9. ANSWERS TO PROBLEMS

1. Male fruit fly life spans

a. Visually inspecting the histograms of the five distributions shows that they are approximately normally distributed. Tests of normality bear this out.
Ho: Distribution of males with treatment (i) is normal
Ha: Distribution of males with treatment (i) is not normal
1 - 1 pregnant female. W=0.949, n=25, P=0.2494: do not reject Ho.
2 - 1 virgin female. W=0.969, n=25, P=0.6323: do not reject Ho
3 - 8 pregnant females. W=0.937, n=25, P=0.1331: do not reject Ho
4 - 8 virgin females. W=0.962, n=25, P=0.4867: do not reject Ho
5 - no females. W=0.9569, n=25, P=0.3700: do not reject Ho

b. The 8 virgin females group appears to be the most different from all of the other groups. The other groups have similar means (the 1 virgin females group has a slightly lower mean than the other 3 groups).

c. ANOVA
Ho: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$
Ha: At least one population mean is not equal
reject Ho. $F = 13.6120$, $P < 0.0001$. At least one of the means is different.

d. This is a fixed effects ANOVA. The treatments are repeatable, and the treatments were specifically chosen to see which particular ones are different.

e. It is invalid to test multisample hypotheses by applying two-sample tests to all possible pairs of samples because the probability of a type I error increases.

f. The ANOVA requires that the variances of each group be equal.
Ho: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2$
Ha: Not all variances are equal
Bartlett's $F = 0.6049$, df=4, $P = 0.6591$ (very sensitive to assumption of normality)
Levene's $F = 0.5405$, df=4,120 $P = 0.7062$
Do not reject Ho.

g. The group with the lowest lifespan is the one exposed to 8 virgin females.
The group with the highest lifespan is the one exposed to one pregnant female. A Tukey test reveals that the 8 virgin female group is significantly different from all the other groups. This can be seen by clicking on the comparison ring for each group, which highlights all other groups whose means are not significantly different from it. None of the other groups were highlighted along with the 8 virgin females group.

h. The Tukey test assumes the same things as the ANOVA: normal distributions and equal variances in all populations.

i. A visual inspection of the distributions of thorax length in each of the five groups shows that the normality assumption may not be met. Indeed tests of normality for three of the five distributions – no females, one virgin female and one pregnant female – reject the null hypotheses of normality. Transforming the data with a log transformation does not improve the situation. Therefore a non-parametric test should be used – the Wilcoxon ranked test.

   Ho: Thorax lengths of the five fruit fly populations are the same
   Ha: Thorax lengths in at least one population is not the same as in the others
   $\chi^2 = 5.9277$, $df = 4$, $P = 0.2046$. Do not reject Ho. There is no evidence that the populations are different.

2. Mimicry in leatherjackets

a. The assumptions of the ANOVA are: random sampling, normal distributions of each population being compared, and equal variance of all populations being compared. Shapiro-Wilks tests for normality do not reject the null hypothesis that the populations are normally distributed. Our sample sizes are very small, however, and visual inspection of the data shows that the distributions look somewhat non-normal. Variances must be tested for equality:
   Ho: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$
   Ha: Not all variances are equal
   Bartlett's F = 2.0442, $P = 0.1053$ (highly sensitive to normality assumption)
   Levene's F=3.2730, df=3,16, P=0.0486. Reject Ho. This assumption of ANOVA does not appear to be met.

b. The data are counts so the square root transformation is a good choice to start. Using $\sqrt{(X+0.5)}$ or $\sqrt{(X)}$ no longer rejects Ho: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$ when using the Levene's test (so some improvement there). Normality is still difficult to assess given the small sample size.

c. A log transformation, $\log(X+1)$, yielded similar results as the square root, and the variability of the standard deviations among groups was even less after transformation than was the case with the square root transformation. Tests of equal variance yielded the largest P-values using this transformation. Perhaps the log is a better transformation, therefore, than the square root, although the improvement may be minor

d. You may decide to be brave and do a parametric test (ANOVA). I used the log-transformed data (you may get different results if you used a different transformation).
   Ho: $\mu_1 = \mu_2 = \mu_3 = \mu_4$
   Ha: At least one population mean is not equal
   $F = 3.7455$, $P = 0.0327$. Reject Ho. At least one of the means is different.

   Or you may to decide to be conservative, given the uncertainty about normality, and use a nonparametric test.

Wilcoxon/Kruskal Wallis test
Ho: Predator approaches were not different between the four colour-pattern groups.
Ha: Predator approaches in at least one of the four populations is different from the rest.
$\chi^2 = 8.1338$, df = 3, $P = 0.0433$. Reject Ho. The means are different.

e. The two most extreme differences between means was significant using the Tukey test (2-step and leatherjacket), with 2-step being most attractive and leatherjacket least attractive. The grouping of the other means was ambiguous: none is significantly different from either of the extremes.

f. This is a fixed effects ANOVA. The treatments applied to the predators (plastic models) is repeatable, and the treatments themselves are of interest.

3. Mother-offspring rat relations

a. With such small sample sizes, the results of the Shapiro-Wilks tests are suspect (power of tests very low). A test of unequal variances
Ho: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_{15}^2$
Ha: Not all variances are equal
fails to reject the null hypothesis (Levene test, F=1.6453, df=14,46, P=0.1025)
Since the data are weight gains, a log transformation might improve matters, but in this case the results of the Levene test was about the same (F=1.5487, df=14,46, P=0.1318).

You may decide to be brave and do a parametric test (ANOVA). I used the log-transformed data.
Ho: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_{15}$
Ha: At least one population mean is not equal
$F = 1.4323$, df=14,46, $P = 0.1770$. Do not reject Ho. There is no evidence of a difference in mean weight gain of offspring between foster mothers.

Or you may to decide to be conservative, given the uncertainty about normality, and use a nonparametric test.
Wilcoxon/Kruskal Wallis
Ho: Weight gains of offspring does not differ between foster mothers
Ha: Weight gains of offspring differs between foster mothers
$\chi^2 = 20.999$, df = 14, $P = 0.1017$. Do not reject Ho. There is no evidence of a difference between foster mothers.

b. The distributions of offspring weight gains of the true mothers have the same problem meeting the normality assumption (i.e., difficult to test when sample size is so low for each mother). There is also a hint of unequal variances, although not a very strong one:
Ho: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_{15}^2$
Ha: Not all variances are equal
Do not reject Ho (Levene test, F=1.6598, df=14,46, P=0.0987).

You may decide to be brave and do a parametric test (ANOVA). I used the untransformed data (there is stronger evidence of unequal variances with the log-transformed data)

Ho: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_{15}$

Ha: At least one population mean is not equal

$F = 2.2298$, df=14,46, $P = 0.0209$. Reject Ho. Mean weight gain of offspring differs between true mothers.

Or you may to decide to be conservative, given the uncertainty about normality and unequal variances, and use a nonparametric test (this is probably the best approach)

Wilcoxon/Kruskal Wallis

Ho: Weight gains of offspring does not differ between true mothers

Ha: Weight gains of offspring differs between true mothers

$\chi^2 = 24.8642$, $df = 14$, $P = 0.0359$. Reject Ho. Weight gains of offspring differs between true mothers.

c. A difference between true mothers in the weight gain of her offspring might be explained by genetics: offspring are similar genetically (same true mother) and they might consequently be more similar to one another in rate of weight gain. Or, the rate gain of offspring might be strongly determined by a non-genetic, maternal effect, if weight gain of offspring is influenced by factors in the womb (e.g., competition for nutrients, or shared womb environment). The inability to detect an influence of foster mothers may be explained by the same factors (genetics or maternal effects might outweigh the foster mother's effect on offspring weight gain).

d. This is a random effects design. True mothers, and also foster mothers, were randomly sampled from a population of mothers. We are not interested in the effect that specific mothers have on their offspring weight gain, but only whether there is an overall difference among mothers.