		6.314	12.706	31.821	63.657	318.309	636.619
2		2.920	4.303	6.965	9.925	22.327	31.599
3		2.353	3.182	4.541	5.841	10.215	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
6		1.943	2.447	3.143	3.707	5.208	5.959
7		1.894	2.365	2.998	3.499	4.785	5.408
8		1.860	2.306	2.896	3.355	4.501	5.041
9		1.833	2.262	2.821	3.250	4.297	4.781
10		1.812	2.228	2.764	3.169	4.144	4.587
11		1.796	2.201	2.718	3.106	4.025	4.437
12		1.782	2.179	2.681	3.055	3.930	4.318
13		1.771	2.160	2.650	3.012	3.852	4.221
14		1.761	2.145	2.624	2.977	3.787	4.140
15		1.753	2.131	2.602	2.947	3.733	4.073
16		1.746	2.120	2.583	2.921	3.686	4.015
17		1.740	2.110	2.567	2.898	3.646	3.965
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
21		1.721	2.080	2.518	2.831	3.527	3.819
22		1.717	2.074	2.508	2.819	3.505	3.792
23		1.714	2.069	2.500	2.807	3.485	3.768
24		1.711	2.064	2.492	2.797	3.467	3.745
25		1.708	2.060	2.485	2.787	3.450	3.725
26		1.706	2.056	2.479	2.779	3.435	3.707
27		1.703	2.052	2.473	2.771	3.421	3.690
28		1.701	2.048	2.467	2.763	3.408	3.674
29		1.699	2.045	2.462	2.756	3.396	3.659
30		1.697	2.042	2.457	2.750	3.385	3.646
32		1.694	2.037	2.449	2.738	3.365	3.622
34		1.691	2.032	2.441	2.728	3.348	3.601
36		1.688	2.028	2.434	2.719	3.333	3.582
38		1.686	2.024	2.429	2.712	3.319	3.566
40		1.684	2.021	2.423	2.704	3.307	3.551
42		1.682	2.018	2.418	2.698	3.296	3.538
44		1.680	2.015	2.414	2.692	3.286	3.526
46		1.679	2.013	2.410	2.687	3.277	3.515
48		1.677	2.011	2.407	2.682	3.269	3.505
50		1.676	2.009	2.403	2.678	3.261	3.496
60		1.671	2.000	2.390	2.660	3.232	3.460
70		1.667	1.994	2.381	2.648	3.211	3.435
80		1.664	1.990	2.374	2.639	3.195	3.416
90		1.662	1.987	2.368	2.632	3.183	3.402
100		1.660	1.984	2.364	2.626	3.174	3.390
120		1.658	1.980	2.358	2.617	3.160	3.373
150		1.655	1.976	2.351	2.609	3.145	3.357
200		1.653	1.972	2.345	2.601	3.131	3.340
300		1.650	1.968	2.339	2.592	3.118	3.323

(e) (30 puntos) Explique y verifique los supuestos de correlación y normalidad en los residuales. Utilice un gráfico de los residuos para complementar el último punto.

```
> dwtest(model)

        Durbin-Watson test

data:  model
DW = 1.9762, p-value = 0.4335
alternative hypothesis: true autocorrelation is greater than 0
```

Se obtiene un estadístico de 1.9762, por lo que al ver la tabla:

33	1.383	1.508
34	1.393	1.514
35	1.402	1.519
36	1.411	1.525
37	1.419	1.530
38	1.427	1.535
39	1.435	1.540
40	1.442	1.544
45	1.475	1.566
50	1.503	1.585
55	1.528	1.601
60	1.549	1.616
65	1.567	1.629
70	1.583	1.641
75	1.598	1.652
80	1.611	1.662
85	1.624	1.671
90	1.635	1.679
95	1.645	1.687
100	1.654	1.694
150	1.720	1.747
200	1.758	1.779

Obtenemos un valor de dU = 1.779, por lo que se acepta la hipótesis nula, es decir que con un nivel de significancia del 5%, **los errores no están relacionados**.

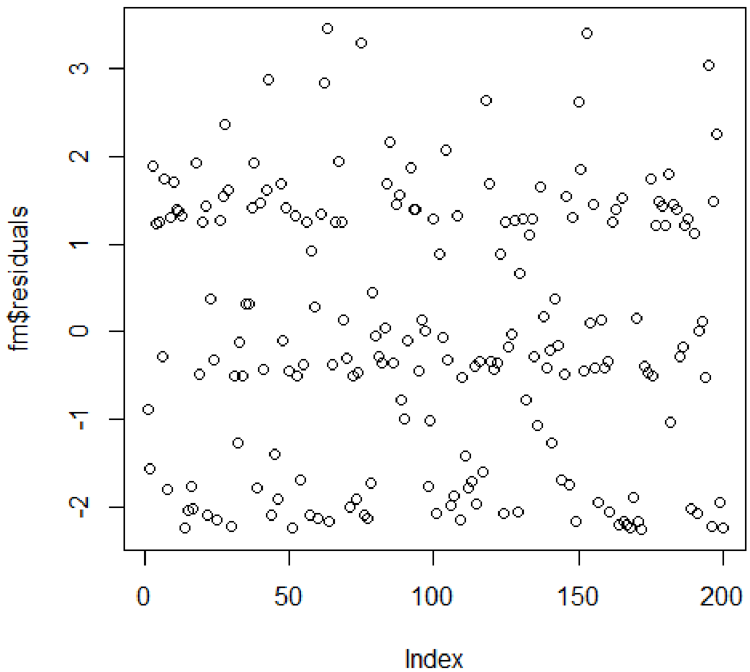
```
> ksnormTest(fm$residuals)

Title:
 One-sample Kolmogorov-Smirnov test

Test Results:
  STATISTIC:
    D: 0.2423
  P VALUE:
    Alternative Two-Sided: 1.257e-10
    Alternative      Less: 6.285e-11
    Alternative   Greater: 2.885e-07
```

Se obtiene un valor estadístico de 0.2423. Ahora si buscamos en la tabla, encontramos que se necesita calcular el valor como:

```
> 1.35810/sqrt(200)
[1] 0.09603217
```

n\ α	0.001	0.01	0.02	0.05	0.1	0.15	0.2
1		0.99500	0.99000	0.97500	0.95000	0.92500	0.90000
2	0.97764	0.92930	0.90000	0.84189	0.77639	0.72614	0.68377
3	0.92063	0.82900	0.78456	0.70760	0.63604	0.59582	0.56481
4	0.85046	0.73421	0.68887	0.62394	0.56522	0.52476	0.49265
5	0.78137	0.66855	0.62718	0.56327	0.50945	0.47439	0.44697
6	0.72479	0.61660	0.57741	0.51926	0.46799	0.43526	0.41035
7	0.67930	0.57580	0.53844	0.48343	0.43607	0.40497	0.38145
8	0.64098	0.54180	0.50654	0.45427	0.40962	0.38062	0.35828
9	0.60846	0.51330	0.47960	0.43001	0.38746	0.36006	0.33907
10	0.58042	0.48895	0.45662	0.40925	0.36866	0.34250	0.32257
11	0.55588	0.46770	0.43670	0.39122	0.35242	0.32734	0.30826
12	0.53422	0.44905	0.41918	0.37543	0.33815	0.31408	0.29573
13	0.51490	0.43246	0.40362	0.36143	0.32548	0.30233	0.28466
14	0.49753	0.41760	0.38970	0.34890	0.31417	0.29181	0.27477
15	0.48182	0.40420	0.37713	0.33760	0.30397	0.28233	0.26585
16	0.46750	0.39200	0.36571	0.32733	0.29471	0.27372	0.25774
17	0.45440	0.38085	0.35528	0.31796	0.28627	0.26587	0.25035
18	0.44234	0.37063	0.34569	0.30936	0.27851	0.25867	0.24356
19	0.43119	0.36116	0.33685	0.30142	0.27135	0.25202	0.23731
20	0.42085	0.35240	0.32866	0.29407	0.26473	0.24587	0.23152
25	0.37843	0.31656	0.30349	0.26404	0.23767	0.22074	0.20786
30	0.34672	0.28988	0.27704	0.24170	0.21756	0.20207	0.19029
35	0.32187	0.26898	0.25649	0.22424	0.20184	0.18748	0.17655
40	0.30169	0.25188	0.23993	0.21017	0.18939	0.17610	0.16601
45	0.28482	0.23780	0.22621	0.19842	0.17881	0.16626	0.15673
50	0.27051	0.22585	0.21460	0.18845	0.16982	0.15790	0.14886
OVER 50	$\frac{1.94947}{\sqrt{n}}$	$\frac{1.62762}{\sqrt{n}}$	$\frac{1.51743}{\sqrt{n}}$	$\frac{1.35810}{\sqrt{n}}$	$\frac{1.22385}{\sqrt{n}}$	$\frac{1.13795}{\sqrt{n}}$	$\frac{1.07275}{\sqrt{n}}$

Debido a que el valor de la tabla es menor al de nuestro estadístico, se rechaza la hipótesis, es decir, con un nivel de significancia del 5% **los errores no siguen una distribución normal**.



Esto se puede observar en el gráfico, ya que los valores no se encuentran distribuidos de forma aleatoria, existen grupos de valores residuales entre [1,2], [-1,0] y [-2,-1].