7. ANSWERS TO PROBLEMS

1. Song Sparrows

a. The two distributions look approximately normal and are similar in shape. Beak length distribution of 'died' birds is slightly right-skewed and has a lower mean.

b. Ho: Beak length of 'died' birds follows a normal distribution
Ha: Beak length of 'died' birds does not follow a normal distribution
do not reject Ho (Shapiro-Wilks W=0.9758, n=74, P=0.4323).

   Ho: Beak length of 'survived' birds follows a normal distribution
Ha: Beak length of 'survived' birds does not follow a normal distribution
do not reject Ho (Shapiro-Wilks W=0.9812, n=71, P=0.6845).

   This does not mean the distributions are truly normal, but we have no evidence to the contrary.

c. Ho: $\mu_{died} - \mu_{survived} = 0$
Ha: $\mu_{died} - \mu_{survived} \neq 0$
t=-0.2812, ν=143, P=0.0056 Reject Ho, conclude that there is a difference in mean beak length between birds that survived and those that died.

d. The assumption of equal variance must be met.
Ho: $\sigma_1^2 = \sigma_2^2$
Ha: $\sigma_1^2 \neq \sigma_2^2$
The Bartlett test is equivalent to the F-ratio test of equal variances for two samples, and application of it fails to reject Ho (Bartlett's F=0.4947, P=0.4818, df=1 [Bartlett's F has a chi-square distribution with 1 df here, not an F-distribution, which is why only one value for degrees of freedom is provided]). This test is vulnerable to the assumption of normality, so consider the Levene test as a backup ($F_{1,143}$=0.4202, P=0.5179), although it is generally less powerful. An alternative test if the assumption of equal variance is not met is the Welch's approximate t-test, or the non-parametric Wilcoxon (Mann-Whitney U) test.

e. 95% CI: $-0.2418 < \mu < -0.04218$

f. see table (next page)

Means	*Hyptheses*	*statistic*	*df*	*P-value*	*Conclusion*
Mass (grams) (survived birds had non-normal dist'n) : W=0.9557, n=71, P=0.0373					
surv 22.846 died 22.808	Ho: $\mu_{died} = \mu_{survived}$ Ha: $\mu_{died} \neq \mu_{survived}$	Z=-0.039	145	0.9684	do not reject Ho
Wing length (mm) (died birds had non-normal dist'n) W=0.9317, n=74, P=0.0006					
surv 64.642 died 64.456	Ho: $\mu_{died} = \mu_{survived}$ Ha: $\mu_{died} \neq \mu_{survived}$	Z=1.179	145	0.2382	do not reject Ho
Tarsus (mm) Assumptions of normality and equal variances met					
surv 19.440 died 19.806	Ho: $\mu_{died} = \mu_{survived}$ Ha: $\mu_{died} \neq \mu_{survived}$	4.025	143	0.0001	reject Ho
Beak depth (mm) (died non-normal dist'n) W=0.9572, n=71, P=0.0410					
surv 5.8478 died 5.8405	Ho: $\mu_{died} = \mu_{survived}$ Ha: $\mu_{died} \neq \mu_{survived}$	Z=0.305	145	0.7603	do not reject Ho
Beak width (mm) (died non-normal dist'n) W=0.9299, n-71, P=0.0007					
surv 6.7169 died 6.6581	Ho: $\mu_{died} = \mu_{survived}$ Ha: $\mu_{died} \neq \mu_{survived}$	Z=1.646	145	0.0997	do not reject Ho

g. A visual inspection of trait distributions of birds that survived and died reveals that beak width, beak depth, beak length, and mass distributions of the 'died' birds had more outliers and higher standard deviation than those of the survivors, indicating possible stabilizing selection. The tarsus distributions appear similar, and the wing length distributions show more outliers for the birds that survived, indicating possibly disruptive selection.

   Only beak width showed a significant difference in variance between surviving and dead birds (Levene test, $F_{1,143}$=7.2112, P=0.0081)

h. The F-test is sensitive to departures from normality.

2. Taxi driver spatial memory

a. A sensible way to proceed is to choose the most powerful test (two-sample t-test for differences in means) but first we must check normal distribution and equal variances of two groups of taxi drivers' posterior hippocampus (<15 years and >15 years).

   Normality:
   Ho: Volume of posterior gray matter in taxi drivers follows a normal distribution.
   Ha: Volume of posterior gray matter in taxi drivers does not follow a normal distribution
   The null hypothesis was not rejected for either groups, and the histograms also do not indicate serious departures from normality.

   Variance:
   Ho: $\sigma_1^2 = \sigma_2^2$
   Ha: $\sigma_1^2 \neq \sigma_2^2$
   do not reject Ho (Bartlett's F=1.1502, df=1, P=0.2835; Levene $F_{1,13}$ = 1.4803, P=0.3246).
   Proceed with caution.

Two-sample t-test:
Ho: $\mu_{<15years} = \mu_{>15years}$
Ha: $\mu_{<15years} \neq \mu_{>15years}$
t=-2.169, v=13, P=0.0492. Therefore, reject Ho, conclude that there is a difference in gray matter in the posterior hippocampus of taxi drivers with < or > 15 years experience.

b. To use a two-sample t-test, test for normality in both groups of taxi drivers' anterior hippocampus (<15 years and >15 years), and for equal variance.

   Normality:
   Ho: Volume of anterior gray matter in taxi drivers follows a normal distribution.
   Ha: Volume of anterior gray matter in taxi drivers does not follow a normal distribution
   The null hypothesis was not rejected for either groups, and the histograms also do not indicate serious departures from normality.

   Variance:
   Ho: $\sigma_1^2 = \sigma_2^2$
   Ha: $\sigma_1^2 \neq \sigma_2^2$
   Reject Ho (F=0.4683, df = 1,13, P=0.4938) Variances equal.

   Two-sample t-test:
   Ho: $\mu_{<15years} = \mu_{>15years}$
   Ha: $\mu_{<15years} \neq \mu_{>15years}$
   t=3.796, P=0.0022. Therefore, reject Ho, conclude that there is a difference in gray matter in the anterior hippocampus of taxi drivers with < or > 15 years experience.

c. There is a *relationship* between gray matter volume and experience, but as there are no controls in the experiment to isolate driving experience, it cannot be said that there is a causal relationship between driving experience and gray matter volume.

d. The two-sample t-test has a higher probability of rejecting the null hypothesis when it is indeed false than does the U-test.

e. For power of at least .5, a sample size of ~16 is needed. For power of at least .9, a sample size of ~ 40 is needed.

f. Reducing the significance level alpha to 0.01 reduces power to 0.2496. Reducing alpha increases beta, the chance of a type II error, and therefore power is reduced (power = 1−beta).

g. No, these data do not confirm that gray matter changes with number of years carrying out difficult feats of spatial memory. In order to determine whether gray matter changes with time individual subjects would need to be followed over a long period. Also, a control group would be needed, made up of individuals not carrying out complicated feats of spatial memory.