

Segundo Trabajo Estadística II

Pruebas de Hipótesis

Integrantes:

1. Juan David Mena Gamboa
2. Luis Angel Bolaño Lopez
3. Heyner

24 de septiembre de 2025

Ejercicio 1: Prueba de hipótesis para la media de una muestra con varianza conocida

- *Planteamiento del problema:* La Gran Encuesta Integrada de Hogares (GEIH) 2024 del DANE recopila información sobre las actividades económicas de los hogares colombianos, incluyendo el ahorro generado por actividades agropecuarias. La variable de ahorro por criar animales (P3095S3) mide los ingresos netos obtenidos por esta actividad. Considerando que esta actividad representa una fuente importante de ingresos para muchas familias rurales y periurbanas, se cree que el ahorro promedio por criar animales en la población es igual a 100,000 pesos.

- Hipótesis: $H_0 : \mu = 100,000$ $H_1 : \mu \neq 100,000$

- Nivel de significancia: $\alpha = 0,030$

- Estadístico de Prueba: Z

Estadístico Empleado	Z para media con varianza conocida
Fórmula del estadístico	$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$
Nombre Parámetro	Valor
Media poblacional bajo H_0 (μ_0)	100,000
Desviación estándar poblacional (σ)	500,000
Tamaño de muestra (n)	4,080
Media muestral (\bar{X})	159,471.37
Valor Estadístico	7.598
Valor P	$3,02 \times 10^{-14}$
Valor Crítico	$\pm 2,170$

- Regla de decisión: Rechazar H_0 si $|Z| > 2,170$ o si p-valor $< 0,030$
- Decisión: Rechazar H_0
- Conclusión: Con un nivel de significancia de 0.030, la evidencia sugiere que el ahorro promedio por criar animales en la población no es igual a 100,000 pesos.

Ejercicio 2: Prueba de hipótesis para la proporción de una muestra

- *Planteamiento del problema:* La GEIH 2024 incluye información sobre la participación social y familiar de los individuos. La variable de asistencia a reuniones familiares (P3087S1) captura la participación en eventos familiares, lo cual es un indicador importante del capital social y la cohesión familiar en Colombia. Dado que las reuniones familiares son tradicionalmente valoradas en la cultura colombiana, se cree que la proporción de personas que asistieron a reuniones familiares es igual al 50 %.
- Hipótesis: $H_0 : p = 0,5$ $H_1 : p \neq 0,5$
- Nivel de significancia: $\alpha = 0,070$
- Estadístico de Prueba: Z

Estadístico Empleado	Z para proporción
Fórmula del estadístico	$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$
Nombre Parámetro	Valor
Proporción bajo H_0 (p_0)	0.5
Tamaño de muestra (n)	166,341
Proporción muestral (\hat{p})	0.0175
Error estándar (SE)	0.0012
Valor Estadístico	-393.54
Valor P	< 0,001
Valor Crítico	$\pm 1,812$

- Regla de decisión: Rechazar H_0 si $|Z| > 1,812$ o si p-valor < 0,070
- Decisión: Rechazar H_0
- Conclusión: Con un nivel de significancia de 0.070, la evidencia sugiere que la proporción de personas que asistieron a reuniones familiares no es igual al 50 %.

Ejercicio 3: Prueba de hipótesis para la media de una muestra con varianza desconocida

- *Planteamiento del problema:* La GEIH 2024 recopila información detallada sobre el mercado laboral colombiano, incluyendo las prácticas y pasantías como forma de vinculación laboral. La variable P3094S3 registra el valor mensual de estas actividades, las cuales son fundamentales para la transición de los jóvenes al mercado laboral formal. Considerando que las prácticas representan una oportunidad de formación y experiencia laboral, se cree que el valor promedio mensual por prácticas o pasantías es igual a 200,000 pesos.
- Hipótesis: $H_0 : \mu = 200,000$ $H_1 : \mu \neq 200,000$
- Nivel de significancia: $\alpha = 0,025$

- Estadístico de Prueba: t

Estadístico Empleado	t para media con varianza desconocida
Fórmula del estadístico	$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$
Nombre Parámetro	Valor
Media poblacional bajo H_0 (μ_0)	200,000
Tamaño de muestra (n)	336
Grados de libertad (df)	335
Media muestral (\bar{X})	998,708.33
Desviación estándar muestral (s)	621,977.40
Valor Estadístico	23.539
Valor P	$< 0,001$
Valor Crítico	$\pm 2,252$

- Regla de decisión: Rechazar H_0 si $|t| > 2,252$ o si p-valor $< 0,025$
- Decisión: Rechazar H_0
- Conclusión: Con un nivel de significancia de 0.025, la evidencia sugiere que el valor promedio mensual por prácticas no es igual a 200,000 pesos.

Ejercicio 4: Prueba de hipótesis para la diferencia de medias de dos muestras con varianzas conocidas

- *Planteamiento del problema:* La GEIH 2024 documenta las actividades agropecuarias de los hogares colombianos, incluyendo tanto el cultivo (P3095S1) como la cría de animales (P3095S3). Estas actividades representan estrategias económicas diferentes: el cultivo requiere inversión en semillas y fertilizantes, mientras que la cría de animales implica costos de alimentación y cuidado. Ambas actividades son fundamentales para la seguridad alimentaria y los ingresos rurales. Se cree que no existe diferencia entre el ahorro promedio por cultivar y por criar animales.
- Hipótesis: $H_0 : \mu_1 - \mu_2 = 0$ $H_1 : \mu_1 - \mu_2 \neq 0$
- Nivel de significancia: $\alpha = 0,045$
- Estadístico de Prueba: Z

Estadístico Empleado	Z para diferencia de medias con varianzas conocidas
Fórmula del estadístico	$Z = \frac{(X_1 - X_2) - 0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$
Nombre Parámetro	Valor
Media muestra 1 (cultivar)	126,937.56
Media muestra 2 (criar)	159,471.37
Desviación estándar 1 (σ_1)	300,000
Desviación estándar 2 (σ_2)	500,000
Tamaño muestra 1 (n_1)	2,892
Tamaño muestra 2 (n_2)	4,080
Diferencia de medias	-32,533.81
Valor Estadístico	-3.385
Valor P	0.0007
Valor Crítico	$\pm 2,005$

- Regla de decisión: Rechazar H_0 si $|Z| > 2,005$ o si p-valor $< 0,045$
- Decisión: Rechazar H_0
- Conclusión: Con un nivel de significancia de 0.045, la evidencia sugiere que hay diferencia entre el ahorro promedio por cultivar y por criar animales.

Ejercicio 5: Prueba de hipótesis para la diferencia de medias de dos muestras con varianzas iguales y desconocidas

- *Planteamiento del problema:* La GEIH 2024 permite comparar diferentes fuentes de ingresos: las prácticas laborales (P3094S3) representan ingresos del sector formal/educativo, mientras que el cultivo (P3095S1) representa ingresos del sector agropecuario. Esta comparación es relevante para entender las diferencias entre ingresos urbanos/educativos versus rurales/agropecuarios en Colombia. Se cree que no existe diferencia entre el valor promedio de prácticas y el ahorro promedio por cultivar, asumiendo varianzas iguales.
- Hipótesis: $H_0 : \mu_1 - \mu_2 = 0$ $H_1 : \mu_1 - \mu_2 \neq 0$
- Nivel de significancia: $\alpha = 0,001$
- Estadístico de Prueba: t (pooled)

Estadístico Empleado	t para diferencia de medias con varianzas iguales
Fórmula del estadístico	$t = \frac{(X_1 - X_2) - 0}{s_p \sqrt{1/n_1 + 1/n_2}}$
Nombre Parámetro	Valor
Media muestra 1 (prácticas)	998,708.33
Media muestra 2 (cultivar)	126,937.56
Tamaño muestra 1 (n_1)	336
Tamaño muestra 2 (n_2)	2,892
Grados de libertad (df)	3,226
Varianza combinada (s_p^2)	190,877,666,859
Diferencia de medias	871,770.77
Valor Estadístico	34.620
Valor P	< 0,001
Valor Crítico	$\pm 3,294$

- Regla de decisión: Rechazar H_0 si $|t| > 3,294$ o si p-valor < 0,001
- Decisión: Rechazar H_0
- Conclusión: Con un nivel de significancia de 0.001, la evidencia sugiere que hay diferencia entre el valor promedio de prácticas y el ahorro promedio por cultivar.

Ejercicio 6: Prueba de hipótesis para la diferencia de medias de dos muestras con varianzas desiguales y desconocidas

- *Planteamiento del problema:* La GEIH 2024 permite analizar la variabilidad de ingresos entre diferentes actividades económicas. Las prácticas laborales (P3094S3) y la cría de animales (P3095S3) representan sectores económicos distintos con diferentes niveles de riesgo e incertidumbre. Las prácticas están más reguladas y estandarizadas, mientras que la cría de animales depende de factores climáticos y de mercado más variables. Se cree que no existe diferencia entre el valor promedio de prácticas y el ahorro promedio por criar animales, sin asumir varianzas iguales.
- Hipótesis: $H_0 : \mu_1 - \mu_2 = 0$ $H_1 : \mu_1 - \mu_2 \neq 0$
- Nivel de significancia: $\alpha = 0,030$
- Estadístico de Prueba: t de Welch

Estadístico Empleado	t de Welch para diferencia de medias
Fórmula del estadístico	$t = \frac{(X_1 - X_2) - 0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$
Nombre Parámetro	Valor
Media muestra 1 (prácticas)	998,708.33
Media muestra 2 (criar)	159,471.37
Tamaño muestra 1 (n_1)	336
Tamaño muestra 2 (n_2)	4,080
Grados de libertad (df)	373.42
Diferencia de medias	839,236.96
Error estándar	34,867.44
Valor Estadístico	24.069
Valor P	< 0,001
Valor Crítico	$\pm 2,178$

- Regla de decisión: Rechazar H_0 si $|t| > 2,178$ o si p-valor < 0,030
- Decisión: Rechazar H_0
- Conclusión: Con un nivel de significancia de 0.030, la evidencia sugiere que hay diferencia entre el valor promedio de prácticas y el ahorro promedio por criar animales.

Ejercicio 7: Prueba de hipótesis para la diferencia de proporciones de dos muestras

- *Planteamiento del problema:* La GEIH 2024 permite analizar la relación entre el nivel socioeconómico y la participación social familiar. Se han creado grupos según el valor de prácticas (P3094S3): bajo y alto valor, para examinar si el nivel de ingresos por prácticas se asocia con la participación en reuniones familiares (P3087S1). Esta relación es importante para entender cómo los ingresos laborales pueden influir en el capital social familiar. Se cree que no existe diferencia en la proporción de asistencia a reuniones familiares entre grupos de bajo y alto valor de prácticas.
- Hipótesis: $H_0 : p_1 - p_2 = 0$ $H_1 : p_1 - p_2 \neq 0$
- Nivel de significancia: $\alpha = 0,070$
- Estadístico de Prueba: Z

Estadístico Empleado	Z para diferencia de proporciones
Fórmula del estadístico	$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}$
Nombre Parámetro	Valor
Proporción grupo bajo (\hat{p}_1)	0.0138
Proporción grupo alto (\hat{p}_2)	0.0085
Tamaño grupo bajo (n_1)	145
Tamaño grupo alto (n_2)	118
Proporción combinada (\hat{p})	0.0119
Diferencia de proporciones	0.0053
Valor Estadístico	0.427
Valor P	0.6697
Valor Crítico	$\pm 1,812$

- Regla de decisión: Rechazar H_0 si $|Z| > 1,812$ o si p-valor $< 0,070$
- Decisión: No rechazar H_0
- Conclusión: Con un nivel de significancia de 0.070, la evidencia sugiere que no hay diferencia en la proporción de asistencia a reuniones familiares entre grupos de bajo y alto valor de prácticas.

Ejercicio 8: Prueba de hipótesis para muestras dependientes

- *Planteamiento del problema:* La GEIH 2024 permite analizar hogares que realizan múltiples actividades agropecuarias simultáneamente. Muchos hogares rurales colombianos combinan el cultivo (P3095S1) con la cría de animales (P3095S3) como estrategia de diversificación de ingresos y reducción de riesgos. Esta diversificación es común en la agricultura familiar colombiana. Se cree que no existe diferencia sistemática entre el ahorro por cultivar y por criar animales en las mismas personas.
- Hipótesis: $H_0 : \mu_d = 0$ $H_1 : \mu_d \neq 0$
- Nivel de significancia: $\alpha = 0,025$
- Estadístico de Prueba: t pareado

Estadístico Empleado	t para muestras pareadas
Fórmula del estadístico	$t = \frac{d - 0}{s_d / \sqrt{n}}$
Nombre Parámetro	Valor
Número de pares (n)	1,296
Grados de libertad (df)	1,295
Media de diferencias (d)	9,776.68
Desviación estándar de diferencias (s_d)	575,782.40
Error estándar de diferencias	15,993.96
Valor Estadístico	0.611
Valor P	0.5411
Valor Crítico	$\pm 2,244$

- Regla de decisión: Rechazar H_0 si $|t| > 2,244$ o si p-valor $< 0,025$
- Decisión: No rechazar H_0
- Conclusión: Con un nivel de significancia de 0.025, la evidencia sugiere que no hay diferencia entre el ahorro por cultivar y por criar animales en las mismas personas.

Ejercicio 9: Prueba de bondad de ajuste

- *Planteamiento del problema:* La GEIH 2024 recopila información sobre la participación social familiar a través de la variable P3087S1 (asistencia a reuniones familiares). En Colombia, las reuniones familiares son eventos culturalmente importantes que reflejan la cohesión social. Una distribución uniforme (50 %-50 %) representaría una participación equilibrada, pero factores como la urbanización, migración y cambios en la estructura familiar pueden influir en esta participación. Se cree que la distribución observada de asistencia a reuniones familiares se ajusta a una distribución uniforme (50 %-50 %).
- Hipótesis: H_0 : La distribución observada se ajusta a la esperada H_1 : La distribución observada no se ajusta a la esperada
- Nivel de significancia: $\alpha = 0,045$
- Estadístico de Prueba: χ^2

Estadístico Empleado	Chi-cuadrado de bondad de ajuste
Fórmula del estadístico	$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$
Nombre Parámetro	Valor
Número de categorías (k)	2
Grados de libertad (df)	1
Frecuencia observada "No" (O_1)	163,423
Frecuencia observada "Sí" (O_2)	2,918
Frecuencia esperada "No" (E_1)	83,170.5
Frecuencia esperada "Sí" (E_2)	83,170.5
Valor Estadístico	154,873.75
Valor P	$< 0,001$
Valor Crítico	4.019

- Regla de decisión: Rechazar H_0 si $\chi^2 > 4,019$ o si p-valor $< 0,045$
- Decisión: Rechazar H_0
- Conclusión: Con un nivel de significancia de 0.045, la evidencia sugiere que la distribución observada de asistencia a reuniones familiares no se ajusta a una distribución uniforme.

Ejercicio 10: Prueba de independencia

- *Planteamiento del problema:* La GEIH 2024 permite examinar la relación entre variables socioeconómicas y participación social. Se ha categorizado el valor de prácticas (P3094S3) en terciles (bajo, medio, alto) para analizar su asociación con la asistencia a reuniones familiares (P3087S1). Esta relación es relevante para entender si los ingresos laborales se asocian con la participación en actividades familiares, lo cual puede reflejar diferencias en capital social según el nivel socioeconómico. Se cree que no existe asociación entre la asistencia a reuniones familiares y el nivel de valor de prácticas (agrupado en terciles).
- Hipótesis: H_0 : Las variables son independientes H_1 : Las variables están asociadas
- Nivel de significancia: $\alpha = 0,001$
- Estadístico de Prueba: χ^2

Estadístico Empleado	Chi-cuadrado de independencia
Fórmula del estadístico	$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
Nombre Parámetro	Valor
Número de filas (r)	3
Número de columnas (c)	2
Grados de libertad (df)	2
Frecuencias observadas	Bajo: 125,0; 2,1
	Medio: 101,0; 1,1
	Alto: 106,0; 1,1
Frecuencias esperadas	Bajo: 125.5, 1.5
	Medio: 100.8, 1.2
	Alto: 105.7, 1.3
Valor Estadístico	0.257
Valor P	0.8793
Valor Crítico	13.816

- Regla de decisión: Rechazar H_0 si $\chi^2 > 13,816$ o si p-valor $< 0,001$
- Decisión: No rechazar H_0
- Conclusión: Con un nivel de significancia de 0.001, la evidencia sugiere que no hay asociación entre la asistencia a reuniones familiares y el nivel de valor de prácticas.

Ejercicio 11: Prueba de signos

- *Planteamiento del problema:* La GEIH 2024 permite analizar la consistencia en el rendimiento de diferentes actividades agropecuarias a nivel de hogar. Se han creado rankings para el ahorro por cultivar (P3095S1) y por criar animales (P3095S3) para examinar si los hogares que tienen mejor rendimiento en una actividad también lo tienen en la otra. Esta relación es importante para entender si existe complementariedad o competencia entre estas actividades en los hogares rurales colombianos. Se

crea que no existe tendencia sistemática en las diferencias entre rankings de ahorro por cultivar y por criar animales.

- Hipótesis: H_0 : No hay diferencia sistemática de signos H_1 : Existe diferencia sistemática de signos
- Nivel de significancia: $\alpha = 0,030$
- Estadístico de Prueba: Prueba de signos

Estadístico Empleado	Prueba de signos para muestras pareadas
Fórmula del estadístico	Estadístico = min(pos, neg)
Nombre Parámetro	Valor
Número total de pares (n)	1,295
Diferencias positivas	679
Diferencias negativas	616
Empates	1
Estadístico (min)	616
Valor P	0.0849

- Regla de decisión: Rechazar H_0 si p-valor $< 0,030$
- Decisión: No rechazar H_0
- Conclusión: Con un nivel de significancia de 0.030, la evidencia sugiere que no hay tendencia en las diferencias entre rankings de ahorro por cultivar y por criar animales.