Universidad de Costa Rica

Escuela de Ciencias de la Computación e Informática

Plantilla de Entregables CI-0115

Sigla y Curso **CI-0115 Probabilidad y Estadística** **Grupo** **02**

Título del Trabajo **Aprendizaje Colaborativo 3 Grupo 3**

Nombre del Equipo **"Los Compas"**

Fecha de Realización **25/11/22** **Fecha Entrega** **30/11/22**

Profesor(a) **Diego Mora Jiménez** **Asistente** **Nayeri Azofeifa**

Tardía: (________)

Estudiante: **José Antonio Mora M** **Nota:** ________ **/100**

Estudiante: **Marcelo Delgado M**

Estudiante: **Jean Paul Chacón G**

Estudiante: ____________________

Ejercicios

1.8.3. Los tiempos de reacción de una muestra aleatoria de 9 personas sujetas a un estimulante fueron registrados como: 2.5, 3.6, 3.1, 4.3, 2.9, 2.3, 2.6, 4.1, y 3.4 segundos. Calcule:

(a) La media
(b) La mediana

Respuesta:

(a) $\frac{2.5+3.6+3.1+4.3+2.9+2.3+2.6+4.1+3.4}{9} = 3.2$

(b) 2.3, 2.5, 2.6, 2.9, 3.1, 3.4, 3.6, 4.1, 4.3 → Ya ordenados el elemento del medio es decir el número 5 es 3.1.

1.8.16. En la temporada de 2004-05 de fútbol americano, la universidad de California del Sur contaba con las diferencias de puntajes de los 13 partidos jugados.

11 49 32 3 6 38 38 30 8 40 31 5 36

Encuentre

(a) La media de la diferencia de puntajes
(b) La mediana de la diferencia de puntajes

Respuesta:

(a) $\frac{11+49+32+3+6+38+38+30+8+40+31+5+36}{13} = 25.15$

(b) 3 5 6 8 11 30 31 32 36 38 38 40 49 al ser 13 datos el número del medio será el número 7 el cual en este caso es 31.

2.8.25. La vida promedio de una máquina para hacer pan es 7 años con una desviación estándar de 1 año. Asumiendo que las vidas de estas máquinas siguen una distribución normal, encuentre:

(a) La probabilidad de que la vida media de una muestra aleatoria de 9 máquinas está entre los 6.4 y ls 7.2 años
(b) El valor de x a la izquierda en el que el caiga 15% de las medias calculadas de una muestra aleatoria de 9 máquinas.

Respuesta:

(a) Utilizando los datos que nos da el problema en el que nos encontramos, para encontrar la probabilidad de que la vida media de una muestra aleatoria de 9 máquinas se encuentre entre los 6.4 y los 7.2 años debemos averiguar el valor de Z cuando X=7.2 y restarle el valor de Z cuando X= 6.4.
Para ellos utilizamos la fórmula de Z y el teorema del limite central de la siguiente manera:
$Z = \frac{X-\mu}{s}$ siendo X lo que queremos averiguar, μ la vida media, y por medio de la desviación estándar obtenemos $s = \frac{1}{\sqrt{9}} = 0,33$ con estos datos obtenemos:
$Z = \frac{7.2-7}{0.333} = 0.6$
Y
$Z = \frac{6.4-7}{0.333}$ = -1.8
Utilizando la tabla de distribución normal obtenemos que:
Z= 0.6 = 0.7257
Z= -1.8 = 0.035
Por lo que basta con restar estos dos valores y obtenemos la respuesta 0.7257-0.035 = 0.6898 = 68.98% es la probabilidad de que la media de vida de las 9 máquinas caiga entre 6.4 y 7.2.

(b) Para encontrar el valor de x a la derecha en el que el 15% de las medias calculadas de 9 máquinas se encuentre basta con tomar el percentil 85 ya que 100-15 = 85 lo cual es lo mismo que decir encontrar X cuando Z toma un valor p= 0.85.
Al ver la tabla de distribución normal vemos que cuando p=0.85, Z = 1.037 solo basta con encontrar X:
$Z = \frac{X-\mu}{s}$
$1.037 = \frac{X-7}{0.333}$
X-7 = 1.037*0.333
X= 7.346 es el valor de X buscado.

3.8.42. Los puntajes de un examen de nivelación de estudiantes de nuevo ingreso de una universidad en los últimos 5 años están aproximadamente distribuidos de manera normal con una media de μ=74 y una varianza de $\sigma^2 = 8$. Considera que $\sigma^2 = 8$ es un valor válido para la varianza si una muestra aleatoria de 20 estudiantes que tomen el examen este año obtienen una nota de $s^2 = 8$?

Respuesta:
Para esto podemos utilizar la prueba de chi cuadrado en el que nuestras hipótesis son:

$H_0 : \sigma^2 = 8$

$H_1 : \sigma^2 \neq 8$

Y podemos utilizar el estadístico

$$X^2 = \frac{(n-1)\, s^2}{\sigma^2} = \frac{(20-1)\, 20^2}{8^2} = 47.5$$

Si vemos en la tabla el dato para $X^2_{0.01} = 36.191$ por lo que al ver que son los datos diferentes podemos concluir que el valor $\sigma^2 = 8$ de varianza no es válido.

4.9.12. Una muestra aleatoria de 10 barras energéticas de chocolate de cierta marca tiene, en promedio, 230 calorías por barra, con una desviación estándar de 15 calorías. Construya un intervalo de confiabilidad de 99% para la media verdadera de calorías de esta marca de barritas energéticas. Asuma que la distribución de las calorías es aproximadamente normal.

Respuesta:
Al tener la media y desviación estándar de la muestra podemos calcular el intervalo de confiabilidad de esta forma:

$\bar{x} - t * \frac{s}{\sqrt{n}} < \mu < \bar{x} + t * \frac{s}{\sqrt{n}}$ donde s se trata de la desviación estándar y donde t se trata de los grados de libertad los cuales obtenemos con n-1 = 10-1 = 9 y el grado de confiabilidad que lo obtenemos de la siguiente manera:

99% = 100(1-α)%

1-α = 0.99

α = 0.01

$\frac{\alpha}{2}$ = 0.005

y obtener el valor de t por medio de la tabla t-student y t = 3.2498

Sustituyendo los valores en el intervalo obtenemos:

$230 - 3.2498 * \frac{15}{\sqrt{10}} < \mu < 230 + 3.2498 * \frac{15}{\sqrt{10}}$

230 - 15.4152 < μ < 230 + 15.4151

214.58 < μ < 245.41 es el intervalo de confiabilidad de un 99%.

4.9.18. De acuerdo al ejercicio 9.13 construya un intervalo de tolerancia de 95% que contenga el 90% de las medidas.

Respuesta:

Gracias a los datos que nos da la pregunta podemos calcular el intervalo de tolerancia de la siguiente manera:

$\overline{x} - y_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + y_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$

Podemos obtener el valor de α de la siguiente manera:

Debido a que queremos el 90% de las medidas tenemos:

90% = 100(1-α)%

0.9 = 1-α

α = 0.1

Ahora sustituimos en el intervalo:

$48.50 - y_{0.05} * \frac{1.5}{\sqrt{12}} < \mu < 48.50 + y_{0.05} * \frac{1.5}{\sqrt{12}}$

Para encontrar $y_{0.05}$ podemos utilizar la tabla t-student para 95% ya que 1- 0.05 = 0.95 y 12-1 = 11 grados de libertad y conseguimo $y_{0.05}$ = 1.796, sustituimos:

$48.50 - 1.796 * \frac{1.5}{\sqrt{12}} < \mu < 48.50 + 1.796 * \frac{1.5}{\sqrt{12}}$

47.72 < μ 49.28

5.9.44. Siguiendo con el ejercicio 9.43, encuentre un intervalo de confiabilidad para μ_1 - μ_2 si las llantas de dos marcas diferentes son asignadas aleatoriamente a las llantas derecha e izquierda traseras de 8 taxis y las siguientes distancias en kilómetros fueron apuntadas:

Taxi	Brand A	Brand B
1	34,400	36,700
2	45,500	46,800
3	36,700	37,700
4	32,000	31,100
5	48,400	47,800
6	32,800	36,400
7	38,100	38,900
8	30,100	31,500

Asuma que las diferencias en las distancias están distribuidas normalmente.

Respuesta:

Para encontrar la media de las diferencias entre las distancias primero debemos encontrar la diferencia entre las distancias de la siguiente manera:

d_1= 34400 - 36700 = -2300

d_2= 45500 - 36800 = -1300

d_3= 36700 - 37700 = -1000

d_4= 32000 - 31100 = 900

d_5= 48400 - 47800 = 600

d_6= 32800 - 36400 = -3600

d_7= 38100 - 38900 = -800

d_8= 30100 - 31500 = -1400

Y podemos obtener la media de las diferencias como:

$\frac{-2300+-1300+-1000+900+600+-3600+-800+-1400}{8}$ = -1112.5

Por otro lado podemos obtener la desviación estándar de la siguiente manera:

Sd = $\sqrt{\frac{1}{n-1}\sum_{1}^{n}(di\ -\ \overline{d})^2}$ = $\sqrt{\frac{1}{8-1}\sum_{1}^{8}(di\ -\ -\ 1112.5)^2}$ = 1454

Por último debemos encontrar el valor de t para el 99%

Primero debemos obtener $\frac{\alpha}{2}$ lo obtenemos :

99% = 100(1-α)%

1-α = 0.99

α = 0.01

$\frac{\alpha}{2}$ = 0.005

Por lo tanto podemos tomar el valor de $t_{0.005}\ o\ t_{0.995}$ y 8-1 = 7 grados de libertad lo cual nos da 3.499.

$$\overline{d} - y_{\frac{\alpha}{2}} * \frac{Sd}{\sqrt{n}} < \mu_1 - \mu_2 < \overline{d} + y_{\frac{\alpha}{2}} * \frac{Sd}{\sqrt{n}}$$

Sustituimos:

$$-1112.5 - 3.499 * \frac{1454}{\sqrt{8}} < \mu_1 - \mu_2 < -1112.5 + 3.499 * \frac{1454}{\sqrt{8}}$$

$-2911.2 < \mu_1 - \mu_2 < 686.2$.

5.9.50. Dos niveles de dosis de insulina (bajo y alto) se les da a dos grupos de ratas diabéticas para revisar su resistencia a la insulina, y se proporcionan los siguientes datos:

Low dose:	$n_1 = 8$	$\bar{x}_1 = 1.98$	$s_1 = 0.51$
High dose:	$n_2 = 13$	$\bar{x}_2 = 1.30$	$s_2 = 0.35$

Asuma que las varianzas son iguales. Dé un intervalo de confianza del 95% para la diferencia en el promedio real de resistencia a la insulina entre las dos muestras.

R/ Debido a que se trata de encontrar el intervalo de confianza de la diferencia de dos medias, en el caso en que las **varianzas son iguales pero desconocidas**, se utilizan las fórmulas:

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$S_p^2 = \frac{(8-1)*0.51^2+(13-1)*0.35^2}{8+13-2} = S_p^2 = 0.173$, como se necesita S_p^2 sin estar elevado al cuadrado, se hace: $\sqrt{0.173} = S_p = 0.416$.

Luego, como se necesita calcular el intervalo de confianza del 95%, se despeja la fórmula: $95\% = 100(1 - \alpha)\% = \frac{95}{100} = (1 - \alpha) = 0.95 = (1 - \alpha) = \alpha = 0.05$.

Ahora, calculando el grado de libertad: $v = n_1 + n_2 - 2 = 8 + 13 - 2 = 19$.

Con esto se puede calcular la función t = $t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025}$, utilizando la tabla de la distribución *t*, se puede observar que $t_{0.025} = 2.093$.

v	0.40	0.30	0.20	0.15	0.10	0.05	0.025
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093

Con todos los valores ya calculados, se pueden sustituir los valores en la fórmula original:

$$(1.98 - 1.30) - 2.093 * 0.416 * \sqrt{\frac{1}{8} + \frac{1}{13}} = 0.2887$$

$$(1.98 - 1.30) + 2.093 * 0.416 * \sqrt{\frac{1}{8} + \frac{1}{13}} = 1.0713$$

Por lo que, el intervalo de confianza de 95% de la diferencia del promedio real de resistencia a la insulina entre las dos muestras es de 0.2887 – 1.0713.

6.9.60. Qué tan grande debe de ser la muestra si deseamos estar 99% seguros que la proporción de la muestra del Ejercicio 9.51 estará dentro del 0.05 de la proporción verdadera de casas en la ciudad que son calentadas con aceite.

R/ Para este problema se debe aplicar el Teorema 9.4:

If $\hat{p}$ is used as an estimate of p, we can be $100(1-\alpha)\%$ confident that the error will be less than a specified amount e when the sample size is approximately

$$n = \frac{z^2_{\alpha/2}\hat{p}\hat{q}}{e^2}.$$

Se indica en el enunciado que la muestra debe estar dentro del $e = 0.05$ de la proporción verdadera de casas calentadas con aceite.

En el ejercicio 9.51 se indica que de n = 1000 casas, x = 228 son calentadas por aceite, y con estos datos se puede averiguar $\hat{p}$ y $\hat{q}$.

$$\hat{p} = \frac{x}{n} = \frac{228}{1000} = 0.228. \qquad \hat{q} = 1 - p = 1 - 0.228 = 0.772.$$

Como se debe de calcular un nivel de confianza de 99% se despeja la fórmula: $99\% = 100(1 - \alpha)\% = \frac{99}{100} = (1 - \alpha) = 0.99 = (1 - \alpha) = \alpha = 0.01$. Utilizando la tabla de distribución Normal se puede averiguar el valor de Z para $\alpha = 0.01$.

En este caso $Z_{\frac{0.01}{2}} = Z_{0.005}$, por lo que hay que buscar en la tabla de distribución normal el valor de Z que corresponde a 1 – 0.005 = 0.995 del área bajo la curva.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952

En este caso, el valor más cercano corresponde a $Z_{\frac{0.01}{2}} = 2.57$.

Al ya tener todos los valores, se pueden sustituir en la fórmula original para averiguar *n* = $\frac{2.57^2 * 0.228 * 0.772}{0.05^2} = 465.03$. Entonces, se necesita de una muestra de aproximadamente 465 para estar 99% seguros de lo que se indica en el enunciado.

6.9.70. De acuerdo con el USA Today (17 de Marzo, 1997), el 33.7% de los editores en las estaciones de TV locales de Estados Unidos eran mujeres en 1990, y el 36.2% de los editores eran mujeres en 1994. Asuma que se contratan 20 nuevos editores.

A) Estime el número de estos que hubieran sido mujeres en 1990 y en 1994 respectivamente.

B) Calcule un intervalo de confianza del 95% para averiguar si hay evidencia de que la proporción de mujeres contratadas como editores fue más alta en 1994 que en 1990.

A) El número de editores contratados en 1990 corresponde a n_1 = 20 y el número de editores contratados en 1994 corresponde a n_2 = 20.

En 1990, si 20 personas corresponden al 33.7% de los editores, entonces el total de editores corresponde a = $\frac{20 * 100}{33.7} = 59.34 = 59$ personas.

En 1994, si 20 personas corresponden al 36.2% de los editores, entonces el total de editores corresponde a = $\frac{20 *100}{36.2} = 55.25 = 55$ personas.

Ahora, se puede calcular la proporción de mujeres contratadas en 1990 = $\hat{p}_1 = \frac{x}{n} = \frac{20}{59} = 0.338$.

Y, la proporción de mujeres contratadas en 1994 es = $\hat{p}_2 = \frac{x}{n} = \frac{20}{55} = 0.363$.

Con estos datos, ahora se puede realizar la estimación:

Cantidad de mujeres contratadas en 1990: $x_1 = n_1 * \hat{p}_1 = 20 * 0.338 = 6.76 \approx 7$.

Cantidad de mujeres contratadas en 1994: $x_2 = n_2 * \hat{p}_2 = 20 * 0.363 = 7.26 \approx 7$.

Entonces, aproximadamente, de las 20 personas contratadas, 7 de ellas hubieran sido mujeres, tanto en 1990 como en 1994.

B) Se utiliza la fórmula para averiguar el intervalo de confianza de la diferencia entre dos proporciones poblacionales:

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

Como se necesita calcular el intervalo de confianza del 95%, se despeja la fórmula:

$95\% = 100(1 - \alpha)\% = \frac{95}{100} = (1 - \alpha) = 0.95 = (1 - \alpha) = \alpha = 0.05$.

En este caso $Z_{\frac{0.05}{2}} = Z_{0.025}$, por lo que hay que buscar en la tabla de Distribución normal el valor de Z que corresponda a 1 – 0.025 = 0.975 del área bajo la curva.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

En este caso, el valor corresponde a $Z_{\frac{0.05}{2}}$ = 1.96.

Sabiendo estos valores, ya se pueden sustituir en la fórmula original:

$$(0.338 - 0.363) - 1.96 * \sqrt{\frac{0.338*(1-0.338)}{20} + \frac{0.363*(1-0.363)}{20}} = -0.3206$$

$$(0.338 - 0.363) + 1.96 * \sqrt{\frac{0.338*(1-0.338)}{20} + \frac{0.363*(1-0.363)}{20}} = 0.2706$$

Entonces, el intervalo de confianza de 95% de la diferencia entre dos proporciones poblacionales de mujeres contratadas en 1990 y en 1994 es de: -0.3206 < $p_1 - p_2$ < 0.2706.

Respondiendo la pregunta inicial, de si hay evidencia de que la proporción de mujeres contratadas fue más alta en 1994, como 1994 incluye al cero, entonces no se puede rechazar la hipótesis nula, debido a que existe muy poca evidencia, por lo que no existe evidencia de que esta proporción fue mayor en 1994.

7.9.80. Construya un intervalo de confianza del 95% para σ_A^2 / σ_B^2 del ejercicio 9.49 de la página 295. Debería de usarse la suposición de igual varianza?

En el ejercicio 9.49 se indica que se prueba el tiempo que toman dos tipos de pintura en secarse, y se presenta en un cuadro el tiempo que tomó (en horas) 15 elementos seleccionados en secarse:

Paint A					Paint B				
3.5	2.7	3.9	4.2	3.6	4.7	3.9	4.5	5.5	4.0
2.7	3.3	5.2	4.2	2.9	5.3	4.3	6.0	5.2	3.7
4.4	5.2	4.0	4.1	3.4	5.5	6.2	5.1	5.4	4.8

Además, se indica que el tiempo de secado está distribuido normalmente, con $\sigma_A = \sigma_{B.}$

Entonces, como son 15 elementos para cada tipo de pintura, se puede establecer:

$Pintura\ A = n_1 = 15$ y $Pintura\ B = n_2 = 15.$

Primero, en este caso, hay que sacar la media y varianza de los dos tipos de pintura:

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}.$$

Media de la pintura A: $\overline{X} = \frac{3.5+2.7+3.9+4.2+3.6+2.7+3.3+5.2+4.2+2.9+4.4+5.2+4.0+4.1+3.4}{15} = 3.82.$

Varianza de la pintura A: $S^2 = \frac{(3.5-3.82)^2+(2.7-3.82)^2+(3.9-3.82)^2+(4.2-3.82)^2+(3.6-3.82)^2+(2.7-3.82)^2+(3.3-3.82)^2+(5.2-3.82)^2+(4.2-3.82)^2+(2.9-3.82)^2+(4.4-3.82)^2+(5.2-3.82)^2+(4.0-3.82)^2+(4.1-3.82)^2+(3.4-3.82)^2}{15-1} = 0.607429.$

Media de la pintura B: $\overline{X} = \frac{4.7+3.9+4.5+5.5+4.0+5.3+4.3+6.0+5.2+3.7+5.5+6.2+5.1+5.4+4.8}{15} = 4.94.$

Varianza de la pintura B: $S^2 = \frac{(4.7-4.94)^2+(3.9-4.94)^2+(4.5-4.94)^2+(5.5-4.94)^2+(4.0-4.94)^2+(5.3-4.94)^2+(4.3-4.94)^2+(6.0-4.94)^2+(5.2-4.94)^2+(3.7-4.94)^2+(5.5-4.94)^2+(6.2-4.94)^2+(5.1-4.94)^2+(5.4-4.94)^2+(4.8-4.94)^2}{15-1} = 0.568286.$

Para calcular el intervalo de confianza para σ_A^2 / σ_B^2 cuando se tienen dos varianzas de muestras independientes se debe de aplicar la fórmula:

$$\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(v_1,v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha/2}(v_2,v_1)$$

Donde $f_{\alpha/2}(v_1,v_2)$ es un valor de la distribución f con $v_1 = n_1 - 1$ y $v_2 = n_2 - 1$ grados de libertad, que deja área de $\alpha/2$ a la derecha, y $f_{\alpha/2}(v_2,v_1)$ es un valor de la distribución f con $v_2 = n_2 - 1$ y $v_1 = n_1 - 1$ grados de libertad, que deja área de $\alpha/2$ a la derecha.

Por lo que, primero se debe calcular el valor de $\alpha = 95\% = 100(1 - \alpha)\% = \frac{95}{100} = (1 - \alpha) = 0.95 = (1 - \alpha) = \alpha = 0.05$.

Después, los valores de los grados de libertad $V_1\, y\, V_2\, para\, f =$

$$V_1 = n_1 - 1 = V_1 = 15 - 1 = 14.$$

$$V_2 = n_2 - 1 = V_2 = 15 - 1 = 14.$$

Con todo esto, se pueden reemplazar los valores en las fórmulas de la distribución *f*, y buscar el valor en la tabla de los valores críticos de la distribución *f* (Lenoir, J., 2013):

Numerator Degrees of Freedom

	10	11	12	13	14	15	16	17	18
1	968.6274	973.0252	976.7079	979.8368	982.5278	984.8668	986.9187	988.7331	990.3490
2	39.3980	39.4071	39.4146	39.4210	39.4265	39.4313	39.4354	39.4391	39.4424
3	14.4189	14.3742	14.3366	14.3045	14.2768	14.2527	14.2315	14.2127	14.1960
4	8.8439	8.7935	8.7512	8.7150	8.6838	8.6565	8.6326	8.6113	8.5924
5	6.6192	6.5678	6.5245	6.4876	6.4556	6.4277	6.4032	6.3814	6.3619
6	5.4613	5.4098	5.3662	5.3290	5.2968	5.2687	5.2439	5.2218	5.2021
7	4.7611	4.7095	4.6658	4.6285	4.5961	4.5678	4.5428	4.5206	4.5008
8	4.2951	4.2434	4.1997	4.1622	4.1297	4.1012	4.0761	4.0538	4.0338
9	3.9639	3.9121	3.8682	3.8306	3.7980	3.7694	3.7441	3.7216	3.7015
10	3.7168	3.6649	3.6209	3.5832	3.5504	3.5217	3.4963	3.4737	3.4534
11	3.5257	3.4737	3.4296	3.3917	3.3588	3.3299	3.3044	3.2816	3.2612
12	3.3736	3.3215	3.2773	3.2393	3.2062	3.1772	3.1515	3.1286	3.1081
13	3.2497	3.1975	3.1532	3.1150	3.0819	3.0527	3.0269	3.0039	2.9832
14	3.1469	3.0946	3.0502	3.0119	2.9786	2.9493	2.9234	2.9003	2.8795

$$f_{\frac{\alpha}{2}}(V_{1,}V_2) = f_{\frac{0.05}{2}}(14, 14) = f_{0.025}(14, 14) = 2.98.$$

$$f_{\frac{\alpha}{2}}(V_{2,}V_1) = f_{\frac{0.05}{2}}(14, 14) = f_{0.025}(14, 14) = 2.98.$$

Ahora que se tienen todos los valores, se pueden reemplazar los valores en la fórmula original:

$$\frac{0.607429}{0.568286} * \frac{1}{2.98} < \frac{\sigma_A^2}{\sigma_B^2} < \frac{0.607429}{0.568286} * 2.98 = 0.359 < \frac{\sigma_A^2}{\sigma_B^2} < 3.185.$$

Debido a que este intervalo de confianza si permite que la razón σ_A^2 **/** σ_B^2 sea igual a 1, se puede asumir que σ_A^2 **=** σ_B^2 con poco grado de error, por lo que sí debería de usarse la suposición de igual varianza en este caso.

8.10.1. Suponga que un alergólogo quiere probar la hipótesis de que al menos 30% del público es alérgico a algún producto que contenga queso. Explique cómo el alergólogo puede cometer:

A) Un Error de Tipo I

B) Un Error de Tipo II

Suponiendo que la hipótesis nula H_0 se refiere a que 30% o más del público general es alérgico a algún producto con queso.

Y la hipótesis alternativa H_1 se refiere a que menos del 30% del público general es alérgico a algún producto con queso.

A) Un error de Tipo I se comete cuando se rechaza la hipótesis nula, o H_0 a favor de H_1, cuando H_0 es verdadera. En este ejemplo específico, un error de Tipo I que puede cometer el alergólogo es que concluya que menos del 30% del público es alérgico a algún producto con queso cuando en realidad 30% o más del público es alérgico.
B) Un error de Tipo II se comete cuando no se puede rechazar la hipótesis nula H_0, cuando la hipótesis alternativa H_1 es correcta, o en otras palabras, cuando se cree que H_0 es correcta cuando no. En el ejemplo específico, un error de Tipo II que puede cometer el alergólogo es concluir que 30% o más del público es alérgico a algún producto con queso cuando menos del 30% es alérgico.

8.10.12. Una muestra aleatoria de 400 votantes en una ciudad se les pregunta si están a favor de un aumento del 4% de impuestos a la gasolina para conseguir fondos necesarios para reparar calles. Si más de 220 pero menos de 260 votantes están a favor del impuesto, se debe concluir que 60% de los votantes están a favor.

A) Encuentre la probabilidad de cometer un error de Tipo I si 60% de los votantes están a favor del impuesto agregado.

B) ¿Cuál es la probabilidad de cometer un error de Tipo II usando este procedimiento si solamente 48% de los votantes están a favor del impuesto agregado?

A) Para calcular la probabilidad de cometer un error, hay que sacar la media y la desviación estándar de la muestra. El enunciado indica que *n* = 400 y que la *p* = 0.60.

μ = $n * p = 400 * 0.60 = 240.$

σ = $\sqrt{n * p * q} = \sqrt{400 * 0.60 * (1 - 0.60)} = 9.80.$

Con estos datos, ahora se puede encontrar la probabilidad de que se cometa el error de Tipo I:

$$\alpha = P(Error\ de\ tipo\ I) = P(X < 220\ cuando\ p = 0.60) + P(X > 260\ cuando\ p = 0.60)$$

Para calcular esto, se necesita el valor del área bajo la curva de la Distribución Normal Z, tomando en cuenta si se está calculando el valor mayor o menor al valor crítico:

Cuando X debe ser mayor a 220(valor crítico): $Z = \frac{X-\mu}{\sigma} = \frac{220.5-240}{9.80} = -1.99$

Cuando X debe ser menor a 260(valor crítico): $Z = \frac{X-\mu}{\sigma} = \frac{259.5-240}{9.80} = 1.99$

$$P(Z < -1.99) + P(Z > 1.99) = P(Z <- 1.99) + (1 - P(Z < 1.99))$$

Utilizando la tabla de Distribución Normal, se puede averiguar el valor del área bajo la curva para el valor Z establecido:

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

$= (0.0233) + (1 - 0.9767) = (0.0233 + 0.0233) = 0.0466.$

Por lo que, la probabilidad de cometer un error de tipo I cuando el 60% de los votantes están a favor del impuesto, es de 0.0466.

B) Nuevamente, para calcular la probabilidad de cometer un error, hay que sacar la media y la desviación estándar de la muestra. El enunciado indica que *n* = 400 y en este caso la *p* = 0.48.

μ = $n * p = 400 * 0.48 = 192.$

σ = $\sqrt{n * p * q} = \sqrt{400 * 0.48 * (1 - 0.48)} = 9.992.$

Ahora se puede calcular la probabilidad de que se cometa un error de Tipo II:

$\beta = P(Error\ de\ tipo\ II) = P(220 < X < 260\ cuando\ p = 0.48)$.

Para calcular esto, se necesita el valor del área bajo la curva de la Distribución Normal Z, tomando en cuenta si se está calculando el valor mayor o menor al valor crítico:

Cuando X debe ser mayor a 220(valor crítico): $Z = \frac{X-\mu}{\sigma} = \frac{220.5-192}{9.992} = 2.85.$

Cuando X debe ser menor a 260(valor crítico): $Z = \frac{X-\mu}{\sigma} = \frac{259.5-192}{9.992} = 6.76.$

$P(2.85 < Z < 6.76)$. Nuevamente, utilizando la tabla de Distribución Normal, se puede averiguar el valor del área bajo la curva para el valor Z establecido:

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981

6.76 no se encuentra dentro de la tabla, debido a que cualquier valor mayor/menor a 3.49/-3.49 es equivalente a 1.0000 / 0.0000 respectivamente.

$P(Z < 6.76) - P(Z < 2.85) = 1.0000 - 0.9978 = 0.0022.$

Por lo que, la probabilidad de cometer un error de tipo II cuando el 48% de los votantes están a favor del impuesto, es de 0.0022.

9.10.37. En el ejercicio 9.42 de la página 295, pruebe la hipótesis de que el ahorro de combustible del Volkswagen mini-truck, en promedio, excede la de un Toyota mini-truck similar por 4 kilómetros por litro. Use un valor de significancia de 0.10.

$Se\ tiene$:

$H_0: \mu_1 - \mu_2 = 4$

$H_1: \mu_1 - \mu_2 \neq 4$

$Gracias\ al\ ejercicio\ 9.42,\ se\ tiene\ que\ n_1 = 12,\ n_2 = 10,\ s_1 = 1,\ s_2 = 0.8,$

$Se\ tiene\ un\ nivel\ de\ significancia\ de\ 0.1,\ y\ se\ calculará\ la\ varianza\ mediante$:

$$S_p = \sqrt{\frac{(n_1-1)s_1^2 - (n_2-1)s_1^2}{n_1+n_2-2}}$$

$$S_p = \sqrt{\frac{(12-1)(1)^2 - (10-1)(0.8)^2}{12+10-2}}$$

$S_p = 0.91$

$Se\ utiliza\ el\ resultado\ anterior\ para\ utilizar\ la\ fórmula\ de\ la\ prueba\ t$:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{0.91\sqrt{\frac{1}{12}+\frac{1}{10}}}$$

$$t = \frac{(16-11)-4}{0.91(0.428)}$$

$t = 2.61,\ con\ 20\ grados\ de\ libertad.$

$Se\ tiene\ un\ nivel\ de\ significancia\ de\ 0.1,\ y\ 20\ grados\ de\ libertad,\ lo\ cual\ resulta$
$en\ un\ valor\ crítico\ de$:

$t_{20,0.1} > 1.725$

$Como\ se\ puede\ observar\ el\ valor\ crítico\ es\ menor\ al\ valor\ obtenido,\ por\ lo\ que$
$se\ rechaza\ la\ hipótesis\ nula.\ Por\ lo\ que\ no\ se\ puede\ decir\ que\ la\ economía\ de$
$gasolina\ de\ de\ los\ vehículos\ de\ Volkswagen\ supera\ a\ la\ de\ Toyota\ por\ 4$
$kilómetros\ el\ litro.$

9.10.54. Nueve sujetos fueron utilizados en un experimento para determinar si la exposición al monóxido de carbono tiene un efecto en la capacidad para respirar. La información fue recolectada por personal del Departamento de Salud y Educación Física de Virginia Tech, y fue analizada en el Centro de Consultas Estadísticas de Hokie Land. Los sujetos fueron expuestos a cámaras de respiración, en la que una contenía una alta concentración de CO. Se tomó la frecuencia de respirado para cada sujeto en cada cámara. Los sujetos fueron expuestos a las cámaras de respiración de manera aleatoria. La información da la frecuencia de respirado, en cantidad de respiros tomados por minuto. Haga un test de una cola de la hipótesis de que la media de frecuencia de respirado es la misma en los dos ambientes. Use *a* = 0.05. Asuma que la frecuencia de respirado es aproximadamente normal.

Subject	With CO	Without CO
1	30	30
2	45	40
3	26	25
4	25	23
5	34	30
6	51	49
7	46	41
8	32	35
9	30	28

El nivel de significancia que nos da el problema tiene un valor de 0.05.
Se utilizará la fórmula de la prueba T

$$t = \frac{\overline{d}-d_0}{s*\frac{1}{\sqrt{n}}}$$

$$\overline{d} = \frac{\Sigma d_i}{n}$$

$$\overline{d} = \frac{18}{9}$$

$$\overline{d} = 2$$

Ahora se calcula la desviación estándar de las diferencias:

$$s = \sqrt{\frac{\Sigma(d_i-\overline{d})^2}{n-1}}$$

$$s = \sqrt{\frac{(0-2)^2+(5-2)^2+(1-2)^2+(2-2)^2+(4-2)^2+(2-2)^2+(5-2)^2+(-3-2)^2+(2-2)^2}{8}}$$

$$s = 2.54$$

El valor de la prueba estadística es:

$t = \frac{2}{2.54 * \frac{1}{\sqrt{9}}}$

$t = 2.353$

El valor de los grados de libertad es:

$df = n - 1$

$df = 9 - 1$

$df = 8$

El valor P es:

$p = P(t > 2.35)$

$p = 0.023$

El valor p es menor que el valor de significancia, por lo tanto se rechaza la hipótesis nula, la cual aceptaba que la frecuencia de respiración era más alta con presencia de CO.

10.10.65. Una comunidad urbana quiere mostrar que la incidencia de cáncer de mama es mayor en su área que en una zona rural cercana. (Se detectó que los niveles de PCB eran más altos en el suelo del área urbana). Si se encuentra que 20 de 200 mujeres adultas en la comunidad urbana tienen cáncer de mama y que 10 de 150 mujeres adultas en la comunidad rural tienen cáncer de mama, se puede concluir con un nivel de significancia del 0.05 que el cáncer de mama es más predominante en la comunidad urbana?

Establecemos que la hipótesis nula simboliza que la probabilidad urbana es igual a la probabilidad rural, y H_1: *probabilidad urbana mayor que rural.*

Procedemos a calcular el valor de $\hat{p}$

$\hat{p} = \frac{mujeres\ con\ cáncer\ urbanas + mujeres\ con\ cáncer\ rurales}{cantidad\ total\ mujeres\ urbanas + cantidad\ total\ mujeres\ rurales}$

$\hat{p} = \frac{20 + 10}{200 + 150}$

$\hat{p} = 0.085$

Ahora se procede a calcular el valor z

$z = \frac{(20/200)-(10/150)}{\sqrt{(0.85)(1-0.85)(1/200+1/150)}}$

$z = \frac{0.0333}{0.0302}$

$z = 1.1$

Finalmente calculamos el valor p

$P = P(Z > 1.1)$

$P = 1 - P(Z < 1.1)$
$P = 1 - 0.8643$
$P = 0.1357$
Como este valor p es mayor que el nivel de significancia establecido, entonces no se puede rechazar la hipótesis nula, por lo tanto no se puede concluir que el cáncer es más predominante en la comunidad urbana.

11.10.73. Se realiza un estudio para comparar la cantidad de tiempo que le toma a los hombres y mujeres ensamblar un producto. Experiencia pasada indica que la distribución de los tiempos tanto para hombres como mujeres es aproximadamente normal pero la varianza de los tiempos para mujeres es menor que la de los hombres. Una muestra aleatoria de tiempos para 11 hombres y 14 mujeres resultó en la siguiente información:

Men	Women
$n_1 = 11$	$n_2 = 14$
$s_1 = 6.1$	$s_2 = 5.3$

Pruebe la hipótesis de que $\sigma_1^2 = \sigma_2^2$ contra una hipótesis alternativa de que $\sigma_1^2 > \sigma_2^2$. Use un P-value en su conclusión.

Nivel de significancia $= 0.05$
La hipótesis nula indica que $\sigma_1^2 = \sigma_2^2$, *por otro lado* $H_1: \sigma_1^2 > \sigma_2^2$
Calculamos el valor de $f = \frac{s_1^2}{s_2^2}$

$f = \frac{(6.1)^2}{(5.3)^2}$
$f = 1.33$
Ahora calculamos los grados de libertad:
$11 - 1 = 10$
$14 - 1 = 13$
Finalmente calculamos el $p - value$:
$p = P(F > 1.33)$
$p = 0.3095$
Nuestro valor p es mayor que 0.05, *por lo que no se puede rechazar* H_0 *y se puede decir que la variabilidad de tiempo para generar el producto no es mayor para el grupo de los hombres.*

11.10.78. Se sabe que la emisión de hidrocarburos de los carros se redujo drásticamente durante 1980. Se realizó un estudio para comparar la emisión de hidrocarburos de velocidad en ralentí, en partes por millón (ppm), para automóviles entre 1980 y 1990. Veinte carros de cada modelo del año fueron seleccionados aleatoriamente, y se anotó los niveles de emisión de hidrocarburos. La información se condensa de la siguiente manera:

```
1980 models:
141 359 247 940 882 494 306 210 105 880
200 223 188 940 241 190 300 435 241 380
1990 models:
140 160  20 20 223  60  20  95 360 70
220 400 217 58 235 380 200 175  85 65
```

Pruebe la hipótesis de que $\sigma^1 = \sigma^2$ contra una hipótesis alternativa de que $\sigma^1 \neq \sigma^2$. Asuma que ambas poblaciones son normales. Use el P-value.

$H_0: \sigma^1 = \sigma^2 \quad H_1: \sigma^1 \neq \sigma^2$

$Se\ utilizará\ la\ siguiente\ fórmula\ para\ calcular\ la\ varianza\ de\ los\ valores\ de\ ambos\ modelos:$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$El\ resultado\ de\ la\ varianza\ para\ los\ modelos\ de\ 1980:$

$S_1^2 = 78998.52$

$El\ resultado\ de\ la\ varianza\ para\ los\ modelos\ de\ 1990:$

$S_2^2 = 14255.08$

$A\ continuación,\ calculamos\ el\ valor\ f\ mediante\ la\ siguiente\ fórmula:$

$f = \frac{78998.52}{14255.08}$

$f = 5.54$

$Ahora,\ calculamos\ el\ valor\ de\ p$

$p = 2P(f > 5.54)$

$p = 2 * 0.0002$

$p = 0.0004$, 19 grados de libertad

$Obteniendo\ el\ valor\ anterior\ se\ puede\ decir\ que\ las\ emisiones\ de\ hidrocarburo$
$son\ más\ consistentes\ en\ los\ modelos\ del\ año\ 1990.$

12.10.86. En un experimento para estudiar la dependencia de la hipertensión a los hábitos de fumado, se tomó la siguiente información de 180 individuos.

	Non-smokers	Moderate Smokers	Heavy Smokers
Hypertension	21	36	30
No hypertension	48	26	19

Pruebe la hipótesis de que la presencia o ausencia de hipertensión es independiente a los hábitos de fumado. Use un nivel de significancia de 0.05.

Del enunciado se define la hipótesis nula como: la presencia o falta de hipertensión es independiente de si la persona fuma o no. H_1*: presencia o falta de hipertensión no es independiente de que la persona fume o no.*

El enunciado nos indica un nivel de significancia de 0.05 *y obtenemos* $(2-1)(3-1) = 2$ *grados de libertad.*

Mediante ambos valores anteriores, se tiene una región crítica $X^2 > 5.99$ *y* 2 *grados de libertad.*

Se debe calcular el valor esperado para cada valor de la tabla, lo cual se hará mediante: la fórmula:

$$e = \frac{valor\ total\ de\ columna * valor\ total\ de\ fila}{total\ de\ la\ tabla}$$

$$e = \frac{69 * 87}{180} = 33.35$$

$$e = \frac{69 * 93}{180} = 35.65$$

$$e = \frac{62 * 87}{180} = 29.97$$

$$e = \frac{62*93}{180} = 32.03$$

$$e = \frac{49 * 87}{180} = 23.68$$

$$e = \frac{49 * 93}{180} = 25.32$$

Ahora se procede a calcular el valor de la prueba X^2 *mediante la fórmula:*

$$X^2 = \sum_{i=1}^{n} \frac{(f_i - e_i)^2}{e_i}$$

$$X^2 = \frac{(21-33.5)^2}{33.35} + \frac{(48-35.65)^2}{35.65} + \frac{(36-29.97)^2}{29.97} + \frac{(26-32.03)^2}{32.03} + \frac{(30-23.68)^2}{23.68} + \frac{(19-25.32)^2}{25.32} = 14.79$$

Este valor sobrepasa el valor crítico estimado por lo que se rechaza H_0 *y se concluye que fumar y la hipertensión no son independientes.*

12.10.96. En un estudio para estimar la proporción de esposas que miran telenovelas de manera regular, se encuentra que 52 de 200 esposas en Denver, 31 de 150 esposas en Phoenix, y 37 de 150 esposas en Rochester miran al menos una telenovela. Utilice un nivel de significancia de 0.05 para probar la hipótesis de que no hay diferencia entre las verdaderas proporciones de esposas que miran telenovelas en estas tres ciudades.

H_0: *No hay diferencia entre las proporciones de esposas*

H_1: *Hay diferencia entre las proporciones de esposas*

El enunciado nos indica un nivel de significancia de 0.05 *y obtenemos* $(2-1)(3-1)=2$ *grados de libertad.*

Mediante ambos valores anteriores, se tiene una región crítica $X^2 > 5.99$ *y* 2 *grados de libertad.*

Denver:

Ven telenovelas: $52,\ \frac{200*120}{500} = 48$ *valor de frecuencia esperada*

No ven telenovelas: $148,\ \frac{200*380}{500} = 152$ *valor de frecuencia esperada*

Phoenix:

Ven telenovelas: $31,\ \frac{150*120}{500} = 36$ *valor de frecuencia esperada*

No ven telenovelas: $119,\ \frac{150*380}{500} = 114$ *valor de frecuencia esperada*

Rochester:

Ven telenovelas: $37,\ \frac{150*120}{500} = 36$ *valor de frecuencia esperada*

No ven telenovelas: $113,\ \frac{150*380}{500} = 114$ *valor de frecuencia esperada*

De manera homogénea a ejercicios anteriores se utilizará la siguiente fórmula de X^2:

$$X^2 = \sum_{i=1}^{n} \frac{(f_i - e_i)^2}{e_i}$$

$$X^2 = \frac{(52-48)^2}{48} + \frac{(31-36)^2}{36} + \frac{(37-36)^2}{36} + \frac{(148-152)^2}{152} + \frac{(119-114)^2}{114} + \frac{(113-114)^2}{114} = 1.39$$

El valor anterior es menor que la región crítica, por lo que no se rechaza H_0, *no hay diferencia entre las proporciones.*

Referencias:

Lenoir, J. (2013). *Critical Values of the F-Distribution (α = 0.025)*. tables-fisher.pdf. Recuperado de: https://jonathanlenoir.files.wordpress.com/2013/12/tables-fisher.pdf.