

Pauta de Corrección Tarea N°1

Parte 1: Test de Hipótesis

Se trata de evaluar los efectos del ruido sobre la capacidad de aprendizaje. Se distribuyeron aleatoriamente 2 grupos de $(220 - n)$ estudiantes cada uno. Al primer grupo se le enseñó una habilidad en condiciones de ruido. Al segundo grupo se le enseñó la misma habilidad, con el mismo profesor, pero sin ruido. Al final del experimento se administró a cada estudiante una prueba para medir su nivel de dominio en la habilidad:

Grupo	Tamaño Muestral	Puntaje Promedio
Con ruido	$220 - n$	80
Sin ruido	$220 - n$	89

¿Proporcionan estos datos evidencia suficiente como para concluir que el ruido es un factor que perjudica el aprendizaje? (Se supone Normalidad y se supone que las varianzas poblacionales son 60 y 45 respectivamente. Considere un nivel de significación del 5%).

El problema se plantea como un Test de Hipótesis sobre la igualdad de dos medias con varianzas conocidas.

1) *Se tiene dos poblaciones de interés: (1) estudiantes sin ruido y (2) estudiantes en condiciones de ruido. La primera con media nivel de dominio de la habilidad μ_1 y varianza conocida $\sigma_1^2 = 45$, y la segunda con media μ_2 y varianza conocida $\sigma_2^2 = 60$. Interesa saber si las dos medias son iguales. (10 puntos)*

2) *Se plantean las hipótesis:*

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 > \mu_2$$

Se escoge tal H_1 porque se busca evidencia fuerte que indique que el ruido perjudica el aprendizaje (nivel de dominio de la habilidad menor). (20 puntos)

3) *El estadístico de prueba es $Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}}$ y bajo H_0 , $\mu_1 - \mu_2 = 0$. (10 puntos)*

4) *Como $\sigma = 0,05$, entonces se rechazará H_0 si Z observado en las muestras es mayor a $Z_{0,05} = 1.645$ (se trata de un test unilateral). (10 puntos)*

5) *Sustituyendo los datos, se tiene $Z = \frac{(89 - 80)}{\sqrt{\left(\frac{45}{200 - n} + \frac{60}{200 - n}\right)}} = \frac{9}{\sqrt{\frac{105}{200 - n}}}$. (5 puntos)*

[Nótese que en los casos extremos de n (el máximo valor entre los números de lista de los integrantes del grupo) 3 y 73, se observa $Z = 12,328$ y $Z = 9,898$ respectivamente]

- 6) *Conclusión: Como $Z > 1.645$ se rechaza H_0 con un nivel de significancia $=0.05$, de donde se concluye que el ruido perjudica el aprendizaje.(5 puntos)*
[Nótese que para todos los valores de n , se rechazará H_0]

Parte 2: Estadística Descriptiva Univariada con SPSS

Considere el fichero SPSS “Datos_Tarea_1.sav”, con datos y variables seleccionadas generadas a partir de la Encuesta Casen 2009 en la Región del Bío Bío.

- a. Clasifique cada una de las variables de acuerdo a su escala (nominal, ordinal, etc.) y corrija la Medida en el fichero SPSS, si corresponde.

(10 puntos)

Nominales: Región, Zona, Sexo, Condición de actividad, Rama de Actividad Económica

Ordinales: Nivel Educacional

De Razón/Numéricas: Edad, Escolaridad, Ingreso ocupación principal.

- b. Realice un análisis unidimensional de cada variable indicando las medidas de dispersión y de tendencia central que correspondan. Grafique características de cada variable.

(20 puntos)

Graficar las variables nominales/ordinales usando gráficos de barra o circulares, histogramas para las de razón. Mostrar promedios y desv. típica / mediana / moda y frecuencia para variables de razón/ ordinales /nominales, respectivamente.

Es importante que los estadísticos se calculen con los datos ponderados por el facto de expansión. Así, por ejemplo las frecuencias de la variable Zona corresponden a:

		zona			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	urbano	1318959	83,1	83,1	83,1
	rural	267317	16,9	16,9	100,0
	Total	1586276	100,0	100,0	

[Hay que ponderar por el Factor de Expansión para los análisis de b. y e. Sin embargo, hay que desactivar la ponderación de los datos para las actividades del punto d.]

- c. A partir de lo anterior, indique claramente cuál es la población caracterizada en la base.

(10 puntos)

Se trata de los habitantes de la Región del Bío-Bío de 15 años o más, es decir la Fuerza de Trabajo de la región.

- d. Use la variable *Activ* para seleccionar la submuestra correspondiente a los individuos ocupados, aplicando la opción de eliminar los casos no seleccionados. Elimine de la base resultante los últimos n casos (ver observación al final). Verifique el resultado de las eliminaciones de casos mostrando las frecuencias la variable *Activ* y guarde la nueva base como *Ocupados_n.sav* (reemplazando n del nombre del archivo por el valor que corresponda).

(10 puntos)

*La base de datos *Ocupados_n* tienen tamaño distinto para cada grupo, con un número de casos entre 13272 y 13199 casos, inclusive.*

- e. Realice nuevamente el análisis unidimensional de b), sobre la submuestra. No es necesario graficar.
(10 puntos)
Mostrar promedios y desv. típica / mediana / moda y frecuencia para variables de razón/ ordinales /nominales, respectivamente.

Parte 3: Estadística Descriptiva Bivariada con SPSS

- a. Averigüe la definición de *Ingreso de la Ocupación Principal* que se utiliza en la Encuesta CASEN. ¿En qué difiere esta definición de las definiciones de *sueldo* y *salario*?

(10 puntos)

Ingreso Ocupación Principal

Ingreso que obtienen los ocupados en su ocupación principal, ya sea por concepto de trabajo dependiente en el caso de los asalariados (empleados y obreros del sector público, de empresas públicas y del sector privado, servicio doméstico puertas adentro y puertas afuera y miembros de las fuerzas armadas y de orden), o por concepto de trabajo independiente en el caso de los patrones o empleadores y trabajadores por cuenta propia.

(Ver <http://www.ministeriodesarrollosocial.gob.cl/casen/definiciones/ingresos.html#1>)

Tanto sueldo como salario refieren a la remuneración de trabajadores dependientes, es decir que tienen un patrón o empleador (difieren en periodicidad del pago y, en algunos casos, por tipo de trabajador), en cambio los ingresos incluyen la remuneración de trabajadores independientes (patrones o empleadores y trabajadores por cuenta propia).

- b. Realice un análisis descriptivo de las variables de *Ingreso de la Ocupación Principal* para las distintas categorías *Sexo, Nivel Educativo, Zona Urbano-Rural y Rama de Actividad Económica*.

(15 puntos)

Se pide estadísticas descriptivas de YOPRAJ para cada categoría de las variables nominales, como en el ejemplo siguiente:

		nivel educacional									Total
		sin educación formal	básica incompleta.	básica completa	media humanista incompleta	media técnico profesional incompleta	media humanista completa	media técnico completa	técnica o universitaria incompleta.	técnica o universitaria completa	
ingreso ocupación principal	Media	168,942	172,285	195,700	220,956	301,685	255,101	298,380	346,269	611,077	306,871
	Desviación típica	149,300	131,983	148,015	177,252	421,212	264,531	296,561	347,423	608,737	368,148
	Mínimo	10,270	4,108	4,108	6,162	1,105	4,108	4,752	16,432	10,270	1,105
	Máximo	1,848,600	4,108,000	3,697,200	4,108,000	3,867,500	10,270,000	4,108,000	3,081,000	11,050,000	11,050,000
N válido		9544	90189	81822	76290	11915	171628	61394	49228	120424	672434

Se pide que los cuadros sean analizados. Por, ejemplo, en el cuadro anterior se observa que el ingreso medio crece conforme más alto es el nivel educacional.

[Los resultados deben estar ponderados por el Factor de Expansión]]

- c. Mediante la opción “Agrupación Visual” recodifique cada variable de *Ingreso de la Ocupación Principal* en cinco categorías de intervalos generados automáticamente¹.

(10 puntos)

Se pide aplicar el módulo SPSS señalado en el pie de página. Como resultado se obtiene una nueva variable, para cual se debe presentar, al menos, sus frecuencias y asignarlo etiquetas a las categorías. Se consigne una variable similar a la del ejemplo siguiente:

¹ Usando SPSS: Transformar->Agrupación Visual

ingreso ocupación principal (agrupado)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Bajo	135101	19,2	20,1	20,1
	Medio-bajo	146732	20,8	21,8	41,9
	Medio	128702	18,3	19,1	61,1
	Medio-alto	127621	18,1	19,0	80,0
	Alto	134278	19,1	20,0	100,0
	Total	672434	95,5	100,0	
Perdidos	Sistema	31571	4,5		
Total		704005	100,0		

Se debe indicar cuál de las dos opciones factibles para 5 categorías del módulo “Agrupación visual” se ha utilizado. En el caso del ejemplo anterior de usó quintiles.

- d. Usando tablas de contingencia y justificando la utilización de la medida η (Eta), analice la asociación entre la variable de *Ingreso de la Ocupación Principal* (recodificada) como dependiente y las variables *Sexo*, *Nivel Educativo*, *Zona* y *Rama de Actividad Económica*².

(15 puntos)

Se pide cruzar la variable de ingreso agrupada con **cada una** de las variables nominales, como en el ejemplo siguiente:

Tabla de contingencia ingreso ocupación principal (agrupado) * zona

% del total		zona		Total
		urbano	rural	
ingreso ocupación principal (agrupado)	Bajo	15,1%	5,0%	20,1%
	Medio-bajo	17,6%	4,3%	21,8%
	Medio	16,8%	2,3%	19,1%
	Medio-alto	17,3%	1,6%	19,0%
	Alto	18,7%	1,2%	20,0%
Total		85,6%	14,4%	100,0%

Para el análisis del coeficiente Eta, hay que notar que es un coeficiente de correlación que no supone “linealidad” de la asociación entre las variables, y que su cuadrado puede interpretarse como la proporción de la variable dependiente que es explicada por la variable independiente. Como Eta mide asociación entre una variable de intervalo y una nominal, se justifica la utilización de Eta considerando la variable ingreso original como dependiente. Para ello, se puede calcular la tabla de contingencia de YOPRAJ con la nominal, con la opción de “suprimir tablas” para solamente obtener el coeficiente. Siguiendo el ejemplo:

Medidas direccionales

			Valor
Nominal por intervalo	Eta	ingreso ocupación principal dependiente	,101
		zona dependiente	,361

En el ejemplo, la asociación entre ingreso es baja, pues el 0,1% ($0,01 * 0,1$) de la varianza de la variable YOPRAJ se explica por la zona urbano-rural.

Parte 4: Estadística Multivariada con SPSS

- a. A partir de la variable o20, calcule la nueva variable $Exper = 2009 - o20$, y etiquétela como *Experiencia*; y también calcule las variables $Exper2 = Exper^2$ y $Lyopraj = \text{Log}(yopraj)$.

(10 puntos)

Comprobar la creación de variables en la base de datos *Ocupados_n.sav*

² Usando SPSS: Analizar->Estadísticos Descriptivos-> Tablas de Contingencia (ve Estadísticos)

- b. Estime la ecuación de Mincer $\text{Log}(\text{Ingreso}) = \beta_0 + \beta_1 \text{Educa} + \beta_2 \text{Exper} + \beta_3 \text{Exper}^2 + \varepsilon$ mediante una Regresión Lineal.
(20 puntos)

Se pide la presentación de una salida como la siguiente:

Variables introducidas/eliminadas^b

Modelo	Variables introducidas	Variables eliminadas	Método
1	exper2, escolaridad (años de educación), exper	.	Introducir

- a. Todas las variables solicitadas introducidas.
b. Variable dependiente: lyopraj

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,519 ^a	,269	,269	,68079

- a. Variables predictoras: (Constante), exper2, escolaridad (años de educación), exper

ANOVA^b

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	113460,312	3	37820,104	81600,719	,000 ^a
	Residual	308045,276	664639	,463		
	Total	421505,588	664642			

- a. Variables predictoras: (Constante), exper2, escolaridad (años de educación), exper
b. Variable dependiente: lyopraj

Coefficientes^a

Modelo		Coefficients no estandarizados		Coefficients tipificados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	11,122493	,003		4320,413	,000
	escolaridad (años de educación)	,088801	,000	,443	420,567	,000
	exper	,041116	,000	,491	180,659	,000
	exper2	-,000692	,000	-,282	-103,709	,000

- a. Variable dependiente: lyopraj

Se requiere analizar estos resultados, destacando por ejemplo que los coeficientes de la regresión son significativos; positivos en el caso de escolaridad y experiencia, es decir que el ingreso (su logaritmo) crece al crecer ambas variables, y negativo en el caso de la experiencia al cuadrado. (Como antes, los datos deben estar ponderados).

- c. De una interpretación para los coeficientes estimados β_1 y β_2 .
(30 puntos)

El coeficiente β_1 se interpreta como tasa promedio de rendimiento de un año adicional de escolaridad, y β_2 una interpretación similar pero para cada año adicional de experiencia en el trabajo.

El modelo de Mincer postula que β_1 y β_2 son positivos. Sin embargo, el efecto de la experiencia no es lineal sino cóncavo (cada año de experiencia tiene un efecto sobre los ingresos, pero menor que el anterior), por lo cual se espera que β_3 sea significativo y negativo.

[No es necesario indicarlo, pero como modelo teórico, el de Mincer tiene algunos supuestos: (1) los costos de inversión en educación son solamente de oportunidad (los ingresos que dejan de percibirse); (ii) el tiempo que un individuo permanece en el mercado laboral es independiente del nivel de escolaridad alcanzado, y su permanencia es continua; (iii) las personas comienzan a trabajar inmediatamente después de finalizar sus estudios.]

Nota T1 = 1 + Promedio Puntajes (Parte1, Parte 2, Parte 2, Parte 4) / 10.