Final Exam Semester 1, 2021 – Solutions

Obstructive Sleep Apnea

1. The standard error is $\text{se}(\bar{x}) = \frac{12.34}{\sqrt{30}} = \mathbf{2.253}$.

2. The margin of error is $t^*\text{se}(\bar{x}) = 2.045 \times 2.253 = \mathbf{4.608}$. The value of t^* comes from `qt(.975,df=`~~#~~`29)`.   *qt(0·975, df ¿ 29)*

3. They are looking to "compare the efficacy" so will test $H_0: \boldsymbol{\mu_M = \mu_T}$ vs $H_1: \boldsymbol{\mu_M \neq \mu_T}$.

4. The standard error is $\sqrt{\frac{s_M^2}{n_M} + \frac{s_T^2}{n_T}} = \sqrt{\frac{9.30^2}{15} + \frac{10.65^2}{15}} = \mathbf{3.651}$.

5. The difference in scores is $8.69 - 7.80 = 0.89$, so the statistic is $t = \frac{0.89 - 0}{3.651} = \mathbf{0.244}$.

6. The degrees of freedom are $\min(15 - 1, 15 - 1) = \mathbf{14}$.

7. The p value is `2*(1-pt(0.244,df=14))` = **.8108**, giving **no evidence**.

8. A **rank-sum test** could be used to compare the two groups.

9. The sample proportion is $\hat{p}_M = \frac{12}{15} = .8$, with standard error $\text{se}(\hat{p}_M) = \sqrt{\frac{0.8(0.2)}{15}} = \mathbf{.1033}$.

10. The margin of error is $z^*\text{se}(\hat{p}_M) = 1.960 \times .1033 = \mathbf{.2024}$. The value of z^* comes from `qnorm(.975)`.

11. The sample proportion is $\hat{p}_T = \frac{6}{15} = .4$, with standard error $\text{se}(\hat{p}_M - \hat{p}_T) = \sqrt{\frac{0.8(0.2)}{15} + \frac{0.4(0.6)}{15}} = \mathbf{.1633}$

12. Test $H_0: p_M = p_T$ vs $H_1: p_M \neq p_T$ using $z = \frac{(.8 - .4) - 0}{.1633} = 2.45$, so p-value is `2*(1-pnorm(2.45))` = **.0142**, giving **moderate evidence**.

MDMA and Post-traumatic Stress Disorder

1. **Randomised comparative experiment**

2. Using `aggregate(Before ~ Dose, mdma, mean)`, we find the mean score for the low dosage group is **77.6**.

3. The mean change in CAPS-IV score **is the same** for all dosage levels.

4. With $n = 54$ and $k = 3$ groups, residual degrees of freedom are $54 - 3 = $ **51**.

5. Using `summary(aov(Change~Dose, data=mdma))`, the total sum of squares is $38.93 + 221.17 = $ **260.1**.

6. From the ANOVA table in Q5, the p-value is 0.016, giving **moderate evidence**.

7. The R^2 value is $R^2 = \frac{38.93}{260.1} = $ **.1497**.

8. One-way ANOVA does not require **a linear relationship**.

9. Normal probability plot of **the residuals**.

10. Using `TukeyHSD(aov(Change~Dose, data=mdma))`, the only significant difference is "**Low-High**".

11. Using `table(mdma$Drop)` shows **23** subjects experienced a drop.

12. Using `prop.table(table(mdma$Dose,mdma$Drop))` shows **.093** of participants received a low dosage of MDMA and experienced a drop.

13. Using `addmargins(table(mdma$Dose,mdma$Drop))` gives the two-way table of counts with row and column totals. The expected count for subjects in the low dosage group experiencing a drop is $\frac{18 \times 23}{54} = $ **7.67**.

14. Using `chisq.test(table(mdma$Dose,mdma$Drop))` gives the χ^2 statistic **2.424**.

15. The degrees of freedom are $(\#\text{rows} - 1)(\#\text{columns} - 1) = (3 - 1)(2 - 1) = 2$.

16. The chi-square test in Question 14 shows $p = $ **.298**, giving **no evidence**.

Age Estimation from Blood Cells

1. Using `sd(dna$ELOVL2)` gives standard deviation **13.64 %**.

2. Using `bwplot(~ ELOVL2, dna)` and/or `histogram(~ ELOVL2, dna)` we see that the distribution of elevation is **symmetric**.

3. Using `cor(dna$Age,dna$ELOVL2)` gives $r = 0.88$, suggesting a **strong positive linear relationship**.

4. Using `summary(lm(Age ~ ELOVL2, data=dna))` gives a slope of **1.1842 years/%**.

5. The regression summary shows the standard error for the ELOVL2 estimate is 0.07401. With df = 71 the critical value for 95% confidence is qt(.975, 71) = 1.9939, so the margin of error is $1.9939 \times 0.07401 = \textbf{0.1476 years/\%.}$

6. Substitute 41.2 in the regression equation from Question 4, or use `predict(lm(Age ~ ELOVL2, data=dna), newdata=data.frame(ELOVL2=41.2))`, to get **55.67 years**.

7. Use `predict(lm(Age ~ ELOVL2, data=dna), newdata=data.frame(ELOVL2=41.2), interval="predict")` to obtain the interval (38.47, **72.86**) years.

8. `summary(lm(Age ~ ELOVL2+PDE4C, data=dna))` gives regression coefficient estimates. For PDE4C the estimate is **0.3790 years/%**.

9. Use `predict(lm(Age ~ ELOVL2+PDE4C, data=dna), newdata=data.frame(ELOVL2=41.2,PDE4C=45.8))`, to get **57.504 years**.

10. Person 1 in the data has an age of 54, giving a residual of $54 - 57.504 = \textbf{−3.504 years}$.

11. Using `plot(lm(Age ~ ELOVL2+PDE4C, data=dna),which=2)` shows that the distribution of the residuals appears **roughly normal**.

12. The Normal probability plot does **not** assess linearity.

13. The regression summary in Question 8 shows $p = 0.0413$, **moderate evidence**.

14. Both slopes are positive, so age is associated with a **high percentage of DNA methylation at both ELOVL2 and PDE4C**.

Parasites

1. The expected value is $E(X) = .18(0) + .34(1) + .23(2) + .25(3) = \mathbf{1.55}$.

2. The variance is $\text{var}(X) = .18(0 - 1.55)^2 + .34(1 - 1.55)^2 + .23(2 - 1.55)^2 + .25(3 - 1.55)^2 = 1.1075$. Thus the standard deviation is $\sqrt{1.1075} = \mathbf{1.052}$.

3. The probability of at least one worm is $1 - .18 = \mathbf{.82}$.

4. **Binomial(10,.82)**

5. Using `sum(dbinom(0:6,10,.82))`, or `pbinom(6,10,.82)`, gives **.0883**.

6. $X \sim \text{Binomial}(100, .82)$ has approximately a Normal distribution with $E(X) = np = 100 \times .82 = \mathbf{82}$ and $\text{sd}(X) = \sqrt{100(.82)(1 - .82)} = \mathbf{3.84}$.

7. 82% of fish are infected and 80% of those will be identified, giving $.82 \times .80 = \mathbf{.656}$.

8. Of the 82% infected, 20% appear to have no parasites, with $.82 \times .20 = .164$. Of the 18% not infected, we assume none will test positive for parasites. So if a fish is identified to have no parasites, the probability it is in this latter group is $\frac{.18}{.164 + .18} = \mathbf{.523}$.