

## ERRORES MÁS COMUNES EN EL MANEJO DE CONCEPTOS ESTADÍSTICOS EN LOS PROCESOS DE EXPERIMENTACIÓN AGROPECUARIOS.

Dr. José R. Binns, PhD

Ingeniero Agrónomo, PhD. Ciencias Agrícolas con énfasis en Estadística. Departamento de Desarrollo Agropecuario. Facultad de Ciencias Agropecuarias, Universidad de Panamá.

El interés de esta conferencia es concienciar a los participantes acerca de errores comunes que se cometen a diario en el manejo de conceptos y/o procedimientos estadísticos en los procesos de experimentación agropecuaria. De igual forma dar a conocer conceptos y técnicas no convencionales, pero de amplia aplicación en la Estadística Experimental Agrícola. La conferencia es nutrida por mi experiencia de más de 31 años al frente de la coordinación de Biometría y Métodos Cuantitativos de la Facultad de Ciencias Agropecuarias de la Universidad de Panamá.

### 1. Hablar de Regresión antes de Correlación

Es un tremendo error. No tiene sentido tratar de encontrar una ecuación de regresión que establezca una relación funcional entre "y" y "x" si no hemos comprobado primero si la variable están relacionadas.

### 2. Pensar que co-varianza es algo diferente al análisis de varianza

En realidad, son conceptos bastante parecidos. El análisis de co-varianza es un análisis auxiliar al de varianza. Pretende ajustar las medidas de los tratamientos debido al posible efecto de una "covariable" o "x", asumiendo que existe una relación lineal entre "x" y nuestra variable de respuesta. De hecho, si se detecta en la **ANOVA** que "x" es significativa, se procede a ajustar las medias por mínimos cuadrados.

### 3. Diseño de Parcelas Divididas considerado como un "Diseño Experimental"

Los diseños experimentales son DCA, DBCA, DCL Y DGL. (Los diseños clásicos).

El DPD es un "Diseño de Tratamiento", es decir, una forma de estructurar las combinaciones factoriales generadas; donde un factor va en parcelas "grandes" o "principales" y el otro en parcelas "pequeñas" o "sub-parcelas".

### 4. Siempre es necesario una transformación Angular o Arcoseno si los datos son porcentuales. Falso.

La experiencia ha demostrado que si los porcentajes fluctúan entre 0 y 20% o entre 80% y 100% funciona mejor una transformación cuadrada.

### 5. Si no existe correlación entre 2 variables, se concluye que no están relacionadas. Falso.

Se descarta solo relación de tipo lineal. Podría haber una relación no-lineal. Ejemplo: recíproca, exponencial o asintótica.

6. Tengo  $n = 3$  observaciones y deseo obtener una ecuación múltiple con 2 variables independientes. Imposible. " $n$ " debe ser una unidad mayor al # de parámetros estimados. Ejemplo: si son 2 variables independientes, los parámetros son  $b_0, b_1, b_2$  por lo tanto se necesitan  $(p + 1)$  observaciones =  $3+1=4$

7. Utilizar métodos no paramétricos sin verificar si los datos son o no normales.

**Incorrecto:** estos métodos solo son aplicables cuando no existe normalidad.

8. Transformar datos con la Raíz Cuadrada sin el uso de una constante(c).

Incorrecto: Sobre todo cuando hay muchos ceros (0) en la base de datos debe utilizarse constantes  $(c) = 0.5$  o  $1.0$

9. Pensar que el Diseño Cuadrado Latino solo se puede utilizar en situaciones de campo.

Falso: Se puede adaptar a cualquier situación donde se tengan 2 criterios de bloque o diferentes.

10. Pensar que los modelos de regresión Polinomiales:  $y = b_0 + b_1x + b_2x^2 + b_3x^3$  son no-lineales.

Falso: Son modelos lineales ya que los coeficientes de regresión ( $b_0, b_1$  etc)son funciones lineales de las observaciones.

11. Pensar que los modelos de Regresión Lineal Múltiple no pueden incluir variables cualitativas.

Falso: Si pueden. En este caso se denominan variables "indicadoras" o "dummy" variables. Se selecciona un símbolo: (H) y se les da valores de 0, 1,...2, etc. Ejemplo: Experimento en Divala  $Z=0$ , en Sortova  $Z=1$  en Alanje  $Z=2$  .....etc.

12. Incluir variables independientes en un modelo de regresión múltiple cuando las mismas están correlacionadas entre sí.

Incorrecto: Al hacer esto se incurre en un problema denominado "Multicolinealidad".

13. Tratar de encontrar simetría en intervalos de confiabilidad para la varianza poblacional y a través de la distribución "t".

Incorrecto: estos intervalos son asimétricos con respecto al estimador puntual y se trabajan con la distribución Chi-Cuadrado.

14. Concluir luego de una prueba de Chi-Cuadrado, que existen o no diferencias significativas entre tratamientos.

Incorrecto: Chi-Cuadrado es un procedimiento no-paramétrico y no es un Diseño Experimental, por tanto, no acepta ni incluye comparaciones entre tratamientos.

15. Asumir que una muestra que representa el 20% del total de la población es óptima.

Incorrecto: Existen métodos para determinar el tamaño óptimo de una muestra. Sobresalen el Método de Sampien y el Método de Altamirano.

16. Asumir que, si no se efectúa observaciones, el valor es "0".

Incorrecto: si no se efectuó observación, no se puede asignar un "0". No existe forma de saber si realmente fue cero (0).

17. Solo es posible analizar una variable de respuesta a la vez en un Diseño.

Falso: Utilizando análisis multivariado se puede hacer un MANOVA (Análisis de Varianza Multivariado) en el cual se analizan varias variables de respuesta simultáneamente.

18. Los gráficos de interacción se deben elaborar después del análisis de varianza de un experimento factorial.

Falso: Deben elaborarse antes de la ANOVA para detectar la posible presencia de una interacción. Posteriormente la ANOVA confirma si dicha interacción es lo suficientemente fuerte para ser considerada significativa.

19. En la ANOVA de un experimento factorial se deben analizar los efectos principales antes que las interacciones.

Falso: las interacciones se deben analizar primero. Si son significativas no tiene mucho sentido interpretar efectos principales.

20. En un análisis de regresión el ajuste del modelo solo se puede medir a través del  $r^2$ .

Falso: Un indicador aún más importante y confiable es el  $r^2$  ajustado.

21. Si un análisis por SAS genera SCI y SC III, cualquiera se puede utilizar.

Incorrecto: Si el experimento es balanceado se puede usar cualquiera. Si existe desbalance se debe utilizar la ANOVA con SC III.

22. Solo existen modelos de efectos fijos, es decir el investigador escoge los tratamientos por anticipado y deliberadamente.

Falso: Existen los modelos aleatorios y los modelos mixtos (que combinan efectos fijos con aleatorios).

23. Cuando se transforman datos y se efectúan pruebas de Tukey o Duncan se utilizan las medias transformadas.

Falso: hay que hacer una retro-transformación de las medias (transformation back) a fin de retornar las medias a su escala original.

24. Si se pierden observaciones en un diseño, ya no se puede hacer nada.

Falso: Se pueden utilizar procedimientos matemáticos para estimar los datos perdidos.

25. En la ANOVA, en los diseños experimentales el Modelo Lineal Aditivo es un modelo matemático.

Falso: Es un modelo estadístico porque incluye o incorpora un término de error.

26. Los modelos, de regresión lineal simple donde la variable independiente es el tiempo, deben manejarse como cualquier otro.

Falso: Este tipo de modelo da origen a una técnica conocida como: Análisis de Serie de Tiempo y deben manejarse de manera especial.

27. Los modelos de regresión múltiple son estáticos, si se empieza con 5 variables independientes no se puede añadir ni quitar.

Falso: Con el procedimiento “Eliminación Retroactiva” se puede eliminar y con el procedimiento “Selección Progresiva” se puede agregar.

28. Solo es posible incluir una covariable en un Diseño Experimental.

Falso: Pueden incluirse varias covariables y probar la significancia de cada una.

29. La única media conocida es la media aritmética ( $\bar{x}$ ).

Falso: Existen medias geométricas, armónicas, cuadráticas, ponderadas, y medias para datos agrupados.

30. Los tamaños de parcelas experimentales son invariables y fijos.

Falso: Existen métodos para optimizar el tamaño de la unidad experimental.

31. El “efecto de borde” es un concepto que se aplica solo a experimentos de cielo abierto.

Falso: También se aplica a experimento de invernadero.

32. Cambiar las unidades de medida altera los resultados de la ANOVA.

Falso: La significancia de los efectos son independientes de las unidades de medidas.

33. Se puede efectuar un ANOVA sin verificar supuestos antes.

Falso: Se deben verificar los supuestos antes de realizar la ANOVA.

34. No existe relación alguna entre la Estadística y la Trigonometría.

Falso: De hecho, la transformación angular se fundamenta en una función trigonométrica inversa.

35. No son importantes los Cuadrados Medios Esperado, (CME) en la ANOVA.

Falso: Son esenciales a la hora de estimar componentes de varianza en estudios de Genética Cuantitativa.

36. Los Diseños Anidados (Nested-designs) son complicados e inoperantes.

Falso: Son sencillos y aplicables a escenarios experimentales muy particulares.

37. Solo es posible utilizar Diseños Experimentales Paramétricos.

Falso: Ante la falta de Normalidad existen versiones no-paramétricas de los Diseños como la prueba de Kruskal-Wallis y la prueba de Friedman.

38. La única opción para correlacionar variables es el coeficiente de Person.

Falso: existe el coeficiente de correlación no-paramétrico de Spearman basado en asignación de “rangos” a los valores de las variables.

39. Si el número de tratamientos factorial es muy alto, deben necesariamente utilizarse todas las combinaciones factoriales en la ANOVA.

Falso: Se pueden utilizar los llamados “factoriales fraccionados” y seleccionar un número reducido de tratamientos para facilitar el análisis sin sacrificar información estadística.