SUTD ESD

Statistics 2017

Assignment 4 Solutions

Refer to the spreadsheet for calculations.

Question 1. (a) Let m be the median of the differences between the treatment and control groups. We want to test $H_0 : m = 0$ vs $H_1 : m > 0$. We ignore the value $d_{11} = 0$ since it equals the hypothesized median. Thus $n = 11$ and $s_+ = 9$. The p-value according to the sign test is

$$P(S_+ \geq 9) = \sum_{i=9}^{11} \binom{11}{i} 2^{-11} \approx 0.0327 < \alpha.$$

(b) We rank the d_i's by their absolute values, again ignoring the 0. In case of ties, we assign the average rank: for instance, $|d_7| = |d_8| = |d_{10}|$; if they were slightly different then they would receive the ranks of 2, 3 and 4 in some order, but here they each receive a rank of $(2+3+4)/3 = 3$. Proceeding this way, we find that $w_+ = 47$.

Using $n = 11$ and the formulas given in the Week 12 slides, we have $\mu = 33$ and $\sigma^2 = 126.5$. With the continuity correction, w_+ gets a z-score of 1.2 and hence the p-value is $0.115 > \alpha$.

So H_0 is rejected using the sign test but not rejected using the signed rank test. An explanation for this is that the negative d_i's, -7 and -12, have large absolute values and 'balance out' the more numerous but smaller positive d_i's. This may suggest that the d_i's are *not* drawn from a symmetric population, and hence the assumption of the signed rank test is not satisfied.

Question 2. If MST = MSA + MSE, then we would have

$$\frac{\text{SST}}{N-1} = \frac{\text{SSA}}{k-1} + \frac{\text{SSE}}{N-k} = \frac{\text{SSA}+\text{SSE}}{(k-1)+(N-k)}.$$

Therefore, after renaming the variables, we are asking if it is possible to have

$$\frac{a}{c} + \frac{b}{d} = \frac{a+b}{c+d}.$$

After multiplying both sides of this equation by $c + d$, we get

$$a \times \frac{d}{c} + b \times \frac{c}{d} = 0,$$

which is *impossible* since all the quantities involved are positive. (The only exception is when $a = b = 0$, i. e. SSA = SSE = 0, which does not occur in practice.)

Question 3. (a) F is much larger than the critical value – in fact the p-value is 4×10^{-11}; hence this is very strong evidence that the mean salinity at the three sites are different.

(b) We first compute $\bar{\bar{y}} = (1118.66 + X)/30$, $\bar{y}_1 = (408.91 + X)/12$, $\bar{y}_2 = 40.10375$, and $\bar{y}_3 = 38.892$.

Now SSA = 38.8008825 from the table; on the other hand, $\text{SSA} = 12(\bar{y}_1 - \bar{\bar{y}})^2 + 8(\bar{y}_2 - \bar{\bar{y}})^2 + 10(\bar{y}_3 - \bar{\bar{y}})^2$ by the formula. After making the substitutions, we obtain a quadratic in X which has two solutions, 38.85 and 89.66.

We can check that the larger solution is invalid (e. g. by constructing an ANOVA table based on that solution), hence we must have $X = 38.85$.

It is also possible to solve for X using SSE or SST. You can use *Excel*'s Solver or Goal Seek function (see spreadsheet for Goal Seek using SSE); however, *Excel* may return the wrong root of the quadratic.

Question 4. See spreadsheet. Here $\alpha = 0.05$, $m = \binom{3}{2} = 3$, $s^2 = 3.996$, and the cut-off difference, computed using the formula, is 1.559.

We conclude that shelf 2 is significantly different from shelf 1, and also from shelf 3. However, there is not enough evidence to show that shelf 1 and shelf 3 are different.

Question 5. (a) See *Excel* for the interaction plot. Since the two lines are almost parallel, there does not appear to be any interaction.

(b) An ANOVA table is produced using *Excel*'s Anova: Two-Factor With Replication option, with 8 as the 'Rows per sample'. The p-value for F_{AB} is 0.93, which is greater than any α we would pick, thus confirming our answer to part (a) (that we cannot reject the hypothesis of no interaction).

The p-value for F_A is 0.0098, and the p-value for F_B is 0.00059. Thus we can conclude that IQ seems to depend on both the biological and the adoptive parents' socioeconomic status, if α is set at a reasonable value, such as 0.05 or 0.01.

Question 6. $L(\theta) = \dfrac{1}{2^{2n+1}} \exp\left(-\sum_{i=1}^{2n+1} |x_i - \theta|\right)$. Maximizing $L(\theta)$ is equivalent to minimizing $\sum_{i=1}^{2n+1} |x_i - \theta|$. In Week 2 lecture 1, we learned that the minimum for this sum is achieved when θ is the *median*, $\tilde{x}$, of the x_i's. Thus the maximum likelihood estimate is $\hat{\theta} = \tilde{x}$.

Note 1: we cannot differentiate $L(\theta)$ in this question, because the absolute value signs make the function not everywhere differentiable.

Note 2: to be more precise, in Week 2 we proved that, when there are an even number of data values, then the minimum is achieved when θ is anywhere between the middle two data values. Here, there are an odd number of data values, but we can slightly modify the proof (by adding $|x_{(n+1)} - \theta|$ to both sides) to show that the minimum is only achieved at $\tilde{x}$.

Question 7. (a) (Critical strikes) There are $n = 402$ normal strikes and $m = 70$ critical strikes. So $\hat{p} = m/(m+n) = 70/472 \approx 0.148$. Using the formula for the CI on a proportion, we find that

$$p \in \left[\hat{p} - z_{0.975}\sqrt{\frac{\hat{p}(1-\hat{p})}{472}},\ \hat{p} + z_{0.975}\sqrt{\frac{\hat{p}(1-\hat{p})}{472}}\right] \approx [0.116, 0.180]$$

with 95% confidence, where p is the true proportion of critical strikes. The purported value, 15%, lies inside the CI.

(Bashes) There are $n = 690$ normal strikes and $m = 130$ bashes. So $\hat{p} \approx 0.159$ and the 95% CI is $[0.134, 0.184]$, which contains the purported value, 17%.

(b) (Critical strikes) The question is asking for the probability of two strikes in a row both being critical; under the independence assumption, this is given by $0.15^2 = 0.0225$. In 472 strikes, we expect this to occur 10.6 times. In the data it occurs only 4 times, which is rather low.

To show that the expected number is 10.6, represent the ith strike by X_i, where $X_i = 1$ if it is a critical and 0 otherwise. Now, define the random variables Y_i, $i = 1, 2, \ldots, 471$, by

$$Y_i = \begin{cases} 1 & \text{if } X_i = X_{i+1} = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then the expected number of double criticals (where a triple is counted as two doubles, etc) is

$$E\Big(\sum_{i=1}^{471} Y_i\Big) = \sum_{i=1}^{471} E(Y_i) = 0.15^2 \times 471 \approx 10.6.$$

(Bashes) The probability of two bashes in a row is $0.17^2 = 0.0289$. In 820 strikes, this is expected to occur $0.0289 \times 819 \approx 23.7$ times. In the data it occurs 14 times, which also seems low (but proportionally, closer to the expected value than the critical strikes).

(c) (Critical strikes) There are in total 133 runs. The distribution of runs is approximately normal with $\mu \approx 120.2$, $\sigma^2 \approx 29.93$ (computed using the above n, m, and the formulas taught in class; see spreadsheet). With the continuity correction, the p-value is $P(|Z| > (132.5 - \mu)/\sigma) \approx 0.025 < \alpha$, thus we conclude that critical strikes are *not* independent.

(Bashes) There are in total 232 runs. Using the same procedure as above, we find that $\mu \approx 219.8$, $\sigma^2 \approx 58.18$, and the p-value is around $0.124 > \alpha$. Therefore we cannot reject the hypothesis that the bashes are independent.

Note: many games use a 'pseudo-random' distribution, where the probability of (say) a critical strike is set to be z% immediately after a successful critical (and after learning the skill), and increases to $(2z)$% if it fails to occur upon the next strike, then increases to $(3z)$% if it fails to occur again, etc. To ensure an average of 15%, z is set to be around 3.222. This distribution also ensures that a critical strike is most likely to occur after 5–6 normal strikes. The data gathered (from a game of *Dota 2*) for critical strikes are highly consistent with what one would expect from a pseudo-random distribution, while the data for bashes are more consistent with a truly random (iid Bernoulli) process.