

# Evaluación 3: Distribuciones de probabilidad. Parte Computacional

## (60 puntos)

Carlos M. Lopera-Gómez, Ph.D. Elementos de Probabilidad y Estadística (3008450).

05/06/2020

### Distribuciones de Probabilidad (Fecha de entrega: junio 8 de 2020)

En este trabajo el estudiante utilizará herramientas numéricas y gráficas disponibles en  $R$  para analizar un problema de control de calidad en el proceso de fabricación de pañuelos faciales. Para ello usará el proceso para revisar las propiedades básicas de las distribuciones binomial y normal a través de un experimento computacional y simulación. En particular, verificará los parámetros y el supuesto de distribución normal del proceso.

En las preguntas de este trabajo el estudiante deberá usar funciones disponibles en  $R$  para las distribuciones normal y binomial.

### Examen de la resistencia de pañuelos

Las grandes empresas solicitan inconformidades a sus consumidores con el fin de identificar y tratar de corregir los problemas en el proceso de fabricación que producen las inconformidades. Una de las quejas más comunes ante un gran fabricante de pañuelos faciales es que las hojas se rompen al retirarlas de la caja. Como la resistencia a la tracción es el factor que determina cuándo se rasgará una hoja, la gerencia decidió re-examinar una variable muy importante del proceso que afecta la resistencia a la tracción, que es la resistencia en la dirección transversal (RDT). La RDT es la fuerza en la dirección en la que el pañuelo se saca de la caja. Un valor de la RDT que sea demasiado alto crea una hoja rígida, mientras que un valor de la RDT que sea demasiado bajo puede hacer que las hojas se rasguen al retirarlas de la caja.

Según las especificaciones, se supone que la RDT sigue una distribución normal con una media de 300 lb/resma y una desviación estándar de 4 lb/resma. Los valores críticos para la RDT son 290 y 315. Una hoja con un valor de RDT inferior a 290 lb/resma o superior a 315 lb/resma se considera inaceptable. (una resma está formada por 480 hojas individuales o pañuelos faciales)

En este trabajo se examinará la distribución de la RDT en  $R$ .

1. (14pts) En esta parte, se analiza cómo los cambios en la media y en la desviación estándar afectan la fracción de pañuelos que no cumple con las especificaciones de RDT. Suponga en esta pregunta que la RDT sigue una distribución normal. Responda las siguientes preguntas utilizando funciones de  $R$ . Los cálculos de probabilidades deben estar acompañados de una gráfica que represente la probabilidad pedida.
  - a. (3pts) Suponga que  $\mu = 300$  y  $\sigma = 4$ . ¿Qué fracción de pañuelos tendrá una RDT inaceptable (RDT por debajo de 290 o por encima de 315 lb/resma)? ¿Qué fracción tendría una RDT inaceptable si el valor de  $\mu$  fuera de 298 con el mismo valor de  $\sigma$ ?
  - b. (2pts) Ahora suponga que la RDT media es 300 con el valor de  $\sigma$  igual a 6. ¿Qué fracción de pañuelos será inaceptable?

- c. (3pts) ¿Cuál es la fracción de pañuelos con el valor de la RDT dentro de 1 desviación estándar de la media de 300? ¿Cuál es la fracción de pañuelos dentro de 2 desviaciones estándar de la media? Suponga acá que  $\sigma = 4$  lb/resma.
  - d. (3pts) Suponga  $\mu = 300$  y  $\sigma = 4$ . ¿Cuál es el valor de la RDT que es superado por el 95% de los pañuelos? ¿Cuál sería el valor de la RDT que es superado por el 99% de los pañuelos?
  - e. (3pts) Una manera en que se puede reducir la fracción de hojas inaceptables es reducir la desviación estándar del proceso. ¿Cuál debe ser la desviación estándar  $\sigma$  para que solo el 1% de los pañuelos tenga un valor de RDT menor que 294? Suponga  $\mu = 300$ . Para esto, se deberá que variar la desviación estándar hasta llegar al percentil deseado. Informe el valor aproximado de  $\sigma$  con dos decimales.
2. (20pts) Suponga que el proceso de fabricación funciona correctamente y la RDT sigue una distribución normal con una media  $\mu = 300$  lb/resma y una desviación estándar  $\sigma = 4$  lb/resma.
- a. (3pts) Investigue como obtener en  $R$  una muestra aleatoria de una distribución normal con las características anteriores y úsela para simular 200 mediciones de la RDT (use el número de la cédula como semilla en la función `set.seed(#CC)`). Esto corresponderá a seleccionar aleatoriamente 200 pañuelos y obtener una medición de RDT para cada uno de ellos.
  - b. (3pts) Cuente el número de hojas inaceptables (RDT por debajo de 290 o por encima de 315) en la muestra. ¿El número observado de hojas inaceptables es coherente con la predicción teórica obtenida en la pregunta 1. literal a.?
  - c. (6pts) Determine el número de observaciones dentro de 1, 2 y 3 desviaciones estándar de la media para la muestra. Compare sus resultados con la regla empírica vista en clase.
  - d. (3pts) Use  $R$  para obtener un gráfico de probabilidad normal (QQplot) de las 200 observaciones. Describa el patrón que se muestra en la gráfica. ¿Es el gráfico consistente con la teoría? Explicar brevemente.
  - e. (2pts) Obtenga los valores para la variable estandarizada  $Z = (RDT - \mu)/\sigma$  para cada observación de la muestra de 200 observaciones de RDT. ¿Cuál es la distribución de los valores estandarizados? Consulte la teoría para responder la pregunta.
  - f. (3pts) Utilice  $R$  para calcular la media y la desviación estándar de las observaciones estandarizadas obtenidas en el punto anterior. Informe los dos valores con cuatro decimales. ¿Los valores son consistentes con los parámetros de distribución proporcionados por la teoría? Explique.
3. (10pts) El fabricante ha implementado recientemente algunos cambios en el proceso de fabricación que se espera que mejoren las propiedades de separación de los pañuelos. Para evaluar el efecto de los cambios en la distribución de la RDT, se tomó una muestra aleatoria que constaba de 200 pañuelos, y las mediciones de la RDT se registraron en el archivo `new_rdt.txt` disponible en <https://drive.google.com/open?id=1C-RVb4R-VcKqLVCgog-Lz4lgo7azwaD>
- a. (4pts) Utilice  $R$  para calcular la media y la desviación estándar de las 200 mediciones de RDT. ¿Son consistentes los valores con los parámetros objetivo (media  $\mu = 300$  lb/resma y una desviación estándar  $\sigma = 4$  lb/resma)? Comente brevemente.
  - b. (3pts) Use la muestra para realizar un gráfico de probabilidad normal para verificar el supuesto de normalidad en la distribución de la RDT. Describa el patrón que se muestra en la gráfica. ¿Hay alguna indicación fuerte de que los datos no siguen una distribución normal? Explicar brevemente.
  - c. (3pts) Use sus hallazgos en las partes a. y b. para preparar un informe breve para la administración que indique si los cambios han mejorado o no las cualidades de separación de los pañuelos. En particular, comente sobre los cambios en los parámetros del proceso (media y desviación estándar) y la normalidad de la distribución de la RDT.
4. (16pts) En esta parte, se utilizarán las funciones de  $R$  para las distribuciones binomial y normal.
- a. (5pts) El objetivo del fabricante es no exceder el límite del 5% en la proporción de pañuelos inaceptables (por debajo de 290 o por encima de 315). Si no se cumple el requisito, el proceso de producción se interrumpe mientras se realizan los ajustes. Una forma reconocida de controlar

la producción es tomar una muestra aleatoria de la producción de cada hora y determinar la fracción de pañuelos inaceptables. El fabricante desarrolló un criterio según el cual se interrumpe el proceso de fabricación para realizar ajustes cuando una muestra aleatoria de 200 pañuelos da como resultado más de 15 pañuelos inaceptables.

¿Qué posibilidades hay de que el proceso de fabricación se interrumpa erróneamente en función de los datos de la muestra? En otras palabras, ¿cuál es la probabilidad de obtener más de 15 pañuelos inaceptables en una muestra aleatoria de 200, dado que la verdadera proporción de pañuelos inaceptables no excede de 0.05?

- b. (6pts) El fabricante afirma que la RDT media objetivo es de 300 lb/resma con un valor de  $\sigma$  igual a 4 lb/resma. Suponiendo que esto sea cierto, calcule la probabilidad de que se obtengan más de 6 pañuelos inaceptables (RDT por debajo de 290 o por encima de 315) en una muestra aleatoria de 200 pañuelos. Consejo: use la probabilidad de obtener un pañuelo inaceptable calculada en el punto 1(a) como el segundo parámetro de la binomial.
- c. (5pts) Use la aproximación normal de la distribución binomial para calcular las probabilidades en a. y b. Compare los resultados y comente acerca de la validez de la aproximación.