



Universidad de Chile
Facultad de Ciencias Sociales
Profesores: Catalina Canals y Eduardo Toro
Apoyo docente: Cristóbal Moya
Ayudante: Juan Carrasco

Preguntas de Repaso Prueba Estadística III
Unidad III "Uso de modelos predictivos en Ciencias Sociales"

La prueba de la Unidad III contará con 4 preguntas, 1 del tema 10, y 3 de los temas 11 y 12. La prueba podrá realizarse con apuntes. Se sugiere responder las siguientes preguntas a modo de estudio para la prueba.

Preguntas

Tema 10

1. ¿Qué dificultades existen para estudiar la causalidad en Ciencias Sociales? ¿En qué consiste "controlar" y como esto aporta al estudio de relaciones entre variables?
2. ¿Qué es un modelo? ¿De qué depende la calidad de un modelo? Explique utilizando un ejemplo.
3. ¿Qué es un modelo? ¿Cómo se relaciona con la teoría?
4. ¿Qué caracterizan a los modelos estadísticos?
5. ¿Por qué usar modelos en Ciencias Sociales?
6. ¿Qué limitaciones tiene modelar en Ciencias Sociales?
7. ¿Cuál es la diferencia entre correlación y causalidad? Explique mediante un ejemplo no mencionado en clases.

Temas 11 y 12

8. ¿En qué consiste un modelo de regresión lineal? ¿Para qué sirve?
9. Describa cómo debe ser la especificación de un modelo de regresión lineal.
10. Plantee 3 preguntas de investigación que podrían ser respondidas a partir de un análisis de regresión lineal.
11. Explique los siguientes conceptos asociados a la regresión lineal: (1) residuos, (2) error, (3) pendiente.
12. ¿Cuáles son los coeficientes del modelo de regresión lineal? ¿Qué indican? ¿Cómo se estiman?
13. ¿Qué son los supuestos de un modelo? ¿Cómo afecta su incumplimiento al modelo? Ejemplifique refiriéndose a uno de los supuestos de la regresión lineal.
14. Un grupo de investigadores realizan una regresión lineal múltiple para predecir la posición política de las personas (en una escala donde 1 es izquierda y 10 es derecha), a partir de sus ingresos, su educación y su comuna de residencia, obteniendo un R cuadrado ajustado=0,17. Interprete estadísticamente el R cuadrado; ¿Qué podría explicar el bajo ajuste de un modelo?
15. ¿El incumplimiento de qué supuestos de la regresión lineal limitan las posibilidades de hacer inferencia estadística?
16. Un grupo de investigadores realizan una regresión lineal múltiple para predecir la posición política de las personas (en una escala donde 1 es izquierda y 10 es derecha), a partir de sus

ingresos, su educación y su comuna de residencia, obteniendo un R cuadrado ajustado=0,17. ¿Qué aspectos cree que podrían influir en la posición política que no están incluidos en esta regresión, de modo que estarían incluidos en el error del modelo? Considerando esto, ¿Cree usted que se cumple el supuesto de independencia de las variables independientes y el error?

17. ¿Qué podemos hacer para modelar relaciones no lineales entre variables a través de una regresión lineal? Explique con un ejemplo.
18. ¿Qué son los problemas de identificación? ¿Qué podrían generar estos problemas en una regresión (lineal, probit o logit)?
19. Las regresiones (lineal, probit y logit), ¿Permiten afirmar la existencia de relaciones causales entre variables? Explique con un ejemplo.
20. Un grupo de investigadores desarrollan un modelo de regresión lineal para predecir la satisfacción en la vida (en una escala de 1 a 10, donde 1 es nula satisfacción y 10 es máxima). Los resultados son presentados en la siguiente tabla, donde los coeficientes con un, dos y tres asteriscos indican que el valor p asociado al test de Wald es menor que 0,1; 0,05 y 0,01, respectivamente.

Variable	Coef.	EE
Edad	-0.055***	(0.020)
Edad ²	0.000**	(0.000)
Género (ref. Hombre)	-0.038	(0.138)
Pareja Estable	0.564***	(0.144)
Religión: Católica (ref. Ninguna)	0.016	(0.201)
Religión: Evangélica (ref. Ninguna)	0.158	(0.269)
Actividad Religiosa	0.043	(0.034)
Zona: Urbana (ref. Rural)	0.225	(0.187)
GSE: Medio (ref. Alto)	-1.576***	(0.321)
GSE: Bajo (ref. Alto)	-2.115***	(0.337)
Triunfadores Frustrados	-0.252*	(0.145)
Intercepto	9.314***	(0.527)
<i>N</i>		1,012
<i>R</i> ²		0.055

- a) Interprete el ajuste del modelo. En caso de no ser perfecto sugiera brevemente por qué podría no serlo.
- b) Con 95% de confianza, ¿Qué variables inciden en la variable dependiente?
- c) ¿Qué efecto produce tener una pareja estable? ¿Qué efecto produce tener nivel socioeconómico bajo? Explique por qué cree usted que existen tales efectos.
21. ¿En qué consisten los modelos probit y logit? ¿Para qué sirven?
22. Plantee 3 preguntas de investigación que podrían ser respondidas a partir de un modelo probit o logit.
23. ¿En qué se diferencian un modelo probit de un modelo logit?
24. ¿Por qué es preferible utilizar un modelo logit o probit en vez de una regresión lineal para predecir una variable dummy? Explique con un ejemplo. Se sugiere utilizar un gráfico para complementar su respuesta.

25. La especificación de modelos probit y logit, ¿Qué requisitos impone sobre el nivel de medida de las variables dependientes e independientes?
26. ¿Cómo se estiman los coeficientes en los modelos probit y logit?
27. Un grupo de investigadores desarrolló un modelo logit para predecir la probabilidad de ser fumador, a partir de los ingresos, los años de escolaridad, la jornada laboral y la autopercepción de salud de las personas. Una vez estimado el modelo, el porcentaje de casos bien clasificados corresponde al 63%. ¿Cómo se calcula este porcentaje? ¿Cómo se interpreta?
28. Un grupo de investigadores desarrolló un modelo probit para predecir la probabilidad de ser fumador, a partir de los ingresos, los años de escolaridad, la jornada laboral y la autopercepción de salud de las personas. Una vez estimado el modelo, el R cuadrado de McFadden=0,12. Interprete estadísticamente el ajuste del modelo. En caso de no ser perfecto sugiera brevemente por qué podría no serlo.
29. Un grupo de investigadores desarrolla un modelo logit para predecir la probabilidad de haber ido a un concierto en los últimos 3 meses. Los resultados son los siguientes.

Tabla 4. Análisis de regresión logística sobre la probabilidad de haber ido a un concierto

	Coefficiente	Error Std.
Mujer ^a	0,076	(0,043)
Edad	-0,025**	(0,002)
Más de 8 años o bachiller, EGB o ESO ^b	0,681**	(0,092)
Bachiller superior, FPI o FPII	1,081**	(0,098)
Diplomado, licenciado, doctor	1,417**	(0,102)
Parado ^c	-0,054	(0,080)
Jubilado, incapacitado, otros	-0,206*	(0,084)
Estudiante	0,238**	(0,077)
Labores del hogar	-0,185*	(0,078)
Casado o en pareja sin hijos ^d	-0,364**	(0,073)
Casado o en pareja con hijos menores	-0,594**	(0,055)
Casado o en pareja con hijos mayores	-0,188**	(0,068)
Nº CDs comprados	0,085**	(0,009)
Posee reproductor MP3	0,456**	(0,044)
Lee revistas culturales	0,643**	(0,052)
Nº horas lúdicas en Internet	0,005	(0,005)
Nº discos grabados o descargados	0,010**	(0,003)
Constante	0,613**	(0,131)

N 14.241

Bondad del ajuste -7.412,346

Pseudo R² 0,172

Fuente: Encuesta de hábitos y prácticas culturales (2007) y elaboración propia.

* valor $p < 0,05$.

** valor $p < 0,01$.

a. «Hombre» como categoría de referencia.

b. «Menos de primaria» como categoría de referencia.

c. «Trabajando» como categoría de referencia.

d. «Soltero» como categoría de referencia.

- a) Con 95% de confianza, ¿Qué variables inciden en la probabilidad de haber ido a un concierto?
- b) ¿Cómo es el efecto de la edad en la probabilidad de ir a un concierto? ¿Cómo es el efecto de estar casado o en pareja con hijos menores en la probabilidad de ir a un concierto? Explique por qué cree que existen tales efectos.
30. Un grupo de investigadores desarrolla un modelo logit para predecir la probabilidad de identificarse como español. Los resultados son presentados en la siguiente tabla.

Predictor	Wald	Exp. B ⁴
Edad	-2,7**	0,93
Sexo (femenino)	-5,6***	0,72
Ciudad de residencia (Barcelona)	-0,6	
Lugar de nacimiento (España)	5,6***	1,66
Tiempo de residencia en España	3,0**	1,03
Estatus familiar (FSES)	-1,2	—
Conocimiento del español (KSI)	-2,1*	0,88
Lugar de nacimiento del padre (España)	4,7***	1,81
Lugar de nacimiento de la madre (España)	2,6**	1,47
Relaciones intergeneracionales (IntGen)	-0,6	
Experiencias de discriminación	-2,2*	0,89
	Pseudo R ² = 0,019	
	N = 4,577	

1. Coeficientes de regresión logística binomial.
 2. Probabilidades netas calculadas solo para efectos significativos.
 3. Coeficientes ordenados de regresión multinomial.
 4. Probabilidades netas calculadas solo para efectos significativos.
- * $p < 0,05$
** $p < 0,01$
*** $p < 0,001$

- a) Interprete estadísticamente el ajuste del modelo.
b) Con 95% de confianza, ¿Qué variables inciden en la probabilidad de considerarse español?
c) ¿Cómo es el efecto del tiempo de residencia en España en la probabilidad de considerarse español? ¿Cuál es el efecto del lugar de nacimiento del padre en la probabilidad de considerarse español?
31. Un grupo de investigadores desarrolla un modelo logit para predecir la probabilidad de tener pareja (estar casado o conviviendo) a partir de la edad y de una variable dummy donde el valor 1 indica que la persona está ocupada y 0 que no lo está (es decir, está cesante o es inactiva). Los resultados son los siguientes.

	Delta-method				
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]
edad	.0075454	.0000527	143.24	0.000	.0074422 .0076487
ocupal	.1630431	.0021776	74.87	0.000	.1587752 .1673111

- a) Con 95% de confianza, ¿Qué variables inciden en la probabilidad de tener pareja?
b) ¿Cómo es el efecto promedio de la edad en tener pareja? ¿Cuál es el efecto promedio de estar ocupado en tener pareja? Explique por qué cree que existen tales efectos.
32. Un grupo de investigadores desarrolla un modelo probit para predecir la probabilidad de considerarse independiente políticamente en México. Los coeficientes estimados del modelo son presentados en la siguiente tabla.
- a) Interprete el ajuste del modelo.
b) Con 95% de confianza, ¿Qué variables inciden en la probabilidad de considerarse independiente?

b) ¿Cómo es el efecto del número de partidos en la probabilidad de considerarse independiente? Explique por qué cree que existe tal efecto.

Independencia partidaria	Modelo PROBIT (Error estándar robusto)
Variables de desempeño y culturales	
Evaluación del desempeño (empeora)	0.105*** [0.014]
Número efectivo de partidos	0.231*** [0.073]
Escolaridad:	
Primaria	-0.126** [0.058]
Secundaria	-0.019 [0.068]
Preparatoria	0.074 [0.071]
Universidad	0.237** [0.082]
Índice de Desarrollo Humano	-0.054 [0.762]
Constante	-1.178* [0.659]
Observaciones	10932
R ² o Pseudo R ²	0.0285
Porcentaje de clasificación correcta	69.14%

Fuente: Calculado con datos de la base de datos de Parametría 2013. Error estándar entre corchetes *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

33. Un grupo de investigadores desarrolla un modelo probit para predecir la probabilidad de aprobar el gobierno de Piñera. La siguiente tabla muestra los efectos marginales de los predictores.

a) Con 95% de confianza, ¿Qué variables inciden en la probabilidad de aprobar al gobierno de Piñera?

b) ¿Cómo es el efecto promedio de identificarse políticamente con la izquierda en la aprobación al gobierno de Piñera? ¿Cómo es el efecto de considerar que el país está en decadencia en la aprobación del gobierno de Piñera?

Variable	Piñera
Sit. Econ. [2] Mala	0.03 (0.04)
Sit. Econ. [3] Ni buena ni mala	0.137*** (0.04)
Sit. Econ. [4] Buena	0.161*** (0.04)
Sit. Econ. [5] Muy Buena	0.168** (0.06)
Visión país. [2] Estancado	-0.348*** (0.01)
Visión país. [3] Decadencia	-0.380*** (0.02)
Sexo [Femenino]	-0.01 (0.02)
Edad	0.00 (0.00)
Identif. política [2] Centro Derecha	-0.168*** (0.02)
Identif. política [3] Centro	-0.351*** (0.02)
Identif. política [4] ¹ Centro Izquierda	-0.202*** (0.02)
Identif. política [5] Izquierda	-0.252*** (0.03)
Identif. política [6] Independiente	-0.219*** (0.04)
Identif. política [7] Ninguna	-0.186*** (0.03)
IGSE [2] C2	-0.076 (0.04)
IGSE [3] C3	-0.04 (0.04)
IGSE [4] D	-0.67 (0.04)
IGSE [5] E	-0.04 (0.05)
Años escolaridad [2] 4-8 años	-0.02 (0.02)
Años escolaridad [3] 9-12 años	-0.05 (0.04)
Años escolaridad [4] 13 y más años	-0.02 (0.04)

Significancia estadística: *** p<0.001 ** p<0.01 *p<0.05. Errores estándar entre paréntesis.