**PRUEBA ESTANDARIZADA DE BIOESTADISTICA GENÒMICA Y MOLECULAR**

**Nombres y Apellidos: …………...............................................................**

**Expositor:** Student PhD & MSc. Ronald Eleazar Huarachi Olivera

**Organizador y Coordinador:** Dr. Antonio Mateo Lazarte Rivera

**LABORATORIO DE BIOTECNOLOGÍA CELULAR Y MOLECULAR AVANZADA (LAB-BIOTCEMA); ESCUELA PROFESIONAL DE BIOLOGÍA; UNIVERSIDAD NACIONAL DE SAN AGUSTÍN; AREQUIPA-PERÚ**

**INDICACIONES:**

* El estudiante deberá responder eficientemente el control, el cuál se asignará una nota de 0 a 20, donde la mínima nota aprobatoria corresponde a 10.5 y la máxima nota aprobatoria corresponde a 20.
* La nota aprobatoria de 10,5 a 20.
* Durante la búsqueda de información en inglés, se recomienda de que el participante interprete correctamente sus respuestes, no solamente traducir de inglés al español, sino que el participante deberá interpretar su respuesta de manera correcta.
* El estudiante deberá enviar el control (en Word) al correo electrónico: [workshop.biol.molecular@gmail.com](mailto:workshop.biol.molecular@gmail.com) con el siguiente título o código **“I\_Work\_Bioinf\_&\_Bioest\_NOMBRE\_APELLIDO”** a más tardar el sábado 8 de Febrero hasta las 12:00 hrs.

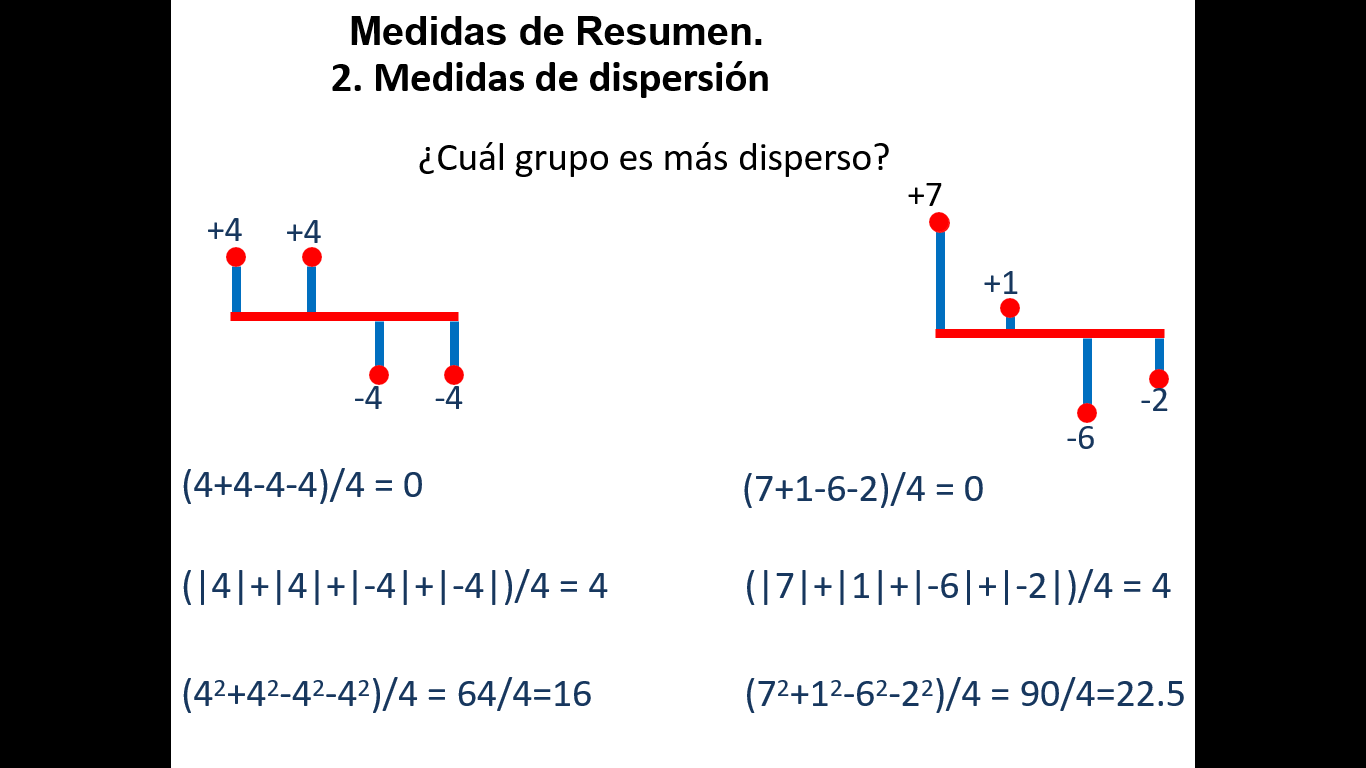
RESPONDA CLARAMENTE LAS SIGUIENTES PREGUNTAS CITANDO REFERENCIAS BIBLIOGRAFICAS (puede basarse en las diapositivas del curso). OJO: NO SERÁN EVALUADAS LAS PRUEBAS CLONADAS. RESUELVA PERSONALMENTE. SI DESEA TAMBIEN PUEDE ENVIAR CAPTURA DE PANTALLA, CON LA FINALIDAD DE SABER QUE UD. RESOLVIO LA PRUEBA ESTANDARIZADA.

**1.- ¿Mencione las diferencias más relevantes entre la estadística clásica y Bayesiana?(1 punto)**

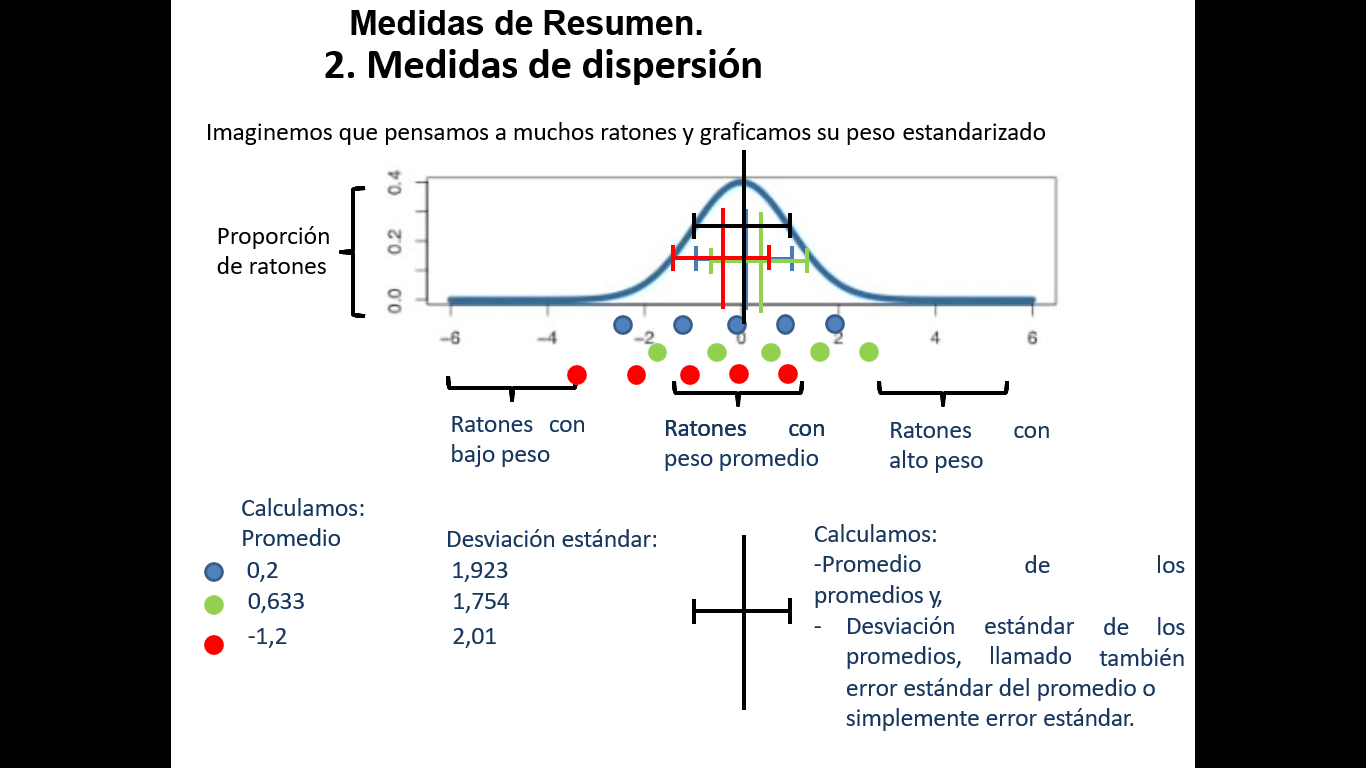
**2.- ¿Por qué decimos que la estadística es una herramienta esencial para la ciencia?. Explique detalladamente. Si desea hacer un esquema propio. Hágalo (1 punto).**

**3.- Con respecto a DATOS e INFORMACIÒN. ¿Cuál es la diferencia entre ambos conceptos desde el punto de vista estadístico y como llegamos a conceptuar las variables?. Mencione los tipos de variables (1 punto)**

**4.- ¿Cuáles son las medidas de posición y dispersión? Y en el siguiente ejemplo cual es el mejor estadígrafo para medir la dispersión y cuál podría ser el límite de ese estadígrafo (1 punto)**



**5.- ¿Cual es la diferencia entre la desviación estándar y error estándar? (1 punto)**



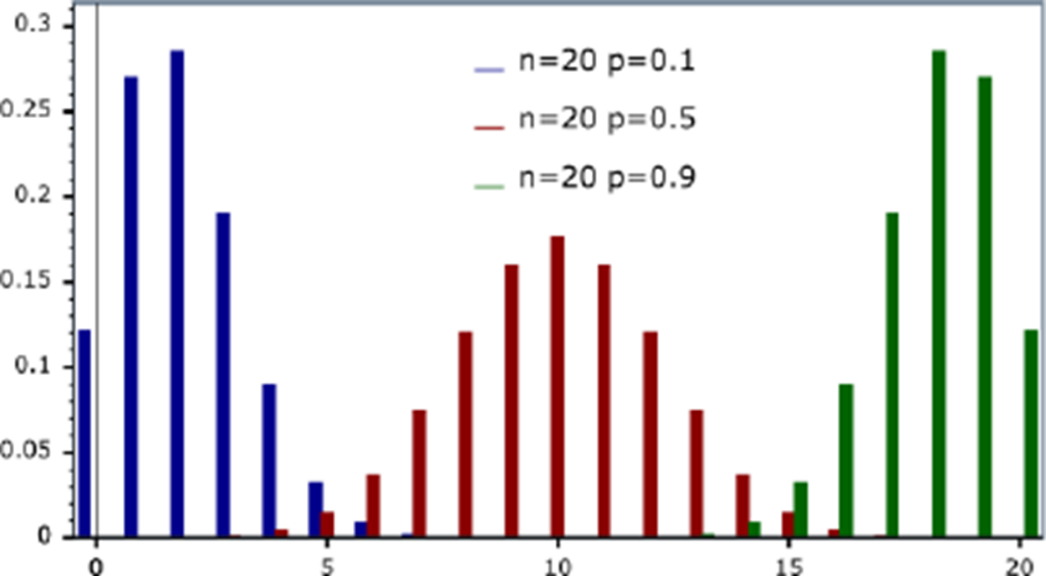
**6.- A que se denomina estimador de parámetros? (1 punto)**

**7.- ¿Explique detalladamente que es RNAseq?. Cree su propio esquema. Use un buen soporte bibliográfico(1 punto)**

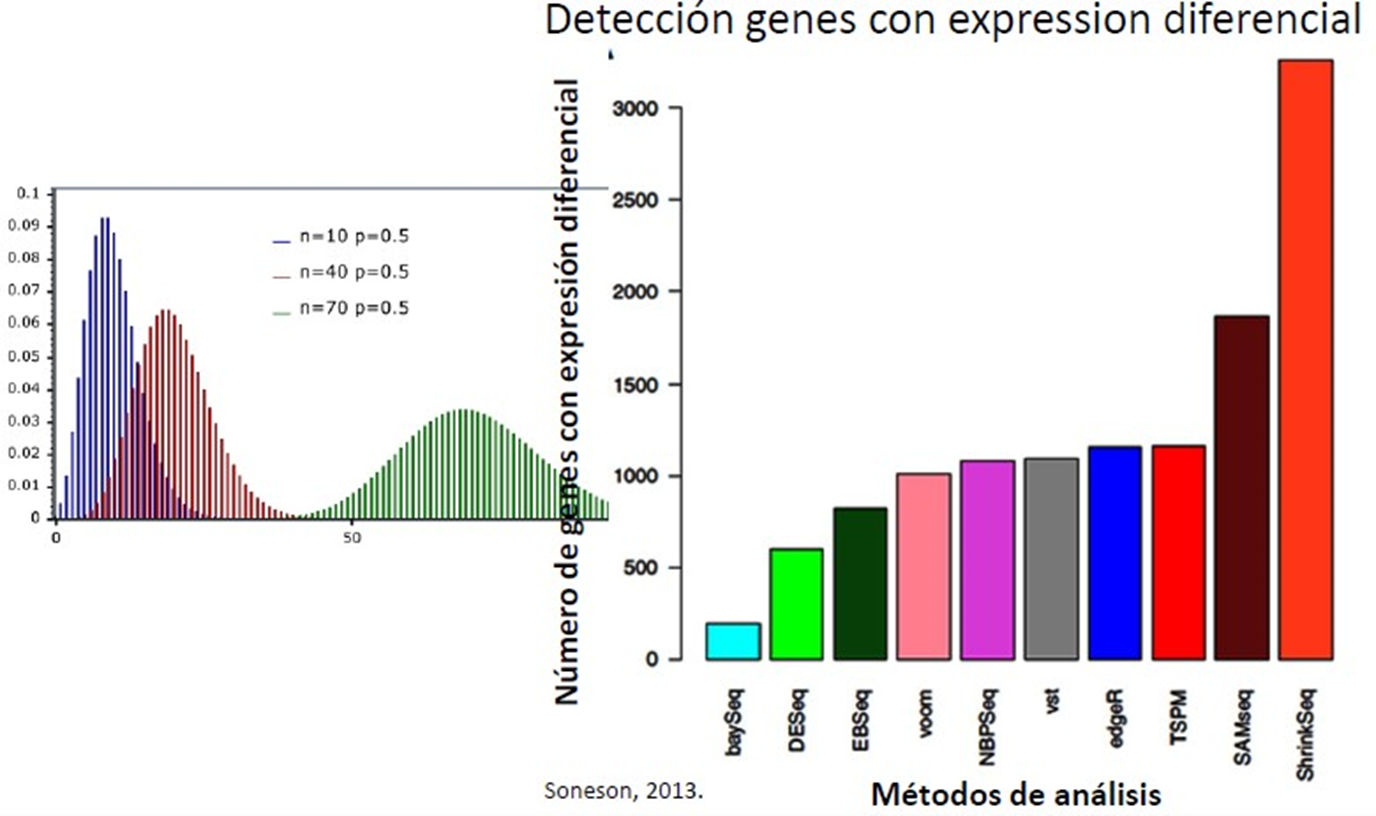
**8.- Para el procesamiento de datos de tablas de conteo de RNAseq ¿Que tipos de distribuciones estadísticas debo usar?. (1 punto)**

**9.- Con respecto a la distribución binomial. En un pequeño Laboratorio de Fisiologia Vegetal, 5 plantas serán expuestas a un paràsito ¿Cuál es la probabilidad de que se infecten 3 plantas de un total de 5 plantas? (1 punto)**

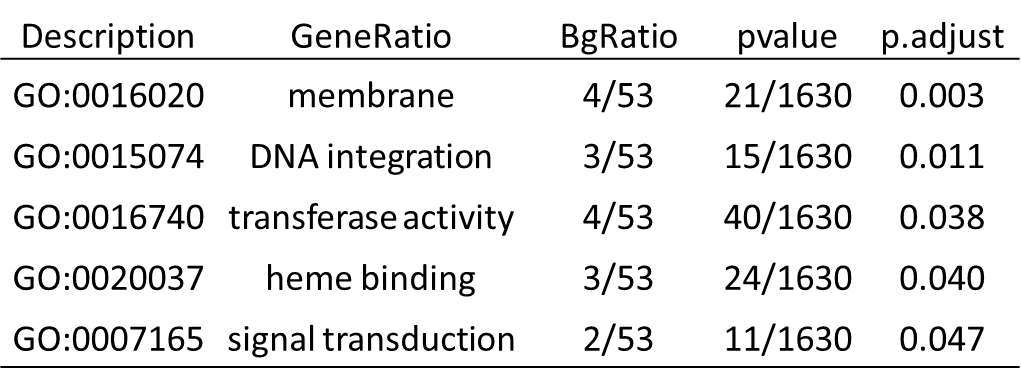
**¿Cuál de estos tres colores de barras se ajustan a una distribución binomial?¿Porque?**

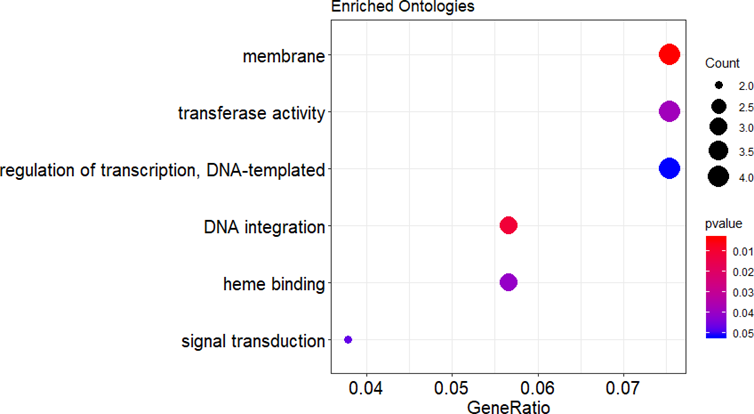
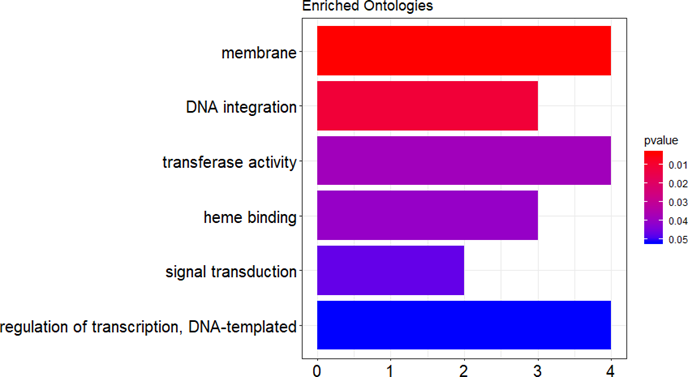
****

**10.- Segùn la Distribución Binomial Negativa (en la siguiente Figura), en el caso de RNAseq estos fueron tomados con datos IN SILICO siendo un total de 12000 genes en donde los autores modificaron la expresión de 12000 genes al 10% para someter a un desafio de diferentes métodos donde se observo que había más de 3000 genes con expresión diferencial el cual nos dice ese método que de la cantidad de genes que está cambiando hay muchos falsos positivos, es decir nos indica que los genes que cambiaron, en realidad nunca cambiaron y si nos vamos al otro extremo tenemos métodos muy conservadores que encuentran muy poquitos genes, debido a que se exige de que la diferencia entre los genes sea muy elevado para que el método sea capaz de clasificarlo como el gen con expresión diferencial. Por tanto ¿Cuáles son los métodos que obtuvieron los mejores rendimientos? ¿A que cantidad de genes corresponde esos metodos? (2 punto)**

****

**11.- En cuanto a la distribución hipergeometrica, hablando de genes, en mi experimento tomo 53 genes de expresión diferencial y en el genoma de mi línea biológica hay 1630 genes del cual tengo 21 genes anotados en membrana y 4 genes anotados en la membrana a partir de 53 genes asi sucesivamente tal como se observa en la siguiente tabla siendo el BgRatio mi listado y pvalue el de referencia. ¿Explique claramente hace la distribución hipergeometrica con estas proporciones? Y que se consigue luego? (1 punto)**

****

****

**12.- EJEMPLO TRIVIAL: DISTRIBUCIÒN NORMAL. En una misión secreta se descubre que en una ciudad ubicada cerca de un reactor nuclear se descubre que toda la población de personas están expuestas a un nivel de radiación peligroso. Ante esta situación el CENTRO DE CRONOBIOLOGIA DE EUROPA(CCE) realiza un estudio de transcriptòmico circadiana en la población, detectando en esa población un promedio de 40 genes acíclicos (no típicos) en su organismo. Por tanto, un grupo de tesistas del laboratorio de LAB-BIOTBEC de la Universidad Nacional de San Agustin de Arequipa realiza un estudio transcriptómico en la población tomando una muestra de 100 personas encontrando que la µ ≠ 40 genes acíclicos (no típicos). Cuestionando seriamente los resultados del CCE. Según la muestra tomada ¿será correcta o incorrecta la información proporcionada por el CCE? Los estudiantes de LAB-BIOTBEC calcularon que X= 50 siendo el nivel de significancia α=0.01 y desviación estándar=10 (2 puntos)**

**13.- A LAB-BIOTBEC viene de visita un investigador de alto impacto de la Universidad de Harvard, y desea otorgar una beca a un tesista para que continúe con su trabajo de investigación en EEUU, para ello el investigador estadounidense trata de evaluar si alguno de los tesistas tienen nociones sobre Bioestadística Genómica y Molecular, y les hace una pregunta. Si llevo a cabo una prueba, la probabilidad de que yo me equivoque es** “**α” siendo la probabilidad de cometer error tipo I ¿Cuál es la probabilidad de no equivocarme si investigo 60000 genes a diferencia de investigar 8 genes en mi línea biológica? (2 puntos)**

**14.-Aplicando la FDR (en español: tasa de los descubrimientos falsos) los siguientes valores p de los 30000 genes de mi línea biológica. Realizar lo siguiente:** **(2 puntos)**

**A) Ordenar de menor a mayor valores p y asignar ranking.**

**B) p ajustado = (i\*α) / m i: ranking; α: 0.05; m: # pruebas**

**C)Luego rechazar H0 cuando p ajustado < 0.05**

**valores p: Valor p valor p ajustado**

**0,04 0,04 ≤ \*0,05/n =**

**0,003 0,003 ≤ \*0,05/n =**

**0,02 0,02 ≤ \*0,05/n =**

**0,008 0,008 ≤ \*0,05/n =**

**0,002 0,002 ≤ \*0,05/n =**

**0,004 0,004 ≤ \*0,05/n =**

**0,0004 0,0004≤ \*0,05/n =**

**0,0006 0,0006 ≤ \*0,05/n=**

**0,0002 0,0002 ≤ \*0,05/n=**

**15.- En la población expuesta a la radiación nuclear. Ante esa situación los tesistas de LAB\_BIOTBEC no disponen de muchos fondos para investigación, por lo que aplicarán la estadística bayesiana usando βeta(29,29). Por tanto, para hacer un estudio de transcriptómica circadiana sólo se tomará una muestra de 10 personas, de las cuales se observa que solo 4 personas tienen genes arrítmicos no típicos. (2 puntos)**

**¿Cuál es la distribución aposteriori que modela la proporción de la población que se detecta genes arrítmicos no tìpicos?**

**Hallar los valores típicos (media, moda y mediana a partir de la distribución beta que se obtuvo)**

**¿Cuál es la probabilidad de pertenecer a un intervalo?**

**¿Cuál es la probabilidad que el mínimo el promedio hallado tenga genes arrítmicos?**

