1. (a) For full siblings ($k_0 = 1/4$, $k_1 = 1/2$, and $k_2 = 1/4$) we have expected value

$$\mathrm{E}\left(\bar{\theta}^*\right) = \mathrm{E}\left(\frac{1}{2000}\sum_{m=1}^{2000}\theta_m^*\right) = \frac{1}{2000}\sum_{m=1}^{2000}\mathrm{E}\left(\theta_m^*\right) = \frac{1}{2000}\sum_{m=1}^{2000}\left(\frac{1}{2}k_2 + \frac{1}{4}k_1\right)$$

$$= \frac{1}{2}k_2 + \frac{1}{4}k_1 = \frac{1}{2}\cdot\frac{1}{4} + \frac{1}{4}\cdot\frac{1}{2} = \frac{1}{8} + \frac{1}{8} = \frac{1}{4},$$

and variance

$$\mathrm{Var}\left(\bar{\theta}^*\right) = \mathrm{Var}\left(\frac{1}{2000}\sum_{m=1}^{2000}\theta_m^*\right) = \left(\frac{1}{2000}\right)^2\sum_{m=1}^{2000}\mathrm{Var}\left(\theta_m^*\right)$$

$$= \frac{1}{2000^2}\sum_{m=1}^{2000}\left(\frac{1}{4}k_2(1-k_2) + \frac{1}{16}k_1(1-k_1) - \frac{1}{4}k_2k_1\right)$$

$$= \frac{1}{2000}\left(\frac{1}{4}k_2(1-k_2) + \frac{1}{16}k_1(1-k_1) - \frac{1}{4}k_2k_1\right)$$

$$= \frac{1}{2000}\left(\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{3}{4} + \frac{1}{16}\cdot\frac{1}{2}\cdot\frac{1}{2} - \frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{2}\right)$$

$$= \frac{1}{2000}\left(\frac{3}{64} + \frac{1}{64} - \frac{2}{64}\right) = \frac{1}{2000}\left(\frac{2}{64}\right) = \frac{1}{64000} \approx 1.563\times 10^{-5},$$

which yields the following 95% confidence interval for $\bar{\theta}^*$:

$$\mathrm{E}\left(\bar{\theta}^*\right) \pm z_{0.975}\cdot\sqrt{\mathrm{Var}\left(\bar{\theta}^*\right)} = \frac{1}{4} \pm z_{0.975}\cdot\sqrt{\frac{1}{64000}}$$

$$\approx 0.2500 \pm 1.960\cdot 0.003953 = (0.2423, 0.2577).$$

(b) For first cousins ($k_0 = 3/4$, $k_1 = 1/4$, and $k_2 = 0$) we have expected value

$$\mathrm{E}\left(\bar{\theta}^*\right) = \frac{1}{2}k_2 + \frac{1}{4}k_1 = \frac{1}{2}\cdot 0 + \frac{1}{4}\cdot\frac{1}{4} = \frac{1}{16},$$

and variance

$$\mathrm{Var}\left(\bar{\theta}^*\right) = \frac{1}{2000}\left(\frac{1}{4}k_2(1-k_2) + \frac{1}{16}k_1(1-k_1) - \frac{1}{4}k_2k_1\right)$$

$$= \frac{1}{2000}\left(\frac{1}{4}\cdot 0\cdot 0 + \frac{1}{16}\cdot\frac{1}{4}\cdot\frac{3}{4} - \frac{1}{4}\cdot 0\cdot\frac{1}{4}\right)$$

$$= \frac{1}{2000}\left(\frac{3}{256}\right) = \frac{3}{512000} \approx 5.859\times 10^{-6},$$

which yields the following 95% confidence interval for $\bar{\theta}^*$:

$$\mathrm{E}\left(\bar{\theta}^*\right) \pm z_{0.975}\cdot\sqrt{\mathrm{Var}\left(\bar{\theta}^*\right)} = \frac{1}{16} \pm z_{0.975}\cdot\sqrt{\frac{3}{512000}}$$

$$\approx 0.0625 \pm 1.960\cdot 0.002421 = (0.05776, 0.06724).$$

For double first cousins ($k_0 = 9/16$, $k_1 = 3/8$, and $k_2 = 1/16$) we have expected value

$$\mathrm{E}\left(\bar{\theta}^*\right) = \frac{1}{2}k_2 + \frac{1}{4}k_1 = \frac{1}{2}\cdot\frac{1}{16} + \frac{1}{4}\cdot\frac{3}{8} = \frac{1}{32} + \frac{3}{32} = \frac{1}{8},$$

and variance

$$\begin{aligned}
\mathrm{Var}\left(\bar{\theta}^*\right) &= \frac{1}{2000}\left(\frac{1}{4}k_2(1-k_2) + \frac{1}{16}k_1(1-k_1) - \frac{1}{4}k_2k_1\right) \\
&= \frac{1}{2000}\left(\frac{1}{4}\cdot\frac{1}{16}\cdot\frac{15}{16} + \frac{1}{16}\cdot\frac{3}{8}\cdot\frac{5}{8} - \frac{1}{4}\cdot\frac{1}{16}\cdot\frac{3}{8}\right) \\
&= \frac{1}{2000}\left(\frac{15}{1024} + \frac{15}{1024} - \frac{6}{1024}\right) = \frac{1}{2000}\left(\frac{24}{1024}\right) = \frac{3}{256000} \approx 1.172 \times 10^{-5},
\end{aligned}$$

which yields the following 95% confidence interval for $\bar{\theta}^*$:

$$\begin{aligned}
\mathrm{E}\left(\bar{\theta}^*\right) \pm z_{0.975}\cdot\sqrt{\mathrm{Var}\left(\bar{\theta}^*\right)} &= \frac{1}{8} \pm z_{0.975}\cdot\sqrt{\frac{3}{256000}} \\
&\approx 0.1250 \pm 1.960 \cdot 0.003423 = (0.1183, 0.1317).
\end{aligned}$$

2. Note: All of the following numbers were computed using `PEDSTATS`.

(a) The estimate of correlation in BMI for full siblings is $\hat{\rho}_{FS} = 0.2720$, which leads to the following estimate for heritability:

$$\hat{h}^2_{FS} = \frac{\hat{\rho}_{FS}}{2\theta_{FS}} = \frac{0.2720}{2\cdot\frac{1}{4}} = 2\cdot 0.2720 = 0.5440.$$

This estimate assumes that we have both negligible dominance and negligible gene-environmental interaction for BMI.

(b) In these data, BMI correlation can be calculated using only 126 half-sibling pairs compared to 19,528 full sibling pairs and 14,032 parent-offspring pairs. Because of this, our estimator $\hat{\rho}_{HS}$ will be quite variable. Furthermore, since the kinship coefficient for half-sibling pairs is half that of full sibling and parent-offspring pairs, the estimator $\hat{h}^2_{HS}$ will have four times the variance as the estimators resulting from those closer relatives. Therefore, the heritability estimate from half-sibling pairs for these data is not very reasonable compared to those from full sibling and parent-offspring pairs.

(c) The estimate of correlation in BMI for full siblings is $\hat{\rho}_{PO} = 0.2092$, which leads to the following estimate for heritability:

$$\hat{h}^2_{PO} = \frac{\hat{\rho}_{PO}}{2\theta_{PO}} = \frac{0.2092}{2\cdot\frac{1}{4}} = 2\cdot 0.2092 = 0.4184.$$

(d) Stratifying the data by sex results in the following estimates for correlation and heritability for full sibling and parent-offspring pairs:

|          | Full Sibling |        | Parent-Offspring |        |
|----------|:------------:|:------:|:----------------:|:------:|
|          | $\hat{\rho}$ | $\hat{h}^2$ | $\hat{\rho}$ | $\hat{h}^2$ |
| Female   | 0.2898       | 0.5796 | 0.2415           | 0.4830 |
| Male     | 0.2767       | 0.5534 | 0.1941           | 0.3882 |

For the full sibling pairs, the differences between the estimates are quite small. For the parent-offspring pairs, these differences are a bit larger but still rather small. There does not appear to be strong evidence of a sex effect with BMI in these data. If we knew the variances of these estimates, we could perform a formal hypothesis test. However, the dependencies inherent in the pedigree data make this difficult.

3. Note: All of the following numbers were computed using R.

   (a) The estimate of correlation in BMI for MZ twins is $\hat{\rho}_{MZ} = 0.6986$, which leads to the following estimate for heritability:

   $$\hat{h}^2_{MZ} = \frac{\hat{\rho}_{MZ}}{2\theta_{MZ}} = \frac{0.6986}{2 \cdot \frac{1}{2}} = 0.6986.$$

   (b) The estimate of correlation in BMI for DZ twins is $\hat{\rho}_{DZ} = 0.3517$, which leads to the following estimate for heritability:

   $$\hat{h}^2_{DZ} = \frac{\hat{\rho}_{DZ}}{2\theta_{DZ}} = \frac{0.3517}{2 \cdot \frac{1}{4}} = 2 \cdot 0.3517 = 0.7034.$$

   (c) Using both MZ and DZ correlations, we can estimate heritability by

   $$\hat{h}^2_{Both} = 2\left(\hat{\rho}_{MZ} - \hat{\rho}_{DZ}\right) = 2\left(0.6986 - 0.3517\right) = 2 \cdot 0.3469 = 0.6939.$$

   (d) Using both MZ and DZ correlations, we can also estimate the shared environmental variance by
   $$\hat{\sigma}^2_C = 2\hat{\rho}_{DZ} - \hat{\rho}_{MZ} = 2 \cdot 0.3517 - 0.6986 = 0.004759.$$

   Based on this estimate, less than half of 1% of the variance of BMI is due to shared environment. Hence, there is not very strong evidence of a shared environmental effect in these data.

   (e) Stratifying the data by sex results in the following estimates for correlation, heritability, and shared environmental variance for MZ and DZ twins:

3

|  | MZ Twins | | DZ Twins | | Both | |
| --- | --- | --- | --- | --- | --- | --- |
|  | $\hat{\rho}$ | $\hat{h}^2$ | $\hat{\rho}$ | $\hat{h}^2$ | $\hat{h}^2$ | $\hat{\sigma}_C^2$ |
| Female | 0.7111 | 0.7111 | 0.3541 | 0.7082 | 0.7141 | -0.00294 |
| Male | 0.6406 | 0.6406 | 0.2982 | 0.5964 | 0.6847 | -0.04417 |

For each set of twins, the differences between the correlation estimates are larger than we saw in Problem 2, indicating a possible sex effect with BMI. However, when both sets of twins are used to estimate heritability, the difference between the estimates is quite small. There may be some evidence that females have higher heritability than males, but as with Problem 2 it is hard to say whether this sex effect is significant. (Note: The negative estimates for shared environmental variance indicate that there is no evidence of a shared environmental effect.)

4. First, recall that the family study from Problem 2 resulted in BMI correlation estimates of 0.2720 for full sibling pairs and 0.2092 for parent-offspring pairs with little evidence of a sex effect. These estimates are relatively consistent with other BMI family studies summarized in Table II of the article. For example, the two large Canada studies of Bouchard et al. (1987) and Perusse et al. (1988) give full sibling correlation estimates of 0.26 and 0.31 and parent-offspring correlation estimates of 0.23 and 0.20, respectively. The Norway study of Tambs et al. (1991) also resulted in similar numbers, sibling correlations that ranged from 0.21 to 0.26 by sex and parent-offspring correlations that ranged from 0.17 to 0.21 by sex. This study also agrees with my conclusion that there isn't much evidence of a sex effect in the family study data.

Next, recall that the twin study from Problem 3 resulted in BMI correlation estimates of 0.6986 for MZ twins and 0.3517 for DZ twins with a combined heritability estimate of 0.6939 and some possible evidence of higher heritability among females. As with the family study, these estimates (and their corresponding heritability estimates) are relatively consistent with other BMI twin studies summarized in Table I of the article. For example, the large and highly published NAS-NRC sample studied by Feinlieb et al. (1977), Fabsitz et al. (1980, 1992), Stunkard et al. (1986), and Selby et al. (1991) results in heritability estimates between 0.64 and 0.82 at the time of induction into the sample. Another large sample from the Finnish Twin Registry used by Turula et al. (1990), Korkeila et al. (1991), and Korkeila et al. (1995) gives heritability estimates between 0.61 and 0.72.

Many of the these BMI studies also stratify by age, which results in correlation (and heritability) estimates that decrease with increasing age. This makes sense because as people get older, environmental effects will begin to dominate genetic effects. When stratifying by sex, many of the twin studies appear to show higher heritability estimates for males, which is the opposite of the results from Problem 3. However, the Danish twin study of Herskind et al. (1996) shows higher heritability estimates for females, but this sample contains twins of age 46-76, which is older than most of the other studies. We were not provided with the ages of the subjects for our twin study, so it's difficult to know how to make a proper comparison. This shows the importance of stratifying by or controlling for covariates like sex and age.