Soluciones a los ejercicios propuestos del Tema 4

4.1. Se trata de hacer el test para contrastar la hipótesis alternativa $H_1 : \mu > 50$, con σ desconocida y $\alpha = 0.05$ (entonces, $t_{1-\alpha}^{n-1} = t_{0.95}^{13} = 1.771$ es el valor que nos da la tabla de la t de Student con el que comparar). El estadístico es $t = \frac{\bar{x}-50}{s/\sqrt{n}} = 1.247216129$, ya que $\bar{x} = 53$, $n = 14$ y $s = 9$. Como no es mayor que el valor de la tabla, no podemos aceptar H_1. La respuesta es NO, no podemos.

4.2. Como $n = 140$ y $\hat{\hat{p}} = \frac{49}{140} = 0.35$, tenemos que $n\,\hat{\hat{p}}\,(1-\hat{\hat{p}}) = 31.85 \geq 18$ y se puede hacer el test asintótico para p. Para contrastar $H_1 : p < 0.4$, el estadístico es $\frac{\hat{\hat{p}}-0.4}{\sqrt{\frac{0.4\times 0.6}{140}}} = -1.207614729$; si $\alpha = 0.05$, $Z_\alpha = -1.645$ y como el estadístico no es menor, no se acepta H_1; si $\alpha = 0.10$, $Z_\alpha = -1.282$ y tampoco. Para contrastar $H_1 : p > 0.3$, el estadístico es $\frac{\hat{\hat{p}}-0.3}{\sqrt{\frac{0.3\times 0.7}{140}}} = 1.290994449$; si $\alpha = 0.05$, $Z_{1-\alpha} = 1.645$ y como el estadístico no es mayor, no se acepta H_1; si $\alpha = 0.10$, $Z_{1-\alpha} = 1.282$ y ahora el estadístico sí que es mayor, y se acepta H_1.

4.3. Sí, podemos decir que difiere, ya que si planteamos el contraste de la hipótesis alternativa $H_1 : \mu \neq 111$, la aceptaremos, puesto que el estadístico es $t = \frac{|\bar{x}-111|}{s/\sqrt{n}} = 3.375 > t_{0.975}^{4} = 2.776$ (la muestra es de tamaño $n = 5$, la media es $\bar{x} = 116.4$ y la desviación $s = 3.577708764$).

4.4. A partir de la muestra de tamaño $n = 10$, de media $\bar{x} = 896.1$ y desviación $s = 3.842742077$, con $\alpha = 0.01$, hemos de contrastar la hipótesis alternativa $H_1 : \mu < 900$ (que corresponde a que la empresa engaña a los consumidores). Primero, si $\sigma = 4$, el estadístico es $z = \frac{\bar{x}-900}{\sigma/\sqrt{n}} = -3.083220719 < Z_{0.01} = -2.326$, luego se acepta H_1. Si ahora σ es deconocida, el estadístico es $t = \frac{\bar{x}-900}{s/\sqrt{n}} = -3.209396476 < t_{0.01}^{9} = -2.821$, y también se acepta H_1.

4.5.(**a**) Estadístico para contrastar $H_1 : \mu < 3.20$: $t = \frac{\bar{x}-3.20}{s/\sqrt{n}} = -3.119588741 < t_{0.05}^{49}$, que está entre -1.684 y -1.671 . Sí se puede decir que el grosor medio es inferior al deseado.

4.5.(**b**) Estadístico para contrastar $H_1 : \sigma > 0.25$: $\frac{(n-1)\,s^2}{0.25^2} = \frac{49\times 0.34^2}{0.25^2} = 90.6304 > \chi^{2,49}_{0.99}$, que está entre 63.69 y 76.15 . Sí se puede decir que las lentes son de baja calidad.

4.6. Estadístico para contrastar $H_1 : \mu > 70$ (que corresponde a tener que prohibir la pesca): $t = \frac{\bar{x}-70}{s/\sqrt{n}} = 2.132007162$, a partir de la muestra de tamaño $n = 9$, media $\bar{x} = 71.\overset{\frown}{6}$ y desviación $s = 2.345207883$. Si $\alpha = 0.05$, $t^8_{0.95} = 1.860$, como el estadítico es mayor, sí que se debería prohibir la pesca. Con $\alpha = 0.01$, $t^8_{0.99} = 2.896$ y el estadístico no es mayor que este valor, así que no debería prohibirse la pesca.

4.7. Hemos de contrastar la hipótesis $H_1 : \sigma > 3$, con $\mu = 100$ conocida (es una situación muy poco común en la práctica). Aunque **no** usaremos el valor para nada más, hemos de calcular la desviación de la muestra, s, para ver si es > 3 (pues en caso contrario no tendría sentido plantear el test). Efectivamente, $s = 5.20718552 > 3$. El estadístico del test es $\frac{\sum_{i=1}^{10} (x_i-100)^2}{3^2} = \frac{\sum_{i=1}^{10} x_i^2 + 10\times 100^2 - 2\times 100\times \sum_{i=1}^{10} x_i}{9} = \frac{96216.33+100000-200\times 979.5}{9} = \frac{316.33}{9} = 35.14\overset{\frown}{7} > \chi^{2,n}_{1-\alpha} = \chi^{2,10}_{0.95} = 18.31$, así que sí se puede decir que la desviación es mayor que 3.

4.8. Estadístico para contrastar $H_1 : \mu < 15$: $t = \frac{\bar{x}-15}{s/\sqrt{n}} = -2.573390829 < t^{19}_{0.01} = -2.539$, luego con $\alpha = 0.01$ sí que se puede decir que es inferior (aunque por poco). Entonces, con $\alpha = 0.05$ lo podremos decir con mayor motivo (no es necesario, pero si buscamos en la tabla vemos que $t^{19}_{0.05} = -1.729$, y el estadístico también es inferior a este valor, de manera más holgada).

4.9. Para contrastar $H_1 : \sigma > 14$, usamos el estadístico $\frac{(n-1)\,s^2}{14^2} = \frac{46\times 23^2}{14^2} = 124.1530612 > \chi^{2,46}_{0.95}$, que está entre 55.76 y 67.50, luego sí que se podemos decir que la desviación es superior a 14 mg. Para contrastar $H_1 : \mu > 200$, usamos el estadístico $t = \frac{\bar{x}-200}{s/\sqrt{n}} = 4.471079087 > t^{46}_{0.95}$, que está entre 1.671 y 1.684, así que también podemos decir que la media es superior a 200 mg.

4.10. En primer lugar contrastamos $H_1 : p \neq 0.75$ a partir de una muestra de tamaño $n = 250$, usando como estimación de p, $\hat{p} = \frac{212}{250} = 0.848$ (podemos, pues $250 \times 0.75 \times 0.25 = 46.875 \geq 18$), con el estadístico $\frac{|\hat{p}-0.75|}{\sqrt{0.75\times 0.25/250}} = 3.578454042 > Z_{0.975} = 1.96$. Por tanto, sí que podemos decir que la proporción es diferente (de hecho, podemos incluso decir que $p > 0.75$). Ahora buscamos I.C. para $p - 0.75$, que es el I.C. para p, restando a ambos extremos 0.75 : $(\hat{\hat{p}} - 0.75) \pm 1.96\sqrt{\frac{\hat{\hat{p}}\,(1-\hat{\hat{p}})}{n}} = (0.848 - 0.75) \pm 0.04450469071 = (0.0534953093, 0.1425046907)$. Por tanto, con una confianza aprox. de 0.95 podemos decir que $p > 0.75$ (pues el intervalo es positivo), y también que $p - 0.75 < 0.15$ (pues no llega a 0.15).

4.11. No podemos decir que sea menor el contenido de calcio de los hijos de madres fumadoras, ya que si planteamos la hipótesis $H_1 : \mu_1 < \mu_2$, el estadístico de contraste es $\frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{1/n_1+1/n_2}}$, siendo

$$s_p^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2} = \frac{76 \times 0.026^2 + 160 \times 0.02^2}{236} = 0.000488813559$$

la variancia ponderada y $s_p = 0.02211066159$ su raíz cuadrada. Entonces, el estadístico queda $\frac{-0.003}{s_p\sqrt{1/77+1/161}} = -0.9792404373$ que no es menor que $t_{0.05}^{236} \cong Z_{0.05} = -1.645$.

4.12.(**a**) No hay evidencia de diferencias puesto que si planteamos el test para $H_1 : \mu_1 \neq \mu_2\, (\mu \neq 0)$ para las dos muestras con datos emparejados, a partir de la muestra de la diferencia obtenida restando (método A $-$ método B), que es

$$5,\ -5,\ 0,\ -2,\ -5,\ 6,\ 2,\ 0\,,$$

obtenemos el estadístico $\frac{|\bar{d}|}{s_d/\sqrt{n}} = \frac{0.125}{4.12093956/\sqrt{8}} = 0.08579436448$ que no es mayor que $t_{0.975}^{7} = 2.365$.

4.12.(**b**) Buscamos entonces el I.C. para $\mu = \mu_1 = \mu_2$ a partir de la muestra formada por los $n = 16$ datos (σ desconocida), cuya media es $\bar{x} = 20.4375$ y desviación $s = 2.220172666$: $\bar{x} \pm t_{0.975}^{15} \times \frac{s}{\sqrt{16}} = 20.4375 \pm 2.131 \times \frac{s}{4} = 20.4375 \pm 1.182796988 = (19.25470301, 21.62029699)$.

4.13. Es un test de comparación de medias con datos emparejados para $H_1 : \mu_1 - \mu_2 = \mu > 0$. Restando los valores de las parejas (antes $-$ después), otenemos una muestra de tamaño $n = 10$ para la diferencia:

$$9,\ -7,\ 22,\ 32,\ 22,\ 21,\ 15,\ 28,\ 13,\ 16$$

cuya media es $\bar{d} = 17.1$ y desviación $s_d = 10.91838409$. El estadístico de contraste es $\frac{\bar{d}}{s_d/\sqrt{n}} = 4.952651194 > t_{0.99}^{9} = 2.821$. Por tanto, sí podemos decir que es significativo el descenso.

4.14. Sí hay suficiente evidencia, porque considerando la hipótesis alternativa $H_1 : \mu_1 < \mu_2$, como tenemos muestras independientes y las variancias poblacionales son desconocidas pero iguales, el estadístico es

$$\frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{1/n_1 + 1/n_2}} = \frac{59.1 - 65.2}{\sqrt{3.805}\sqrt{2/40}} = -13.98516203 < t^{78}_{0.05},$$

que está entre -1.671 y -1.658 (hemos calculado s_p como la raíz cuadrada de s_p^2, que es la media ponderada, aritmética en este caso en que los tamaños muestrales son iguales, de las variancias de las dos muestras, $s_p^2 = \frac{1.9^2+2.0^2}{2} = 3.805$).

4.15. No presentan evidencia, pues si se considera la hipótesis alternativa $H_1 : \sigma_1^2 \neq \sigma_2^2$, el estadístico de contraste es $\frac{s_1^2}{s_2^2} = 2.904255319$, que se encuentra entre $F^{9,9}_{0.025} = 1/4.03 = 0.2481389578$ y $F^{9,9}_{0.975} = 4.03$.

4.16. Sí podemos decirlo, porque tenemos dos muestras independientes del mismo tamaño, $n_1 = n_2 = 17$, de medias y desviaciones respectivas $\bar{x}_1 = 1.758823529$, $s_1 = 0.1460257834$, $\bar{x}_2 = 1.423529412$ y $s_2 = 0.1562426469$, y planteamos el test para la hipótesis alternativa $H_1 : \mu_1 > \mu_2$, que aceptamos puesto que el estadístico es $\frac{\bar{x}_1-\bar{x}_2}{s_p\sqrt{1/n_1+1/n_2}} = 6.464346939 > t^{32}_{0.95}$, que está entre 1.684 y 1.697 (s_p es la raíz de s_p^2, la media ponderada de las variancias, $s_p^2 = \frac{s_1^2+s_2^2}{2} = 0.2286764707$).

4.17. Planteamos el test para la hiptótesis alternativa $H_1 : \mu \neq 548$, con σ desconocida. El estadístico del test es $t = \frac{|\bar{x}-548|}{s/\sqrt{n}} = \frac{39}{10/\sqrt{11}} = 12.93483668 > t^{10}_{0.975}$ ($= 2.228$), luego sí que hay evidencia significativa de que difiere. De hecho, la evidencia es en el sentido de que es mayor que 548 (esto es, que las lecturas con el nuevo método tienen sesgo positivo), puesto que $\bar{x} > 548$.

4.18. No, puesto que si se considera la hipótesis alternativa $H_1 : \mu > 29$, ésta no será aceptada. En efecto, el estadístico del test es $t = \frac{\bar{x}-29}{s/\sqrt{n}} = 0.7742408478$ (con $n = 8$, $\bar{x} = 30.7875$ y $s = 6.530026799$), que no es mayor que $t^7_{0.99} = 2.998$.

4.19. No podemos decirlo, porque no podemos aceptar la hipótesis nula $H_1 : \mu \neq 1.75$, ya que el estadístico es $t = \frac{|\bar{x}-1.75|}{s/\sqrt{n}} = \frac{0.14}{0.42/\sqrt{26}} = 1.699673171$, que no es mayor que $t^{25}_{0.975} = 2.060$.

4.20. No podemos estar de acuerdo con ellos, puesto que aceptamos $H_1 : \mu_1 \neq \mu_2$, ya que el estadístico de contraste es $\frac{|\bar{x}_1-\bar{x}_2|}{\sqrt{s_1^2/n_1+s_2^2/n_2}} = \frac{50}{\sqrt{241}} = 3.220783132 > t^{\nu=16}_{0.975} = 2.120$ (hemos realizado un test de comparación de medias con muestras independientes y variancias desconocidas, y hemos obtenido el número de grados de libertad ν como el natural más próximo al valor ser obtiene al aplicar la fórmula, que es $\frac{1}{0.06422452169} = 15.57037676$).

4.21. Sí, ya que el valor del estadístico es

$$\frac{|\bar{x}_1 - \bar{x}_2|}{s_p\sqrt{1/n_1 + 1/n_2}} = \frac{0.68}{s_p\sqrt{1/8 + 1/12}} = 3.551072092 > t^{18}_{0.995}\,(= 2.878)\,,$$

con s_p la raíz cuadrada de $s_p^2 = \frac{7\times 0.51^2 + 11\times 0.35^2}{18} = 0.1760\overset{\frown}{1}$.
4.22. No podemos, porque el estadístico que se utiliza para contrastar $H_1 : \mu_1 \neq \mu_2$ cuando las muestras son independientes y las variancias desconocidas pero iguales es $\frac{|\bar{x}_1 - \bar{x}_2|}{s_p\sqrt{1/n_1+1/n_2}} = 1.28194405$, que no es mayor que $t^{12}_{0.975} = 2.179$ (hemos usado que las dos muestras tienen tamaño 7, que las medias y desviaciones respectivas son $\bar{x}_1 = 772.5714286$, $\bar{x}_2 = 780.8571429$, $s_1 = 13.56290462$ y $s_2 = 10.41519043$, y que la variancia ponderada es $s_p^2 = \frac{s_1^2+s_2^2}{2} = 146.2142867$ y s_p su raíz cuadrada).
4.23. Sí podemos decir que la variabilidad es mayor para 75 minutos que para 30, porque aceptamos la hipótesis alternativa $H_1 : \sigma_1^2 < \sigma_2^2$, ya que el estadístico es $\frac{s_1^2}{s_2^2} = \frac{0.6964285714}{77.92857143} = 0.008936755271 < F^{7,7}_{0.01} = \frac{1}{F^{7,7}_{0.99}} = \frac{1}{6.99} = 0.14306$.
4.24. No podemos decir que difieran significativamente, porque no aceptamos la hipótesis alternativa $H_1 : \mu_1 \neq \mu_2$. El estadístico del test es $\frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{s_1^2/n_1+s_2^2/n_2}} = 0.53705428$, que no es mayor que $t^{\nu=4}_{0.975} = 2.776$. El valor de ν se obtiene como el natural más próximo al valor que nos da la fórmula, que es $\frac{1}{0.2523972211} = 3.962008756$.
4.25. Sí se puede decir que se ha incrementado, porque se acepta $H_1 : p_1 < p_2$, a partir de los datos: primera muestra de tamaño $n_1 = 759$, con estimación de p_1 que es $\hat{\hat{p}}_1 = \frac{46}{759} = 0.\overset{\frown}{06}$, segunda muestra de tamaño $n_2 = 838$ y estimación de p_2, $\hat{\hat{p}}_2 = \frac{109}{838} = 0.130071599$, y media ponderada de las estimaciones, $\bar{\bar{p}} = \frac{n_1\hat{\hat{p}}_1 + n_2\hat{\hat{p}}_2}{n_1+n_2} = 0.09705698184$. El estadístico del test (asintótico) es $\frac{\hat{\hat{p}}_1 - \hat{\hat{p}}_2}{\sqrt{\bar{\bar{p}}\,(1-\bar{\bar{p}})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = -4.682915081 < Z_{0.05} = -1.645$ (se puede hacer porque $n_1\hat{\hat{p}}_1\,(1-\hat{\hat{p}}_1) = 43.\overset{\frown}{21} \geq 18$ y $n_2\hat{\hat{p}}_2\,(1-\hat{\hat{p}}_2) = 94.8221957 \geq 18$).
4.26. No podemos decirlo porque no aceptamos $H_1 : \sigma_1^2 > \sigma_2^2$ (la población 1 corresponde a la especie A y la población 2 a la B). Las variancias muestrales de las dos muestras de tamaños $n_1 = n_2 = 12$ son $s_1^2 = 0.8824242437$ y $s_2^2 = 0.53787879$. El estadístico del test es $\frac{s_1^2}{s_2^2} = 1.640563376$, que no es mayor

que $F_{0.95}^{11,11}$ (valor que se encuentra entre 2.79 y 2.85).
4.27. Tenemos dos muestras independientes con $n_1 = n_2 = 150$. Si la población 1 corresponde al tratamiento tradicional y la 2 al nuevo, tenemos que $\hat{\hat{p}}_1 = \frac{102}{150} = 0.68$ y $\hat{\hat{p}}_2 = \frac{120}{150} = 0.8$. Para contrastar $H_1 : p_1 < p_2$ usamos el estadístico $\frac{\hat{\hat{p}}_1 - \hat{\hat{p}}_2}{\sqrt{\bar{\bar{p}}\,(1-\bar{\bar{p}})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = -2.369241569 < Z_{0.05} = -1.645$, usando que $\bar{\bar{p}} = \frac{n_1 \hat{\hat{p}}_1 + n_2 \hat{\hat{p}}_2}{n_1+n_2} = \frac{\hat{\hat{p}}_1 + \hat{\hat{p}}_2}{2} = 0.74$. Por tanto, con el 5 % sí podemos decir que el nuevo tratamiento es mejor ($\Rightarrow$ con el 10 % también). También lo podemos decir con el 1 % porque el estadístico también es $< Z_{0.01} = -2.326$, aunque "por poco".

Se puede llevar a cabo el test asintótico porque $n_1 \hat{\hat{p}}_1 (1 - \hat{\hat{p}}_1) = 32.64 \geq 18$ y $n_2 \hat{\hat{p}}_2 (1 - \hat{\hat{p}}_2) = 24 \geq 18$.
4.28.(**a**) Se trata de hacer un test de comparación de medias con muestras independientes, variancias desconocidas, para $H_1 : \mu_1 > \mu_2$ (la población 1 corresponde al fármaco A y la 2 al B). A partir de las observaciones podemos calcuar $n_1 = 9$, $\bar{x}_1 = 22.5\overset{\frown}{3}$, $s_1^2 = 0.1975000012$, $n_2 = 12$, $\bar{x}_2 = 20.791\overset{\frown}{6}$ y $s_2^2 = 0.8771969718$. El estadístico del test es $\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} = 5.649398992 > t_{0.95}^{\nu=17}$, donde el valor de ν se obtiene como el natural más próximo al que nos da la fórmula, que es $\frac{1}{0.060043951296} = 16.54546754$ (también se puede tomar $\nu = 16$ para que el test sea algo más conservador, pero en todo caso hay poca diferencia y, además, el nivel de significación es aproximado porque la distribución t del estadístico tampoco es exacta). Por tanto, sí se puede decir que el fármaco A es mejor.
4.28.(**b**) Ahora tenemos los datos emparejados. Restando (fármaco A $-$ fármaco B) tenemos una muestra de tamaño $n = 10$ de la diferencia, que es

$$0.9,\ 0.9,\ 0.3,\ 0.6,\ 1.1,\ 1.4,\ 0.3,\ 0.6,\ -0.4,\ 0.3$$

La media y la desviación de esta muestra son $\bar{d} = 0.6$ y $s_d = 0.5099019514$, y el estadístico para contrastar $H_1 : \mu_1 > \mu_2\,(\mu > 0)$ es $\frac{\bar{d}}{s_d/\sqrt{n}} = 3.721042038 > t_{0.95}^{9} = 1.833$. Así que ahora también podemos decir que el fármaco A es mejor.
4.29.(**a**) Tenemos que $n = 8$, $\bar{x} = 50.2125$ y $s = 0.8007808689$. I.C. para σ:

$$\left(\frac{\sqrt{7}\times s}{\sqrt{16.01}}, \frac{\sqrt{7}\times s}{\sqrt{1.69}}\right) = (0.5295013151,\ 1.629743872)\,.$$

I. C. para μ:

$$\bar{x} \pm 2.365 \times \frac{s}{\sqrt{n}} = 50.2125 \pm 0.6695759415 = (49.54292406,\ 50.88207594)\,.$$

4.29.(**b**) Para contrastar $H_1 : \sigma < 1.5$ (el método nuevo es más preciso) se usa el estadístico $\frac{(n-1)\,s^2}{1.5^2} = 1.995 < \chi^{2,7}_{0.05} = 2.17$. Luego se acepta H_1, es decir, que el nuevo método es más preciso.
4.29.(**c**) Se plantea la hipótesis alternativa $H_1 : \mu > 50$, que no se acepta puesto que el estadístico $\frac{\bar{x}-50}{s/\sqrt{n}} = 0.7505683357$ no es mayor que $t^{7}_{0.95} = 1.895$. Por tanto **no** podemos decir que las lecturas de nuevo método presenten sesgo positivo significativo.
4.30.(**a**) Sí, puesto que aceptamos la hipótesis alternativa $H_1 : \mu_1 < \mu_2$ (la población 1 corresponde a la marca A y la 2 a la B), ya que el estadístico es $\frac{\bar{x}_1-\bar{x}_2}{\sqrt{s_1^2/n_1+s_2^2/n_2}} = \frac{-4}{\sqrt{4/50+3.24/50}} = -10.51176662 < t^{\nu=97}_{0.01}$, que está entre $t^{60}_{0.01} = -2.390$ y $t^{120}_{0.01} = -2.358$. El valor de ν, los grados de libertad, se obtiene como el natural más próximo al que da la fórmula, que es $\frac{1}{0.01031652245} = 96.93188813$.
4.30.(**b**) I. C. para $\mu_1 - \mu_2$ con $\gamma = 0.95$:

$$(\bar{x}_1 - \bar{x}_2) \pm t^{\nu=97}_{0.975} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = -4 \pm t^{\nu=97}_{0.975} \times 0.3805259518\,,$$

y este intervalo está incluido en

$$-4 \pm 2 \times 0.3805259518 = (-4.761051904,\ -3.238948096)\,,$$

que corresponde a tomar $t^{60}_{0.975} = 2.000$ como cota superior par $t^{97}_{0.975}$.
4.30.(**c**) Planteamos la hipótesis alternativa $H_1 : \sigma_1^2 > \sigma_2^2$, que no se puede aceptar con $\alpha = 0.01$ puesto que el estadístico es $\frac{s_1^2}{s_2^2} = 1.234567901$ no es mayor que $F^{49,49}_{0.99}$ (cuyo valor que está entre $F^{60,60}_{0.99} = 1.84$ y $F^{40,40}_{0.99} = 2.11$).