Biostats VHM 801 Course
Fall 2019, Atlantic Veterinary College, PEI
Henrik Stryhn

Solution to home assignment III

The solution presents one way in which the statistical and resulting discussion could be done, but other ways are possible as well. Some parts are more detailed than required for a 100% mark. All analyses shown used Minitab 18, but Stata or other programs would give similar figures and results.

1. Confidence interval for weight gain

As discussed in Home assignment I, the distribution of the weight gains among the 1200 fish differs from a normal distribution by having to heavy tails, as indicated for example by a too large of number of suspected outliers both at the lower and upper end of the distribution. Therefore the natural model is that the 1200 observations form a simple random sample, from a distribution with unknown mean (μ) and standard deviation (σ). If we let X_i denote the weight gain for the ith fish, we could say

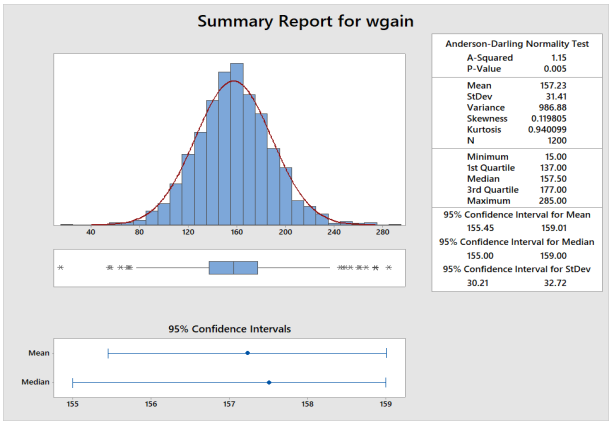
$$X_1, \ldots, X_{1200} \text{ are i.i.d. from some distribution } (\mu, \sigma).$$

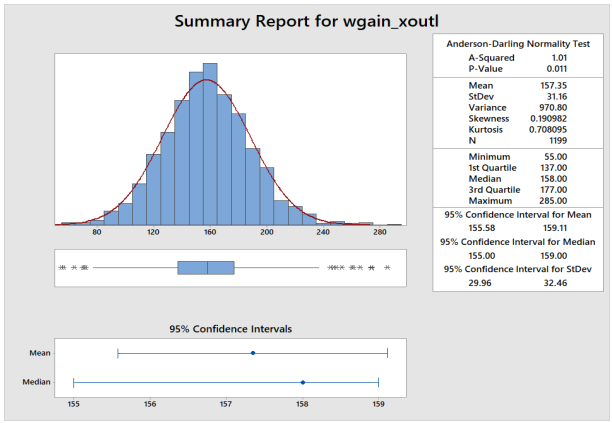
We estimate the parameters by the sample statistics: $\hat{\mu} = \bar{X} = 157.23$ and $\hat{\sigma} = s_X = 31.41$. Even though the distribution is not assumed normal, t-distribution inference for μ may be valid due to the large sample size. According to the guidelines from Lecture 7 (slide 7L−3), as $n \gg 40$ our only concern would be strong outliers. Without a more specific definition of "strong outlier", it is somewhat subjective whether the distribution (displayed below in the left graphical summary box) has any strong outliers. If not, our 95% CI simply becomes simply the interval based on the t-distribution (with df = 1199) displayed in the box:

$$95\% \text{ CI for } \mu: \ (155.45, 159.01).$$

As discussed in the earlier home assignment, only the smallest observation ($15g$) seems to be a candidate for a real outlier and by visual judgement also for a "strong outlier", so it could be of interest to explore its impact on the results. That is, we redo the estimation and 95% CI without this observation (shown in the right graphical summary):

$$\textit{(without outlier):} \quad \hat{\mu} = \bar{X} = 157.35, \quad \hat{\sigma} = s_X = 31.16, \quad 95\% \text{ CI for } \mu: \quad (155.58, 159.11).$$





Unsurprisingly, when the lowest value is excluded the mean increases slightly and the standard deviation increases slightly, and the resulting change in the CI limits amounts to about $0.1g$. With so small changes it seems fair to say that the results are quite insensitive to the potential outlier. It would therefore be most natural to keep it as part of the analysis (unless it could be established that the observation was a real error).

2. Weight gains for different sexes

We will follow the suggested approach in the question to first compare the two male groups. These are naturally considered as two independent samples, and we will assume the samples to represent two distributions with unknown and different parameters: μ_1 and μ_2 for the means and σ_1 and σ_2 for the standard deviations (where 1 $\sim$ normal and 2 $\sim$ precocious males (as coded in the `sex3` variable). There is no pressing need to assume the standard deviations to be equal, although the estimates (below) are rather close. The parameter estimates are the respective sample statistics:

$$\hat{\mu}_1 = \bar{X}_1 = 155.76, \;\; \hat{\sigma}_1 = s_1 = 31.96,$$
$$\hat{\mu}_2 = \bar{X}_2 = 162.06, \;\; \hat{\sigma}_2 = s_1 = 29.76.$$

The comparison of the two groups uses t-distribution inference for two independent samples. The statistical hypotheses of interest are:

$$\mathrm{H}_0 : \mu_1 = \mu_2, \;\; \mathrm{H}_a : \mu_1 \neq \mu_2,$$

where we in absence of any prior information about the difference between the groups use a two-sided alternative hypothesis. The Minitab display shows the 95% CI for $\mu_1 - \mu_2$ to be entirely below 0, and the 2-sample t-test statistic equals $t = -2.93$ with a corresponding $P = 0.004$ from $t(667)$. We conclude there is clear evidence of a difference in weight gains, whereby the precocious males have grown more than the non-precocious males, an average difference of about $6.3g$.

Two-Sample T-Test and CI: wgain, sex3

Method

μ₁: mean of wgain when sex3 = 1
μ₂: mean of wgain when sex3 = 2
Difference: μ₁ - μ₂

Equal variances are not assumed for this analysis.

Descriptive Statistics: wgain

sex3	N	Mean	StDev	SE Mean
1	584	155.8	32.0	1.3
2	309	162.1	29.8	1.7

Estimation for Difference

Difference	95% CI for Difference
-6.29	(-10.51, -2.07)

Test

Null hypothesis H₀: μ₁ - μ₂ = 0
Alternative hypothesis H₁: μ₁ - μ₂ ≠ 0

T-Value	DF	P-Value
-2.93	667	0.004

The two weight gain distributions show the same characteristics as the combined distribution (reasonably symmetric, heavy tails, some evidence against normality by the A-D test), but by the same reasoning as for the combined sample our only concern is with strong outliers. The $15g$ weight gain is for a non-precocious male, so to address this concern we may rerun the analysis without this observation. The results are quite similar:

$$\textit{(without outlier):} \;\; 95\% \text{ CI for } \mu_1 - \mu_2 : \; (-10.25, -1.86), \;\; t = -2.83, \;\; P = 0.005.$$

This shows that the results and certainly the conclusion is not sensitive to inclusion of this observation. Another possibility would be to compare the two distributions by a non-parametric test, the Mann-Whitney-Wilcoxon test. It gives $P = 0.007$, thus also providing evidence of a difference between the distributions (or their medians if we were willing to assume the distributions to be of the same shape).

The second step of our analysis will compare the weight gains to the female group. Because of the clear difference between the two male groups, it seems little meaningful to compare the females to the two male groups combined. It turns out that the mean weight gain for the females ($155.15g$) is close to that of the non-precocious males, and it is perhaps also the biologically most meaningful comparison (of normally developing fish). Therefore, this will be our first focus; we summarize the results briefly as follows:

$$\begin{aligned} \text{estimation :} \quad & \hat{\mu}_0 = \bar{X}_0 = 155.15, \;\; \hat{\sigma}_0 = s_0 = 31.57, \\ 95\% \text{ CI for } (\mu_0 - \mu_1) : \quad & -0.61 \pm 4.39 = (-5.00, 3.78) \\ \text{2-sample } t\text{-test :} \quad & t = -0.27, \;\; P = 0.785. \;\; (\mathrm{H}_0 : \mu_0 = \mu_1, \;\; \mathrm{H}_a : \mu_0 \neq \mu_1) \end{aligned}$$

The results show that there is no statistical evidence to support a difference in mean weight gains between females and non-precocious males. By similar analyses as above, it can be established that this conclusion is not affected by the fish with lowest weight gain.

An additional comparison between females and precocious males may be conducted as well, along similar lines as above. As indicated earlier, such a comparison may not be of real biological interest. It can also be argued that because we already concluded there was a clear distinction between non-precocious and precocious males and there was no evidence whatsoever to distinguish the females from the non-precocious males, this comparison seems of little interest any more. Generally speaking such reasoning should be used with caution, but because our two previous conclusions were very clear it is easy to guess that a 2-sample comparison between females and precocious males will show statistical significance (indeed, $t = -2.79$ and $P = 0.005$). We conclude that the precocious males have higher average weight gains than non-precocious males and females, which in turn have similar average weight gains. Our findings are strong enough for the conclusions not to be affected by adjustment for multiple (3) comparisons, if desired; this issue was however considered beyond the scope of the home assignment.

3. Comparison of sex distributions across vaccine groups

The `sex3` variable is categorical with three categories, and a cross-tabulation of the counts in these categories with the vaccine groups leads to a two-way table. For the statistical model behind a two-way analysis we have a choice between a model (I) to compare several multinomial distributions and a single multinomial distribution model (II) to assess independence between two variables. Our choice of model is often guided by the number of response variables in the data obtained. Here `vaccine` is a controlled variable (with 200 fish per vaccine group, and vaccines administered early in the study), and the randomness would have been in the sex of the selected fish or in the potential development of a precocious condition in the selected fish. This way of looking at the randomness seems most natural considering that the the fish were probably part of a much larger population at the hatchery. This would to a model I for comparing multinomial distributions across vaccine groups. Even if a model II can possibly also be justified, the wording of the question towards comparing sex distributions within vaccine groups would effectively direct us to look at conditional distributions for each vaccine. These conditional distributions are effectively the multinomial distributions involved in the model I across vaccine groups.

We denote by N_{ij} the count of fish in sex category i ($i = 0, 1, 2$) and vaccine group j ($j = 1, \ldots, 6$). For vaccine group j, we assume the counts (N_{0j}, N_{1j}, N_{2j}) to follow a multinomial distribution with total n_j, where all $n_j = 200$, and probabilities (p_{0j}, p_{1j}, p_{2j}). Our interest is in estimating the parameters p_{ij} and to assess the null hypothesis of equal sex distributions across vaccine groups, i.e.

$$\mathrm{H}_0 : p_{01} = p_{02} = \ldots = p_{06}, \quad \text{and} \quad p_{11} = p_{12} = \ldots = p_{16}, \quad \text{and} \quad p_{21} = p_{22} = \ldots = p_{26},$$

against the two-sided alternative (H_a) that H_0 does not hold. The statistical analysis uses a Pearson X^2 (chi-square) test, and the following Minitab display (on the next page) contains the observed counts and the estimated parameters (i.e., the within-column proportions).

The test statistic is seen to be totally non-significant: $X^2 = 8.940$, df $= 10$ and $P = 0.54$. All expected counts (not shown) are much larger than 5, e.g. $e_{11} = 307 \cdot 200/1200 = 51$. The estimated proportions for the three sex categories are seen to be very similar across vaccine groups (as also indicated by the test). We conclude there is nothing to indicate that the sex distributions are not essentially identical between vaccine groups.

Tabulated Statistics: sex3, vaccine

Rows: sex3 Columns: vaccine

	1	2	3	4	5	6	All
0	47	51	59	46	53	51	307
	23.50	25.50	29.50	23.00	26.50	25.50	25.58
1	103	93	92	99	107	90	584
	51.50	46.50	46.00	49.50	53.50	45.00	48.67
2	50	56	49	55	40	59	309
	25.00	28.00	24.50	27.50	20.00	29.50	25.75
All	200	200	200	200	200	200	1200
	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Cell Contents
Count
% of Column

Chi-Square Test

	Chi-Square	DF	P-Value
Pearson	8.940	10	0.538
Likelihood Ratio	9.049	10	0.527

4. Comparing weight gains across vaccine groups

Although as indicated in the wording of the question it would have been valid for the home assignment to conduct a number of pairwise comparisons between vaccine groups (in order to avoid utilizing very recently introduced material in the course), the natural analysis involves a simultaneous comparison of all six vaccine groups, in a typical one-way ANOVA layout.

If we let X_{ij}, $i = 1, \ldots, 6 \sim$ vaccines and $j = 1, \ldots, 200 \sim$ fish, denote the weight gains, the one-way ANOVA model assumes all observations to be independent and $X_{ij} \sim \mathrm{N}(\mu_i, \sigma)$, or written in an equation:

$$X_{ij} = \mu_i + \varepsilon_{ij}, \quad \text{where} \quad \varepsilon_{ij} \sim \mathrm{N}(0, \sigma).$$

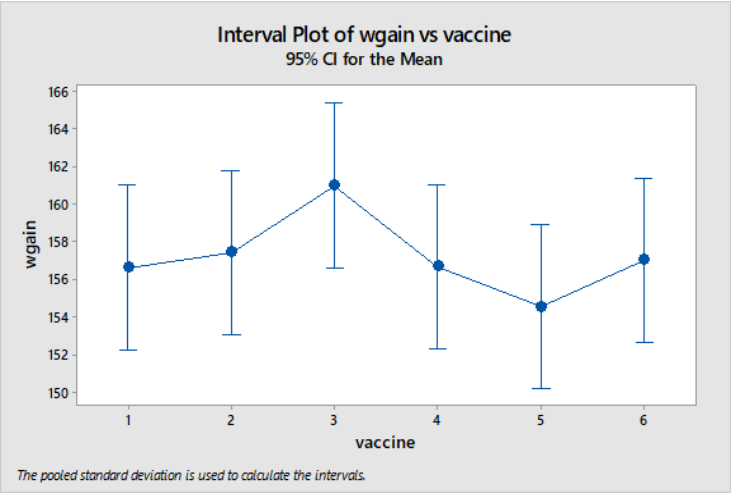
We previously discussed concerns with assuming the weight gains to be normally distributed, but here there is no choice. The assumption of equal standard deviations across the six vaccine groups should be fine with the sample standard deviations very similar across groups (table below; specifically, the ratio of the largest to lowest s_i equals $33.54/30.13 \ll 2$).

The analysis involves estimation of the mean parameters of the different vaccine groups (i.e., by the sample means) and of the pooled standard deviation. From these estimates confidence intervals for the mean parameters are obtained, shown in the table below and in the graph on the next page.

Means

vaccine	N	Mean	StDev	95% CI
1	200	156.64	33.54	(152.28, 161.00)
2	200	157.43	31.30	(153.07, 161.79)
3	200	161.00	30.83	(156.64, 165.35)
4	200	156.70	31.45	(152.34, 161.06)
5	200	154.56	30.13	(150.20, 158.92)
6	200	157.04	31.18	(152.68, 161.39)

Pooled StDev = 31.4217



Next, we set up the ANOVA table for testing the null hypothesis of same means across the vaccine groups ($H_0 : \mu_1 = \ldots = \mu_6$) against the non-specific alternative (H_a) that some of the means differ.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
vaccine	5	4402	880.4	0.89	0.486
Error	1194	1178864	987.3		
Total	1199	1183266			

It is seen that the F-statistic is totally non-significant: $F(5, 1194) = 0.89$ and $P = 0.49$. There is no evidence whatsoever to indicate that there are substantial differences between the mean weight gains across vaccine groups. The plot of means with CIs tells the same story, with all confidence intervals clearly overlapping. We have no reason to further conduct pairwise comparisons between vaccine groups, and if such comparisons were to be conducted they would need to take into account a large number of pairwise comparisons ($= 6 \cdot 5/2 = 15$).

One may speculate that this apparently very clear-cut conclusion is robust to any deviations from the model assumptions; we would certainly need to be more concerned with less clear findings. The assumed equal standard deviation across vaccine groups seems very reasonable when the individual group standard deviations are so close. We are nowhere near a violation of the guideline for when this should be considered a concern (i.e., the ratio between the most extreme standard deviations exceeding 2). It is more difficult to assess the impact of violating the assumed within-group normal distributions. Normality probability plots and normality tests for each vaccine group show the same problems with the tails as previously discussed, and the A-D test shows significance for vaccines 1 and 6 (the latter including our low outlier of $15g$). We can rerun the analysis without the low outlier, but as expected it barely changes the F-value and certainly not the conclusion ($F = 0.91$, $P = 0.47$). We can also try the non-parametric Kruskal-Wallis test to compare the vaccines, and it is also clearly non-significant ($P = 0.34$). In summary, it seems our finding is quite robust to the assumptions, and we may conclude that the data show no evidence of (substantial) differences weight gains across vaccine groups.